



*Predictors of SDG Disclosures in Chinese Public Companies:
A Machine Learning Approach*

Muchen Yu

Discipline of Accounting, Regulation, and Governance

The University of Sydney Business School

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

2025

Statement of Originality

I affirm that the content of this thesis is my original work. This thesis does not contain materials submitted for any other degree.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been properly acknowledged and cited.

I affirm that I did not use generative AI for writing. I only used AI tools to check my grammar, such as Grammarly. After I used the tool, I also reviewed and edited it to ensure the accuracy of my thesis.

Student: Muchen Yu

Signature:

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Associate Professor Geoff Frost. Without his guidance, expertise, and support, I could not have finished this journey. His insightful feedback consistently challenged me to push my boundaries of thinking and explore new directions. His inspiring approaches have shaped my own way of approaching research, my skills (i.e. problem-solving and critical thinking) and developing my PhD graduate qualities. His patience, encouragement and mentorship are instrumental to shape my research and personal growth. Words cannot fully capture my gratitude for the impacts he has had on my academic journey, future career development and my growth.

I am also extremely grateful to my supervisor, Professor Stewart Jones, for his profound expertise and invaluable guidance throughout my PhD journey. His criticism and academic feedback are crucial for me to refine my research design, my skills (i.e. critical thinking), improve empirical analysis and PhD write up. His patience, and encouragement have kept me motivated during the challenging part of the research, which also have great impacts on my personal and professional growth. This endeavour would not have been possible without his guide and support.

I also want to express special thanks to the panel members in the Accounting, Governance, and Regulations Discipline, Professor John Robert and Associate Professor Demetris Christodoulou. I appreciate Professor John Robert for his support and guide that challenged me to think beyond my limits. Their support made my whole PhD journal pleasant and smooth.

I would like to express special thanks to my parents for their love, patience, and support. Their belief in me has been the power throughout this journey. Their understanding and unconditional encouragement are a source of motivations to keep going in this journey.

I would like to acknowledge the University of Sydney for its financial support, the University of Sydney International Scholarship. I also want to express special thanks to the Accounting, Governance, and Regulation discipline, and the University of Sydney Business School for offering opportunities and financial support for international conferences, which are invaluable for my career development.

The PhD journey has been an enjoyable, transformative, and unique experience. I am deeply grateful to everyone who has been part of it to support, guide, and encourage me.

Authorship Attribution Statement

I acknowledge that this thesis contains materials previously published in the following conferences:

Yu, M. (2023). What factors may influence the extent of voluntary SDG disclosures by Chinese companies? *2023 Accounting and Finance Association of Australia and New Zealand (AFAANZ), PhD Symposium (June-July)*.

Yu, M., & Frost, G. (2024). SDG disclosure: Automated data capture of Chinese companies' reporting. *2024 Accounting and Finance Association of Australia and New Zealand (AFAANZ) conference (30 June- 2July)*.

Yu, M. (2023). What factors may influence the extent of voluntary SDG disclosures by Chinese companies? *2024 A-CSEAR Early Scholar Colloquium (4 Dec)*.

Yu, M., & Frost, G. (2024). Poverty alleviation disclosure: Automated data capture of Chinese companies' reporting. *2024 A-CSEAR Conference (5 Dec-6 Dec)*.

The materials comprise Abstract, Chapter 1 Introduction, Chapter 2 Sustainable Development in China, Chapter 3 The variability of SDG disclosures, Chapter 4 An overview of content analysis, and Chapter 6 Data analysis. I drafted these conference papers and my supervisor, the co-author, edited them.

In addition to the authorship attribution statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Name: Muchen Yu

Signature:

Date:

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Name: Geoff Frost

Signature:

Date:

Abstract

The United Nations' call for companies to adopt SDGs presents a challenge to report on SDG performance to external stakeholders. Despite emerging discussion by standard setters and within the literature, there remains ambiguity as to the nature and extent of SDG disclosures; hence, a core challenge for future corporate SDG disclosures is identifying quantifiable and consistent measures for performance. This is particularly prevalent in China, where the government embraced the UN 2030 Agenda for Sustainable Development within the 13th National Five-Year Plan, through which they have encouraged companies to become more proactive in sustainable development. Therefore, this thesis focuses on the growing use of SDG disclosures in Chinese companies and the factors that may predict Chinese companies' SDG disclosure levels.

To address this challenge, this thesis, drawing upon existing reporting standards, academic literature and Chinese government publications, has developed a content analysis tool of 27 indicators to capture Chinese corporate SDG disclosures. A sample of 60 companies listed on the Shanghai and Shenzhen Stock Exchanges were examined during the period of the 13th National FYP (2016-2020) to provide analysis of Chinese corporate SDG disclosures, particularly on SDG 1, 7 and 13.

A review of existing literature suggests that a wide range of factors can affect the variation in the level of SDG disclosures; however, these studies rely on traditional statistical methods, which are limited to in-sample analysis and overlook the complex non-linear associations or interaction effects among predictors. To address these challenges, this thesis employs the gradient boosting machine to **1.** identify the predictive power of 159 predictors from government control, financial performance, corporate governance, and ESG performance aspects; **2.** offer a more reliable out-of-sample analysis. This thesis observed that the gradient boosting machine outperforms traditional methods to interpret the predictive power of government control, financial performance, corporate governance, and ESG performance on the level of SDG disclosures. However, the predictive power and direction of the prediction results widely vary due to variable selection and variable measurements.

Table of Content

Statement of Originality	i
Acknowledgement	ii
Authorship Attribution Statement.....	iii
Abstract.....	iv
Table of Content	v
List of Figures.....	viii
List of Tables	ix
A list of Abbreviations.....	xi
Chapter 1 Introduction.....	1
1.1 Chapter introduction	1
1.2 Research Problem	2
1.3 Motivation.....	2
1.4 Research setting	3
1.5 Overview of thesis	5
1.6 Chapter conclusion	7
Chapter 2 Sustainable Development in China	8
2.1 Chapter introduction	8
2.2 The UN Sustainable Development Goals	9
2.2.1 The purposes of the SDGs	9
2.2.2 An introduction and classification of SDGs at goal level.....	9
2.2.3 Challenges of SDGs.....	11
2.2.4 The United Nations encourages corporate engagement in SDGs.....	12
2.2.5 The UN encourages corporate engagement in SDG disclosures	13
2.3 SDG implementation in China.....	13
2.3.1 National regulatory frameworks on sustainable development in China.....	14
2.3.2 Corporate sustainable development regulations in China	15
2.3.3 SDGs as part of the national regulatory system.....	16
2.4 Current state of corporate engagement in SDGs	17
2.5 Why SDGs may matter to Chinese companies.....	18
2.5.1 SDGs and building a community of shared future of mankind.....	18
2.5.2 SDGs and building a ‘Moderately Prosperous Society’	19
2.5.3 SDGs and resolving environmental risks in the value chain	20
2.6 Why SDG disclosures may matter to Chinese companies.....	20
2.6.1 SDG disclosures as a corporate response to the Chinese government’s information needs	20
2.6.2 SDG disclosures as a means of meeting information needs of stakeholders.....	21
2.7 Guidelines for observing the nature and extent of corporate SDG disclosures.....	21

2.7.1 What constitutes an SDG disclosure?.....	21
2.7.2 Guidelines for observing corporate SDG disclosures.....	25
2.8 Chapter conclusion	37
Chapter 3 Literature review on the variability of SDG disclosures.....	38
3.1 Chapter introduction	38
3.2 The variability of sustainability disclosures in Chinese companies	39
3.3 Literature review of factors that may predict corporate SDG disclosures.....	39
3.3.1 Contextual factor—government control	39
3.3.2 Corporate characteristics on the level of SDG disclosures.....	53
3.4 Model selection.....	67
3.4.1 Literature review of model selection	67
3.4.2 Hypothesis development.....	69
3.5 Chapter conclusion	69
Chapter 4 An overview of automated content analysis	71
4.1 Chapter introduction	71
4.2 An overview of content analysis.....	72
4.3 Sample	74
4.4 Timeframe.....	77
4.5 Data Source.....	77
4.6 The process of the automated content analysis on capturing SDG disclosure text	78
4.6.1 Obtaining text	78
4.6.2 Text cleaning	116
4.7 Dependent variable—SDG disclosures	119
4.8 Chapter conclusion	120
Chapter 5 Research Design.....	122
5.1 Chapter introduction	122
5.2 Empirical framework—Gradient Boosting Machine.....	123
5.2.1 An introduction of the gradient boosting machine	123
5.2.2 The limitation of the model	124
5.2.3 Why gradient boosting machine is adopted in my research?	124
5.2.4 Model setup and validation.....	128
5.2.5 Result interpretation tools.....	129
5.3 Potential factors that may predict the level of SDG disclosures	131
5.3.1 Contextual factors.....	131
5.3.2 Corporate characteristics	144
5.3.3 Control variables.....	150
5.4 Chapter conclusion	153

Chapter 6 Data analysis	154
6.1 Chapter introduction	154
6.2 Descriptive analysis of the level of SDG disclosures	155
6.2.1 Box plots for visualising the level of SDG disclosures	155
6.2.2 Descriptive statistics of SDG disclosures	158
6.2.3 Kruskal-Wallis test for SDG disclosures across years	160
6.3 Empirical analysis of factors that may predict the level of SDG disclosures.....	161
6.3.1 Descriptive statistics	161
6.3.2 Empirical result interpretation	162
6.4 Discussion of the results	194
6.5 Chapter conclusion	199
Chapter 7 Conclusion	201
7.1 Chapter introduction	201
7.2 Contributions	202
7.3 Implications	204
7.4 Limitations	205
7.5 Future research.....	208
7.6 Chapter conclusion	210
Reference	211
Appendix.....	237
Appendix 1 A manual comparison of GRI, IR, and SASB to evaluate how these reporting frameworks can help measure SDG disclosures.....	237
Appendix 2 A list of sampled companies	239
Appendix 3 Keyword lists generated based on TF-IDF	241
Appendix 4 Irrelevant data collected for climate change spending or resource commitments due to less representative keywords	278
Appendix 5 Irrelevant data for poverty alleviation through ecological protection	279
Appendix 6 Less representative keywords or misclassified data for energy spending or resource commitment	280
Appendix 7 SDG disclosure data for Chinese companies	281
Appendix 8 Descriptive statistics	288
Appendix 9 Interaction effect	292
Appendix 10 Definition of predictors.....	295
Appendix 11 Post hoc tests.....	306
Appendix 12 Relative Variable Importance Scores.....	309

List of Figures

Figure 2-1 Extracted from Ordonez-Ponce et al. (2021, p.56)	24
Figure 2-2 Extracted from UNGC and GRI (2018, p. 23).....	27
Figure 5-1 K-fold cross-validation, extracted from Scikit-Learn (2019)	129
Figure 6-1 Total SDG disclosure scores by year	156
Figure 6-2 Poverty alleviation disclosure scores	157
Figure 6-3 Energy disclosure scores.....	157
Figure 6-4 Climate change disclosure scores	158
Figure 6-5 MSE for learn, test and learn sampled set for gradient boosting model.....	163
Figure 6-6 R-Sq for learn, test and learn sampled set for gradient boosting model.....	163
Figure 6-7 Partial dependency plot for the number of politically connected CEO & executives	170
Figure 6-8 Partial dependency plot for the number of politically connected directors	171
Figure 6-9 Partial dependency plot for party committee.....	172
Figure 6-10 Partial dependency plot for total debts.....	177
Figure 6-11 partial dependency plot for cash flow from financing	178
Figure 6-12 partial dependency plot for cash flow from operations	179
Figure 6-13 partial dependency plot for cash flow from investing activities	179
Figure 6-14 partial dependency plot for ROE	180
Figure 6-15 partial dependency plot for ROA.....	181
Figure 6-16 partial dependency plot for Tobin's Q.....	182
Figure 6-17 partial dependency plot for Enterprise value	182
Figure 6-18 partial dependency plots for institutional shareholding	185
Figure 6-19 partial dependency plots for shares owned by top 5 shareholders.....	186
Figure 6-20 partial dependency plot of ESG performance.....	187
Figure 6-21 partial dependency plot of governance score.....	188
Figure 6-22 partial dependency plot for ESG management scores	189
Figure 6-23 partial dependency plot for shares owned by analyst recommendation	192
Figure 6-24 partial dependency plot for firm size	192

List of Tables

Table 2.1 UN 17 SDG classifications, extracted from United Nations (2015, pp. 14-28)	11
Table 2.2 Extracted from Consolandi et al. (2020, p.518).....	23
Table 2.3 Extracted from KPMG (2018, pp. 4-5).....	33
Table 2.4 Extracted from Hummel (2019, p. 48).....	34
Table 2.5 Extracted from Subramaniam et al. (2020, p. 10).....	34
Table 2.6 Extracted from Zhang and Guan (2020)	35
Table 3.1 Extracted from the State Council of People's Republic of China (2015, pp. 1-10).....	44
Table 3.2 A summary of political connection measurements/	47
Table 3.3 Model selection and model robustness for testing the determinants of SDG disclosures	68
Table 4.1 A list of largest Chinese investors in the BRI in2023, extracted from https://greenfdc.org/china-belt-and-road-initiative-bri-investment-report-2023/	75
Table 4.2 Indicators for assessing explicit references to SDGs.....	82
Table 4.3 Indicators for assessing SDG 1 Poverty alleviation disclosures.....	87
Table 4.4 Indicators for assessing SDG 7 energy disclosures	88
Table 4.5 Indicators for assessing SDG 13 Climate change disclosures	89
Table 4.6 SDG disclosure framework for Chinese companies	90
Table 4.7 TF-IDF keyword results for poverty alleviation through education	95
Table 4.8 Examples of renewable vs non-renewable energy extracted from GRI 302 (2016, p.14).....	97
Table 4.9 irrelevant data for poverty alleviation through employment	98
Table 4.10 irrelevant data for renewable energy consumption	99
Table 4.11 Similarity test based on the simple proportion method.....	103
Table 4.12 An example of inconsistent results	105
Table 4.13 An example of poverty alleviation through ecological impacts.....	106
Table 4.14 Score-level consistency test	109
Table 4.15 Consistency test-- poverty alleviation.....	111
Table 4.16 Consistency test-- Energy	111
Table 4.17 Consistency test-- Climate change.....	112
Table 4.18 Consistency test-- Poverty alleviation	113
Table 4.19 Consistency test--Energy	113
Table 4.20 Consistency test-- climate change.....	114
Table 4.21 An example of the semantics of keywords	114
Table 4.22 An example of the surrounding context	115
Table 4.23 An example of the formatting issue	117
Table 4.24 An example of word frequency test based on Air China 2019 ESG report	118
Table 5.1 Political connection keyword.....	136
Table 5.2 Examples of irrelevant political connection data based on the keyword 'government'	136
Table 5.3 Similarity test for the first-round data collection at senior management level	138
Table 5.4 Similarity test for the first-round data collection at CEO & chairman level	138
Table 5.5 Similarity test for the second-round data collection at senior management level	139
Table 5.6 Similarity test for the second-round data collection at CEO & chairman level.....	139
Table 5.7 Irrelevant data based on the keyword 'government'	139
Table 5.8 Comparison between data collection methods.....	140
Table 5.9 results of Spearman's rank correlation and Kendall's Tau.....	141
Table 5.10 Keywords for party committees.....	142

Table 5.11 Consistency test results	144
Table 6.1 Descriptive statistics of the number of observations involving in each category over the sampled period.....	160
Table 6.2 Kruskal-Wallis results.....	161
Table 6.3 Model error summary	164
Table 6.4 Relative Variable Importance Scores for government control predictors	168
Table 6.5 Relative Variable Importance Scores for financial performance predictors	174
Table 6.6 Relative Variable Importance Scores for corporate governance predictors.....	184
Table 6.7 Relative Variable Importance Scores for ESG performance predictors.....	186
Table 6.8 Relative Variable Importance Scores for control predictors	190
Table 6.9 Top Variable interaction pairings	194
Table 8.1 How GRI, SASB, and IR can be used to operationalise SDGs	238
Table 8.2 Post hoc test for overall SDG disclosures.....	306
Table 8.3 Post hoc test for poverty alleviation disclosure	307
Table 8.4 Post hoc test for energy disclosures	307
Table 8.5 Post hoc test for climate change disclosures.....	308

A list of Abbreviations

13th National Five-Year Plan	13 th National FYP
2030 Agenda for Sustainable Development	Agenda 2030
Akaike information criterion	AIC
Bayes information criterion	BIC
Belt and Road Initiative	BRI
China Securities Regulation Commission	CSRC
Chinese People’s Political Consultative Conference	CPPCC
Communist Party of China	CPC
corporate social responsibility	CSR
corrected Akaike Information Criterion	AICc
earnings per share	EPS
Elaboration Likelihood Model	ELM
environmental, social and governance	ESG
Global Reporting Initiative	GRI
Innovation Demonstration Zone	IDZ
Integrated Reporting	IR
International Integrated Reporting Council	IIRC
Latent Dirichlet Allocation	LDA
Mean absolute deviation	MAD
Mean absolute percentage error	MAPE
mean squared error	MSE
Median Absolute Deviation	MAD
National People’s Congress	NPC
National Science, Technology, and Innovation plan	National STI Plan
Ordinary least-squares	OLS
People’s Bank of China	PBOC
photovoltaic	PV
property, plant and equipment	PPE

R squared	R-sq
Random Forests	RF
Relative variable important score	RVIs
return on asset	ROA
return on equity	ROE
Root Mean Squared Error	RMSE
Shanghai Stock Exchange	SSE
Shenzhen Stock Exchange	SZSE
Social and environmental accounting	SEA
Support vector machine	SVM
Sustainability Accounting Standard Board	SASB
Sustainable Development Goals	SDGs
Targeted Poverty Alleviation	TPA
Toxic Release Inventory	TRI
United Nations	UN

Chapter 1 Introduction

1.1 Chapter introduction

This chapter provides an overview of the thesis. It starts with an introduction of the research problems in Section 1.2. This is followed by a discussion of the motivations of this thesis in Section 1.3 and an outline of the research setting in Section 1.4. Section 1.5 introduces the structure of this thesis, and Section 1.6 concludes this chapter.

1.2 Research Problem

The Chinese government introduced its 13th National Five-Year Plan (13th National FYP) in 2016, which emphasises ‘innovative, coordinated, green, open, and shared’ development (Communist Party of China, 2016b, p. 3). Following its publication by the United Nations (UN) in 2015, the Chinese government signed on to the 2030 Agenda for Sustainable Development (Agenda 2030) in 2016 (Communist Party of China, 2016a; UNDP, 2016). The government, subsequently, has ‘actively implemented’ Agenda 2030 and utilised it to operationalise the plan of ‘All-around Opening up’ within the 13th National FYP (Communist Party of China, 2016a, p. 136; National Development and Reform Commission, 2016, p. 153). According to ‘China’s Progress Report on Implementation of Agenda 2030’ (Communist Party of China, 2017), the Chinese government is working on assessment mechanisms and appropriate indicators for evaluating Sustainable Development Goal (SDG) performance in China. While there are no existing requirements for SDG disclosures, the number of Chinese companies that provide SDG disclosures in their corporate social responsibility (CSR) reports has increased continuously, rising from 174 companies in 2018 to 254 companies in 2020 (Zhang & Guan, 2020). PwC (2020) finds China among the top 5 countries with corporate SDG disclosures. The focus of this thesis is on the growing use of SDG disclosures by Chinese companies as part of their reports. The broad research questions that are addressed in this thesis are as follows:

1. What is the extent of SDG disclosures in Chinese companies, following China’s adoption of Agenda 2030?
2. What factors may predict the level of SDG disclosures in Chinese companies based on the gradient boosting machine?

1.3 Motivation

The United Nations introduced the Agenda 2030 in 2015 as a holistic framework designed for global actions towards sustainable development (United Nations, 2015, p. 4). However, as an inter-governmental initiative, SDGs are designed to promote sustainable development rather than as a corporate reporting framework. Therefore, SDG disclosures are poorly understood (Bebbington & Unerman, 2018; KPMG, 2018a). Existing frameworks are only partially aligned with the SDGs. For example, Global Reporting Initiative (GRI) indicators provide little emphasis on innovation-related actions, which are prioritised in the SDGs. Integrated reporting (IR) only provides guiding principles without standardised indicators, while the Sustainability Accounting Standard Board (SASB) only focuses on a small number of accounting issues. Therefore, there remains a gap in our understanding of what elements we can use to observe SDG disclosures in reports.

The institutional and regulatory settings in China offer an opportunity to explore SDG disclosures in Chinese companies. With the introduction of Agenda 2030, the SDGs have become part of China’s

national regulatory system. Chinese regulatory authorities have introduced requirements and guidelines to guide corporate engagement and disclosure practices towards sustainable development. However, there are no reporting frameworks to guide corporate SDG disclosures in China. Therefore, there remains a gap in understanding corporate SDG disclosures that take into account the unique circumstances of the Chinese context.

This thesis also identifies several factors that can potentially predict the level of SDG disclosures in Chinese companies. Social and environmental accounting (SEA) research calls for an emphasis on contextual differences, particularly in developing countries, in motivating their SEA disclosures (Qian et al., 2021, p. 3). This could include their social-political environment and institutional focus (Qian et al., 2021). Additionally, prior research does not provide conclusive evidence regarding the associations between SDG disclosures and corporate characteristics due to the variables selected, various contexts, the value relevance of indicators, and the measurement of SDG disclosures. This thesis attempts to accommodate a greater number of variables and identify useful ones in predicting the level of SDG disclosures.

1.4 Research setting

SDG goals selected. The UN has introduced 17 goals to mobilise countries' efforts towards ending poverty and hunger, fighting inequalities and promoting decent work, improving economic growth and mitigating environmental concerns (United Nations, 2015); however, not all SDG goals are equally relevant and important to China and Chinese companies. According to Permanent Mission of the People's Republic of China to the UN (2022), China has made significant progress towards SDG 1 No Poverty, 7 Clean Energy, 13 Climate Action, 14 Life Below Water and 15 Life on Land when the 13th National FYP was effective. However, SDG 14 and 15 focus on biodiversity, which are the least prioritised goals by Chinese companies. Therefore, this thesis focuses on SDG 1, 7, and 13, which are prioritised by the Chinese government and Chinese companies.

SDG reporting framework. Given that there are no existing SDG datasets, this thesis develops an SDG framework for data capture, particularly taking into account the unique circumstances of the Chinese context. Existing studies are limited to international guidelines, such as GRI and IR, or boilerplate disclosure frameworks without evaluating specific SDG actions that a company has taken. They often do not consider regulatory enforcement and institutional background when measuring SDG disclosures. To fill the gap, this thesis develops a framework of 27 indicators to capture SDG disclosures, from corporate explicit references to SDGs, such as mentioning SDG activities in reports, to implicit SDG performance related to SDG 1, 7 and 13. Indicators are drawn from the SDG framework, the GRI guideline, the Chinese government policy documents and prior studies.

Sample. To analyse the impact of the 13th National FYP, this thesis sampled the top 60 Chinese companies listed on the Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchanges (SZSE) from

2016 to 2020. These companies were selected because large companies are generally more motivated to make SDG disclosures based on SDG 12.6 and the Chinese government policy document. Larger Chinese companies are the leaders in the innovation of SDG disclosures in China and the realisation of SDGs among the BRI international cooperation.

Data source. To focus analysis on SDG disclosures, data is collected from CSR, sustainable development, and other non-financial reports. These reports are focused reports of corporate SDG disclosures and provide standardised data annually. This also offers an opportunity to conduct a cross-sectional analysis.

Automated content analysis. Given that there is no pre-existing dataset to provide this data, this thesis developed an automated tool for SDG disclosure data capture. Prior studies widely adopt manual content analysis, existing databases or machine learning techniques to capture environmental, social and governance (ESG) disclosure data (Huang et al., 2023; Ordonez-Ponce & Khare, 2021). However, manual content analysis is often criticised for human coders' subjective understanding and data collection efficiency (Kasperuniene et al., 2020). Existing databases provide limited transparency and comparability across data vendors (Lin et al., 2024), while machine learning techniques have different focus, such as sentiment analysis or topic modelling (Huang et al., 2023). To overcome these challenges, this thesis will develop a dictionary-based automated content analysis tool to capture SDG disclosure data. Given that the automated tool is unable to process the data as humans, this thesis develops a dictionary to allow the tool to classify text based on pre-defined categories and keyword lists. To train the automated tool, the whole process involves several steps: obtaining text, keyword development, refining and validation, and text cleaning.

Gradient boosting machine. Given that this thesis includes a sample of 60 Chinese Public companies and 159 predictors (the number of predictors is greater than the sample size), this research applies the gradient boosting machine to identify factors that may predict the level of SDG disclosures. Existing studies typically employ traditional statistical methods, such as OLS, logit models or structural equation modelling, which heavily rely on significance testing (García-Meca & Martínez-Ferrero, 2021; Orazalin & Mahmood, 2020). Given that prior studies are limited to selecting a small number of variables based on theoretical framing or other methodological constraints (Frost et al., 2023), this thesis employs the gradient boosting machine, a machine learning technique. It is an ensemble approach to resolving regression and classification problems (Friedman, 2001). This method is selected as it can identify the predictive power of a wide range of variables and offer a reliable out-of-sample analysis. Specifically, it matches my research setting, including a small sample size, and a greater number of variables, including mixed-type variables. It can also generate more accurate prediction outcomes compared than other machine learning methods. It also takes into account the complex non-linear relationships and interaction effects across different variables.

1.5 Overview of thesis

The remainder of the thesis is set out as follows:

Chapter 2 provides an overview of the nature of SDG disclosures and the variation in SDG disclosures in Chinese companies. This chapter starts with an overview of Agenda 2030 and SDGs. The UN has called on companies and governments to participate in SDGs, which are considered an international unified framework for addressing the most pressing sustainable development challenges. To see how the Chinese government responds to this call, this thesis discusses how SDGs become part of the regulatory system in China. Given that the 13th National FYP highlighted the important role of Chinese companies in achieving national goals, this thesis also discusses why SDG and SDG disclosures matter to Chinese companies and the current trend of SDG disclosures in Chinese companies. However, as discussed above, SDGs are an aspirational performance framework, and there remains a gap in understanding SDG disclosures. Therefore, this thesis also reviews prior studies to observe the nature and extent of SDG disclosures, serving as the basis for developing an SDG disclosure framework that takes into account the unique circumstances of the Chinese context in Chapter 4.

Chapter 3 provides an overview of the variation of SDG disclosures in Chinese companies, following China's adoption of Agenda 2030. It establishes the relevance of issues to corporate SDG practices, which may finally affect corporate responses, through SDG disclosures. Due to the unique authoritative setting, this chapter explores whether government control may predict the level of SDG disclosures. This thesis also reviews the literature regarding whether corporate characteristics, such as financial performance, corporate governance, and ESG performance, may predict the level of SDG disclosures. This chapter also reviews the existing literature on model selection, focusing on their model fit in different contexts and research designs. This will provide the foundation for the empirical model employed in Chapter 5.

Chapter 4 develops a framework to observe the nature and extent of SDG disclosures. 27 indicators were created, which were identified from different sources. To collect data, an automated content analysis tool is developed, and the process is introduced, such as obtaining text, identifying unit of analysis, text location, keyword development, refining and validation, and text cleaning. This is followed by a discussion of the limitations of this method. This thesis also describes the selected SDG goals, sample size, time period, and data source. The analysis aims to observe what information is disclosed by Chinese companies. The company will receive a score of one for a particular indicator, if it discloses relevant information. This thesis will finally use SDG disclosure data as the dependent variable in the predictive analysis in Chapter 6.

Chapter 5 focuses on the research design for RQ2. Specifically, this chapter describes the model used to predict the level of SDG disclosures. As discussed above, this thesis will employ the gradient boosting machine as the primary statistical learning technique for the analysis. This chapter also describes

variable selection for factors that may predict the level of SDG disclosures, from government control, financial performance, corporate governance, and ESG performance aspects. Specifically, it will include proxy selection, data source, and data collection process of variables.

Chapter 6 interprets the findings of RQ1 and RQ2. This thesis adopts box plots to visualize the variation in SDG disclosures between 2016 and 2020. This thesis also visualises the variation in SDG disclosures across different categories. Regarding RQ2, this thesis employs relative variable importance scores (RVIs) to visualize the predictive power of variables on the level of SDG disclosures. Partial dependence plots are also used to visualize the direction of their contributions on the level of SDG disclosures. This thesis also evaluates the interaction effects of variables used in the analysis.

This thesis suggests that financial performance, corporate governance, ESG performance, government control, and other variables have predictive power on the level of SDG disclosures by Chinese companies. However, the predictive power of variables and the direction of their contributions on the variation of the level of SDG disclosures widely vary due to variable selection, different research contexts, and the heterogeneity of disclosure practices.

Chapter 7 provides the conclusion of the thesis from contributions, implications, limitations of this research, and future research.

Five major contributions have been identified:

1. This thesis develops a novel SDG disclosure framework to explore SDG disclosures in Chinese companies, following China's adoption of Agenda 2030.
2. This thesis develops an automated content analysis tool to capture SDG data to overcome the challenges of inter-coder reliability, which can be used in other accounting research.
3. This thesis offers insights into the nature and extent of SDG disclosures in Chinese companies. As discussed above, this thesis draws upon several sources, to develop the SDG disclosure framework. This thesis observes the variation of SDG disclosures from corporate references to SDGs and SDG performance related to SDG 1, 7 and 13 aspects.
4. This thesis contributes to a multifaceted framework for better understanding government control in China. Specifically, this thesis goes beyond government shareholding and includes regulatory, political and incentivising influence. It develops several novel methods to measure different types of government control tools.
5. This thesis also employs a modern machine learning method to explore factors that may predict the level of SDG disclosures. This thesis extends existing studies by accommodating 159 predictors in one single model, to understand factors that may predict the level of SDG disclosures in Chinese companies.

Implications of this thesis can be classified into three areas: implications for regulators, researchers and potential investors. Limitations of this thesis are classified into machine learning and methodological aspects. This thesis also offers several suggestions for future research:

1. Use discourse analysis to explore what the company claims they have done in ESG reports and whether SDG disclosures in Chinese companies are symbolic, greenwashing, or substantive.
2. Use statistical inference analyses to develop an explanatory model for identifying factors that may influence the level of SDG disclosures.
3. Compare the performance of the gradient boosting machine with that of other machine learning techniques, such as Random Forests and deep learning.
4. Focus on developing a theoretical framework for explaining the variation in SDG disclosures.
5. Explore alternative measures for predictors. This thesis developed novel methods for different government controls.
6. The generalisability of the reporting framework and prediction results to other contexts.

1.6 Chapter conclusion

This chapter provides an overview of the research problem, motivations of this thesis, the research setting, contributions and an overview of this thesis. Specifically, Section 1.2 introduces the research problem, while Section 1.3 describes potential research motivations. Since SDGs were introduced, regulatory bodies and governments have been working on assessment mechanisms to evaluate their SDG progress and disclosures. However, SDGs are an inter-governmental initiative, rather than a corporate reporting framework. Therefore, there remains a gap in observing the nature of SDG disclosures, especially in the Chinese context.

Section 1.4 outlines the research setting of this thesis. Specifically, this thesis samples 60 Chinese Public Companies between 2016 and 2020. This thesis develops a reporting framework and an automated tool to capture SDG disclosure data. This thesis also applies the gradient boosting machine in this thesis. Section 1.5 offers a brief discussion of the remaining chapters.

Chapter 2 Sustainable Development in China

2.1 Chapter introduction

Chapter 2 provides an overview of SDG disclosures in China. Specifically, Section 2.2 starts with an overview of the Agenda 2030. Section 2.3 discusses sustainable development in China and how SDGs become part of the regulatory system in China. This is followed by introducing the current state of corporate engagement in SDG and SDG disclosures in China in Section 2.4. Section 2.5 focuses on why SDGs matter to Chinese companies, while Section 2.6 discusses why SDG disclosures matter to Chinese companies. Section 2.7 reviews the literature regarding the nature and extent of SDG disclosures. Section 2.8 concludes this chapter.

2.2 The UN Sustainable Development Goals

In 2015, the United Nations introduced the Agenda 2030, consisting of ‘actions for people, planet, and prosperity’ (United Nations, 2015, p. 4). As part of the Agenda 2030, the UN launched SDGs, containing 17 global goals (see Table 2.1). The goals ‘independently consider economic, environmental, and social concerns, and are broadly interdependent with other goals’ (United Nations, 2015, p. 1). Over the next fifteen years, these goals aim to ‘universally mobilise countries’ efforts to end all forms of poverty, reduce hunger, fight inequalities and tackle climate change, create economic growth, map decent work, and address other social and environmental needs’, to ensure that ‘no individual is left behind’ (United Nations, 2015, p. 1).

2.2.1 The purposes of the SDGs

SDGs are designed to provide a ‘more sustainable future for all’ (United Nations, 2015, p. 1). While nations worldwide face massive ESG challenges, the SDGs offer an ambitious framework for people to prioritise ESG strategies and implement them (United Nations, 2015).

SDGs were also developed to carry out the Millennium Development Goals¹(MDGs), which ended in 2015. Unlike the MDGs, which highlight poverty reduction and hunger eradication, SDGs emphasise environmental sustainability, economic growth, and other non-material aspects. In contrast to the MDGs primarily promoted by the UN, the achievement of the SDGs relies on collaborations by all actors: the UN, all 193 countries, companies, and stakeholders. Specifically, the UN has convened all 193 member states to sign this agreement, which acts as the Post-2015 Development Agenda². Therefore, Agenda 2030 is commonly considered an inter-governmental initiative to provide a universal guide for countries to implement SDGs; it also relies on a bottom-up approach to involve various stakeholders including companies who are the UN calls upon to ‘apply innovation to pursue sustainable development’ (United Nations, 2015, p. 4).

2.2.2 An introduction and classification of SDGs at goal level

SDGs include 169 targets (outcome targets and means of achieving targets) and 232 indicators. To simplify SDG implementation and assessments, each target has up to 3 indicators that can be used to

¹ The Millennium Development Goals (MDGs) were established following the UN Millennium Summit in 2000. All 191 countries and over 20 international companies are committed to helping achieve the MDGs by 2015, related to poverty alleviation, education, gender equality, child health, maternal health, diseases, environmental protection, and global partnership (Faye et al., 2012). These goals primarily aimed to meet social concerns and human development in developing countries (low-income/ the poorest countries). MDGs were measured by 21 targets and 60 indicators. However, they were often criticised for overlooking environmental sustainability (Faye et al., 2012). UNDP (2015) also pointed out that MDG targets are not proper to measure actual actions taken by ‘off-track’ countries. Therefore, MDGs offered limited insights and guidance for low-income countries to pursue sustainable development.

² The Post-2015 Development Agenda was a process from 2012 to 2015 led by the United Nations to define the future global development framework that would succeed the Millennium Development Goals. The new framework, starting from 2016 is called Sustainable Development Goals.

measure progress toward reaching the target. This thesis combines the traditional three-pillar view and Huber et al. (2018) to provide an overarching overview of 17 SDGs (See Table 2.1).

Social objectives consider human wellbeing and social issues. Goals include no poverty (SDG 1), zero Hunger (SDG2), good health & well-being (SDG3), quality education (SDG4), gender equality (SDG5) and reduce inequality (SDG10). These goals focus on people's quality of life and on social inequality.

The environmental dimension includes both resource accessibility and environmental sustainability. Goals include clean water and sanitation (SDG6), affordable and clean energy (SDG7), climate action (SDG13), life below water (SDG14), and life on land (SDG 15). This dimension overlaps with the 'Planet' dimension in Huber et al. (2018). SDG6 and SDG7 consider water and clean energy as resources to support daily life – accessibility, while SDG13-15 illustrate environmental issues with regards to climate, ocean, and land.

Economic objectives are linked to private sectors to promote sustainable economic growth, innovation, consumption, and production. Goals include decent work and economic growth (SDG8), industry, innovation & infrastructure (SDG 9), sustainable cities and communities (SDG11), and responsible consumption & production (SDG12).

Governance and other dimensions include SDG16 Peace and Justice Strong Institutions and SDG17 Partnerships to Achieve the Goal.

Table 2.1 UN 17 SDG classifications, extracted from United Nations (2015, pp. 14-28)

Social dimension	SDG1 No Poverty	‘End poverty in all its forms everywhere’
	SDG2 Zero Hunger	‘End hunger, achieve food security and improved nutrition and promote sustainable agriculture’
	SDG3 Good Health and Well-being	‘Ensure healthy lives and promote well-being for all at all ages’
	SDG4 Quality Education	‘Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all’
	SDG5 Gender Equality	‘Achieve gender equality and empower all women and girls’
	SDG10 Reduced inequality	‘Reduce inequality within and among countries’
Environmental dimension	SDG6 Clean Water and Sanitation	‘Ensure availability and sustainable management of water and sanitation for all’
	SDG7 Affordable and Clean Energy	‘Ensure access to affordable, reliable, sustainable, and modern energy for all’
	SDG13 Climate Action	‘Take urgent action to combat climate change and its impact’
	SDG14 Life Below Water	‘Conserve and sustainably use the oceans, seas, and marine resources for sustainable development’
	SDG15 Life on Land	‘Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss.’
	SDG12 Responsible Consumption and Production	‘Ensure sustainable consumption and production patterns’
Economic dimension	SDG8 Decent work and Economic Growth	‘Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all’
	SDG9 Industry, Innovation and Infrastructure	‘Build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation’
	SDG11 Sustainable Cities and Communities	‘Make cities and human settlements inclusive, safe, resilient, and sustainable’
Governance and other dimensions	SDG16 Peace and Justice Strong Institutions	‘Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels’
	SDG17 Partnerships to Achieve the Goal	‘Strengthen the means of implementation and revitalise the global partnership for sustainable development’

2.2.3 Challenges of SDGs

While 232 indicators have been created to monitor SDG progress, however, there are considerable challenges in developing practical indicators that can be applied to various levels (i.e. company level and governmental level).

Not all the indicators are measurable. All 232 indicators could be classified into three tiers regarding their methodology and data availability (Dang & Serajuddin, 2020). According to United Nations

(2018), 101 indicators could be classified into Tier 1 with sufficient data and well-established measures, and 91 indicators belong to Tier 2 having an established measure but a limited amount of data available, while there are 34 Tier 3 indicators, which have neither proper methods nor data. Only 6 indicators are covered in multiple tiers. According to the latest survey by United Nations (2021), the number of indicators that can be classified into Tier 1 increased to 130, while 97 indicators belong to Tier two and 4 indicators are in multiple tiers. Overall, a limited number of SDG indicators can practically monitor SDG progress.

The universal set of SDG indicators does not account for cross-country differences, such as culture, institutional backgrounds, and political and economic structures (Huan et al., 2019). SDGs are a general and aspirational guideline, prioritising the most complex sustainable development issues without taking into account the unique circumstances in different countries.

2.2.4 The United Nations encourages corporate engagement in SDGs

Despite these challenges in developing appropriate and actionable indicators for SDGs, the UN has encouraged companies to participate in SDGs for the following reasons:

SDGs may serve as a value-creating tool for companies to ‘pursue inclusive and sustainable economic growth’ (United Nations, 2015, p. 7). The United Nations (2015, p. 1) has called on companies to ‘apply their innovation and creativity’ to pursue sustainable development. SDGs provide innovative solutions for companies to identify opportunities for their long-term survival, by innovating their technologies, promoting productive employment, and pursuing sustainable consumption. SDGs may also serve as the basis for improving corporate sustainability values (SDG Compass, 2015). For example, SDG 7, SDG 12, and SDG 13 offer various approaches to promote renewable energy, encourage effective use of resource, and mitigate climate change issues. This may subsequently improve corporate economic development.

SDGs can serve as risk management tools in corporate value chains (Deloitte, 2019 ; SDG Compass, 2015). The SDG Compass (2015) suggests that SDGs can be mapped onto the entire value chain, from logistics to production and consumption, the lifecycle of products, as well as distributions. Risks can be reduced by either improving a company’s competence and products or reducing negative ecological impacts in the value chain. For example, the SDG Compass (2015) suggests that a company’s long-term resilience may be improved by aligning with SDG 8 and SDG 13. Negative impacts across the value chain can be mitigated through actions aligned with SDG 6, SDG 11, and SDG 12 (SDG Compass, 2015, p. 12). Specifically, empirical studies have suggested that SDGs may serve as the basis for corporate resource planning management systems to improve corporate environmental and financial performance of Ukrainian companies (Lyulyov et al., 2023). Zhilkina et al. (2022) also argue that SDGs could be socially strategic approaches for corporate financial risk management during the COVID-19 pandemic.

SDGs may provide a common language for companies to pursue global sustainable development (KPMG, 2017; SDG Compass, 2015). SDGs can be used as a template for shifting companies' attention from micro- to macro-level business sustainable development and may allow companies to identify some overlooked issues in existing sustainability standards (SDG Compass, 2015). For instance, SDGs include some specific targets for innovation-driven sustainable development which are not included in current sustainability frameworks like the GRI or IIRC (KPMG, 2017).

2.2.5 The UN encourages corporate engagement in SDG disclosures

The UN has encouraged companies to develop new disclosures to reflect their SDG progress (UNGC & GRI, 2018). Disclosure may indicate a company's capability to pursue sustainable business practices to achieve long-term success (PwC, 2016; UNDP (China), 2017). As the SDGs may allow companies to obtain more sustainable alternatives to facilitate their sustainable business growth, disclosure may indicate whether companies have sufficient funds to implement more sustainable alternatives (PwC, 2016; UNDP (China), 2017). Finally, this may lower the costs of capital for companies.

SDG disclosures may also indicate the effectiveness of corporate management in dealing with risks in their value chain (GRI & United Nations Global Compact, 2018). As SDGs may guide companies to reduce environmental risks in the value chain, SDG disclosures may reveal how corporate management aligns their sustainable decisions with SDGs and how they address risks in their value chain (GRI & United Nations Global Compact, 2018). This may, in turn, affect investors and other stakeholders' evaluation on a company's opportunities and risks, which affect their investment decisions and capital allocation (CDSB & WBCSD, 2017). Overall, SDG disclosures may indicate the efficiency of corporate management and serve as the basis for a dialogue with the broader community.

SDG disclosures may also provide a language for the internal evaluation and understanding companies' contributions to macro-level SDGs (GRI et al., 2018; GRI & United Nations Global Compact, 2018). As SDGs can provide a template for companies in implementing sustainable development, disclosing SDG information may enable companies to understand how they have engaged with specific SDGs and understand the opportunities and risks in implementing them (GRI & United Nations Global Compact, 2018). This may guide companies in optimising corporate practices in sustainable development, with disclosures serving as an internal evaluation tool.

Overall, the UN and other proponents of the SDGs suggest that they can serve as an action-oriented framework to guide companies and countries to participate in sustainable development activities.

2.3 SDG implementation in China

China has 'experienced rapid economic growth' and become the second-largest economy worldwide; however, this has also left China with severe environmental consequences, such as carbon pollution and energy overconsumption (Qian et al., 2021, p. 15; Situ & Tilt, 2018). As a response to its sustainability

crisis, the Chinese government has increasingly ‘recognised the importance of and made a transition to a green economy as its long-term development strategy’ (Yi & Liu, 2015, p. 11). The government has actively incorporated sustainable development into national strategies over the last decade. For example, following its publication by the UN in 2015, the Chinese government signed on to Agenda 2030 in 2016 (Communist Party of China, 2016a; UNDP, 2016).

This section starts with an overview of sustainable development regulations in China, followed by a discussion of how the Chinese government responded to the Agenda 2030 and SDGs.

2.3.1 National regulatory frameworks on sustainable development in China

Given the hierarchical and complex regulatory system in China, the central government not only acts as a regulator but also plays a role in Chinese corporate systems (Lin, 2011). Since 2013, it has launched a series of initiatives, programs and guidelines, including the Belt and Road Initiative, 13th National FYP, and the green bond market, to support the country’s economic growth without compromising its sustainable development (National Development and Reform Commission et al., 2021; Zhu et al., 2022). These regulatory enforcements also highlight a more significant role of Chinese companies in achieving national goals and political ideologies (Bin-Feng et al., 2023). While a series of political ideologies and initiatives have been introduced to ‘politicalize’ sustainable development in China (Qian et al., 2024), they are broad and aspirational, rather than providing actionable guidelines for corporate engagement.

2.3.1.1 13th National Five-Year Plan

In 2015, the Communist Party of China (CPC) introduced the 13th FYP, which provided pathways for China’s development from 2016 to 2020, by prioritising economic growth at a rate of under 6.5%–7% of GDP per year, resolving environmental issues, and improving efficiency. The core slogan of the plan is to build a ‘Moderately Prosperous Society’, while the principles of the 13th FYP include ‘innovation, coordination, green, open, and shared development’ (Communist Party of China, 2016b, p. 3). It informed China’s priority in promoting green transition and sets the direction of regulations to operationalise sustainable development from 2016 to 2020. It also called on companies to play a role in sustainable development (Yi et al., 2020). For example, in 2016, the 13th Five-Year Plan for Poverty Alleviation was issued, which emphasised the important role of state-owned and private companies in poverty alleviation (Bin-Feng et al., 2023).

2.3.1.2 Belt and Road Initiative

The Chinese government launched the Belt and Road Initiative (BRI) to support its ‘Going Global’ strategy in 2013, which specifically aimed at international trade, economic growth, infrastructure development, and regional integration (Xinhua, 2015). The Chinese government intends to create an international trade route for accelerating China’s economic globalisation. However, negative voices and concerns surrounding the BRI as an unsustainable initiative were rife (Teo et al., 2019). The 13th National FYP also outlines the ‘All Around Opening Up’ strategy in 2015. As a response, in 2017,

‘Guidance on Promoting Green Belt and Road Initiative’ was released by Chinese ministries, aiming to ‘incorporate the principles of being resource efficient and environment friendly into the whole process of international cooperation’ (Ministry of Ecology and Environment the People’s Republic of China, 2017). By 2021, China had formed contractual partnerships and invested in 144 countries (Zhu et al., 2022). It has also directly invested over US\$59.5 billion overseas in exchange for international trade and openness to its global partnership (Zhu et al., 2022). Overall, this initiative serves as a basis for achieving economic globalisation along a sustainable path.

2.3.1.3 Green Bond Market

In 2015, the government also launched its green bond market to fully implement the ‘Guidelines for Establishing the Green Financial System’. Since the People’s Bank of China (PBOC) issued the first green bond in December 2015, the Chinese green bond market has developed rapidly and has become the largest worldwide. In contrast to the West, the Chinese green bond market has been directed by the central government, with multiple official authorities setting rules for the market. The aims of the green bond market include promoting ecological civilisation in Chinese companies, providing sustainable economic development and reaching the 30/60 goal (carbon peak in 2030 and carbon neutralisation in 2060). Specifically, the market intends to provide capital to green projects related to energy saving, environment pollution, resource-saving, recycling and reuse, ecological protection, and climate change, which are consistent with the goals outlined in the 13th National FYP (National Development and Reform Commission et al., 2021).

Overall, an increasing number of national strategies and initiatives have been introduced to call on corporate engagement in sustainable development, highlighting the important roles of companies in achieving these goals. However, they are broad and aspirational rather than corporate guidelines.

2.3.2 Corporate sustainable development regulations in China

Since the 1990s, the Chinese government has complied with some international reporting standards to evaluate corporate sustainable development. Some widely adopted international sustainability reporting standards include the International Standard Organisation’s ISO 26000, GRI, SASB, and IR. However, these Western reporting frameworks appear to be partially aligned with the Chinese unique circumstances (Liu & Anbumozhi, 2009). For instance, the contents of sustainability reports generated by Chinese companies may have substantive differences from the required reporting frameworks (Ervtits, 2021). One of the reasons for explaining this is the “dynamic ‘crossvergence’” (Ervtits, 2021, p. 4). That is, the Chinese government contributes to sustainable development while these international reporting guidelines target corporate sustainability. Another reason could be the institutional complexity (Luo et al., 2017). Given that Chinese companies face pressure from governments and stakeholders, this may increase the diversity of their reports rather than standardisation. Additionally, the differences may also

attribute to the requirements: coercive, mimetic or normative (Ervits, 2021). Therefore, international reporting regulations may be partially aligned with corporate sustainable development in China.

The Chinese government also introduced localised mandatory ESG reporting requirements for Chinese companies. In 2008, the State-owned Assets Supervision and Administration Commission of the State Council issued ‘Notification on the Issuance of the Guideline on Fulfilling Social Responsibility by Central Enterprises’, which was the first mandatory reporting requirement in China. It requires state-owned companies to make CSR disclosures. In 2015, the Chinese government introduced the Environmental Protection Law that mandated and guided Chinese companies to disclose their environmental performance. China Securities Regulatory Commission also regulated the contents and structure of annual reports issued by Chinese companies. However, these regulations are limited to environmental disclosures or guidelines for state-owned enterprises.

Chinese stock exchanges, including SSE and SZSE, and the CSR Research Centre have also introduced some CSR reporting regulations and guidelines. In 2006, the SZSE began encouraging companies to voluntarily report their sustainability information, while the SSE has required companies to report their sustainability information since 2008 (Zheng et al., 2014). In the same year, SSE also issued Notice for better preparing the 2008 Annual Reports of Listed Companies which requires companies to disclose ESG information in their annual reports. In addition, the CSR Research Centre of the Chinese Academy of Social Sciences also introduced China CSR Report Preparation Guide in 2009 and this guideline was updated several times to capture the latest CSR concerns (Wu & Håbek, 2021). However, these regulations target a limited number of Chinese companies and focus on corporate sustainability rather than national prioritised sustainable development. How companies make sustainable development disclosures is still poorly understood.

Overall, existing guidelines and policies offer limited insights into the current national strategies and initiatives. Given that there are no unified frameworks for guiding and observing corporate sustainable development practices, how companies respond to these aspirational and complex national initiatives and strategies remains unclear.

2.3.3 SDGs as part of the national regulatory system

Following its publication by the UN in 2015, the Chinese government signed on to Agenda 2030 in 2016 (Communist Party of China, 2016a; UNDP, 2016), and the SDGs helped to inform the development and operationalisation of the 13th National FYP (Communist Party of China, 2016b, p. 3). According to the UNDP (2016, p. 6), SDGs are ‘closely aligned with the 13th FYP’ in promoting ‘sustainable economic growth in the long term, driving innovation and coordinated development, improving environmental quality, and resolving social challenges’. Furthermore, Communist Party of China (2016a, p. 153) also stated that ‘we will actively implement the 2030 Agenda for Sustainable

Development' in fulfilling China's international responsibilities. SDGs can be considered a means to achieve China's international cooperation.

The Chinese government is also working on the development of assessment mechanisms and appropriate indicators for evaluating SDG performance in the country. Internationally, China has participated in the UN's SDG monitoring program to assess its overall SDG progress based on the 17 goals and 169 targets, and selected Deqing county in Zhejiang Province as the pilot case (UN Department of Economic and Social Affairs Sustainable Development, 2017). However, this program is a pilot study to monitor SDG progress, which may not be generalisable to other cities. In addition, Ministry of Foreign Affairs of The People's Republic of China (2017) is also working on assessment criteria at the company level; for example, the 2017 implementation report states that the Chinese government will work on water-related evaluation criteria for water-saving companies based on SDG 6. Overall, the Chinese government is seeking to formulate evaluation criteria based on SDGs.

The Chinese government has established several initiatives to support the implementation of the Agenda 2030. As SDGs encourage countries to shift their attention to innovation-driven development, the Chinese government has established the plan of building Innovation Demonstration Zones (IDZ) on December 3, 2016, seeking to build approximately 10 national zones to pursue sustainable economic growth (Xinhua, 2018). Shenzhen, Taiyuan and Guilin have been approved to be the pilot cities for this program. In addition, the Chinese government also formulated National Science, Technology and Innovation (STI) plan, introduced China Health 2030 initiative, and built the Green Technology Bank. All these initiatives seek to support Agenda 2030 in driving innovation and development in China.

However, SDGs, an inter-governmental initiative, are a framework of goals, targets and indicators, designed to promote sustainable development rather than serving as a corporate reporting framework. Therefore, how companies make SDG disclosures is very varied and poorly understood (KPMG, 2018a).

2.4 Current state of corporate engagement in SDGs

Although no standardised SDG disclosure framework has been created for measuring Chinese companies' SDG performance, since 2016 an increasing number of Chinese companies have begun to voluntarily disclose SDG information in their reports. The number of companies that disclosed their SDG information rose from 174 in 2018, to 254 in 2020 (Zhang & Guan, 2020) (See Section 3.2).

PwC (2020) surveyed 89 largest Chinese companies regarding their SDG engagement, finding that Chinese companies have some understanding of SDGs. Of the sampled 89 companies, 89% of these companies have some knowledge about SDGs, and 69% of these companies mention SDGs in their reports, while around 42% of these companies are unclear about how they can evaluate their SDG performance and 38% of these companies work on it without reporting any information. The next steps for companies will target incorporating SDGs into their sustainability-related projects and sharing their

SDG achievements. Overall, Chinese companies are aware of SDGs, have some understanding of SDGs, and are willing to engage in these goals. However, what remains unclear in relation to the growing number of Chinese companies that make SDG disclosures is the extent and nature of these disclosures, and whether these represent a change in the nature of company disclosures.

2.5 Why SDGs may matter to Chinese companies

Since the introduction of the 13th National FYP, the government has increasingly promoted ‘building a community of shared future of mankind’, building a ‘Moderately Prosperous Society’³, patriotism (Jia Guo Qing Huai), low-carbon development, and the Beautiful China plan (Green Mountains, Clean water, and Blue-sky plan (‘Qing Shan, Lv Shui, and Lan Tian Plan’)) (Communist Party of China, 2016a; Wan et al., 2021). However, existing performance regulations may not cover the latest issues in periodic national plans (i.e. the 13th National FYP). With China’s adoption of Agenda 2030, SDGs may serve as a basis for companies to engage in national priorities, such as building a community of shared future of mankind and building a moderately prosperous society, to pursue sustainable business growth, and to reduce risks in the value chain.

2.5.1 SDGs and building a community of shared future of mankind

SDGs may serve as a basis for corporate engagement in the goal of ‘building a community of shared future of mankind’, where the Green BRI is considered an instrument for achieving this goal (The State Council Information Office of the People's Republic of China, 2019; Xinhua, 2017). In 2017, President Xi announced the establishment of the BRI international cooperation coalition to promote environmental protection and build ‘a sound ecosystem’ to achieve the goals in the Agenda 2030 (Xinhua, 2017). The State Council Information Office of the People's Republic of China (2017) also states that the Green BRI is a practical tool for the government to build ‘a community of shared future of mankind’. Therefore, we argue that companies that involve in BRI-related SDG goals may help the government to achieve the goal of ‘building a community of shared future of mankind’ to some extent. Prior research has suggested that SDG goals are closely aligned with the BRI (Lewis et al., 2021; Yin, 2019). Lewis et al. (2021) explore the evolving synergies between BRI and SDGs in mobilising policies and corporate decisions in pursuing sustainable development in China. This study suggests that SDG 7, 8, 9, 10, 11, 14, and 17 are closely aligned with the goals of BRI; specifically, SDG 7 and 14 involve ecological and environmental protection, SDG 9 emphasises sustainable infrastructure, and SDG 8, 10 and 17 concern global trade, green finance, and partnership. Similarly, Yin (2019) suggests that SDG 6,

³ It is worth noting that ‘building a community of shared future of mankind’ as China’s foreign policy has finally been written into the *Constitution of People's Republic of China* in 2018 (<https://www.elegislation.gov.hk/hk/A7%21sc.assist.pdf>). Recently, the government also claimed patriotism (Jia Guo Qing Huai) as one of the most crucial qualities of entrepreneurs. Patriotism refers to the idea that business success is closely linked with national prosperity. For example, the Chinese government consistently emphasises improving the quality of products that are made in China in order to create a globally recognised Chinese brand and economic development (Zhang, 2022).

7, 9, 11 and 17 directly connect to the Green BRI to facilitate sustainable infrastructure development, regional connectivity and economic development.

2.5.2 SDGs and building a ‘Moderately Prosperous Society’

SDGs may serve as a means of corporate engagement in the national strategy of building a ‘Moderately Prosperous Society’. In 2016, the 13th Five-Year Plan for Poverty Alleviation was issued, emphasising the important role of state-owned and private companies in poverty alleviation (Bin-Feng et al., 2023). The CPC established ‘building a Moderately Prosperous Society’ as the prioritised goal within the 13th National FYP (Chang et al., 2020). To achieve this, the government launched a national campaign highlighting the ‘targeted method’ to eradicate poverty, aiming to mobilise government, people and organisations from all corners of the society, including leveraging corporate capabilities to achieve poverty alleviation (Chang et al., 2020). By 2020, President Xi announced that we had achieved the goal of ‘Moderately Prosperous Society’.

Interestingly, since 2015, Chinese companies have actively engaged in the Targeted Poverty Alleviation (TPA) programs (Qiao et al., 2021). For example, a program called ‘A Hundred Counties and Ten Thousand Villages’ has been launched by SOEs, while the project ‘Ten Thousand Enterprises supporting Ten Thousand Villages’ has been introduced and implemented by non-SOEs (Qiao et al., 2021). As part of the 13th FYP Plan of Poverty Alleviation, Chinese companies are also encouraged to provide poverty alleviation disclosures. The China Securities Regulatory Commission (CSRC) introduced the "Opinions of the China Securities Regulatory Commission on the Role of the Capital Market in Serving National Poverty Alleviation", which requires SSE and SZSE to format corporate poverty alleviation disclosures. In December 2016, the “Notice on Further Improving the Information Disclosure of Poverty Alleviation of Listed Companies” was issued by SSE and SZSE to further guide listed companies to report poverty alleviation; for example, companies are required to report their plan, overview, input and output, and follow-plan of TPA in the ‘Important Events’ section in annual reports. Non-compliant companies need to acknowledge their failure in poverty alleviation practices. Companies are also encouraged to disclose their poverty alleviation information in their ESG reports, provided that these disclosures do not impose additional costs on companies. Since 2015, Chinese companies have been actively engaging in TPA programs and disclosing relevant information (Qiao et al., 2021). Corporate commitments to poverty alleviation are not limited to an economic task, with increasing recognition as having social and political implications. In China, it is argued that the state-initiated ideology has symbolic power in influencing corporate ESG practices, which is 'an imprinting process of ideological prioritisation' exerted by CPC (Chua et al., 2024; Zhu, Spence & Ezzamel, 2021). It finally becomes the 'habitus' for companies to engage in ESG reporting practices (Situ, Tilt & Seet, 2021). Therefore, to survive and grow, companies tend to react to the political ideology to ensure that companies can operate in a broader value system that encompasses the Chinese government and stakeholders. When the 13th National FYP was effective, building a 'Moderately Prosperous Society' became the new political ideology.

By 2020, over 2100 companies have engaged in TPA programs. According to China's Progress Report on the Implementation of the 2030 Agenda for Sustainable Development (2019), SDG 1 has been widely adopted to help China eradicate extreme poverty. Overall, SDGs may guide Chinese companies to engage in poverty alleviation programs, which may help achieve the goal of 'building a Moderately Prosperous Society'. This provides a unique opportunity for exploring corporate involvement in SDG 1.

2.5.3 SDGs and resolving environmental risks in the value chain

SDGs may provide a multi-dimensional framework for resolving risks in the value chain (SDG Compass, 2015). They are closely aligned with the mission to develop an ecological civilisation within the 13th National FYP. The government has introduced multiple strategies, such as the Green Mountains, Clean water and Blue-sky Plan (Qingshan, Lvshui, and Lantian Plan) and low-carbon development, to achieve this goal. As discussed above, China has made great progress in ecological civilisation through implementing Agenda 2030 (Permanent Mission of the People's Republic of China to the UN, 2022). Recent studies also primarily provide a conceptual framework to identify how SDGs can be used to promote a sustainable supply chain and reduce relevant risks (Russell et al., 2018; Zimon et al., 2020). Therefore, we argue that SDGs may assist Chinese companies in reducing environmental risks.

Overall, SDGs play an important role in corporate engagement in sustainable development and can serve as a basis for companies to involve in national priorities in China. However, what remains unclear is the extent to which Chinese companies have involved in SDGs.

2.6 Why SDG disclosures may matter to Chinese companies

The government, investors and the public have paid more attention to SDG disclosures (CDSB & WBCSD, 2017; Qian et al., 2021). The increase in disclosure can be viewed as a corporate response to 'public calls for sustainable development' (Qian et al., 2024, p. 2). An increasing number of studies explored SDG disclosure practices in large companies and argue that 'these companies are more prone to face legitimacy and institutional pressure' (Datta & Goyal, 2022, p. 3).

2.6.1 SDG disclosures as a corporate response to the Chinese government's information needs

SDG disclosures may serve as a corporate response to the national strategies in China, demonstrating their political legitimacy. Given the unique authoritative and regulatory setting in China, the government can substantially influence businesses to achieve its political aims (Situ et al., 2018). Since 2015, political aims include 'building a moderately prosperous society' or developing an ecological civilisation. Under a 'responsive' authoritative setting, disclosure may be a visible instrument for companies to remain responsive to the government's expectations (Situ & Tilt, 2012). This may subsequently assist companies in achieving political legitimacy (Marquis & Qian, 2014; Situ & Tilt, 2012). Larger companies may have even stronger motivations to comply with the latest national

strategies and policies and report on it, due to increased public, regulatory, and stakeholder pressure (Situ & Tilt, 2012). Failing to do so may result in political costs (Beck et al., 2018).

2.6.2 SDG disclosures as a means of meeting information needs of stakeholders

SDG disclosures may also serve as a basis for an ongoing dialogue with stakeholders, which may create opportunities for companies to attract investment capital (Adams et al., 2020; Rosati & Faria, 2019a, 2019b). Companies are receiving increasing recommendations from analysts regarding their participation in SDG activities (García-Sánchez et al., 2020). Disclosures may allow investors to evaluate what corporate management has done to pursue SDGs and enable them to recognise material risks and opportunities related to corporate sustainable development (Adams et al., 2020; Rosati & Faria, 2019a, 2019b). Larger companies are more prone to facing institutional pressure 'for having to communicate with various stakeholders', which may lead to increased transparency in SDGs (Datta & Goyal, 2022, p. 3; Elalfy et al., 2021).

However, the SDGs are an agenda rather than a reporting framework (Grainger-Brown & Malekpour, 2019). Although the UN does provide companies with some standardised indicators to measure their SDG performance, how companies make SDG disclosures is very varied and poorly understood.

2.7 Guidelines for observing the nature and extent of corporate SDG disclosures

Over the past decades, an increase in ESG disclosures has been observed (Bose & Khan, 2022; Pizzi et al., 2020); however, existing reporting frameworks are partially aligned with the measurements of corporate sustainable development; issues that are inadequately addressed include 'risks and opportunities regarding sustainable development, the implication of value creation, and the impact of the achievement of SDGs' (Adams et al., 2020, p. 6). Despite the evolution of reporting worldwide, there remains a need for evolving better ESG disclosures. Since 2015, SDG disclosures have gradually become a part of ESG disclosures to supplement existing reporting frameworks (Bose & Khan, 2022).

2.7.1 What constitutes an SDG disclosure?

Target 12.6 'encourages companies, especially large and transnational companies to adopt sustainable practices and to integrate sustainability information into their reporting cycle' (United Nations, 2015, p. 27). By improving reporting practices, companies will be able to gain a better understanding of corporate practices in SDGs, improve their effectiveness of their communication with stakeholders and enhance their management capacities (GRI & United Nations Global Compact, 2018). Since SDGs are an inter-governmental initiative and a performance guideline, rather than a reporting framework, there has been a growing number of studies exploring what counts as an SDG disclosure at both definition and observable levels.

Definition and purpose

To measure corporate contributions towards SDGs, the United Nation Global Compact argues that SDG disclosures can be viewed as

'an opportunity for businesses to demonstrate to investors, competitors and customers how and why the SDGs are fundamentally shaping the way they do business responsibly and sustainably. Businesses are increasingly expected to transparently report how they are managing and mitigating sustainability matters, and the SDGs provide a powerful universal framework through which to do this' (<https://unglobalcompact.org.au/sustainable-development/sdg-reporting/>).

Rosati and Faria (2019a, p. 1313) define an SDG disclosure as 'the practice of reporting publicly on how an organisation addresses the SDGs'. Similarly, Nechita et al. (2020) view SDG disclosures as a framework that may guide companies to communicate their performance and impacts with stakeholders, and increase their potential to collaborate with other companies. Companies need to find their approaches to identify and communicate their contributions towards SDGs.

Izzo et al. (2020a) assume SDG disclosures as the latest innovation in sustainability. By sampling 40 Italian companies on FTSE MIB Index in 2016, this study concludes that

'SDGs represent—at least for the analysed companies—an extension to, or evolution of, corporate responsibility reporting, influencing their reporting practices and allowing them to implement the information provided.' (Izzo et al., 2020, p. 11).

Adams (2017) on behalf of the IIRC, introduced a conceptual framework for SDG disclosures, which is aligned with the IR framework. It looks at how SDGs can be embedded into

'an organization's decision making, strategies and business model by considering risks and opportunities presented by the external environment; and, acknowledging that creating value over time requires the organisation to acknowledge and address the risks and opportunities associated with social and environmental stewardship' (Adams, 2017, p. 6).

Academic studies also provide some insights into what an SDG disclosure is. Bebbington and Unerman (2018, p. 18) conceptually argue that SDG disclosures can be seen as a holistic reporting framework that can capture environmental and human development issues, in order to achieve 'transformational global development'. It may provide effective communication across stakeholders, further facilitating collaboration among sectors (Florini & Pauli, 2018). Particularly, SDGs serve as a unified language for companies to communicate their social and environmental performance towards various parties, such as governments, regulators and social media (Schonherr et al., 2017).

Overall, prior studies conceptually provide insights into what counts as an SDG disclosure; however, no consensus has been reached on the nature of SDG disclosures.

Observable level

An increasing number of studies have mapped SDG disclosures onto existing reporting frameworks, such as GRI, SASB, and IR, to observe corporate SDG disclosures.

Consolandi et al. (2020) investigate whether SDGs can be mapped onto the SASB reporting framework to avoid further profusion of ESG standards. This study sampled US-headquartered companies in the healthcare sector from January 2017 to February 2018 (See Table 2.2). By linking ‘30 generic ESG issues’ with the SDG 3 targets related to ‘health and wellbeing’, including ‘air quality, energy management, fuel management, water and wastewater management, waste and hazardous materials management, access and affordability, customer welfare, fair disclosure and labelling, fair marketing and advertising, labour relations, fair labour practices, employee health, safety, and wellbeing, lifecycle impacts of products and services, environmental, social impacts on assets and operations, product quality and safety, accident and safety management, materials sourcing, and supply chain management’ (Consolandi et al., 2020, p. 517). However, SASB is an industry-based guideline, and the reporting template is customised in different industries, which may increase the difficulty in demonstrating corporate contributions to sustainability in a consistent way.

Table 2.2 Extracted from Consolandi et al. (2020, p.518)

SASB's GICs relevant to SDG 3 for the health care sector	SDG 3 targets affected
Air Quality	3.1,3.2, 3.4, 3.9
Energy management	3.1, 3.2, 3.4, 3.9
Fuel management	3.1, 3.2, 3.4, 3.9
Water and wastewater management	3.3, 3.4, 3.9
Waste and hazardous materials management	3.4, 3.9
Access and affordability	3.1, 3.2, 3.3, 3.4, 3.5, 3.7, 3.8
Customer welfare	3.1, 3.2, 3.4, 3.5, 3.8, 3.9
Fair disclosure and labeling	3.1, 3.2, 3.4, 3.5
Fair marketing and advertising	3.1, 3.2, 3.4, 3.5
Labor relations	3.3, 3.4, 3.8, 3.9
Fair labor practices	3.3, 3.4, 3.5, 3.9
Employee health, safety, and well-being	3.1, 3.3, 3.4, 3.5, 3.7, 3.8, 3.9
Lifecycle impacts of products and services	3.4, 3.5, 3.9
Environmental, social impacts on assets and operations	3.3, 3.8
Product quality and safety	3.1, 3.2, 3.4, 3.5, 3.8, 3.9
Accident and safety management	3.6, 3.9
Materials sourcing	3.9
Supply chain management	3.1, 3.2, 3.4, 3.5, 3.8, 3.9

Note. SDG = Sustainable Development Goal; SASB = Sustainability Accounting Standard Board; GIC = general issue categories.

Ordonez-Ponce and Khare (2021) sampled 30 largest carmakers from the Fortune 500 to investigate business contributions towards sustainability, particularly environmental sustainability, based on GRI indicators. This study argues that SDGs may introduce ‘a useful framework towards new business models for sustainability’ to companies (Ordonez-Ponce & Khare, 2021, p. 47). This study suggests that a broad range of SDGs, including SDG 1, 2, 3,6, 7, 9, 11, 12, 14, and 15, can be incorporated into

business models to help companies to build environmental sustainability and GRI 300 Environmental Standards⁴ can be a useful tool to track SDG progress in the environmental dimension (See Figure 2.1). Specifically, GRI 300 can be used to operationalise SDGs in the following aspects: ‘materials, energy, water and effluents, biodiversity, emissions, and effluents and waste’. However, most GRI indicators are quantitative, measuring the volume, weight, usage, and proportion. Corporate actions, such as intentions, plans, and future actions, are not adequately measured. Therefore, there remains scope for exploring appropriate indicators for measuring corporate SDG-related disclosures.

Figure 2-1 Extracted from *Ordonez-Ponce et al. (2021, p.56)*

GRI 300		Sustainable Development Goals																
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17
Materials 301	-1: Materials used -2: Recycled input -3: Reclaimed products and packaging																	
Energy 302	-1: Consumption within the organization -2: Consumption outside of the organization -3: Intensity -4: Consumption reduction -5: Requirement reduction of products and services																	
Water and Effluents 303	-1: Interactions as a shared resource -2: Management of discharge-related impacts -3: Withdrawal -4: Discharge -5: Consumption																	
Biodiversity 304 [¶]	-1: Operational sites -2: Impacts of activities -3: Habitats protected or restored -4: IUCN Red List species affected by operations																	
Emissions 305 ^{¶¶}	-1: Direct GHG emissions -2: Indirect GHG emissions -3: Other GHG emissions -4: Intensity -5: Reduction -6: Ozone-depleting substances -7: NOX, SOX and other air emissions																	
Effluents and Waste 306(GRI 2018f)	-1: Discharge -2: Waste -3: Spills -4: Hazardous waste -5: Water bodies affected																	

Michalczuk and Konarzewska (2018) rely on the GRI G4 reporting framework to capture corporate SDG disclosures of stakeholder-oriented sustainable development activities. They argue that among all the G4 indicators, GRI 102-8 (employee and workers), GRI102-16 (standards and principles), GRI 102-41 (collective bargaining), GRI 103-2 (management approaches), GRI 205-3 (incidents of corruption), GRI302-1 (energy consumption), and GRI401-1 (employment and turnover) are well aligned with SDG 4, SDG 8, SDG 12, and SDG16. However, a study by Szennay et al. (2019, p. 9) suggest that GRI indicators for the ‘supply chain, local community, employee training, and customer satisfaction’ are not well aligned with SDGs; existing GRI indicators may not be adequate to measure some SDG-related social and governance issues, as a limited number of GRI indicators can be used to measure SDGs. Some SDG goals and targets are still overlooked since they are not part of the GRI framework. Although

⁴ The selected groups of GRI 300 include Material (GRI 301 1-5), Energy (GRI 302 1-5), Water and Effluents (GRI 303 1-5 and GRI 304 1-3), Emissions (GRI 305 1-7), Effluents and Waste (GRI 306 1-5), Environmental Compliance (GRI 307-1), and Supplier environmental assessment (GRI 308-1).

these frameworks provide some specific suggestions for measuring SDG disclosures in existing reports, they are inadequate for measuring them due to the insufficient number of indicators⁵.

Overall, an SDG disclosure is conceptually defined as an effective framework to communicate with governments, customers, investors, and other stakeholders (UN, 2015). However, existing frameworks are limited in providing understanding of what constitutes SDG disclosures. Particularly, in China, SDGs have been employed by the government to achieve the goals in the 13th National FYP. Therefore, there remains a gap in exploring what SDG disclosures mean to China and Chinese companies.

2.7.2 Guidelines for observing corporate SDG disclosures

Various parties have developed guidelines and measurements to explore corporate SDG disclosure practices, such as evaluating the adoption of SDG disclosures, analysing the frequency of SDG-related terminology in reports, employing generic score-based disclosures, and conducting an in-depth analysis of SDG disclosures. However, existing methods focus more on the breadth of SDG disclosures and the alignment of a company's strategy, KPI, and operation with SDGs, without providing details for a particular goal, or taking into account China's unique circumstances.

2.7.2.1 SDG disclosure measurements developed by standard setters

SDG disclosure measurements developed by the United Nations

In support of Agenda 2030, the UN sets 17 goals, 169 targets, and 232 indicators against which progress in relation to these goals and targets could be monitored, aiming to provide an overview of corporate implementation towards global challenges. Each of the 17 goals has between eight and twelve targets⁶ against which to monitor progress towards the goals. This UN framework of goals, targets and indicators was designed to be relevant to a broad range of stakeholders in different areas and levels of society. Some of the specified targets seek to motivate companies to contribute to the realisation of Agenda 2030 through adopting some SDG goals and including SDG information in the disclosures. For example, one of the targets for SDG 12 'Sustainable Consumption and Production' is to 'encourage companies, especially large and transnational companies, to adopt sustainable practices and to integrate sustainability information into their reporting cycle' (United Nations, 2015, pp. 18-32). In relation to SDG 9 'Sustainable innovation and industrialisation', the suggested indicator 9.3.1 is 'Proportion of small-scale industries in total industry value added', and for SDG 8 'Sustainable economic growth, decent work and productive employment', for which the suggested indicator 8.1.1 is 'Annual growth rate of real GDP per capita' (United Nations, 2015, pp. 18-32). As these examples make clear, SDG goals and targets are aspirational and have a generic focus, and the mixture of broad qualitative and

⁵ I manually compared the extent to which existing reporting frameworks, including GRI, IR, and SASB, can provide guidelines for SDG goals and targets. It is observed that existing reporting frameworks are partially aligned with SDGs (See Appendix 1).

⁶ Targets include both 'outcome targets' and 'means of implementation targets' (United Nations, 2015).

quantitative indicators means that the SDG goals, targets and indicators were neither designed nor are adequate as a corporate SDG disclosure framework. There remains a challenge in developing appropriate SDG disclosures.

Since the launch of Agenda 2030, the UN, in partnership with the GRI, United Nations Global Compact, WBCSD, and PRI has sought to develop guidelines specifically related to business reporting in relation to the SDGs. In practice, this has taken the form of guidelines which show how different SDGs align with these different existing reporting guidelines. The first of these, the SDG Compass (2015)⁷ sought to illustrate how companies can use the SDGs to map and assess opportunities and risks in their entire value chain, in order to achieve materiality analysis. However, this framework tends to provide a general overview of corporate disclosure alignment with SDGs, which does not provide detailed indicators and targets for guiding and evaluating performance in relation to the value chain.

- ‘Why the SDG has been identified as relevant and how (for example, describe the process for defining SDG priorities and any stakeholder engagement used);
- The significant impacts, whether positive or negative, related to the relevant SDG;
- Their goals for the relevant SDG and progress made in achieving them;
- Their strategies and practices to manage impacts related to the SDGs and achieve goals through
- Integration across the business (for example, a description of policies, systems and processes such as due diligence)’ (SDG Compass, 2015, pp. 3-28).

A further step towards a specifically corporate set of SDG disclosure guidance came in 2017 with the publication of ‘Business Reporting on the SDGs: Analysis of the Goals and Targets’ (UNGC & GRI, 2017) which specifies goals, indicators and targets that are particularly relevant to companies. This guide also draws on the existing reporting frameworks of the GRI and UNGC principles to extract a set of existing indicators that are aligned with the SDG goals and targets and can be used to guide business reporting of SDGs. Of the total of 169 UN targets, this guide argues that 128 targets are relevant to companies, with 62 UN targets being ‘highly likely to be relevant’ to companies, 42 targets being ‘likely to be relevant to’ companies, and 24 indicators being ‘less likely to be relevant’ to companies (UNGC & GRI, 2017, pp. 208-209). In 2018, ‘Integrating the SDGs into corporate reporting: A practical guideline’ (UNGC & GRI, 2018) provides a step-by-step guide to help companies identify and prioritise relevant SDG goals and targets, incorporate SDGs into their reports, and generate a checklist of possible information related to SDGs that they need to include in their reports (See Figure 2.2 below). It suggests that SDG disclosures are expected to be a tool to show the impacts of companies on global challenges. While these frameworks do provide some specific guidance for corporate reporting related to the SDGs, some elements of the SDGs are still overlooked since they are not part of existing reporting frameworks.

⁷ ‘SDG Compass was developed through the partnership between GRI, UNGC, and WBCSD’ (United Nations, 2015).

Figure 2-2 Extracted from UNGC and GRI (2018, p. 23)

BOX 8

Checklist – What to set out in your report?

- ✓ Your company’s significant impacts, whether based on risks to people or environment or on beneficial products, services and investments.
- ✓ How your analysis of these impacts has informed your identification of priority SDG targets.
- ✓ How stakeholder feedback informed your conclusions.
- ✓ Your strategy, including objectives (goals) and measurement (indicators) for contributing to the priority SDG targets, recognizing that positive contributions can result from both tackling risks and providing beneficial products or services. This may include a description of relevant company policies, systems and processes, including your engagement with stakeholders.
- ✓ Instances where your company has caused or contributed to actual negative impacts, and the action(s) you have taken to enable effective remedy to anyone whose human rights were harmed.
- ✓ Indicators and data that demonstrate how your company is progressing towards its objectives for contributing to its priority SDG targets and any setbacks it has encountered.
- ✓ Your future plans for achieving further progress.

SDG disclosure frameworks developed by other standard setters

Standard setters and professional bodies, such as GRI and PwC, also provide insights into how companies have committed to SDG disclosures, by introducing checklists for evaluating the overall SDG disclosure levels. However, these frameworks do not take into account, for example, the unique national circumstances or culture.

Adams (2017) argues that indicators for assessing the overall level of SDG information disclosed in reports include several factors: whether companies mention SDGs in their reports, whether companies show the impacts of SDGs on external environment, whether companies identify relationships between SDGs and risk management framework of IR, whether companies report any methods for evaluating corporate SDG performance, whether companies link SDGs with any reporting framework, whether companies involve external parties in SDG activities, and whether companies have incorporated SDGs into integrated thinking and value creation process. This framework also includes indicators related to separate integrated sections related to SDGs, relevant indicators or data regarding SDGs, and effective SDG reporting in existing reports. However, this framework, again, is limited to a guideline rather than providing detailed indicators for measuring SDG disclosures. It did not take into consideration the unique circumstances within China and what elements can be used to observe corporate SDG performance, rather than the generic focus on SDGs in companies.

Adams (2017, p. 7) evaluates

- ‘Understand sustainable development issues relevant to the organisation’s external environment
- Identify material sustainable development issues that influence value creation
- Develop a strategy to contribute to the SDGs through the business model
- Develop integrated thinking, connectivity and governance
- Prepare the integrated report’

PwC (2018) observes overall corporate SDG reporting levels from the following dimensions: mentioning SDGs in reports, identifying SDG-related priorities, KPIs or targets related to SDGs, incorporating SDGs into business strategies or business models, and mentioning SDGs in the CEO's statement. Of the 729 companies from 21 countries, this study suggests that a number of companies are committed to SDGs, with good intentions to implement them; however, there remains a gap between their commitments and actions. PwC (2018) also attempts to help companies to identify how SDGs may create value to businesses, including whether companies link measurement approaches (i.e. the PwC total impact measurement and management framework) to SDGs, whether companies have separate reporting on SDGs, and whether companies identify any material value-creating activities regarding SDGs. These studies treat SDG disclosures as a tool to identify how SDGs may create value to companies; however, these frameworks focus on corporate commitments to SDG disclosures and fall short in offering a better understanding of corporate engagement for each goal or target.

PwC (2018, p. 22) evaluates

- 'The company makes a statement about the SDGs and their important but does not include any specific aspirations or ambitions
- The company makes a statement about the SDGs and includes a qualitative ambition or aspiration on achieving them
- The company identifies quantitative KPIs for its relevant SDGs
- The company identifies quantitative KPIs and targets for its relevant SDGs
- The company links its SDG KPIs to its societal impact'

Overall, standard setters have been working on developing frameworks to measure corporate SDG performance. Reporting frameworks developed by IIRC and PwC, primarily focus on guiding principles, rather than indicators that can help measure SDG performance, or generic disclosure frameworks without providing actual corporate SDG performance. It also does not take into account the unique circumstances within China, such as the government's emphasis on SDGs, and national SDG policies that may influence how companies prioritise SDG implementation and disclosures. Therefore, there remains a gap in our understanding of what elements we can use to observe SDG disclosure levels in reports.

2.7.2.2 SDG disclosure measurements developed by academic research

Academic studies have utilised various approaches to interpret the complexity of SDG disclosures, including adopting measurements developed by standard setters, evaluating the adoption of SDG disclosures, analysing the frequency of SDG-related terminology in reports, employing generic score-based disclosures, and conducting an in-depth analysis of SDG disclosures.

The adoption of SDG disclosures

Rosati and Faria (2019a) and Rosati and Faria (2019b) employ the GRI database to evaluate corporate SDG disclosures in 2016. These studies treat SDG disclosures as a binary variable, by measuring whether companies mention SDGs in their reports. A similar method has been used in Girón et al. (2020), which suggest that the increasing concerns regarding environmental sustainability and ecosystem preservation may encourage companies to integrate long-term sustainable development strategies into their businesses.

Similarly, García-Meca and Martínez-Ferrero (2021) suggest that SDG disclosures are considered a symbolic strategy for companies to maintain their legitimacy; on the other hand, for environmental-sensitive industries, SDG disclosures may serve as a basis for showing how companies become better and more transparent in resolving ethical and environmental issues, as stakeholders value corporate commitments towards SDGs. This study sampled 523 firm-year observations under the European context between 2015 and 2016. However, it is limited to examine whether companies address SDGs in general rather than looking at SDG targets.

What remains unclear is the level of their involvement with these goals and the goals that companies have prioritised in Chinese companies.

Word frequency method in measuring SDG disclosures

Some studies argue that SDG disclosures are a symbolic tool or rainbow washing, and therefore, they simply analyse whether companies disclose SDG information in their reports or adopt word frequency as an indicator of corporate in-depth discussions of SDG in their reports (Izzo, Ciaburri, et al., 2020; van der Waal & Thijssens, 2020). Girón et al. (2020), and Lassala et al. (2021) investigate the adoption of SDGs in a company's sustainability report, to see whether companies improve SDGs in their reports. van der Waal and Thijssens (2020) apply mechanistic content analysis to measure SDG involvement of companies based on word frequency; that is, companies can be classified based on the existence of the word 'SDG' in their reports and the frequency that the word 'SDG' appears. This differentiates the SDG disclosure levels into 'nominal mention' or 'a dominant and substantive piece of SDG related information' (van der Waal & Thijssens, 2020, p. 4). Similarly, Izzo et al. (2020a) observe that corporate awareness of SDGs is high; however, the requirements of the SDGs, and KPI related to SDGs are missing in reports. This study collects SDG information from annual reports, sustainability reports and integrated reports to identify the frequency with which SDGs are mentioned in reports, either in words or phrases. This study also investigates the level of SDG information disclosed, by analysing the frequency with which SDGs occurred in the reports and the number of pages that have shown SDG related information. This method has also been used in Horne et al. (2020). However, it may be hard to differentiate between pages that only reference a number of SDGs and pages that include in-depth discussions of SDGs. Lacking strategic-oriented information may undermine corporate performance

towards sustainable development, which may be considered rainbow-washing activities or philanthropic add-ons. Therefore, there remains scope in identifying appropriate indicators to capture the substantiveness of SDG disclosures.

Generic score-based SDG disclosure measurements

In addition to measuring corporate commitments to SDGs and the discussion on SDGs based on word frequency methods, recent academic studies also make use of score-based disclosure frameworks to measure the quality of SDG disclosures.

Izzo et al. (2020b), for example, suggest that integrated reports can be used to see how companies leverage IR to embed SDG information. However, IR is a ‘principle-based framework’ that seeks to pose questions that companies need to consider when preparing their IR (IIRC, 2021, p. 2). It is not a framework that provides standardised disclosure indicators.

Cosma et al. (2020) adopt a two-point coding system, by classifying SDG disclosures into generic/qualitative reference and quantitative reference to the SDGs two levels. Calabrese et al. (2022) use content analysis to assess corporate contributions towards SDGs, based on 0-4 scale. Specifically, Calabrese et al. (2022, p. 5) analyse ‘whether the company contributes to any SDG, whether any SDG target is mentioned without mentioning ambition, whether SDG targets are mentioned with a qualitative ambition, whether SDG targets are mentioned with a quantitative ambition, and whether SDG targets are mentioned with quantitative achievements’. It is, again, limited to generic disclosures of all goals, rather than assessing SDG performance per se.

Calabrese et al. (2022, p. 5)

- ‘No contribution to SDGs
- SDGs are mentioned without ambitions
- SDG targets are mentioned with a qualitative ambition
- SDG targets are mentioned with a quantitative ambition’

Ionaşcu et al. (2020) use quality scores to compare the depth of sustainability commitments related to SDGs among 78 EU real estate companies. It may not accurately differentiate between companies that only reference a number of goals, and companies that have disclosed substantive SDG-related information.

Ionaşcu et al. (2020, p. 7) states ‘where:

- *The entity declares the importance of the SDGs, but without including certain aspirations.*
- *The entity declares the importance of the SDGs and includes qualitative aspirations for*
- *achieving them. To determine the quality of the SDGs reporting prioritised by real estate entities, the goals were evaluated by a score from 1 to 5*

- *The entity identifies quantitative key performance indicators (KPIs) for the relevant SDGs.*
- *The entity identifies KPIs and quantitative targets for the relevant SDGs.*
- *The entity correlates KPIs assigned priority SDGs with the impact on the company.'*

Some studies also use more articulated methods to capture the level of details that companies have disclosed in their reports, intending to indicate how companies address SDG activities. Ionaşcu et al. (2020) examine to what extent RU real estate companies fulfil their obligations in SDGs to maintain their legitimacy and suggest that companies have shown their intention and interests in SDGs, however, most companies have limited actions or do not have a strategy to turn their commitments into actions. This study follows Nechita et al. (2020, p. 8) to classify SDG disclosures into the following 5 levels: ‘

- The entity declares the importance of the SDGs, but without including certain aspirations.
- The entity declares the importance of the SDGs and includes qualitative aspirations for achieving them.
- The entity identifies quantitative key performance indicators (KPIs) for the relevant SDGs.
- The entity identifies KPIs and quantitative targets for the relevant SDGs.
- The entity correlates KPIs assigned priority SDGs with the impact on the company’.

However, this method fails to provide indicators for capturing SDG performance and for exploring the extent to which companies have turn their commitments into actions.

According to García-Sánchez et al. (2020), SDG disclosures can be observed based on four levels: no information linked to 17 goals, structured SDG information without complying with the GRI or SDG Compass framework, disclosed SDG information aligned with SDG Compass and GRI G4 at an 'accordance-core' level, and SDG information based on SDG Compass strategy. Similarly, Avramou et al. (2019) measure European banks' SDG endorsement against the GRI framework (0-point: no relevant information; 1-point: only generic SDG information; 2-point: limited number of clear information; 3-point: extensive reported information; 4-point: comprehensive and systematic information against GRI framework). However, they neither assess SDG performance per se nor evaluate the comprehensiveness of performance information. Overall, this study focuses on the quality of SDG disclosures, by assessing their alignment levels with the GRI framework. There remains a gap in understanding what elements of SDGs, such as goals or targets, are prioritised in a company.

García-Sánchez et al. (2020, p. 7) evaluate

- ‘Value 0 was given to those companies that, although disclose indicators related to the SDGs, are not specifically linked to some of the 17 objectives established by the UN.
- Value 1 was assigned to those companies that have issued an autonomous SDGs report, in which they structure and organize the information around the relevant SDGs, although it does not comply with the GRI-SDG Compass strategy.

- Value 2 was received by companies that adopted the structure recommended by the SDG Compass but only disclose in their CSR report the number of indicators established in the GRI G4 guide for the level in "accordance-core."
- Value 3 was for those companies that adopted the structure recommended by the SDG Compass, disclosing in their CSR report the number of indicators established in the GRI G4 guide for the "in-accordance-comprehensive" level.'

Manes-Rossi and Nicolo' (2022) explore corporate SDG disclosures in European energy companies between 2017 and 2019, aiming to determine whether companies substantially report SDG information or symbolically report their SDG information to maintain their legitimacy. This study suggests that SDGs gradually become part of corporate sustainability disclosures, but the overall level of SDG disclosures is limited, indicating that companies do not have substantial changes in reports regarding SDGs. Again, this study investigates the quality of SDG disclosures and classifies reporting levels into seven levels:

Extracted from Manes-Rossi and Nicolo' (2022, pp. 9-14)

- 'Does the CEO and/or Chair's message talk about the SDGs?
- Does the materiality matrix include SDGs priorities?
- Does the reporting link corporate strategy or strategic objectives to priority SDGs?
- Does the reporting integrate priority SDGs into corporate business model?
- Does the reporting identify potential risks and opportunities related to the SDGs?
- Does the reporting identify specific SDG targets that are relevant to the business?
- Does the reporting associate specific Key Performance Indicators (KPI) to priority SDGs'

These studies rate SDG information on a scale of 0-n levels, with a higher score indicating better performance in SDGs. However, a limited number of metrics have been adopted in these studies, which may limit a more profound understanding of corporate involvement in SDGs. For example, companies may simply state their intentions and link SDGs to existing CSR activities, without making significant changes or achieving any outcome regarding SDGs. Therefore, there remains scope for differentiating corporate intentions and symbolic actions from actual corporate changes in contributing to SDGs.

An in-depth discussion of SDG disclosures

KPMG (2018) sampled 250 largest companies worldwide to investigate current corporate SDG disclosure state worldwide. This study develops a framework, aiming to provide a comprehensive overview of what a good SDG disclosure framework may look like. This framework includes three dimensions: corporate understanding of SDGs, corporate SDG prioritisation, and SDG measurements. Specifically, the understanding level explores whether the report shows a business case for corporate engagement in SDGs, whether CEO message mention SDGs, and whether the report assesses corporate

impacts on SDGs. The prioritisation level explores the goals that are prioritised by companies, the methods that a company uses to prioritise these goals, and relevant SDG targets. Regarding the measurement level, whether the report includes any SDG performance goal, whether the report set any goals that are SMART, and whether any detail has been provided to measure SDG progress are included.

Table 2.3 Extracted from KPMG (2018, pp. 4-5)

Understanding	<p>‘Does the reporting demonstrate the business case for taking action on the SDGs?</p> <p>Does the CEO and/or Chair’s message talk about the SDGs?</p> <p>Does the reporting assess the business’s impact on the SDGs?’</p>
Prioritisation	<p>‘Does the reporting identify priority SDGs for the company?</p> <p>Does the reporting explain the methodology the company used to prioritize the SDGs?</p> <p>Does the reporting identify specific SDG targets that are relevant to the business?’</p>
Measurement	<p>‘Does the reporting disclose SDG performance goals for the company?</p> <p>Does the reporting set SDG performance goals that are SMART?</p> <p>Does the reporting detail the indicators the company is using to measure the progress of its SDG activities?’</p>

Hummel (2019) combines textual and content analysis to assess the quality of SDG disclosures. Specifically, it uses textual analysis to identify the presence of SDGs in corporate reports. It also adopts content analysis to provide an in-depth analysis, by distinguishing the ‘definition and prioritisation of the SDGs’ from ‘measurement and analysis of the SDGs’, based on 11 reporting items provided by GRI and UN Global Compact.

Table 2.4 Extracted from Hummel (2019, p. 48)

	Disclosure Indicator	Measurement
<i>Definition and prioritization</i>		
DEF1	Does the statement from the senior decision-maker reference the SDGs?	0 (no) 1 (yes)
DEF2	Does the report provide general information on the SDGs?	0 (no) 1 (yes)
DEF3	Does the report provide information on the process of SDG prioritization?	0 (no) 1 (yes)
DEF4	Does the report provide information on the outcome of SDG prioritization (i.e., the SDGs they relate to)?	0 (no) 1 (yes)
DEF5	Does the SDG prioritization relate to negative impacts in addition to positive ones?	0 (no) 1 (yes)
<i>Measurement and analysis</i>		
MEA1	Does the report provide information on <i>qualitative</i> targets related to the SDGs?	0 (no) 1 (yes)
MEA2	Does the report provide information on <i>quantitative</i> targets related to the SDGs?	0 (no) 1 (yes)
MEA3	Does the report provide information on <i>specific (past or ongoing) actions</i> related to achieving the SDGs?	0 (no) 1 (yes)
MEA4	Does the report provide <i>qualitative</i> information on the outcome of these actions?	0 (no) 1 (yes)
MEA5	Does the report provide <i>quantitative</i> information on the outcome of these actions?	0 (no) 1 (yes)
MEA6	Does the report provide information on <i>future actions</i> related to achieving the SDGs?	0 (no) 1 (yes)

Subramaniam et al. (2020) sampled Australia’s top 150 publicly listed companies to explore their SDG disclosures. The disclosures are assessed based on three dimensions: ‘awareness and commitment to SDGs, integration of SDGs into strategic materiality analysis, and performance and measurement’ (Subramaniam et al., 2020, p. 10). Specifically, whether companies mention SDGs in their reports, whether companies mention SDGs in their CEO’s messages, whether board members show expertise in SDGs or sustainability, and SDG prioritisation have been used to indicate corporate awareness and commitment to SDGs. This may illustrate corporate interests in the SDGs presented by the ASX150 and corporate commitments towards SDGs to their stakeholders. While prioritisation of SDGs may not mean that companies have translated their commitments into practical actions, the extent of corporate integration of SDGs into strategic decisions assesses the effectiveness of corporate engagement and the impacts on SDGs. This is followed by designing a set of quantitative and qualitative indicators to measure corporate SDG performance.

Table 2.5 Extracted from Subramaniam et al. (2020, p. 10)

Awareness and commitment to SDGs	(i)	mention or acknowledgement of the SDGs in either the annual or sustainability reports
	(ii)	mention of the SDGs in the Chair or CEO message

	<ul style="list-style-type: none"> (iii) presence of board members with SDG or sustainability knowledge and expertise (iv) presence of a sustainability committee (v) prioritisation of the SDGs
Integration of SDGs into strategic materiality analysis	<p>the integration of SDGs in</p> <ul style="list-style-type: none"> (i) materiality assessment (ii) SDGs linked to materiality assessment at the SDG target level (iii) SDGs incorporated to stakeholder engagement processes (iv) SDGs linked to sustainability initiatives (v) SDGs linked to sustainability initiatives at the SDG target level
SDG and performance measurement	<p>the extent to which SDGs are mapped or linked with</p> <ul style="list-style-type: none"> (i) business performance indicators (ii) business performance indicators at the SDG target level (iii) historic performance data demonstrating SDG contribution A score of one was listed if an item was met.

Zhang and Guan (2020) comprehensively explore the variation of SDG disclosures in Chinese companies during the period 2018-2020. This study analyses the disclosure levels from four levels: the format of SDG information in CSR reports, SDG goals that have been mentioned in reports, corporate SDG management, and corporate SDG reports and communication. Each category includes numerous indicators to help understand the extent of SDG disclosures. This study suggests that an increasing number of companies report SDG information; however, the quantity of information disclosed is limited.

Table 2.6 Extracted from Zhang and Guan (2020)

The format of SDG information in CSR reports	<ul style="list-style-type: none"> Using SDGs as a reference framework Mentioning SDGs in CEO message or Background analysis Information related to how companies implement SDGs SDG benchmark table
SDG goals that have been mentioned in reports	Goal 1 -Goal 17
Corporate SDG management	Incorporating SDGs into corporate development strategies

	<ul style="list-style-type: none"> Developing a plan for SDGs Adopting SDG indicators Developing SDG benchmark and research Building SDG evaluation systems Summarising SDG progress regularly
Corporate SDG reports and communication	<ul style="list-style-type: none"> Planning and implementing SDG-related projects SDG seminars Releasing SDG-related videos Disclosing SDG information Releasing SDG initiatives Using other methods for SDG communication

These studies classify SDG disclosures into multiple levels, such as understanding, commitments, prioritisation, management, and performance measurement, which may provide a relatively holistic overview of corporate SDG disclosures. However, SDGs are treated as a fashionable activity. Companies may simply relabel and rebrand their existing CSR activities as their SDG activities and initiatives, rather than implementing concrete and significant changes in their reports. That is, ‘it is frequent to link existing activities to SDGs-icons as a “design-complement” (Heras-Saizarbitoria et al., 2022, p. 9). Therefore, there remains scope in evaluating whether companies consider SDGs as an aspirational agenda or have substantive changes in their behaviours—in other words, what specific SDG-related actions have companies taken.’⁸

Hence, a core challenge for corporate SDG disclosures is identifying quantifiable and consistent measures for performance.

⁸ Other methods in measuring SDG disclosures.

Some other methods, such as semi-structure interviews, and survey, have also been adopted to observe the substance of SDG disclosures (Curtó-Pagès, F et al, 2021; Hummel, 2019; Topple et al., 2017). Topple et al. (2017) analyses multinational enterprises' sustainability practice against the corporate sustainable assessment framework during the period 2013-2016 (the introduction of SDGs). Applying semi-structured interviews and documental analysis, this study find that international sustainability-related guidelines significantly affect corporate sustainability practice. This study claims materiality analysis as a useful tool to address sustainability challenge. Similar method is also adopted by Ike et al. (2019). Going a step further, Curtó-Pagès et al. (2021) rely on longitudinal analysis to see SDG disclosure variations under the context of Spanish companies during the period of 2016-2019.

2.8 Chapter conclusion

This chapter has sought to provide an overview of the nature and extent of SDG disclosures. Following the publication of Agenda 2030 by the UN in 2015, the Chinese government signed on to it in 2016 (Communist Party of China, 2016a; UNDP, 2016) and subsequently, has ‘actively implemented’ Agenda 2030 within the 13th National FYP (Communist Party of China, 2016a, p. 136; National Development and Reform Commission, 2016, p. 153). The government also subsequently introduced a series of guidelines to encourage Chinese companies, particularly large Chinese companies, to make SDG disclosures.

Although no standardised corporate SDG disclosure framework has been created in China, since 2016, an increasing number of Chinese companies have begun to voluntarily make SDG disclosures in their reports. SDG disclosures may serve as a corporate response to the China’s national strategies. SDG disclosures may also indicate a company’s financial capability and competence in SDG activities. SDG disclosures may also serve as a basis for an ongoing stakeholder dialogue.

A review of existing studies and guidelines developed by standard setters suggests that no consensus has been reached on what constitutes an SDG disclosure. Specifically, prior studies have widely investigated the adoption of SDG disclosures (Girón et al., 2020; Lassala et al., 2021) and SDG word frequency in reports (Izzo et al., 2020a; van der Waal & Thijssens, 2020). However, a lack of strategic information may have limited implications for actions that the company has undertaken. Some studies utilised existing reporting frameworks (Ordonez-Ponce & Khare, 2021), such as GRI and SASB; however, these frameworks are partially aligned with SDGs and may not be applicable to Chinese circumstances. Some studies adopt a limited number of metrics, which may limit a more profound understanding of corporate involvement in SDGs. Therefore, what remains unclear in relation to the growing number of Chinese companies that make SDG disclosures is the extent and nature of these disclosures.

Chapter 3 Literature review on the variability of SDG disclosures

3.1 Chapter introduction

In Chapter 2, this thesis observed that the number of Chinese companies that provide SDG disclosures in their CSR reports has increased continuously; however, there are no existing requirements for companies to make SDG disclosures in their corporate reports in China. Therefore, the focus of this thesis is the growing use of SDG disclosures in Chinese companies.

Chapter 3 proceeds with establishing the relevance of issues to corporate SDG disclosures or SDG performance, which may finally affect corporate responses through disclosures. Specifically, Section 3.2 starts with an overview of the variation in SDG disclosures in Chinese companies. This is followed by a review of factors that may predict SDG disclosures in Section 3.3, which will serve as the basis for proxy selection in Chapter 5. Section 3.4 provides an overview of model selection in SDG disclosure studies. Section 3.5 concludes this chapter.

3.2 The variability of sustainability disclosures in Chinese companies

Over the past several decades, a growing number of corporate ESG disclosure regulations and guidelines have been introduced in China, which has subsequently led to an increase in disclosures among Chinese companies (Shen et al., 2023). The number of Chinese companies that make ESG disclosures rose from 638 in 2012 to 1429 in 2021 (Shen et al., 2023).

After introducing SDGs, Zhang and Guan (2020) suggest that the number of Chinese companies that make SDG disclosures in their CSR reports has increased continuously; rising from 174 companies in 2018, to 254 companies in 2020. However, the level of SDG disclosures widely varied across sampled companies. By the end of 2020, nearly 30% of the sampled companies disclosed that they have incorporated SDGs into their business strategies (Zhang & Guan, 2020). Only 13.81% of the sampled companies disclosed their plans for implementing SDGs, while around 10% of sampled companies disclosed their SDG progress and around 3% of the sampled companies developed their SDG assessment mechanisms (Zhang & Guan, 2020). Overall, the variation in the level of SDG disclosures was observed. However, what remains unclear is the extent and nature of these disclosures in Chinese companies, and the underlying motivations of their SDG disclosures. The research questions are as follows:

- What is the extent of SDG disclosures in Chinese companies, following China's implementation of Agenda 2030?
- What factors may predict the level of SDG disclosures in Chinese companies based on the gradient boosting machine?

3.3 Literature review of factors that may predict corporate SDG disclosures

Section 3.3.1 reviews contextual factors that may predict corporate SDG disclosures. Section 3.3.2 reviews literature on corporate factors that may influence voluntary SDG disclosures, such as financial performance (García-Meca & Martínez-Ferrero, 2021; Muhmad & Muhamad, 2020), corporate governance (Buniamin et al., 2020; Pizzi et al., 2021c), and ESG performance (Hummel, 2019).

3.3.1 Contextual factor—government control

Contextual focus in developing countries has been highlighted as an emerging factor in understanding social and environmental accounting and forming 'unique reporting practices that are still evolving' (Qian et al., 2021, p. 20). Given the authoritative setting in China, the imprinting process of ideological prioritisation exerted by the CPC, and the unique government-company relations, the Chinese government arguably maintains absolute control of national development and economic growth (Chua et al., 2024; Lin, 2011; Zhu & Zhang, 2015) and can substantially and directly influence business practices (Situ et al., 2018). Therefore, even though China's economy is becoming more market-oriented, Chinese government still largely influences corporate decision-making.

The government has introduced tangible and intangible control mechanisms to influence corporate behaviours. For example, fulfilling corporate social responsibility has been entrenched within national ideological prioritisation over the past decade (Qian et al., 2024). As discussed in Section 2.3, the 13th National FYP highlights the important role of state-owned and private companies in achieving national goals, such as poverty alleviation (Bin-Feng et al., 2023), building a community with a shared future of mankind, promoting patriotism (Jia Guo Qing Huai), as well as developing an ecological civilisation and low-carbon development. The government contextualises institutional approaches to a sustainable development agenda and outlines what companies can do to help China transition to a green economy (Parsa et al., 2021). Situ et al. (2021) and Qian et al. (2024)⁹ also argue that China's political ideologies significantly influence corporate environmental disclosures, highlighting that disclosure serves as the basis for Chinese companies to show their obligation to fulfil the 'socially legitimate request', even if there is no mandatory requirement.

To operationalise these aspirational and intangible ideologies across companies, the government subsequently introduced different tools to exercise government control. Government shareholding is widely explored (Zeng et al., 2012; Zhao et al., 2018). However, Situ et al. (2020) suggest that governmental controls can also be exercised through direct and indirect actions, such as regulatory influence and incentivising influence. Therefore, moving beyond shareholding control, this study explores the government control from *1. Regulatory, 2. Shareholding, 3. Political, and 4. Incentivising influence* aspects.

3.3.1.1 Regulatory influence

The first type of government control. Regulatory influence refers to laws, regulations, guidelines, strategies, political ideologies, and other instruments introduced to achieve the objectives of the government and regulatory bodies (OECD, 2021). In China, it is argued that the state-initiated ideology has symbolic power in influencing corporate ESG practices, which reflects 'an imprinting process of ideological prioritisation' exerted by the CPC (Chua et al., 2024; Zhu et al., 2021, p. 4). To operationalise it, an increasing number of CSR regulations have been introduced to guide corporate CSR disclosures, such as, China Corporate Social Responsibility Reporting Guidelines (See Section 2.3). Prior studies suggest mixed results between government regulations and corporate ESG disclosures (Bohnett et al., 2022; Chauvey et al., 2015; Wang & Sarkis, 2017), due to the nature and types of regulations. As suggested, they may have varying requirements in terms of their disclosure content, the level of compliance, or disclosure formats (Coenen et al., 2021, p. 9).

⁹ Situ et al. (2021) investigate the impact of China's political ideology, 'Building a Harmonious Society', on corporate environmental disclosures in China, where 'Building a Harmonious Society' was emphasised during the 12th National FYP being effective. It finds that China's political ideology significantly affects corporate environmental disclosures, as disclosure becomes a habitus, deeply embedded by companies, even if there is no mandatory requirement. Similar results have been found in Qian et al. (2024), which employ a field study to investigate the impact of the socialist ideology in China on corporate environmental disclosures.

Early studies explore the relationship between mandatory regulatory influence and the level of CSR disclosures (Chauvey et al., 2015; Wang & Sarkis, 2017). However, inconsistent results have been found due to different research focuses, such as disclosure adoption, quality or levels. Wang et al. (2017) investigate the impact of mandatory CSR regulations on CSR disclosure quality in the Chinese context between 2009 and 2012. CSR disclosure quality score is based on the RKS rating score. This study suggests that government regulations positively affect the quality of CSR disclosures. Situ et al. (2020) explore the regulatory influence of the SSE mandatory regulation on corporate environmental disclosures. It suggests that mandatory regulatory influence positively affects corporate decisions to make environmental disclosures; however, it does not improve the extent of disclosures. Similar results have been found in Chauvey et al. (2015), indicating that mandatory regulations have a more symbolic impact on corporate disclosures. One of the reasons for explaining this may be the nature of the stock exchange guideline, which tends to have a generic focus, rather than emphasising corporate disclosure quality and quantity (Qian et al., 2024).

Government regulatory influence is not only limited to mandatory regulations; existing studies explored corporate-level reporting regulations and international corporate reporting guidelines. Situ et al. (2020) investigate whether the Chinese government may leverage its political power through the implementation of voluntary regulations, such as the Guidelines on Environmental Information Disclosure by companies listed on the SSE and the Chinese CSR Report Preparation Guide. This study suggests that these state-issued regulations significantly influence corporate compliance with these environmental disclosures; however, they do not affect the extent of environmental disclosures. Surveying the top 50 Nigerian companies between 2016 and 2018, Erin et al. (2022) suggest that a lack of regulatory requirements on SDGs may lead to a lower level of SDG disclosures in these companies. SDG disclosure is measured using the content analysis. The framework is based on PwC, GRI, and IIRC frameworks. This study argues that a lack of regulatory reporting frameworks may lead to weaker enforcement and commitments to their SDG disclosures.

In China, SDGs have been incorporated into national strategies, encouraging Chinese companies to engage in and report on it; however, prior studies have complex theorisation, and heterogenous research designs, which increase the uncertainty of the relationship between national strategies and corporate SDG disclosures. Specifically, these studies are limited to SDG 14 and 15 biodiversity issues, overlooking other social and environmental areas prioritised in the 13th National FYP. Some companies treat it as a tool for international expansion, while some companies consider it political activities, which are non-market tools. This may affect the way how companies respond to the BRI. For example, as suggested, BRI may serve as a political instrument for infrastructure development, economic growth and environmental accountability in China, Bohnett et al. (2022) find that companies participating in BRI tend to disclose limited SDG information in Chinese companies. SDG disclosures are measured based on SDG 14 and 15 to capture corporate biodiversity practices. One of the explanations for this is

the lack of reporting guidelines for corporate biodiversity disclosures and the existing directives that are vague in forcing companies to balance their environmental accountability and their economic development (Bohnett et al., 2022). Zhu et al. (2022) consider the BRI as a non-market strategy for the government to encourage companies to participate in corporate internationalisation. Sampling 261 Chinese companies between 2010 and 2017, this study finds that disclosures can be seen as corporate responses to institutional pressure to show their compliance and participation in non-market activities.

Overall, mixed results have been found between regulatory influence and corporate ESG disclosures, due to the types and requirements of the regulations and political ideologies. While existing studies widely explored the impact of corporate-level regulations and guidelines, there remains scope for understanding whether voluntary regulations and central government initiatives may have regulatory influence on SDG disclosures in Chinese companies.

3.3.1.2 Governmental shareholding

The second type of government control. Government shareholding refers to the state's 'control over the company through ownership arrangements' (Tihanyi et al., 2019, p. 2297). Situ et al. (2021, p. 10) argue that China is unique for 'its linkages and coordination within a party-government-military-economy regime'. Despite transitioning to a more market-oriented economy, government control still substantially influences national development, economic growth, and business practices (Lin, 2011; Situ & Tilt, 2018). Companies with government shareholding may need to engage with the government's political agenda, even if it may preclude their goals of profit maximisation (Ervits, 2021a). Thereby, SOEs in China serve as 'a medium of social and economic policy', such as involving in sustainable development activities and disclosure practices (Ervits, 2021a, p. 7).

Existing studies widely explore the relationship between governmental shareholding and CSR disclosures in the Chinese context, by adopting SOEs as the proxy for governmental shareholding. These studies argue that government shareholding significantly affects corporate issuance of CSR reports to maintain their political legitimacy; however, the government appears to play a weaker role in influencing the quality and substantiveness of CSR disclosures. It is observed that existing studies are limited to exploring whether the company is a SOE or non-SOE. Different-level government control and the nature of SOEs may also affect how companies respond to CSR disclosures. Zeng et al. (2012) argue that SOEs are more likely to make environmental disclosures, as their survival is closely aligned with the government. The 'encouragement' from the government to disclose environmental information appears to be more legitimate for SOEs. SOEs tend to be more cooperative in doing what the government encourages them to do, while disclosure could serve as a basis for manifesting corporate cooperation with the government and showing their fulfilment of social responsibility (Zeng et al., 2012). This study also suggests that SOEs do not influence the substantiveness of environmental disclosures, as a higher level of disclosures may add additional costs to companies, without helping

companies obtain extra credibility from the government. Similarly, Chu et al. (2013) argue that SOEs make less GHG disclosures than private companies, with GHG data collected from annual reports. Since these companies are under the government's protection, disclosure may not be necessary for companies to demonstrate their credibility or support towards government strategies. Marquis and Qian (2014) suggest that Chinese companies that are partially controlled or fully owned by the government may disclose less CSR information, as their ties to the state improve their credibility in the eyes of the government. Interestingly, this study also suggests that SOEs are more likely to issue stand-alone CSR reports to maintain their political legitimacy, as the government controls critical resources¹⁰ that may be important to corporate survival. Similar results have been found in Pattern et al. (2015), which suggest that stand-alone CSR disclosures are a legitimate tool for SOEs. Zhao and Patten (2016) explore the impacts of governmental influence, such as coercive, normative, and mimetic pressure, on the level of social and environmental disclosures in SOEs. This study suggests that social and environmental disclosures may serve as a tool for SOEs to improve their corporate image. Similarly, Situ et al. (2018) argue that governmental shareholding appears to drive the issuance of CSR reports rather than the variation in disclosures, arguing that these companies appear to be more trustworthy in the eyes of the government and thereby may be easier to get support from the government. Disclosure may be treated as a responsive tool for Chinese companies to indicate their support for official signals and national political aims. In return, companies may be more competitive in obtaining official resources. However, Zheng et al. (2014) suggest that state-owned companies are marginally and positively associated with the level of CSR disclosures, due to the potential pressure and punishment for failing to engage in sustainable development.

Under a top-down governance structure, local governments, as law-takers, also play a pivotal role in following the central government's rules and political aims. However, in parallel with sustainable development at local government levels, they may maintain some degree of independence to pursue their economic development and may respond differently (Qian et al., 2022). Therefore, the complex and heterogeneous results may be due to the heterogeneous measurement of government shareholding (SOE vs non-SOE, central SOE vs regional SOE). For example, Luo et al. (2017) argue that SOEs located in regions are less likely to produce their CSR reports, as these companies tend to prioritise their short-term GDP. However, this relationship appears to be inverse when regional governments emphasise ESG activities like the central government does. Companies tend to have a higher level of quality CSR disclosures. Similar results have been found in Parsa et al. (2021) and Situ et al. (2018). These studies explain that different levels of governments may have different attitudes towards sustainable development; specifically, the central government prioritises national strategies for sustainable development, while local governments may treat the economy as its priority.

¹⁰ The critical resources may refer to licenses, government subsidies, tax deductions and other approvals.

It is worth noting that the State Council of People's Republic of China (2015) introduced Opinions of the State Council on the Development of An Economy of Mixed-Ownership of State-Run Enterprise, which classifies SOEs into the following 5 levels: Commercial SOEs in competitive industries fields, Commercial SOEs in natural monopoly industries, Financial SOEs, Cultural SOEs, and Public interest SOEs (See Table 3.1). Therefore, there remains a gap in understanding whether the nature of SOEs may affect the level of SDG disclosures.

Table 3.1 Extracted from the State Council of People's Republic of China (2015, pp. 1-10)

Types	Areas	Role	Oversight
Commercial SOEs in competitive industries fields	‘National security and national economy’	‘For State-run enterprises in the category of commerce whose main business is in important industries or crucial fields concerning State security or lifelines of the national economy and that mainly assume tasks of major exclusive projects, the share controlling status of State-owned capital in them shall be maintained, and acquisition of shares in them by non-State-owned capital shall be supported’.	company-level oversight, particularly the allocation of governmental resources
Commercial SOEs in natural monopoly industries	‘Military industry, grid power, petroleum and petrochemical, telecommunications, coal, civil aviation, and shipping’	‘They follow the strategy of ‘separation of government functions from enterprise management, separation of the government's socioeconomic management function from its function as the owner of State-owned assets, and government regulation as the major contents shall be applied, and separation of networks of State-owned infrastructure with the. nature of natural monopoly from their operation with the nature of market competition’.	Directly overseen by the government
Financial institutions	‘Insurance companies, banks’	Companies that focus on dealing with monetary transactions, such as investments, loans, and deposits	Centrally controlled state ownership
Cultural SOEs	‘Social media’	‘They are an important force for the prosperity and development of the cultural industry, the promotion of socialist core values, and the construction of advanced socialist culture’. ‘State-owned cultural enterprises are divided into news information services, content creation and production, communication channels, investment operations, and comprehensive operations. They are treated differently, and policies are implemented in different categories to ensure the preservation and appreciation of state-owned cultural assets and enhance the core competitiveness of state-owned cultural enterprises.’	Centrally controlled regional controlled
Public interest SOEs	‘Water supply, electric power generation, and gas and heat supply, public transportation and public facilities, taxonomic guidance’	‘The focus of public interest SOEs is on protecting people's livelihood, serving the society, providing public products and services efficiently, and being a good provider of high-quality services.’	The quality and efficiency of public goods and services oversight from the public

3.3.1.3 Political influence

As a **third type of government control**, political influence is widely considered ‘strong government ties on companies, via network connections’ (Marquis & Qian, 2014; Wang et al., 2018a, p. 1186). Considering the political-financial relationship, the political connection is instrumental for companies

to acquire preferential treatments, finally affecting corporate value and performance in China (Qian & Chen, 2021). Unlike government ownership, corporate political connections serve as indirect tools for the government to influence the dynamics of state-company relationships through connections (Qian & Chen, 2021). Disclosure may indicate that companies are strategically building a stable relationship with the government, to acquire government support and maintain corporate political legitimacy (Li et al., 2021, p. 81). Political influence can be classified into political connection individual embedding and party committee embedding two levels.

Political connection individual embedding can be classified into: *1.* directorship role undertaken by former government officials, and *2.* representative embeddedness which involves being a member of the National People's Congress (NPC) or Chinese People's Political Consultative Conference (CPPCC) (L. Wang et al., 2021; Wang et al., 2020).

- 1. Directorship roles undertaken by former government officials** tend to have communication channels and networking connections with existing government officials or regulatory authorities. This may offer opportunities for companies to acquire cross-sector information and scarce resources controlled by the government (Goldman et al., 2013). Having former government officials in the senior management team is a common strategy for companies to establish connections with the government to acquire value and administrative resources that are crucial for corporate growth (Zhang et al., 2017; Qin and Zhang, 2019). According to Li and Guo (2022), 642 independent director positions were occupied by government officials in 2012, accounting for 8.45% of all posts, by exploring 2012 annual reports by Chinese listed companies. However, on 19 Oct 2013, the Organization Department of the Central Committee of the Communist Party of China issued Document No. 18: Opinions on Further Regulation on Party and Political Leaders and Cadres Working Part-Time (Holding Offices) in Enterprises (Xinhua, 2013b). The document states that incumbent government officials and those with a current position but not officially retired from public office are prohibited from working in the enterprises. Officials who have resigned or retired from the public office must strictly follow the rules before joining an enterprise. For example, former government officials within three years of resigning from government positions should not work in an enterprise.
- 2. Representative embeddedness** is typically obtained through government recognition, as a reward for corporate growth and economic contributions (Zhang et al., 2023). NPC and CPPCC are National Two Sessions (全国两会), held annually. These meetings aim to serve as a basis for 'mediating policy differences between the CPC and various parts of Chinese society' (Marquis & Qian, 2014, p. 132). Prior studies suggest that these sessions provide more symbolic benefits, as they can only offer suggestions and are forbidden from intervening the CPC and the government's judicial activities (Marquis & Qian, 2014). Companies will be granted members of NPC or CPPCC

if they have good performance or have made a significant contribution to the local economy (Du & Zhou, 2010).

Diverse methods have been employed to observe corporate political connections with the government, by Western and Chinese papers (See Table 3.2). They can be classified into different types of political connections (government official vs members of NPC or CPPCC) and different levels of political connection (i.e. CEO, chairman, management, or directors). Their impacts on corporate performance and corporate disclosures also vary widely due to the measurements of political connections. However, these studies lack descriptions regarding what are government-related jobs, and which positions can be seen as government officials, which are crucial criteria in exploring corporate political connections.

1. **Western papers:** Wu, Wu and Rui (2012) investigate the CEO and chairman's political connections; however, this study does not provide clear boundaries for what constitutes political connections. Marquis and Qian (2014) argue that CEOs have political connections if they are government officers, particularly officers of central or regional governments; or were members of CPPCC. Cheng et al. (2017) suggest that the level of political connection can be measured based on the number of party members in the corporate executive board. Wang et al. (2018, p. 1189) argue that political connection refers to 'a company's directors, supervisors, or senior officers being (1) a government official, (2) a member of the NPC, or (3) a member of CPPCC'. A similar method has been utilised in Rauf et al. (2021). Li et al. (2021, p. 83) look at 'whether a company's manager has or had any experience as government officials or being a member of NPC or CPPCC'. Yu and Chi (2021) explore the political background of the CEO or chairman—whether they have worked in any government entity or military force. This study further classifies political connections at state/central or regional levels. Going a step further, Dang et al. (2022) classify political connections into symbolic and material connections. A material connection refers to the situation where 'the CEO or board chair is a government official, military officer or the communist party secretary in an administrative jurisdiction' (Dang et al., 2022, p. 7). A company is viewed as having symbolic connections if its CEO or the board chair is a member of NPC or CPPCC (Dang et al., 2022, p. 7). L. Wang et al. (2021, p. 370) collect information from the 'Profile of Directors and senior manager' section in annual reports to explore whether a board member is '(1) a former government official, (2) a current or former member of the CPC, (3) a current or former member of the CPPCC, or (4) a current or former military official'. This study investigates 'the number of politically connected directors on board, divided by board size' (Wang et al., 2021, p. 370). To investigate the impacts of different types of politicians as directors, this study classifies connected directors into local and home connected. Locally connected directors refer to directors connected at or below the provincial level, while home-connected directors refer to directors from the home province.
2. **Chinese papers:** Du and Zhou (2010, p. 127) classify political connections into: '1) whether the senior management team was or is currently in the party committee; 2) whether the senior

management team is or was a member of the Party committee, NPC or CPPCC; 3) the level of position and workplace’. Luo and Liu (2019) focus on the political connections of CEOs and managers—whether they previously worked or currently work in central or regional governments, courts, and procuratorates, and whether they are or were members of NPC or CPPCC. A similar method has been adopted by Zhu and Liu (2022). Additionally, Shen (2023) explores the political connections of independent directors, by investigating the presence of independent directors working in the public office and the number of independent directors working in public office out of the total number of independent directors. Yang et al. (2023) argue that a company has a political connection if its CEO or executives were government officials or members of NPC or CPPCC. This study also investigates the level of political connections, including district, municipal, provincial, and national levels. Similarly, Zhang et al. (2023) look at the chairman or general managers to observe whether they served as government officials in governments, party committees, the National People’s Congress or the Chinese People’s Political Consultative Conference, procuratorates and courts. This study also investigates whether a company’s chairman or general manager is a member of NPC or CPPCC.

Table 3.2 A summary of political connection measurements/

Studies from Western journals	Wu et al. (2012, p. 702)	‘POLCON is a dummy variable that equals one if the firm has a politically connected CEO and/or chairman, and zero otherwise’
	Marquis and Qian (2014, p. 135)	‘CEO NPC/ CPPCC membership is a dummy variable equal to 1 if the CEO was a member of the CPCC, and 0 otherwise. CEO as government official is a dummy variable used to measure bureaucratic embeddedness. It equals 1 when the CEO is or has been an officer of the central or regional governments, and it is 0 otherwise’
	Cheng et al. (2017, p. 215)	‘This paper uses the number of party members in the company’s executive board to measure the degree of political connection’.
	Wang et al. (2018b)	if its directors, supervisors, or senior officers (1) is or was a government official, (National Congress of the Chinese Communist Party) a member of the NPC, and or (3) a member of CPPCC.
	Y. Li, Wong, and Miao (2021, p.7)	‘Political connection is a dummy variable, it is equal to 1 if the chairman is/was (1) a government official, (2.) a member of NPC/ CPPCC, and 0 otherwise’. ‘Political promotion is a dummy variable that is equal to 1 if chairman is politically promoted to be (1) officials in a government position at the same or higher level; (National Congress of the Chinese Communist Party) chairman/Party Secretary of the parent company (3) chairman/Party Secretary of the parent company in larger SOEs’.

	Rauf et al. (2021, p. 6)	'whether a company's manager has or had any experience as government officials or being a member of NPC or CPPCC.'
	Wang et al. (2021, p. 370)	'information from 'Profile of Directors and senior manager' section in annual reports to explore whether a board member is (1) a former government official, (2.) a current or former member of the CPC, (3) a current or former member of the CPPCC, or (4) a current or former military official.'
	Yu and Chi (2021)	whether CEO or chairman have worked in any government entity or military force.
	Dang et al. (2022, p. 7)	if 'the CEO or board chair is a government official, military officer or the communist party secretary in an administrative jurisdiction'
Studies from Chinese journals	Du and Zhou (2010, p. 127)	<ul style="list-style-type: none"> • GOVDUM: Whether senior management was or is currently in the party committee (Including the Discipline Inspection Commission), government, standing bodies of the National People's Congress or the Chinese People's Political Consultative Conference, courts, and procuratorate. (1993, Public Servant law) • DBWYDUM: Whether senior management is or was a member of Party Committee, NPC or CPPCC • GOVMAX: the level of position and workplace.
	Luo and Liu (2019)	Whether CEO or manager was worked or currently works in central or regional governments, courts and procuratorates. Whether CEO or manager is or was a member of NPC or CPPCC
	Zhu and Liu (2022)	<ul style="list-style-type: none"> • Whether CEO or manager 1.) was government officials or (National Congress of the Chinese Communist Party) is or was a member of NPC or CPPCC
	Shen (2023)	<ul style="list-style-type: none"> • Whether independent directors worked in public office • The number of independent directors worked in public office out of the total number independent directors.
	Yang et al. (2023)	<ul style="list-style-type: none"> • Political connection: whether CEO or executive was a government official or a member of NPC or CPPCC • The level of political connection: district, municipal, provincial, and national levels.
	Zhang et al. (2023, p. 23)	'If the chairman or general manager of the enterprise (1) served as a government official in government, party committees, the National People's Congress or the Chinese People's Political Consultative Conference, procuratorates and courts. Or (2.) a member of the National People's Congress or the Chinese People's Political Consultative Conference'

Studies on the relationship between corporate political connections and CSR disclosures are also emerging (Qian & Chen, 2021). Inconsistent results have been found between political connections and disclosures, due to different sample sizes, political connection measurements, and diverse disclosure measurements. The under-theorisation of the concept of political connection individual embedding may increase the difficulty in drawing a generalisable conclusion. Marquis and Qian (2014) explore whether different types of dependency on the government have impacts on CSR disclosures in China, by sampling 1600 listed companies from 2006 to 2009. It argues that companies that COE has membership in political councils are more likely to issue CSR reports. By issuing CSR reports, companies and executives may be able to maintain their legitimacy in the eyes of the government and thereby obtain access to governmental resources. This study also suggests that bureaucratic embeddedness has significant impacts on the substantiveness of CSR disclosures. Bureaucratic embeddedness is proxied by whether the CEO has experience as a government official. Similar results have been found in Wang et al. (2018a). By sampling A-share companies listed on SSE and SZSE between 2008 and 2014, this study, again, suggests companies whose executives have political connections tend to issue their CSR reports. Similar results have been found in Cheng et al. (2017), which explore the impact of senior managers' political connections on the level of environmental disclosures. Senior managers' political connections aim to measure the degree of political connections, by measuring the number of party members among the executive board. This study finds a positive relationship between them; however, this study argues that the quality may be sacrificed, because politically embedded members may mask negative parts to some extent. Therefore, to maintain corporate political legitimacy, companies tend to disclose more environmental information. However, this does not improve disclosure quality or the substantiveness of disclosures.

It is interesting to note that Li et al. (2021) find a negative relationship between top executives' political connections and the level of environmental disclosures. This study argues that politically connected companies can more easily acquire government support and first-hand information from the government and therefore, may be less motivated to disclose more environmental information. Qian and Chen (2021) argue that corporate political connections may positively affect corporate environmental disclosures or only provide limited environmental information to the public to show their compliance with environmental policies. Disclosure may simply serve as a tool to demonstrate their good corporate image, and enjoy a relaxed regulatory environment and government support. Some studies found that political promotion may be related to corporate contributions to the local economy rather than sustainability efforts. As a consequence, companies with executives holding political positions tend to emphasise economic development in line with the goals of their political careers, as they will be promoted based on their economic contributions rather than their sustainability efforts (Li et al., 2021, p. 79).

Party committee embedding is another form of political influence, strongly exercised by the CPC. A party committee, as a formal group with Chinese characteristics, is integrated into micro-level organisations to bridge the gap between enforcing external regulations and developing an internal culture of social responsibility (Yin et al., 2019). Party committees follow the ‘two-way entry, cross-appointment, and pre-discussion’ approach to enhance its involvement in corporate governance (Zhang, 2023, p. 2). The power of party organisations is different due to the nature and scale of the companies. According to Bi (2021), SOEs with more than 100 party members can be approved by their highest-level party organisation (i.e. parental companies) to set up their own party committees. Company Act and Constitution of the Communist Party of China (中国共产党党章) state that the secretary of the party committee in SOEs will also be the board members, particularly the CEO or directors, in the company (National Congress of the Chinese Communist Party, 2022), through which the CPC can exercise direct control over the company. For example, party committees can monitor the daily business operation of SOEs. Business decisions must receive approval from the committee before proceeding, while investment decisions need to be approved by every party committee member (Bi, 2021). Furthermore, SOEs and their party committees are also ‘annually evaluated by central or regional governments in terms of their ability to maintain and create value for state-owned assets’ (Bi, 2021, pp. 1-2).

Various methods have been used to assess the impact of the party organisation embedding on companies. However, existing studies are limited to investigating the impact of party committees on CSR engagement rather than disclosures. Yu et al. (2019) investigate the impact of party organisation embedding on CSR engagement, by sampling companies listed on SSE and SZSE between 2008 and 2016. This study defines that a company has party organisation embedding if its board members, supervisors, or management teams have a party organisation background. This study finds that CSR engagement is positively influenced when members of party organizations also serve on the board of directors or supervisory teams. This is likely because the board of directors plays a key role in corporate decision-making, while supervisory teams are crucial in overseeing the implementation and management of CSR activities. However, this relationship becomes insignificant if the party organisation is embedded at the managerial level, as management teams are responsible for implementing the decision. Similarly, Zhang et al. (2021) found that party committees have a significant impact on other social responsibility activities, such as charitable donations. A party committee is determined by whether any of the board members, CEO, and management members currently also serve as the secretary of the party committee. This study finds that party committee governance has significantly positive impacts on corporate charitable donations in heavy-pollution industries. However, these studies are limited to investigating CSR engagement rather than disclosures.

Some studies investigate individual-level CPC impacts on companies. Li et al. (2020) explore whether any members of the CPC are directors, supervisors or senior executives. Li and Cheng (2021) argue

that corporate party embeddedness is measured based on whether members of CPC are directors, supervisors or senior managers. This study also investigates whether the secretary of the CPC serves as chairman. Gao and Feng (2022) classify party impacts into two levels: (1) the ratio of party members of its directors, executives, or supervisors, to the total number of these senior management team members; (2) whether the chairman is the secretary of the party committee. Following Li et al. (2020), S. Wang et al. (2023) utilise a multidimensional variable to assess the impact of the CPC on enterprise value, including CPCALL, CPCBOD, and CPCSM.

- CPCALL investigates whether at least one member of the director, supervisor, or senior manager is a CPC member.
- CPCBOD assesses the ratio of CPC members to the board of directors.
- CPCSM refers to the overlap ratio of CPC members to the supervisory board.

However, limited studies have explored the impact of party committees on the level of ESG disclosures. Yu et al. (2019) suggest that companies with a party committee tend to have higher environmental disclosures. The result becomes even more pronounced when the party committee is at the board or supervisory team level. However, insignificant results have been found between the party committee at the manager level and environmental disclosures. This indicates that corporate positions of members in the party committee may affect corporate decision-making. Yu et al. (2019) argue that a party committee may encourage companies to participate in environmental activities and comply with environmental strategies to fulfil their social responsibilities. Disclosure may serve as a means for companies to maintain their legitimacy. However, this study is limited to environmentally sensitive industries in China between 2008 and 2016. There remains scope for understanding whether the party committee may also affect the level of SDG disclosures in Chinese companies.

3.3.1.4 Incentivising influence

As **a fourth type of government control**, incentivising influence refers to tax and fee incentives, grants, government subsidies, and other funding offered by the government to encourage corporate participation (Wu et al., 2022). The Chinese government substantially controls critical resources, such as natural resources, bank support, and taxes, which may be linked directly to corporate survival and competitive positions (Marquis & Qian, 2014; Zhu et al., 2022). Failing to adhere to the government's interests may pose additional risks to companies, such as tax penalties, lawsuits, fines, or restrictions on their strategic options (Chiu et al., 2020). These government incentivising strategies may have different requirements or expectations, such as more transparency in their ESG reports, requirements on the issuance of their green bonds, or their progress update, which will affect the way that the company reacts.

In 2015, the government introduced the Chinese green bond market, where green bonds are designed as financial instruments to incentivize companies to participate in sustainable development projects

(Hoang et al., 2022). As a government-led market, green bonds become an instrument for companies to raise capital for their green projects, which may support the goals within the 13th National FYP, such as ecological civilization. The plan also states that ‘we will establish a green finance system, develop green credit and bonds, and launch green development funds’ (p.140). Regulatory authorities have requirements for the issuance of green bonds, rather than disclosure requirements, while only green investors and institutional bodies require companies to report their allocation of funds in their projects (Zhou & Cui, 2019). It is argued that green bonds are aligned with most investors’ investment preferences who often perceive them as ‘medium and long-term financial products based on low-risk environmental protection projects’ (Khurram et al., 2023, p. 5). Companies are willing to participate in these projects, as they may improve corporate long-term value (Khurram et al., 2023). Disclosures are considered a response to investors and other bodies who require the disclosures of project fund allocation to corporate fulfilment of social responsibility (Zhou & Cui, 2019). This will develop a friendly environment for companies to operate and strengthen their social reputation and corporate value in the long run.

Prior studies find that disclosures on green bond issuance can ‘release a green signal to the market, leading other issuers in the same industry to take more actions that are conducive to environmental protection’ and significantly reduce bond financing costs (Wu et al., 2022, p. 3). Hoang et al. (2022) argue that the amount of green bonds issued by companies is positively associated with ESG disclosures in all dimensions. Particularly, the influence is higher for environmental disclosures, while social disclosures are marginally and significantly influenced. Similar results have been found in Wang and Liu (2024).

Prior research has also explored the relationship between governmental financial incentives and sustainability disclosures in the Chinese context, with financial incentives proxied by government-led green financing actions (Wang et al., 2019; Xing et al., 2021). However, these studies fail to provide consistent results on the impact of government-led green financing on disclosures. This may be due to the types of government incentivising policies and the multidimensionality of the ESG disclosures. In 2012, the Chinese government introduced a green credit policy to mitigate carbon pollution in high-polluting companies. Specifically, it requires banks to provide financial support to companies that are evaluated as credible green performers. Wang et al. (2019), however, suggest that no significant relationship has been found between green credit and corporate environmental disclosures, indicating that, in China, environmental disclosures may not provide useful signals to acquire green credit in the market. This study samples 320 Chinese companies from high-polluting industries listed on SSE between 2008 and 2016. Similarly, Xing et al. (2021) also test the effectiveness of China’s green credit policy, by sampling 1086 Chinese companies from the manufacturing industry between 2012 and 2017. This study, again, does not find significant relationships between environmental disclosure quality and green loan financing. However, this study, going a step further, explores the connection between

different types of environmental disclosures, such as soft and hard environmental disclosures, and corporate green loans. It finds that soft environmental disclosures negatively affect corporate access to green loans, while hard environmental disclosures marginally positively affect corporate access towards green loans. Therefore, disclosures may be related to corporate access to green loans. Situ et al. (2018) investigate whether government incentivising influence may affect corporate environmental disclosures. Incentivising influence is proxied by the total government grants that companies have received from the government and the total amount of grants that a company received to invest in environmental activities. This study suggests that companies that receive grants and funding tend to disclose more environmental information, as such information is necessary to obtain support from the government. This relationship becomes weaker when companies disclose environmental information in a stand-alone CSR report, indicating that annual reports are still the primary tool for companies to communicate their financial and non-financial performance with the government.

Overall, mixed results have been found between governmental incentivising influence and corporate sustainable development disclosures, due to the requirements of different types of government grants and incentive policies. There remains scope in investigating the impact of government financial incentivising on SDG disclosures in Chinese companies.

3.3.2 Corporate characteristics on the level of SDG disclosures

SDG disclosure literature mainly focuses on developed-country contexts and has explored corporate factors, such as financial performance, corporate governance and ESG performance, on the extent of SDG disclosures (García-Meca & Martínez-Ferrero, 2021; Pizzi et al., 2021c). However, the research designs, such as variable selection, the measurement and multidimensional nature of disclosures, and contexts, are rather ‘complex and heterogenous’ (Frost et al., 2023; Velte et al., 2020). These uncertainties provide opportunities to investigate corporate characteristics that may predict the level of SDG disclosures in Chinese companies.

3.3.2.1 Financial performance

Economic performance may indicate a company’s capacity and financial conditions for involving in sustainable development activities (Ullmann, 1985). There has been an increasing number of studies seeking to observe the associations between financial performance and voluntary sustainability disclosures (García-Meca & Martínez-Ferrero, 2021; Muhmad & Muhamad, 2020). These studies argue that if financial performance is connected to the extent of SDG disclosures, participating in sustainable development activities may be considered a way to pursue inclusive corporate economic growth (Muhmad & Muhamad, 2020). That is, companies can ‘do well by doing good’ (Porter & Kramer, 2011; SDG Compass, 2015). Disclosure may offer a clear message to stakeholders about a company’s capacity and financial condition for dealing with sustainability risks (García-Meca & Martínez-Ferrero, 2021). Disclosure may also indicate whether companies have sufficient financial resources to develop their

competence and technological capacity to pursue sustainable business growth (EY Global, 2017). This may bring more investment and revenue-generating opportunities for companies. (Izzo, Ciaburri, et al., 2020; Rosati & Faria, 2019b).

Prior studies find mixed results between financial performance and corporate SDG disclosures. Specifically, existing studies widely adopt traditional financial ratios, often overlooking the impacts of raw accounting numbers and market-based accounting indicators on corporate SDG disclosures. Davern et al. (2019) also argue that non-GAAP financial information is more value-relevant than traditional measures and this study encourages to use more forward-looking financial performance measurements. Therefore, this thesis argues that variable selection may affect how we can draw a generalisable conclusion about the ‘relationship’ between financial performance and corporate disclosures. Consistent with Frost et al. (2023), this study considers corporate financial performance from: accounting-based and market-based aspects.

Accounting-based financial performance and SDG disclosures

Prior research widely employs traditional accounting-based financial performance indicators to investigate the associations between financial performance and SDG disclosures. However, mixed results have been found, which may be due to variable selection, different contexts, and the heterogeneity of disclosure measurements. This thesis observed that a limited number of accounting-based indicators have been used, which are heavily on measuring corporate profitability and expenses based on ratios. The same variable also has several measurements/ proxies, which may also affect how these proxies can explain the variation in the level of corporate disclosures. Additionally, prior studies have adopted raw accounting numbers in predicting corporate bankruptcy in the Chinese context, which is often overlooked in ESG disclosure literature. However, unscaled numbers may lead to heteroscedasticity within the database, and they do not take into account firm size and other corporate characteristics. Jones (2017) suggests that this issue may not be sensitive when using the gradient boosting model. Therefore, there remains a gap in understanding the impact of raw accounting numbers on the level of SDG disclosures.

Raw accounting numbers from financial statements. Existing SDG disclosure literature widely employs expense and cost proxies, and it is interesting to note that positive relationships have been found between financial costs and ESG disclosures across different contexts, research periods, different measurements of disclosures. However, existing studies are limited to interest expense and R&D costs, which can only present a small proportion of corporate expenses and costs. Therefore, there remains scope for understanding whether different types of costs and financial outputs may also affect the level of SDG disclosures. Rosati and Faria (2019b) find positive impacts of R&D costs on corporate SDG disclosures. Disclosure may indicate the relationship between ‘research and development advancements and their contribution to innovation and technology’ and subsequently demonstrate corporate capability

in committing to SDG activities (Rosati & Faria, 2019b). Rosati and Faria (2019b) conduct a preliminary analysis of 408 largest companies worldwide. Nechita et al. (2020) investigate the impact of financial performance on the variation of SDG disclosures based on the disclosures of chemical companies in Eastern Europe, where SDG disclosures are mandated. This study also suggests that companies with a higher level of interest expense and R&D costs tend to disclose more SDG information.

Raw accounting numbers from balance sheets, income statements, or cash flow statements have strong predictive values in predicting corporate performance, such as corporate bankruptcy and accounting fraud. These studies offer insights into the value of raw accounting numbers in investors' decision-making. Given that this study will also use the gradient boosting machine, there remains a gap in understanding whether these accounting numbers may also affect the level of SDG disclosures, and uncertainty in understanding the predictive values of these raw numbers and the scaled proxies for measuring the same variables. Besides interest expense and R&D costs, Nechita et al. (2020) also employ other intangibles and property, plant, and equipment as the financial performance proxy. This study suggests that companies with a higher level of property, plant and equipment (PPE) tend to disclose a lower level of SDG information. Sun et al. (2021) selected 28 raw accounting numbers from balance sheets, income statements and cash flow statements, including current assets, cash, short-term investment, current liabilities, debt in current liabilities, total assets, available-for-sale securities, held-to-maturity securities, long-term equity securities, total liabilities, long-term debt, debt in current liabilities, accounts receivables, inventory, PPE, construction in progress, project material, account payable (trade), surplus reserve, retained earnings, income tax payable, sales, cost of goods sold, net income, financial expense, total income taxes, cash received from absorbing investments, and proceeds from issuance of bonds. This study also includes 10 financial ratios. Based on the XGboost model, this study observed that net income, surplus reserve, current liabilities, and PPE are the top 5 strongest variables in detecting accounting fraud across sampled companies, with relative variable important scores (RVIs) at 5.946, 4.381, 4.174, 4.165, and 4.029. This is followed by long-term debts, inventory, accounting payables, retained earnings and debt in current liabilities, with RVIs ranging from 4.029 to 3.625. This study also finds that raw accounting numbers have greater predictive values than financial ratios. However, inconsistent results have been found in Jones (2017). Jones (2017) adopts a number of raw accounting numbers, financial ratios and market-based financial performance indicators to predict corporate bankruptcy in the US context. Raw accounting numbers include total assets, net operating cash flows, total debts, net investing cash flows, net financing cash flows, annual growth in net income, and total bank debt. This study suggests that total revenue has the strongest predictive value across all raw accounting numbers in predicting corporate bankruptcy in US companies, with a RVI score of 33.02. This is followed by total assets, net operating cash flows, total debts, net investing cash flows, and net financing cash flows, with RVIs at 27.79, 22.56, 20.23, 12.88, and 11.95 respectively. It is also observed

that market-based financial indicators and financial ratios also have strong predictive values in predicting corporate bankruptcy.

Financial ratios. Prior studies widely investigate the impact of return on asset (ROA) and return on equity (ROE) on the level of disclosures; however, mixed results have been found, which may be due to different research contexts or disclosure measurements. For example, European companies are mandated to make SDG disclosures, which is largely voluntary in other countries. Rosati and Faria (2019b) suggest that ROE is marginally and negatively associated with SDG disclosures, which is inconsistent with the argument of showing their willingness, resources and flexibility in participating in SDGs and making SDG disclosures. However, Martínez-Ferrero and García-Meca (2020) investigate the adoption of SDG disclosure among 365 European companies from 2016 to 2017 and suggest ROA does not significantly affect the adoption of SDG disclosures. No significant relationship has been found between ROA and SDG disclosures in Nechita et al. (2020) in the European context, Martínez-Ferrero and García-Meca (2020) by sampling 1723 European companies, and Pizzi et al. (2021c) in the Italian context. Moving to the Asian and African continents, Girón et al. (2020) examine whether similar results can be found in the context of a developing country. This study suggests that profit margin and total assets ratios have positive impacts on the level of SDG disclosures, while leverage, ROE, and ROA do not influence SDG disclosure levels. This study is limited to a single-year analysis of 2017. Existing Chinese literature is limited to explaining the extent of CSR disclosures and adopting a limited number of traditional financial performance indicators, such as ROA and ROE. Meng et al. (2013) explore whether the relationship between ROE and environmental disclosures may vary due to the transition of disclosures from voluntary to mandatory. Sampling all manufacturing companies listed on SSE and SZSE in 2006, a positive relationship between economic performance and voluntary social and environmental disclosures has been found. Disclosure is considered an accountability mechanism to improve corporate communications with investors and other stakeholders, showing stakeholders that companies have sufficient resources to invest in environmental activities. With the mandatory environmental reporting standards introduced in China, this relationship has become negative. Environmental disclosures may serve as a legitimate tool when companies face political and social pressure from the government. Zheng et al. (2014) adopt ROA and leverage as financial performance proxies to explore the behaviour of CSR disclosures of a sample of 5334 companies. It suggests that ROA positively affects a company's decision on CSR disclosures between 2008 and 2010, while leverage has an inverse relationship with CSR disclosure levels. However, Liu and Anbumozhi (2009) do not find significant relationships between the level of environmental disclosures and ROE and leverage in Chinese companies, with disclosure data collected from the GRI database.

Some studies also employ ratios concerning cash flow, risk, and capital structure to evaluate different types of accounting-based ratios on corporate ESG disclosure disclosures. Mixed results have been found, due to variable selection, model selection, research contexts, and the heterogeneity of ESG

disclosure measurements. Some studies argue that disclosure may indicate that companies that have sufficient financial resources intend to involve in voluntary sustainability disclosures (Jones et al. 2007), while some studies argue that disclosure may demonstrate corporate willingness to be more legitimate to the government's expectation and their willingness of showing their financial capacity to invest in sustainable development activities (Naheed et al., 2021). Jones et al. (2007) explore the impact of financial performance on sustainability disclosures, with financial indicators proxied by cash position, cash flow return, working capital, profitability and turnover. It finds that financial performance has positive impacts on sustainability disclosures. However, market-based indicators, such as abnormal return and volatility, do not have impacts on sustainability disclosures. Nechita et al. (2020) suggest that companies with a higher level of cash flow tend to disclose more SDG information, with cash flow measured by the ratio of the sum of different types of cash flow activities to total assets. This study also finds that companies with a higher level of leverage tend to disclose a lower level of SDG information. Martínez-Ferrero and García-Meca (2020) do not find significant associations between financial performance and SDG disclosures based on a sample of 1723 European companies, with financial performance proxied by leverage and sustainability risk. Similar results have been found Naheed et al. (2021), with cash flow and leverage using the same measurements in Nechita et al. (2020). Naheed et al. (2021) adopt cash flow and sales revenue ratios as proxies for financial performance. This study suggests that companies with better cash flow and sales revenue tend to disclose more CSR information. This study is limited to 409 non-financial Chinese companies listed on SSE and SZSE from 2009 to 2016. Therefore, we observed that financial performance has significant influence on CSR and environmental disclosures in China. Jones and Wright (2018) quantify a number of financial performance indicators, including traditional financial ratios and market-based indicators. Indicators include current ratio, debt to cash, cash to total assets, EBIT to total assets, free cash flow per share, intangibles to total assets, price to book, profit margin, retained earnings to total assets, ROA, ROE, sales to total assets and working capital ratio. Sampling ASX300 companies between 2008 and 2012, this study finds a positive relationship between financial performance and CSV activities.

Overall, inconsistent findings have been identified between accounting-based financial performance indicators and SDG disclosures, which may be due to differences in variables selected and various contexts. Chinese literature has explored a limited number of traditional accounting-based variables, such as ROA, ROE, and leverage, which offers scope for further exploring whether other financial performance variables may also affect corporate disclosures in the Chinese context. There also remains scope for exploring the value relevance of traditional accounting numbers, such as elements in balance sheets and profit and loss statements, is deteriorating and capital market information becomes more relevant (Davern et al., 2019; Lev, 2018).

Market-based financial performance and SDG disclosures

Given that more forward-looking financial information is available from alternative sources, an increasing number of capital market indicators are gradually becoming useful and have been found to have stronger power in predicting corporate future performance (Lev, 2018). Recent studies show an increase in utilising market-based variables.

Tobin's Q is the most frequently adopted market performance proxy. It is shown to have a positive relationship with SDG disclosures (Sasanti et al., 2020). Using a sample of 16 Indonesian listed companies, Sasanti et al. (2020) argue a positive relationship between SDG disclosures and company values, although this relationship is weakened by the types of industry involved, particularly for high-profile companies. Similar results have been found in García-Meca and Martínez-Ferrero (2021), which investigate whether SDG disclosures are symbolic or substantive among European companies, based on the association between firm value (proxied by Tobin's Q). This study does not find a significant relationship between them; however, this relationship becomes significant in environmentally sensitive industries in the European context in 2015-2016. This indicates that SDG disclosures are symbolic and add little value to stakeholders; however, in environmentally sensitive industries, SDG disclosures are a value-enhancing tool when companies face high social scrutiny, or the company's stakeholders recognise the importance of SDGs. These studies are limited to the *adoption* of SDG disclosures, which motivate studies to test whether these factors may also affect the *level* of SDG disclosures. However, Phan et al. (2020) investigate the potential values that non-financial information disclosures (measured by SDG disclosures) could bring from a stakeholder perspective after the introduction of the EU policy (Directive 95/2004/EU)-- a mandatory non-financial disclosure framework. This study does not find a significant association between SDG disclosures and financial/ market performance. In Martínez-Ferrero and García-Meca (2020), Tobin Q does not appear to influence SDG disclosure levels. This study is limited to a single-year analysis of 2017.

Market value is another widely-used market-based financial performance indicator. Consistently, positive results have been found between market value and corporate disclosures (Weber, 2014; Naheed et al., 2021). Disclosure may also indicate a company's capacity to adhere to the government's political aims and their capability in involving in sustainable development, which may also aim to increase the confidence of relevant stakeholders in companies (Weber, 2014; Zheng et al., 2014). Weber (2014) adopts market return as a proxy for financial performance and investigates ESG disclosures of the top 100 green Chinese companies. This study again suggests a positive relationship between market return and ESG disclosure levels. Similarly, Naheed et al. (2021) adopt market-to-book ratio as the proxy for market-based financial performance and again suggest a positive relationship with CSR disclosure. This study focuses on non-financial listed Chinese companies between 2009 and 2016. Jones and Wright (2018) also include some market-based indicators, such as cash flow returns, stock price returns, earnings per share (EPS), growth in EPS, and earnings management proxy. Overall, a small number of

Chinese studies have investigated the impacts of market-based financial performance on sustainability disclosures.

Corporate risk. Several studies explore the impacts of corporate risks on the level of corporate disclosures; however, mixed results have been found, due to financial risk proxy selection and research contexts. Disclosure, on the one hand, may indicate their healthy financial condition, efficiency, or liquidity in participating in SDGs and making SDG disclosures, while disclosure, on the other hand, may indicate their capability in managing risks. Gambetta et al. (2021) examine whether corporate risk portfolios have impacts on Spanish financial institutions' SDG disclosures. Risk portfolios are measured based on Tier 1 ratio (show capital adequacy), provisions for loan losses, management efficiency ratio, profitability ratio, ROA, loans, liquidity ratio, and interest-bearing liabilities. This study attempts to test the impacts of financial stability and capital adequacy on the variation of SDG disclosures. It suggests that the Tier 1 capital ratio is positively associated with SDG disclosure levels, indicating that companies with healthier financial conditions tend to disclose more SDG information. Management efficiency is measured by the cost-to-income ratio and this study finds that companies with a higher level of management efficiency tend to disclose more SDG information, indicating that when companies are less efficient at managing costs or have a complex business situation, they tend to incorporate SDGs into their business strategies and disclose SDG information to show their capacity in managing SDG activities. Similarly, this study also suggests that companies that face higher market risks tend to disclose more SDG information. However, other variables do not appear to significantly affect the level of SDG disclosures. Phan et al. (2020) find that companies with higher risk (beta or larger size) tend to improve their SDG disclosures in European companies. Similarly, Izzo, Ciaburri, et al. (2020) investigate the level of SDG disclosures in Italian companies and suggest that companies with higher liquidity, like those that are highly capitalised tend to make SDG disclosures.

S3.3.2.2 Corporate governance

Corporate governance is a broad term, which includes multiple components, such as relationships with stakeholders, frameworks, decision-making and responsibility (ASIC, 2019). An increasing number of institutional and individual stakeholders recognise ESG as an important component of opportunities and risks that a company may face (Lagasio & Cucari, 2019; Limkriangkrai et al., 2017). Corporate management plays an important role in balancing between economic and sustainability goals, and between individual and communal expectations, to ensure corporate operations and long-term survival (Khan, 2010). Disclosure may serve as a basis for showing the effectiveness of corporate governance in incorporating sustainable development into business operations; it may also be viewed as a strategy for corporate governance to communicate with stakeholders, to show that companies have fulfilled their expectations (Khan, 2010). Recently, analysts recommend that companies participate in SDG activities and make SDG disclosures as investors are increasingly interested in SDGs (García-Sánchez et al., 2020). Disclosure may indicate the effectiveness of corporate management in pursuing SDG activities,

which may allow investors to identify material risks and opportunities related to sustainable development and thereby may influence investors' final capital allocation decisions (Adams et al., 2020; KPMG, 2018a; Rosati & Faria, 2019a, 2019b). However, the priority and preferences may vary across different owners regarding their risks, growth, and performance (Khlif et al., 2017, p. 377). The complex research designs and 'under-theorisation of concepts' lead to the uncertainty of the role of governance-related predictors on disclosures and difficulty in drawing generalisable conclusions (Frost et al., 2023, p. 1117; Velte et al., 2020).

In China, corporate governance is unique and complex, due to the coexistence of SOEs and non-SOEs, which may affect corporate responses to economic development, social and political objectives. Both types of companies have concentrated ownership structure, either one controlling or multiple large shareholders (Jiang et al., 2018). As suggested by Jiang and Kim (2024), large shareholders, particularly in SOEs, are not the owners of the company and they may act based on their own interests and political backgrounds. It is also interesting to note that institutional shareholders in Chinese companies are minority shareholders, while board of directors in Chinese companies is primarily 'a formality than an effective governance body' (Jiang & Kim, 2024, p. 5). Therefore, this section intends to understand the dynamics of corporate governance in Chinese companies due to their board of director and large shareholders' scale, interests, and social identities. To address these challenges, this study will explore two broad types of corporate governance: board characteristics and ownership concentration.

Board composition

Prior studies widely explore the impacts of board characteristics and corporate SDG disclosures; however, mixed results have been found. One possible explanation for this may be that the roles of governance variables in different jurisdictions are ambiguous and uncertain. China is unique for its concentrated shareholding governance structure and board characteristics may be a less effective governance body. This led to exploring the predictive power of board characteristics on SDG disclosures in Chinese companies.

Board size is a widely used corporate governance proxy in ESG disclosure studies (Buniamin et al., 2020; Pizzi et al., 2021c). It is argued that a larger board size may overcome skill and experience discrepancies in corporate governance (Dyduch & Krasodomska, 2017). Disclosure may serve as a tool to show their compliance with ESG regulations and guidelines to acquire legitimacy and a better corporate image among non-financial Chinese companies. However, as discussed above, Chinese companies are unique for their concentrated governance structure and boards are often less effective in corporate governance. Therefore, there remains scope for understanding whether board size can still have strong predictive values in predicting the variation in the level of SDG disclosures in Chinese companies. By sampling 219 Malaysian listed companies, Buniamin et al. (2020) argue that board size is positively associated with the adoption of SDG disclosures, as larger boards tend to have broader

experience with different backgrounds and workload distributions. They can offer expertise and capability to allow companies to fulfil their sustainable development objectives and cover the expectations of various stakeholders. Disclosure may serve as a tool to show their compliance with sustainable development objectives. European Union introduced Directive 2014/95/EU, which requires large public interest companies to report how they respond to social and environmental challenges. Pizzi et al. (2021c) explore the impact of this EU Directive by investigating the impacts of corporate characteristics on SDG disclosures. Sampling 153 Italian Public Interest Entities, it also finds that board of directors is positively associated with the level of SDG disclosures, showing that if they are interested in SDGs, they can affect the decision to incorporate SDGs with business strategies and their reports. Similar results have been found in Chinese studies. Shahab and Ye (2018) and Xu et al. (2021) argue that board size also has positive impacts on CSR disclosure quality. Xu et al. (2021) argue that a larger board size may indicate a greater diversity of the board members and better governance structure in representing corporate owners. Disclosure may indicate their better management of CSR activities.

Independent directors on board (board independence) are a large component of corporate governance, and they may offer more effective monitoring and greater control over the management team (M. K. Khan et al., 2021). Mixed results have been found between independent directors and corporate disclosures, which may be due to research contexts and the measurements of disclosures. On the one hand, independent directors tend to have financial and non-financial expertise, and they may request more information to evaluate corporate performance. Disclosure may serve as a tool to meet their information needs, particularly in developing countries, where have relatively weaker investor protection systems. Martínez-Ferrero and García-Meca (2020) suggest a higher level of board independence increases the probability of SDG disclosures, by sampling European companies between 2016 and 2017. This indicates that stronger corporate management will encourage companies to participate in SDG activities and attempt to take into account various stakeholders' interests. Similar results are also found in Pizzi et al. (2021c). Buniamin et al. (2020) move to the Malaysian context in 2020, with corporate governance proxied by board independence and the number of women directors on board. However, board independence is not significantly associated with the level of SDG disclosures, which may be due to the research context and the data source (annual reports). Moving to Chinese literature, M. K. Khan et al. (2021) investigate the impacts of internal corporate governance on the level of environmental disclosures and environmental accountability between 2010 and 2019. Again, this study suggests a positive relationship between board independence and corporate environmental disclosures. This study argues that independent directors attempt to motivate companies to seek transparency of environmental information and emphasise stakeholders' interests, by balancing companies' behaviours and societal values. Similar results have been found in Shahab and Ye (2018) and Guping et al. (2020).

Other characteristics of board composition, such as board diversity, board meetings, and CSR committees, have also been used in prior studies. Martínez-Ferrero and García-Meca (2020) also adopt CEO non-duality and board meetings as proxies for internal governance strength to investigate their impacts on SDG disclosures in the European context. A positive relationship has been found between CEO non-duality and SDG disclosures. CEO non-duality refers to the separation of roles of the CEO and the board chairman, which may influence corporate effectiveness. However, no significant relationship has been found between the number of board meetings and SDG disclosures. Similarly, Pizzi et al. (2021c) adopt CSR committee, the number of women directors on board, and board meetings as the proxies for corporate governance.

Ownership concentration

Investment horizontal heterogeneity, such as its scale and the roles of investors, across Chinese companies may also improve our understanding of corporate governance (Jiang & Kim, 2024). Unlike Western countries, China is unique in its large shareholder monitoring, either one single or multiple shareholders. These investors appear to have clear and strong incentives to pursue their self-interests. It is also interesting to note that, unlike Western companies, institutional investors in Chinese companies are minority shareholders, however they have rights in active trading to pursue their interests, rather than direct corporate engagement (Jiang et al., 2018). Prior studies widely focus on unique state ownership, however, the ownership structure has dramatically changed after companies went public and companies are comprised of different types of stakeholders with diverse economic and social expectations in China (Zheng et al., 2014). There remains a gap in understanding whether corporate ownership concentration may influence their corporate SDG disclosures.

Institutional shareholding. Prior studies found mixed results between different types of institutional shareholding and corporate SDG disclosures. Institutional shareholders may have financial and nonfinancial know-how and different information needs to evaluate corporate performance and make investment decisions. Khlif et al. (2017) conduct a meta-analysis to identify the relationship between ownership concentration and voluntary disclosures, suggesting that institutional ownership has positive impacts on voluntary disclosures. The impact is more evident in developing countries, where capital markets are less developed, and investor protection is weaker. Larger institutional shareholders have better expertise and expect the company to meet their information needs to protect themselves. Larger institutional investors also have voting power in taking 'corrective actions'. Disclosure may serve as the basis for meeting the information needs of stakeholders before raising new or additional funds for the company. García-Sánchez et al. (2020), going a step further, investigate the impact of different types of institutional ownership on the adoption of SDG disclosures by sampling 989 international companies worldwide. Institutional ownership is proxied by foreign institutional investors, crossholding, governmental ownership, ownership owned by financial institutions, pension funds and endowments, and other ownership. This study suggests that companies with a higher level of ownership owned by

foreign investors tend to disclose SDG information. This study argues that foreign investors do not have direct relations with companies and this type of investors may be more willing to use internationally recognised guidelines to evaluate corporate sustainable development performance and make their investment decisions. This study also suggests that companies with a higher level of pension funds and ownership owned by other investors may disclose SDG information, as they may be more concerned with Agenda 2030. However, other types of investors, such as crossholding, financial institutions, and the government, do not influence corporate SDG disclosures. Overall, limited evidence has been provided on the impact of ownership concentration on the level of SDG disclosures.

Insider shareholding. Mixed results have been found between insider shareholding and corporate ESG disclosures, due to the measurements of ESG disclosures. It is observed that insider shareholding positively affects the issuance of CSR reports (Zheng et al., 2014) and negatively affect the level/extent of ESG disclosures (Shahab & Ye, 2018; Zheng et al., 2014). One of the possible reasons for explaining this is that managerial shareholders may face greater pressure from the government and society to do ‘politically correct’ activities, by showing their compliance with ESG activities in their reports. Zheng et al. (2014) investigate the impact of ownership concentration from state shareholding, and managerial block-holders. It is found that the existence of state block-holders and managerial block-holders positively affects the issuance of CSR reports. It is argued that this type of companies tends to prioritise the interests of the society, as they are under government and social pressure to participate in ESG activities (Zheng et al., 2014). However, when companies tend to have more shares owned by state or managerial block-holders, the extent of CSR disclosures tends to be negative. This may indicate that ownership concentration negatively affects the extent of CSR disclosures. Similar results have been in Shahab and Ye (2018). Shahab and Ye (2018) test the relationship between corporate governance and CSR disclosures using a neo-institutional theory and longitudinal analysis of CSR disclosure levels. With a sample of 1166 non-financial companies from 2008 to 2015, this study shows that block ownership tends to negatively affect the level of CSR disclosures.

Shares owned by top shareholders. Prior studies, again, found mixed results between shareholders owned by top shareholders and corporate ESG disclosures, due to different research contexts and variable selection. Frost et al. (2023) employ shares owned by the top 5, 10 and 25 shareholders to evaluate corporate ownership concentration on corporate carbon reporting in Australian companies. This study suggests that ownership concentration has strong predictive power on carbon reporting levels. The RVI scores of these proxies range from 41.97 to 30.17, indicating that variable selection may affect the overall prediction results. This study also finds that a higher level of shares owned by the top 5 shareholders increases the probability of a lower level of carbon reporting in Australian companies. The impact may differ in the Chinese context, as larger shareholders may be the government. Li and Zhang (2010) investigate the impacts of ownership structure on CSR, finding that results appear to be different for state-owned and non-state-owned companies. For state-owned companies, shareholding owned by

the largest stakeholder tends to have positive impacts on corporate CSR practices, as the government tends to be the largest stakeholders who have incentives to contribute to their social and environmental stability. For non-state-owned companies, ownership concentration tends to have negative impacts on corporate CSR practice, due to the government's weaker political interference and the diverse interests of multiple large stakeholders.

Other types of ownership concentration. Rosati and Faria (2019a) investigate whether corporate management influences how SDGs can be incorporated into business strategies. This study suggests that companies with female directors owning more shares and companies with a younger board of directors tend to adopt SDG disclosure. However, it again examines the adoption of SDG disclosures rather than the level of SDG disclosures.

S3.3.2.3 ESG performance

ESG performance is defined as 'the aggregate negative or positive bottom line of economic, environmental and social impacts of an entity against a defined baseline' (Büyüközkan & Karabulut, 2018, pp. 253-254). It is argued that divergence may occur between ESG disclosure practices and their ESG performance, as a corporate response to sustainable development may be underpinned by regulatory force, institutional influence, implementation mechanism, financial performance, corporate governance, as well as political and community interests (Frost et al., 2023). In China, companies face pressure from both institutional parties and stakeholders to incorporate sustainable development into businesses. Disclosures may serve as a tool for companies to show how they legitimise their performance in the eyes of the government, to maintain their legitimacy (X. Meng et al., 2014). Due to the multidimensional nature of ESG, companies may respond differently to stakeholder pressure and regulatory demands. Despite an increase in ESG regulations, current ESG performance rating and disclosures remain largely unstandardised and customised across companies. While existing databases have introduced metrics for different observable ESG activities for corporate analysis, there remains a gap in understanding whether different types of ESG performance may predict the level of SDG disclosures.

After the introduction of Agenda 2030, SDGs may be viewed as a novel method for companies to engage in macro-level sustainable development, through improving business strategies (Ghosh & Rajan, 2019; Ordonez-Ponce & Khare, 2021), enhancing business operations (van der Waal et al., 2021), clarifying missing elements in existing standards (P. A. Khan et al., 2021) and providing a comprehensive framework for sustainability assessments (Lu et al., 2021; Topple et al., 2017). Disclosures may reflect the extent to which companies have incorporated SDGs into businesses, which may be useful for investors and other stakeholders to assess corporate SDG performance (Gutterman, 2020; Izzo, Ciaburri, et al., 2020). Schonherr et al. (2017) theoretically argue an association between

CSR and SDGs, emphasising the importance of SDGs as a reference for promoting sustainable development.

Empirical evidence also suggests a positive relationship between sustainability performance and SDG disclosures (Hummel, 2019), indicating that companies with better sustainability performance are more likely to disseminate their SDG information. This study samples listed companies from the STOXX Europe-600 index between 2015 and 2018. Nicolo' et al. (2024) fill the gap of to what extent corporate SDG disclosures may reflect their sustainability performance. This study explores 635 companies from 45 countries during the period between 2016 and 2020 and suggests a positive relationship between sustainability performance and SDG disclosures. This indicates that companies tend to demonstrate their superior performance to investors and other stakeholders. Meanwhile, disclosure costs and litigation risks can be mitigated. Overall, a business case for the relationship between sustainability performance and SDG disclosures has been established; however, limited studies have investigated this relationship.

Prior studies also provide insights into ESG performance-disclosure relationships, focusing on understanding the impact of different types of observable ESG performance. X. Meng et al. (2014) argue a non-linear relationship between environmental performance and disclosures. Extremely good environmental performers tend to disclose more environmental information to differentiate themselves from poor performers, in exchange for the Chinese government's incentives, such as 'environmental awards, refinancing convenience, tradable permits, and a green credit policy' (X. Meng et al., 2014, p. 359). This study also suggests that extremely poor performers also have a higher level of environmental disclosures to legitimise their performance with soft claims. Similarly, Kartal et al. (2024) suggest that companies with greater ESG participation tend to have a higher level of ESG disclosures, by sampling 102 publicly traded Turkish companies in 2022. This study argues that disclosure may serve as the basis for showing corporate management structure and their effective management. Sampling Australian listed companies from 2002 to 2016, Li et al. (2022) analyse corporate sustainability disclosures from the following factors: tone of optimism, certainty, and clarity and readability, as textual characteristics may affect the way that stakeholders interpret corporate sustainability disclosures. Sustainability performance data is collected from Thomson Reuters Asset4 ratings. This study suggests a positive relationship between 'textual characteristics' of sustainability disclosures and sustainability performance, indicating that 'how things are said' contributes to the 'inclusive and inconsistent' research on this relationship (Li et al., 2022, p. 2).

Given that the Chinese government faces significant challenges in addressing environmental crises, an increasing number of Chinese studies focus on environmental performance-disclosure relationships. A U-shape relationship has also been found by Li et al. (2017). This study sampled 475 Chinese listed companies between 2013 and 2014. Environmental performance data is collected from the Toxic

Release Inventory (TRI) database, Council on Economic Priorities, and KLD. This study finds that companies with relatively poor environmental performance tend to disclose more environmental information due to social, political, and stakeholder pressure. Disclosure may serve as the basis for showing corporate capability of involving in environmental activities. This study also finds that companies with superior environmental performance are more likely to disclose more environmental performance. Disclosure may demonstrate their superior performance among their competitors. It is also interesting to note that this study argues that when companies have a fair level of environmental performance, they seem to lack motivation and pressure to involve in ESG activities. Some studies also argue that a positive relationship between environmental performance and disclosures, suggesting that companies disclose environmental information to inform outsiders of their superior performance and their performance related to the commitment to the environment (Al-Tuwaijri et al., 2004; Clarkson et al., 2008; Deswanto & Siregar, 2018; Mahmood et al., 2017). Specifically, Clarkson et al. (2008) use univariate analysis and suggest a positive relationship between environmental performance and disclosures, while Mahmood et al. (2017) use the Ordinary least-squares (OLS) model. However, a negative result has been found in Aboagye-Otchere et al. (2019), arguing that mining and manufacturing companies in Ghana adopt environmental disclosures to legitimise their environmental activities. This study employs a content analysis of 25 companies' annual reports, based on a disclosure framework created by Wiseman (1982) and environmental performance information is based on AKOBEN environmental performance rating scores. Lemma et al. (2019) suggest that companies with a higher level of carbon emissions are more likely to provide a higher level of carbon disclosures. Disclosure may demonstrate corporate fulfilment of social and environmental obligations, mitigating corporate risks.

While existing studies broadly investigate ESG performance-disclosure or environmental performance-disclosure relationships, Frost et al. (2023), going a step further, argue that ESG performance can be classified into different types of observable activities. This study subsequently explores whether ESG performance may predict the level of carbon reporting in top Australian companies, with ESG performance measured from the following four aspects: overall ESG performance, environmental performance, social performance, and governance performance. Given that this study employs the gradient boosting machine, the RVI results suggest that ESG final score has the strongest out-of-sample predictive power in predicting the level of carbon reporting (RVI =100). The remaining ESG performance metrics also have strong predictive power: environmental scores (RVI =56.42), ESG combined scores (RVI=44.10), governance final scores (RVI=37.54), and social scores (RVI=23.75).

Overall, mixed results have been found between ESG performance and disclosures. This can be explained by the fact that prior studies have adopted different rating systems to capture corporate sustainability performance, which may have missing data, lack consistency and standardisation, and the multidimensional nature of sustainability performance (Zhang & Zhang, 2023). These may affect how

the rating system evaluates corporate ESG performance, finally affecting users who employ the rating system. In addition, incomplete ESG disclosure is also a common issue (Kartal et al., 2024). Companies commonly follow different regulations and guidelines to provide their ESG disclosures, such as SASB, GRI, and IR. However, these guidelines are customised in formatting and metrics (Kartal et al., 2024), which will affect the outcome of drawing a generalisable comparison. While SDG disclosures have been emerging in China, what remains unclear is whether ESG performance may also influence the extent of SDG disclosures in China due to institutional and stakeholder pressure.

3.4 Model selection

Model selection is an important step for selecting the optimal model that best fits data and research purposes. Models can be classified into parametric and non-parametric models (machine learning methods). Parametric models, such as linear and logistic regression models, are most widely adopted to test the relationships between variables. They rely on traditional hypothesis testing and need to satisfy all underlying assumptions, such as linearity, no multicollinearity, independent and normal distribution, and the number of variables being smaller than the sample size (Hansen, 2010). However, recent studies argue that traditional analysis fails to provide out-of-sample accurate analysis, has strict rules and assumptions, and is unable to handle complex interaction and data issues. Accounting researchers and practitioners encourage the application of machine learning methods in accounting, which will be able to provide better data-driven predictions of ESG impacts (Zhang & Zhang, 2023).

3.4.1 Literature review of model selection

Existing SDG disclosure literature uses either logistic or linear regression models to investigate potential factors that may influence the extent of SDG disclosures (See Table 3.3). Multivariate and OLS regression models are the widely adopted linear regression models (Gambetta et al., 2021; Nechita et al., 2020). With the highest adjusted R squared at nearly 60%, the model in Nechita et al. (2020) can explain approximately 60% of the variation in SDG disclosures, while Gambetta et al. (2021) obtain a lower adjusted R squared value of 0.1693, which does not seem to be an appropriate model for explaining the variation of SDG disclosures. Similarly, with the highest value of pseudo-R at 0.507 and the lowest value at 0.03, logistic models also have limited power in explaining the variation of SDG disclosures (Girón et al., 2020; Martínez-Ferrero & García-Meca, 2020). One possible reason could be that these models have limited explanatory power for explaining the associations. The assumption of linearity in regression models may not be appropriate for explaining the variation of SDG disclosures. Overall, existing studies widely employ statistical analysis, focusing on detecting linear relationships and providing explanation-based analysis. They often fail to capture complex and non-linear relationships, limit the number of variables and have strict rules and assumptions, which may affect the nuanced patterns.

Recently, accounting researchers and practitioners offer insights into using machine learning techniques to develop a data-driven understanding of ESG disclosures (Zhang & Zhang, 2023). Lin and Hsu (2023) investigate the influence of the COVID-19 pandemic, financial performance and corporate governance on ESG rating prediction, focusing on ‘listed and over-the-counter nonfinancial companies in Taiwan’ that follow ESG regulations between 2018 and 2021. Specifically, this study investigates the following three periods: the entire period (2018-2021), the pre-pandemic period (2018-2019), and the pandemic period (2020-2021). Employing over 27 financial and corporate governance indicators, this study attempts to compare the model performance of random forest (RF), XGBoost, Elaboration Likelihood Model (ELM), and support vector machine (SVM). With R-squared for training and test samples over 0.95, this study demonstrates that machine learning methods provide excellent model performance in making ESG rating predictions. Similarly, Frost et al. (2023) compare machine learning methods with traditional techniques in predicting factors that may influence carbon disclosures in Australian companies. This study, again, found that most variables that are not significant in statistical results have non-zero relative variable importance scores, suggesting greater out-of-sample predictive power in predicting the level of carbon reporting. The ROC scores for learn and test samples based on gradient boosting machine are 0.946 and 0.82 respectively, while ROC scores based on random forests model are 0.878 and 0.799 respectively for learn and test samples. This suggests that gradient boosting machine outperforms random forest and other traditional methods in predicting carbon reporting. This method has also been widely used in financial bankruptcy studies. Therefore, consistent with Frost et al. (2023), this study will adopt the gradient boosting machine.

Table 3.3 Model selection and model robustness for testing the determinants of SDG disclosures

Author	Research context	SDG disclosure measurement	types of models	model fit
Rosati and Faria (2019b)	408 organisations worldwide	dummy variable (yes/no)	logit model	Pseudo $R^2 = 0.373$ (low but valid)
García-Meca and Martínez-Ferrero (2021)	1723 European companies	dummy variable (yes/no)	Multivariate analysis	Kruskal–Wallis test 0.091
van der Waal and Thijssens (2020)	2000 largest stock listed businesses worldwide.	dummy variable (yes/no)	logistic model	Pseudo $R^2 = 0.28$
	30 businesses with extensive SDG disclosures	SDG word frequency	logistic and quantile regressions	Pseudo $R^2 = 0.22$
Martínez-Ferrero and García-Meca (2020)	365 European companies	dummy variable (yes/no)	logit regression	Kaiser–Meyer–Olkin (KMO) 0.5069 (above the threshold of 0.5),
Sasanti et al. (2020)	Indonesia Stock Exchange	SDG disclosure: the ratio of goals disclosed relative to 17 goals	regression model	without moderating effect: adjusted $R^2 = 0.161$ with moderating effect adjusted $R^2 = 0.356$

Phan et al. (2020)	Italy 111 Italian listed companies issued NFD	SDG reporting scores based on content analysis	cross-sectional models multivariate regression analysis	adjusted $R^2 < 24\%$
Elena Nechita et al. (2020a)	47 Central and Eastern European Chemical Companies	SDG reporting scores	multiple linear regression	adjusted R^2 : 0.524 0.533 0.557 0.598
Gambetta et al. (2021)	Spanish companies 2016-2019	SDG reporting scores	OLS regression analysis	adjusted $R^2 = 0.1693$
Pizzi et al. (2021c)	153 Italian Public Interest Entities	SDG Reporting Score	OLS regression analysis	adjusted R^2 : 0.336 0.529 0.32 0.402 0.384
Girón et al. (2020)	366 Asian and African companies	dummy variable (yes/no)	logistic	Pseudo $R^2 = 0.037$
Buniamin et al. (2020)	219 public listed companies in Malaysia	SDG reporting in annual reports	logistic regression	Nagelkerke $R^2 = 0.311$

3.4.2 Hypothesis development

With an overview of potential issues relevant to SDG and SDG disclosure practices, this thesis observed that both contextual and corporate characteristics are widely discussed in prior research. However, existing studies fail to offer generalisable conclusions in terms of predictors that may predict the level of SDG disclosures, due to variable selection, the uncertainty of the definition of potential proxies, complex research designs, and model selection. Therefore, to fill these gaps, this study develops the following propositions to explore factors that may predict the level of SDG disclosures in Chinese companies based on the gradient boosting machine:

H1: There is no predictive relationship between government control and the level of SDG disclosures based on the gradient boosting machine.

H2: There is no predictive relationship between financial performance and the level of SDG disclosures based on the gradient boosting machine.

H3: There is no predictive relationship between corporate governance and the level of SDG disclosures based on the gradient boosting machine.

H4: There is no predictive relationship between ESG performance and the level of SDG disclosures based on the gradient boosting machine.

3.5 Chapter conclusion

Chapter 3 has sought to provide a review of factors that may predict SDG disclosures. Given the authoritative setting in China, the imprinting process of ideological prioritisation exerted by the CPC, and the unique government-company relations (Chua et al., 2024; Lin, 2011; Zhu & Zhang, 2015), the variation in the level of SDG disclosures may be related to the nuances of governmental controls. Prior

research has also widely identified corporate factors, such as financial performance, corporate governance, and ESG performance that may affect SDG disclosures.

Model selection: Existing studies widely employ statistical significance testing, focusing on detecting linear relationships and providing explanation-based analysis. They often fail to capture complex and non-linear relationships, limit the number of variables and have strict rules and assumptions, which may affect the patterns. To investigate factors that may predict the level of SDG disclosures, this thesis adopts the gradient boosting machine, which moves from parameter space to numerical optimisation in function estimation (Chen et al., 2022). This method is selected as a result of the nature of data, prediction accuracy, its validity and generalisability, interaction and complexity, and non-linear and high-dimensional analysis.

Chapter 4 An overview of automated content analysis

4.1 Chapter introduction

As stated in Chapter 2, the United Nations called for corporate SDG engagement and disclosures. Despite discussions within the literature and by standard setters, there remains a core challenge in identifying quantifiable and consistent measures for observing corporate SDG disclosures in China. To facilitate data capture, this chapter develops an SDG disclosure framework, and an automated content analysis tool for data capture.

This chapter begins with an overview of content analysis in Section 4.2. This is followed by the research setting in Section 4.3 Sample, Section 4.4 Timeframe, and Section 4.5 Data Source. The content analysis process is discussed in Section 4.6. The process of turning the collected text into a dependent variable is discussed in Section 4.7, which will be used in the analysis of factors that may predict SDG disclosures in Chinese companies. Section 4.8 concludes this chapter.

4.2 An overview of content analysis

Content analysis is defined as a technique for ‘objective, systematic, and quantitative description of the manifest content of communication’ (Vourvachis & Woodward, 2015, p. 3). According to Vourvachis and Woodward (2015), it can be used to describe the observable nature of text data and code the collected data into different categories. It can also analyse the latent meaning of text to make inferences. This thesis focuses on capturing what information is disclosed by Chinese companies, rather than the latent meaning of the text.

Content analysis has been widely used in social and environmental disclosure research (Guthrie & Abeysekera, 2006; Heichl & Hirsch, 2023; Ordonez-Ponce & Khare, 2021). It has been used to explore the extent of CSR disclosures (Christopher & Filipovic, 2008; Guthrie & Abeysekera, 2006), subsequently extended to the level of disclosures, such as carbon, environmental, or social disclosures (Cheng et al., 2017; F. Lu et al., 2017), the quality of disclosures (Abeysekera et al., 2021), the readability of disclosures (Nilipour et al., 2020), and the development of disclosure frameworks to observe corporate SDG disclosure practices (Ordonez-Ponce & Khare, 2021). However, these studies rely on manual content analysis, which is often criticised for being ‘complex and time-consuming due to its robustness and big volume’ (Chakraborty & Bhattacharjee, 2020, p. 758). Different human coders may also have their subjective understanding and interpretation of text, which will affect the reliability and consistency of data (Ryans, 2021; Short et al., 2018).

To address this limitation, this thesis develops an automated content analysis tool, following Bochkay et al. (2023) and Chakraborty and Bhattacharjee (2020). Automated content analysis outperforms manual content analysis in coder reliability and efficiency. It follows consistent coding schemes and algorithms to systematically search and identify relevant text for each indicator based on keywords across all data points. This will mitigate human coders’ subjective understanding and improve data consistency (Parhankangas & Ehrlich, 2014). Once coding has been created, it will iterate the entire process automatically, one report by one report, without further human intervention. It will enable a large volume of data to be generated promptly (Kasperuniene et al., 2020). Therefore, the automated tool is more reliable, consistent, and efficient in conducting content analysis, while it may be less effective in analysing context-dependent information.

Automated content analysis can be broadly classified into dictionary-based analysis and machine learning two types. Dictionary-based analysis has been employed in various research, such as financial reporting, ESG reporting, analyst reports and earnings press releases (Banner et al., 2019). Specific keyword lists are essential for dictionary-based analyses, as they focus on classifying text based on pre-defined categories and keyword lists. Accounting research has employed this method in sentiment analysis and other more forward-looking analyses (Bochkay et al., 2023, p. 789).

To effectively extract data from reports, this thesis follows Bochkay et al. (2023, p. 774) five-factor model selection to find the optimal method to extract SDG information from ESG reports. The five factors include ‘functionality, simplicity, familiarity, model performance and computing costs’.

- **Functionality** refers to whether the model can perform the intended function (Bochkay et al., 2023). To develop an automated tool, this thesis involves several steps: identifying what information can be included as text, identifying text location, selecting the unit of analysis, keyword selection, refining and validation, and text cleaning to ensure that text can be classified based on the pre-defined categories. According to Guo et al. (2016, p. 335), dictionary-based content analysis is an optimal tool to do so¹¹.
- **Simplicity.** Bochkay et al. (2023) also argue that simplicity should be another criterion for model selection; that is, when multiple models may work for this thesis, the simplest method is optimal. Machine learning methods, such as Naïve Bayes algorithm and FinBERT, can be used to classify text into positive, negative and neutral three categories, which have been adopted in sentiment analysis literature. Huang et al. (2023) suggest that content analysis based on large language models or deep learning algorithms includes a number of parameters, making the whole process more difficult and less interpretable. Therefore, this thesis selects the simple dictionary-based analysis.
- **Familiarity** may refer to the understanding of the context and content, and the competence in operating the automated tool (Bochkay et al., 2023). Once the coding schemes and keywords are set up, the automated content analysis tool can replicate the manual process to a certain level, which can mitigate the discrepancy in inter-coder understanding of the whole process.
- **Model performance.** According to Kasperuniene et al. (2020) and Guo et al. (2016), dictionary-based automated content analysis increases the efficiency of text classification and can be close to the manual process.
- **Computing costs.** This thesis adopts Python to conduct automated content analysis. Python is a simple user-friendly programming language that is excellent for automating tasks, data analysis, and software development (Github, 2023). It includes extensive modules, libraries and packages as open sources for users to customise their coding (Python, 2024). For example, existing libraries have already developed functions and built-in modules and therefore, users do not have to write every single function from scratch (Github, 2023). It also supports users to develop their customised Python scripts to perform specific tasks. Furthermore, according to Github (2023), one of the benefits of Python is to automate the manual analysis and conduct repetitive tasks. It is also a preferred language for data science and academic research, such as

¹¹ Machine learning methods, such as KNN and LDA, are not appropriate for this study. These are unsupervised machine learning techniques that focus on classifying text into different categories automatically without following pre-defined coding schemes (Guo et al. 2016). The BOW cosine similarity model is a text similarity comparison tool (Guo et al. 2016), which is not appropriate for this study.

statistical analysis, and machine learning. For example, Python includes a rich set of libraries, such as NumPy, Pandas, and Matplotlib, for data cleaning, visualisation, and transformation. Given that it has some basic packages for machine learning algorithms, it can also enable data analysis, such as identifying data patterns or making predictions (Github, 2023). Overall, Python is a powerful programming language to perform customised and automated tasks, while is excellent for its open source, uncomplicated nature, and productivity.

Therefore, this thesis employs the dictionary-based automated content analysis to capture SDG disclosure data in Chinese companies.

4.3 Sample

To effectively collect data from Chinese companies, sample selection follows the following criteria: *1.* large Chinese companies, as Section 2.4 discusses that larger companies have stronger motivations in participating in SDG disclosures. *2.* Companies that have English ESG-related reports, due to the accessibility of reports for international stakeholders. Therefore, *the initial sample includes the top 100 Chinese companies listed on SSE and SZSE respectively, based on market capitalisation in 2016.* Now, I will discuss it in detail:

- SDG 12.6 ‘encourages companies, especially large and transnational companies to adopt sustainable practices and to integrate sustainability information into their reporting cycle’ (United Nations, 2015, p. 27). As a response, China’s National Plan on Implementation of the 2030 Agenda for Sustainable Development (2016) also encourages companies, particularly large Chinese companies, to participate in and report on their SDG performance. This provides a unique setting for investigating SDG disclosures in large Chinese companies.
- Larger companies are the leaders in the innovation of SDG disclosures. According to KPMG (2020), China ranks among the top five among the world’s largest companies (G250) in advancing SDG disclosures. Between 2017 and 2020, the percentage of Chinese companies among G250 disclosing SDG information increased from 5% to 48% (KPMG, 2020). This indicates that large Chinese companies actively engage in SDG disclosures.
- Early CSR disclosure and SDG disclosure literature focused primarily on large companies, due to data availability (Brammer & Pavelin, 2006; Moneva & Llana, 2000; Yu et al., 2020). These companies, often first movers, may serve as the basis for exploring corporate early responses to the adoption of SDGs, especially in the context where SDG disclosures are still in an early stage (PwC, 2020; Yu et al., 2020). Specifically, PwC (2020) surveyed and interviewed 89 Chinese companies to explore their incorporation of SDGs, including their practice, potential challenges, and future opportunities, Yu et al. (2020) explore top 100 Chinese companies listed on Shanghai Stock Exchange. As suggested by Zhang and Guan (2020), limited SDG information has been disclosed in Chinese companies between 2018 and 2020. Therefore, this

thesis does not expect small or medium companies to take more or the same level of initiative in their SDG disclosures. A focus on top Chinese companies may offer insights into interpreting and recognising corporate SDG practices.

- In 2017, President Xi announced the establishment of the BRI international cooperation coalition to jointly achieve the goals of Agenda 2030 (Xinhua, 2017). SDGs serve as the basis for cooperating and communicating with BRI partners. Table 4.1 displays that the majority of BRI project investors are the largest companies in China in 2023, indicating that larger companies were the dominant players in BRI and jointly realising SDGs.

Table 4.1 A list of largest Chinese investors in the BRI in 2023, extracted from <https://greenfdc.org/china-belt-and-road-initiative-bri-investment-report-2023/>

Largest Chinese investors in the BRI in 2023 (parent companies)		Largest Chinese construction companies in the BRI in 2023 (parent companies)	
	% of total		% of total
Contemporary Amperex Tech	15.2%	Power Construction Corp. (PowerChina)	20.1%
Zijin Mining	10.8%	China Petroleum and Chemical (Sinopec)	16.8%
Southern Power Grid	9.3%	China National Petroleum Corp. (CNPC)	10.1%
Zhejiang Huayou Cobalt	8.8%	China Railway Construction	9.8%
China Molybdenum	5.4%	China Energy Engineering	8.0%
Minmetals	5.1%	China Communications Construction	7.8%
Alibaba	4.0%	China National Building Material	5.7%
China Communications Construction	3.5%	China National Machinery Industry (Sinomach)	5.5%
China Petroleum and Chemical (Sinopec)	3.5%	State Construction Engineering	4.7%
Shandong Xinhai, BaoWu Steel (Baosteel)	3.4%	Power Construction Corp. (PowerChina), Dongfang Electric	3.0%
GEM	3.3%	China Energy Construction Corporation	2.5%
Hailiang	2.3%	China Railway Engineering	1.6%
Bytedance	2.3%	Minmetals	1.6%
China National Nuclear	2.1%	China National Chemical Engineering	1.5%
West China Cement	2.1%	China Three Gorges	1.3%
Shenzhen Xingyuan	1.9%		
Aviation Industry Corp. (AVIC)-led consortium	1.8%		
Hefei Xinmei Materials	1.7%		
China Communications Construction, Guangxi Beibu	1.6%		
China Railway Construction	1.6%		
Inventronics	1.6%		
Hunan Zhongwei	1.6%		
Zhonghuan International Group	1.4%		
BYD	1.4%		
China National Petroleum Corp. (CNPC)	1.4%		
China Energy Engineering	1.4%		
China National Geological & Mining Corp	1.4%		

- The top 100 companies cover all GICS sectors. This allows a broad and comprehensive exploration of corporate SDG disclosures across all sectors.

Audience: This thesis only focuses on companies that provide reports in English, due to global accessibility, investor interests, and BRI partners.

- (1) **Global accessibility:** According to the United Nations (2015), SDGs are created as a common language and global benchmark to promote international dialogue around sustainable development. Given ‘the prevalence in modern commerce’, English reports may mitigate

language barriers to corporate communications in the international community (Lang & Stice-Lawrence, 2015).

- (2) **Investor interests.** Recently, financial analysts recommended that companies need to participate in SDG activities and make disclosures, as investors are increasingly interested in SDGs (García-Sánchez et al., 2020). It is argued that language is a crucial component of corporate report comparability in the international market (Jeanjean et al., 2010). Bao and Wei (2024) also argue that issuing English CSR reports can assist companies in acquiring more foreign shareholding. This suggests that English-language reports may expose companies to a broader base of potential investors in the market.
- (3) **BRI partner communication.** The Chinese government encourages BRI partners to participate in SDGs. SDGs may act as a means for the Chinese government to establish connections with BRI partners in promoting sustainable development. English reports could serve as the basis for facilitating international dialogue among the Chinese government, Chinese companies, and the global community.

Therefore, providing English reports can be seen as the company's desire to communicate with international stakeholders. To confirm the number of Chinese companies that voluntarily provide their English ESG reports (see Section 4.5 Data Source), this thesis searched multiple sources, including corporate official websites, SSE and SZSE platforms, Juchao Zixun (巨潮资讯) platform, and Baidu and Google search engines. This process was implemented twice to ensure that no English report was overlooked¹².

This thesis found that, of the 200 sampled companies, 82 companies' English reports are currently available online. This thesis also contacted companies whose English reports could not be found online. The phone number was obtained from their Chinese CSR reports—it was the contact number provided for further inquiries and information requests. Only one company told the author that its English ESG reports between 2015 and 2020 can be found on the HongKong Stock Exchange website. Four companies did not start providing their English reports until 2021 and one company only provides reports for their investors.

¹² ESG reports were initially downloaded from the official websites of the selected companies (both English and Chinese version websites), as they are considered the most reliable and easy way to access corporate reports. However, as some companies only post their most recent reports online, historical reports were collected from other databases, such as the SSE and Juchao Zixun websites.

The SSE platform is selected as it requires companies listed on the SSE to release their Chinese annual reports on the SSE website and some companies will upload their ESG reports as well, including their English reports. Therefore, SSE is selected as an alternative source for collecting English ESG reports. However, there is no similar requirement on SZSE. This study also attempted to search for ESG reports from the SZSE for several companies, but no reports were identified.

Juchao Zixun is selected as an additional source for searching English ESG reports. It is a website designated by the China Securities Regulatory Commission (CSRC) to gather reports and other information from listed companies. English ESG reports were also searched from Google and Baidu, the largest search engines.

Of the 82 companies, some have attached their Chinese reports to their English websites or offer English reports with content in Chinese. Therefore, the final sample in my study is 60 Chinese public companies (Appendix 2 shows the list of sampled companies).

4.4 Timeframe

The research period of this thesis is between 2016 and 2020 which covers the period of the 13th National FYP being effective. While Agenda 2030 and SDGs were introduced in 2015, the Chinese government recognised SDGs in 2016, which can be considered an appropriate starting point in my research.

- The 13th National FYP was introduced in 2016 to set the direction for China's development between 2016 and 2020. The Chinese government has adopted SDGs to operationalise and inform the development of 13th National FYP (see Section 2.3). The plan also stated that 'we will actively implement the Agenda 2030 for sustainable development' (Communist Party of China, 2016b, p. 136; National Development and Reform Commission, 2016, p. 153). 2016 can be viewed as the pre-operative reporting time, as the Chinese government started implementing Agenda 2030 in 2016.
- The 13th National FYP ended in 2020. The Chinese government utilised SDGs to operationalise the plan. Although the 14th National FYP continued to promote SDGs, the focus shifted from building a 'Moderately Prosperous Society' to 'Building a Modern Socialist Country'. Therefore, the research period for this thesis ends in 2020.

4.5 Data Source

Prior studies collect ESG data from various sources, such as annual reports, ESG reports, websites and databases (Horne et al., 2020; Sciandra et al., 2021) (see Section 2.7). Annual reports are designed for providing financial impacts of a company, particularly targeting shareholders and investors (Dhaliwal et al., 2011; Friedman & Ormazabal, 2024). In China, corporate annual reports typically include only one section discussing corporate ESG performance, which provides limited information¹³(Dhaliwal et al., 2011; Friedman & Ormazabal, 2024). Websites will also not be selected in this thesis, as they focus on advertising, lack standardised information and typically provide only the latest updates (Wilkerson & Casas, 2017). Therefore, consistent with Christensen et al. (2021) and Dhaliwal et al. (2011), this thesis only focuses on corporate ESG reports to observe their SDG disclosure practices. The justifications are as follows:

1. ***Focused reports of corporate SDG disclosures.*** ESG reports may serve as a valuable source for observing corporate SDG disclosures. Dhaliwal et al. (2011) suggest that CSR reports offer

¹³ This study conducted keyword searching throughout the entire annual reports of the test sample (20 firm-year observations) to observe the extent to which companies make disclosures related to poverty alleviation, energy, and climate change in annual reports. This study investigates the location of each data type in reports by quantifying the occurrence of keywords in every section. However, annual reports provide limited information related to SDG 1, 7, and 13, and thereby may not be the optimal data source for collecting SDG disclosure data.

more useful and comprehensive sustainability information compared with annual reports. These reports tend to act as channels for companies to show their commitment to and involvement in ESG-related national strategies. Similarly, Friedman and Ormazabal (2024) also suggest that ESG reports offer boarder topics, clear objectives, and more measurements. Therefore, ESG reports can provide more details for our analysis.

2. **Standardised data.** The Chinese government and regulatory bodies have introduced a series of guidelines and policies to ensure that ESG information is provided in a standardised, consistent, and comprehensive manner (See Section 2.3).
3. **Cross-sectional analysis.** ESG reports are released by Chinese companies annually and can be obtained from various publicly available sources such as corporate official websites, the SSE, and Juchao Zixun (巨潮资讯) websites. This aligns well with the aim of this thesis to facilitate a cross-sectional analysis of corporate SDG disclosures.

Therefore, this thesis will focus on English ESG reports to explore SDG disclosures in Chinese companies.

4.6 The process of the automated content analysis on capturing SDG disclosure text

4.6.1 Obtaining text

What should be included as text

Determining what information should be included in the analysis is the beginning step of content analysis. ‘Text’ encompasses more than just written information (Steenkamp & Northcott, 2007). For example, visualised components, such as pictures, symbols, tables, figures, and charts, are also important elements of corporate CSR reports and are rich in information (Steenkamp & Northcott, 2007). Recent studies argue that analyses may be incomplete if visualised data is ignored (Steenkamp & Northcott, 2007). Therefore, both narrative and non-narrative reported information will be considered in this thesis.

Non-narrative reporting information may include charts, tables, and pictures (Guthrie & Abeysekera, 2006). Although pictures might impress stakeholders and enable a better understanding of corporate management of ESG issues, they may complicate the analysis process, as ‘a picture may be worth a thousand words’ (Guthrie & Abeysekera, 2006, p. 19). It is also hard for Python to interpret pictures without surrounding text. Therefore, this thesis analyses both narrative and non-narrative disclosed information, excluding pictures without surrounding text.

What SDG information should be collected from text

Given that there remains ambiguity as to what constitutes SDG disclosures, this section develops a disclosure framework for capturing corporate SDG disclosures related to SDG 1, 7, and 13.

SDG goals that are prioritised by Chinese companies

Given that not all SDG goals are equally relevant and important to China, this thesis draws upon SDGs most relevant to Chinese companies. The Permanent Mission of the People's Republic of China to the UN (2022) reported that 'China has made significant progress in meeting sustainable development goals, including eliminating poverty (SDG1), protecting the marine environment (SDG 14), effectively utilizing energy (SDG7), tackling climate change (SDG13), and conserving land ecosystems (SDG15)'.

- Of these five goals, SDG 1 is consistent with the national strategy of building a 'Moderately Prosperous Society', which is the top prioritised goal in the 13th National FYP. The national poverty alleviation goal aims to mobilise the involvement of various sectors of the society and Chinese companies are also the key contributors of this strategy (Xu et al., 2025). Unlike other emerging economies, where companies are less likely to participate in poverty alleviation activities, a significant portion of Chinese companies participate in addressing poverty alleviation problems (Xu et al., 2025). One of the reasons in prior research could be state enforcement and support for corporate engagement in poverty alleviation (Huang et al. 2022). As discussed in Section 2.5.2, TPA programs have been introduced to guide companies to involve in poverty alleviation activities. According to the Permanent Mission of the People's Republic of China to the UN (2022), China has achieved SDG 1 10 years ahead of schedule and achieved the goal of eradicating absolute poverty by 2020. Existing studies have sought to explore the association between government initiatives and corporate poverty alleviation disclosures; however, they are primarily limited to existing databases for collecting poverty alleviation disclosure data and neglect detailed analysis for it in ESG reports. Therefore, this provides a unique opportunity to investigate corporate poverty alleviation disclosures.
- SDG7 and 13 are closely aligned with the 13th National FYP, in promoting green and low-carbon development, and in pursuing ecological civilisation. The Chinese government also announced to the international community China's '3060' plan, which is to achieve a 'carbon peak before 2030 and carbon neutrality before 2060' (Y. Wang et al., 2021, p. 721), to pursue ecological civilisation and 'build the world's largest clean energy system' (Permanent Mission of the People's Republic of China to the UN, 2022). Therefore, SDG 7 and 13 are selected.
- SDG 14 and 15 are consistent with the goal of protecting biodiversity. While the government prioritised these two goals; however, existing research suggests that there remain imbalances in implementing these goals across provinces and the ecosystem (Xie et al., 2021, p. 2). Additionally, PwC (2020) suggested that SDG 14 and SDG 15 received the least attention among all 17 goals across Chinese companies. During the manual data collection, I also observed that limited biodiversity information was disclosed by large Chinese companies. Therefore, SDG 14 and 15 were not considered part of this thesis's SDG disclosure framework.

Overall, this thesis develops an SDG disclosure framework around SDG 1, 7, and 13, which are the most relevant goals for Chinese companies.

A framework for observing SDG disclosures

Drawing from multiple sources, such as the official SDG framework, the GRI, Chinese government documents and prior studies, this thesis starts from over 200 potential indicators and arrives at the final number of 27 indicators for observing SDG disclosures.

1. **SDG framework:** using the UN 17 goals, 169 targets, and 232 indicators is a starting point. Each of the 17 goals has between eight and twelve targets against which to monitor progress towards the goals. Targets include both ‘outcome targets’ and ‘means of implementation targets’ (United Nations, 2015). For example, the suggested indicator 7.1 is ‘By 2030, ensure universal access to affordable, reliable and modern energy services’, and for SDG 13 ‘Take urgent action to combat climate change and its impacts’, for which the suggested indicator 13.1.1 is ‘Number of countries with national and local disaster risk reduction strategies’ (United Nations, 2015, pp. 10-17). However, as the examples make clear, the SDG goals and targets are aspirational and have a generic focus (see Section 2.2). It provides a higher-level framework that needs to be translated into something observable. To operationalise the Agenda 2030, this thesis also draws upon the GRI framework and the Chinese government documents.
2. **GRI framework:** Since the launch of Agenda 2030, the UN, in partnership with the GRI, United Nations Global Compact, WBCSD, and PRI has sought to develop guidelines specifically related to business SDG disclosures (SDG Compass, 2015; UNGC & GRI, 2018). Prior studies have used existing reporting frameworks such as GRI, SASB, and IR to observe corporate SDG disclosures (see Section 2.7). GRI is a leading and widely used framework worldwide to address ESG activities due to its ‘comprehensiveness, continuous improvement, and consensus style’ (GRI, 2017; Ordonez-Ponce & Khare, 2021, p. 51). Recent studies also find that the GRI framework may be the basis for observing SDG disclosures (Michalczyk & Konarzewska, 2018). Ordonez-Ponce and Khare (2021) also suggest that the GRI 300 can be used as an effective tool to operationalise SDG 7 and 13. Therefore, this thesis will also draw upon GRI indicators to observe SDG disclosures¹⁴.

¹⁴ Why the GRI is adopted to observe corporate SDG disclosures in China:

1. The GRI framework provides a standardised framework to guide companies to report their ESG information. This framework covers the most material ESG topics that are relevant to companies, such as GRI 100 (general information), GRI 200 (economic topics), GRI 300 (environmental topics), and GRI 400 (social topics). These standards are designed to guide reporters in disclosing their ESG information in a manner that is both comparable and credible. Overall, GRI indicators clearly describe what should be included in each indicator.

However, this thesis manually compared the alignment and overlap between SDGs and GRI frameworks (see Appendix 1). The GRI framework focuses on quantitative measurements rather than a balanced framework with a mixture of quantitative and qualitative measures. This may lead to the omission of corporate commitments or generic data. Additionally, the GRI framework is only partially applicable to the Chinese context, which may influence the effectiveness of this framework in providing a comprehensive view of corporate SDG disclosures. Therefore, this thesis also uses Chinese guidelines to develop measurable SDG disclosure indicators.

3. **Chinese government documents.** The 'Notice on the 13th Five-Year Plan for Poverty Alleviation', issued by the State Council of People's Republic of China (2016), is a procedural framework under the administrative law. It was established to accelerate the progress towards poverty alleviation between 2016 and 2020. The State Council of People's Republic of China (2016) outlines the overall approach, main targets, tasks, and measures for poverty alleviation within the period that the 13th FYP was effective. Subsequently, local regulations on poverty alleviation offer more actionable targets, clear implementation instructions for this national strategy and a clear description and explanation of their plans. However, these localised regulations on poverty alleviation are customised and localised, which cannot be used as a benchmark framework for measuring corporate behaviours on SDG 1. Following this, the CSRC also introduced the 'Opinions of the China Securities Regulatory Commission on the Role of the Capital Market in Serving National Poverty Alleviation', which requires SSE and SZSE to format corporate poverty alleviation disclosures. However, they only provide notice rather than detailed guides or performance measurements. Therefore, this thesis follows the guideline set by the State Council of People's Republic of China (2016) to look for a means for measuring corporate poverty alleviation performance.

-
2. The GRI framework is an organisational-level framework to observe corporate sustainability performance. According to GRI, 'any organisation, large or small, public or private, from any sector or location, can use GRI standards' (GRI, 2024, p. 2). As SDGs are an inter-governmental initiative, SDG targets and indicators, thereby, are partially applicable to companies. As GRI indicators are partially aligned with SDG goals, GRI indicators may help observe corporate SDG performance.
 3. The GRI indicators overlap with SDGs. In order to enable more integrated reporting on SDGs, the United Nations has cooperated with GRI and UNGC to launch several guidelines to guide companies to report SDG information, such as 'linking SDGs to GRI' and 'Business reporting on the SDGs: Analysis of the Goals and Targets'. These guidelines consider the GRI framework an appropriate instrument for measuring SDG performance. It is interesting to note that Ordóñez-Ponce et al. (2021) argue that GRI 300 is a useful framework to operationalise SDGs to assess corporate environmental sustainability. Therefore, the GRI framework is widely promoted to operationalise SDGs.

4. **Prior studies.** As discussed in Section 2.7, emerging discussions have been made within prior literature and by professional bodies to observe SDG disclosures among companies. These indicators are well-established and tested.

Therefore, this thesis looks at *explicit references to SDGs (the overall level of SDG disclosures)* and *implicit SDG performance (specifically SDG performance related to SDG 1, 7, and 13)*. The proposed framework (see Table 4.6 below) comprises the following components: 3 indicators for observing the overall level of SDG disclosures, 12 indicators related to poverty alleviation, 5 indicators for observing energy disclosures, and 7 indicators related to climate change¹⁵. The process of reducing the number of indicators is discussed as follows:

Indicators for assessing explicit references to SDGs include several factors, based on Clarkson et al. (2008) and Subramaniam et al. (2020) (see Table 4.2). **1. Mentioning SDGs in CSR reports** may offer insights into corporate awareness and their willingness to align with these goals (Clarkson et al., 2008; Subramaniam et al., 2020). **2. Mentioning SDGs in the Chairman or CEO message** may offer a statement about corporate interests in SDGs and the relevance of these goals to the company (Subramaniam et al., 2020). The statements by corporate leaders may also suggest corporate ‘potential to make real commitments and actions’ to SDGs (Subramaniam et al., 2020, p. 12). **3. Acknowledging Agenda 2030 or SDGs in the ‘About the report’ or ‘Preparation Basis’ section** may also indicate that Agenda 2030 plays a role in structuring the content of ESG reports. Overall, these indicators may indicate corporate awareness, willingness, and potential to participate in SDGs.

Table 4.2 Indicators for assessing explicit references to SDGs

No.	Indicators
1	Whether companies explicitly mention SDGs or Agenda 2030 in CSR reports?
2	Whether companies mention SDGs or Agenda 2030 in the Chairman or CEO message?
3	Whether the report acknowledges Agenda 2030 or SDGs in the 'About the report' or 'Preparation Basis' section?

This thesis also draws upon UN 232 SDG indicators, the GRI, and the Chinese government document to develop indicators for measuring implicit SDG disclosures:

¹⁵ The SDG goals and targets are aspirational and have a generic focus. The mixture of broad qualitative and quantitative indicators means that the SDG goals, targets and indicators are neither designed nor adequate as a corporate SDG reporting framework. Therefore, only five indicators are developed based on SDG targets and indicators, while indicators for assessing corporate actions towards poverty alleviation (nine indicators) are theoretically developed based on the State Council of People’s Republic of China (2016), five indicators based on prior studies, and other indicators are drawn from the GRI framework.

Indicators for assessing corporate poverty alleviation intend to capture corporate SDG 1 poverty alleviation disclosures. Despite 7 targets introduced by the official SDG guideline, one of the SDG 1 targets is to assess the proportion of people that ‘live on less than \$1.25 a day’. SDG 1.4 also targets ensuring that women and men ‘have equal rights to economic resources’ (United Nations, 2015, p. 19). However, these targets are aspirational and have a generic focus and the mixture of broad qualitative and quantitative indicators. The SDG goals, targets and indicators were neither designed nor adequate as a corporate SDG disclosure framework.

To facilitate data capture, this thesis draws upon the State Council of People’s Republic of China (2016), SDG 1a, FS7, G4-EC8, and UNGC and GRI (2018) to develop 12 indicators to capture poverty alleviation in Chinese companies. Specifically, one indicator is drawn upon SDG 1a and one indicator is drawn upon UNGC and GRI (2018), while all other 10 indicators are based on the Chinese government document--State Council of People’s Republic of China (2016) (See Table 4.3):

4. *Mentioning or acknowledging activities related to SDG 1* is drawn upon UNGC and GRI (2018).

This indicator explores whether companies acknowledge SDG 1 or poverty alleviation activities that the Chinese government identified as important between 2016 and 2020. Such may indicate corporate commitments and contributions towards poverty alleviation. Information will be collected from non-performance sections in ESG reports, as the preliminary analysis shows that implicit SDG performance details are typically disclosed in ESG performance sections, while limited performance information is disclosed in non-performance (generic information) sections (see *Section* Text location). Therefore, this indicator was developed to capture SDG 1 or poverty alleviation information that is disclosed in non-performance sections in ESG reports.

5. *Corporate spending or resource commitments related to SDG 1* is drawn upon SDG 1a, FS7, and G4-EC8. This indicator may indicate corporate inputs, such as cash flow and resources, in poverty alleviation activities.

6. *Government-related poverty alleviation* identifies funds received from the government or other institutional bodies to support corporate engagement in poverty alleviation programs, based on SDG 1a (United Nations, 2015, p. 1). Defining this indicator includes several steps, such as indicator development based on the official SDG guideline and adjustments through the manual data collection. The process is more complex, as this indicator is drawn from the UN framework, which is an aspirational initiative and may not apply to the Chinese context. Therefore, this indicator is refined during the preliminary analysis, as it is observed that companies participated in government-related poverty alleviation projects in various ways, rather than only the government-sponsored poverty alleviation activities. For example, companies also have been observed donating money to entities, such as the China Foundation for Poverty Alleviation, the China Red Cross Foundation, local governments, and provincial governments to support government poverty alleviation projects.

For example, CNPC (2018, p. 61) also disclosed ‘CNPC has cooperated with the government of Xinjiang Autonomous Region in extensive areas...’ Therefore, this thesis considers whether ‘government-related poverty alleviation activities’ might be more interpretable than ‘government-sponsored or other authority-sponsored poverty alleviation activities’ when understanding corporate participation in government-related activities. Therefore, this indicator was revised to generally capture ‘whether the company discloses any information related to government-related poverty alleviation activities.’

7. **Poverty alleviation through industrial development** identifies industrial development activities to support poverty alleviation in rural areas, based on the State Council of People’s Republic of China (2016). Industrial development focuses on accelerating the development of speciality industries in rural areas, by taking advantage of their local resources (State Council of People’s Republic of China, 2016). It is a market-oriented approach, aiming to develop distinct products and form an industrial chain, to improve poor people’s skills and the stability of operating and property income in rural areas (State Council of People’s Republic of China, 2016). To achieve these, the government highlights that poverty alleviation through industrial development focuses on agriculture, tourism, e-commerce, and technological initiatives (State Council of People’s Republic of China, 2016).
 - **Industrial development through agriculture** identifies corporate activities related to agricultural production, actively develops aquaculture, accelerates the growth of the forestry sector, speeds up the cooperation among industries, and advances agricultural skills in rural areas (State Council of People’s Republic of China, 2016). For example, the government aims to improve agricultural production, by developing high-quality and standardised farmlands. Specifically, it encourages stable and high-yield production through promoting crops that require low water consumption and can be conducive to ecological restoration. The government also introduced several projects in this area, such as One Village One Product (一村一品), the ‘development of the primary secondary and tertiary industries and training in rural areas’ (The State Council of the People's Republic of China, 2018, pp. 1-4).
 - **Industrial development through tourism** attempts to take advantage of natural resources and scenery in rural areas (State Council of People’s Republic of China, 2016). Actions may include working on travel-related infrastructure construction, the development of tourism and highly recommended travel routes in rural areas, which may help ‘develop local economy and offer more job opportunities for local people’ (The State Council of the People's Republic of China, 2018, pp. 1-4).
 - **Industrial development through e-commerce** aims to introduce and integrate e-commerce into rural areas. The government encourages companies to partner with rural areas, helping residents start online ventures. This will provide more opportunities for people in rural areas

to establish and expand the e-commerce market, such as transportation, delivery services, and business-to-business and business-to-customers services (State Council of People's Republic of China, 2021).

- ***The earnings from poverty alleviation funds and assets*** facilitate 'the assets generated from facility agriculture, photovoltaic (PV) power stations, and rural tourism using central government funds for poverty alleviation and other agriculture-related funds' (State Council of People's Republic of China, 2021). 'Funds are converted into shares and allocated proportionately to poor villages to promote their businesses, increase villagers' incomes and more importantly help generate revenue for the villages' collective economy' (State Council of People's Republic of China, 2021).
- ***Industrial development through technological development*** focuses on introducing the latest and advanced scientific and research achievements to rural areas, fostering creativity among rural populations, and accelerating the establishment of innovation platforms in these areas (State Council of People's Republic of China, 2021).

Therefore, any of the above aspects can be considered evidence for poverty alleviation through industrial development: agriculture, farming, e-commerce, power stations, tourism, and technological development in rural areas (State Council of People's Republic of China, 2021).

8. ***Poverty alleviation through health care*** identifies corporate activities related to health care in rural areas, based on the State Council of People's Republic of China (2016). In 2016, the Chinese government introduced the 'Health Poverty Alleviation Project', aiming to reduce and prevent poverty caused by health problems (Lu & Abeysekera, 2021). It covers diverse aspects, such as reducing medical expenses, improving medical and health services, improving Medicare coverage, and strengthening disease prevention and controlling public sanitation (State Council of People's Republic of China, 2021).
9. ***Poverty alleviation through bottom-up protection*** identifies corporate activities related to bottom-up protection, based on the State Council of People's Republic of China (2016). Bottom-up protection in rural areas targets the most vulnerable people in society. It focuses on improving social aid systems, expanding universal social security coverage in the People's Republic of China, and directing attention to the rural 'three left-behind' group and the disabled care service (State Council of People's Republic of China, 2021). The 'three left-behind' group refers to the elderly, left-behind children, and orphanages (State Council of People's Republic of China, 2021).
10. ***Poverty alleviation through social impacts*** identifies corporate activities related to social impacts, based on the State Council of People's Republic of China (2016). Social poverty alleviation includes west-east cooperation and paired-up assistance in rural areas, enterprise assistance, army assistance, social and volunteer assistance, and international cooperation (State Council of People's Republic of China, 2016). Specifically, the paired-up assistance involves activities to help paired-up villages

or counties reduce poverty. State Council of People's Republic of China (2016) suggests that this involves selecting and sending outstanding young and middle-aged cadres to designated poverty alleviation counties to enhance work coordination and supervise corporate paired-up assistance. Enterprise assistance aims to strengthen SOEs' responsibilities towards poverty alleviation and 'to encourage private enterprises to participate in poverty alleviation activities' (State Council of People's Republic of China, 2016). Such can indicate corporate engagement in social protection, ultimately contributing to safeguarding people's livelihoods.

11. ***Poverty alleviation through education*** identifies corporate activities related to education, based on the State Council of People's Republic of China (2016). Reducing poverty by promoting education is a vital component of China's poverty alleviation strategy (Yuan & Ding, 2023). The Chinese government has adopted a series of strategies, such as improving basic education levels, reducing financial burden in rural areas, offering training on vocational skills and improving the quality of education (State Council of People's Republic of China, 2016).
12. ***Poverty alleviation through employment*** identifies corporate activities related to employment, which is based on State Council of People's Republic of China (2016). This encompasses activities related to professional skill improvement, staff training and increasing job opportunities (State Council of People's Republic of China, 2016). Such may indicate corporate contributions towards poverty alleviation through employment transfer, such as improving the employability of poverty-stricken or increasing job opportunities in rural areas.
13. ***Poverty alleviation through ecological protection*** identifies corporate activities related to ecological protection, based on the State Council of People's Republic of China (2016). Ecological protection refers to the mitigation of poverty induced by environmental issues (State Council of People's Republic of China, 2016). It includes strengthening ecological restoration and establishing compensation mechanisms for ecological protection (State Council of People's Republic of China, 2016).
14. ***Poverty alleviation through regional development*** identifies corporate activities related to regional development, based on the State Council of People's Republic of China (2016). Regional development prioritises poverty alleviation 'in old revolutionary areas, ethnic minority areas, frontier areas, and concentrated contiguous destitute areas' (State Council of People's Republic of China, 2016). Simultaneously, improving the quality of life in particular rural areas and strengthening infrastructure construction can also be viewed as regional development to reduce poverty (State Council of People's Republic of China, 2016).
15. ***Poverty alleviation through relocating*** identifies corporate activities related to relocating, based on the State Council of People's Republic of China (2016). The objective of relocation projects is to enhance life quality of relocated individuals, to address problems such as clean drinking water, electricity, and household safety (State Council of People's Republic of China, 2016). Relocating falls under the framework of 'Two Assures and Three Guarantees', aiming to ensure people's living

conditions and quality of life for individuals' post-relocation (State Council of People's Republic of China, 2016).

Table 4.3 Indicators for assessing SDG 1 Poverty alleviation disclosures

5	Does the company disclose corporate spending or resource commitments to SDG 1?
6	Does the company disclose information related government-related poverty alleviation activities?
7	Does the company disclose their industrial development activities to support poverty alleviation in rural areas?
8	Does the company disclose information related to poverty alleviation through health improvement?
9	Does the company disclose information related to poverty alleviation through bottom-up protection?
10	Does the company disclose information related to poverty alleviation through social poverty alleviation?
11	Does the company disclose information related to poverty alleviation through education?
12	Does the company disclose information related to poverty alleviation through employment?
13	Does the company disclose their contributions to poverty alleviation through ecological improvement?
14	Does the company disclose information related to poverty alleviation through regional development?
15	Does the company disclose information related to poverty alleviation through relocating for people?

Indicators for assessing corporate energy disclosures intend to capture corporate SDG 7 energy disclosure data. The official SDG guideline introduces 5 targets to guide companies and countries to pursue affordable and sustainable energy, such as 'increasing substantially the share of the renewable energy in the global energy mix' (SDG 7.2), 'ensuring access to affordable, reliable and modern energy services', and 'doubling the global rate of improvement in energy efficiency' (United Nations, 2015, pp. 19-27). Given that these indicators are broad and have a generic focus, they provide limited guidance for corporate energy disclosures.

Energy is of global concern and included in international reporting guidelines, such as GRI, which provides some guidance for Chinese companies to frame their ESG reports. Therefore, indicators for assessing SDG 7 disclosures include several factors: corporate engagement in renewable energy consumption, energy consumption, and energy efficiency. The indicator in assessing corporate renewable energy consumption is based on SDG 7.2, and corporate spending or resource commitments to SDG 7 are drawn up on SDG 7.3, while other energy indicators are drawn from GRI et al. (2015) (See Table 4.4).

16. Mentioning or acknowledging activities related to SDG 7 is drawn upon UNGC & GRI (2018).

This indicator explores whether companies acknowledge SDG 7 or energy activities that the Chinese government identified as important between 2016 and 2020. Such may indicate corporate commitments and contributions towards energy activities. This indicator is developed to capture SDG 7 or energy information disclosed in non-performance sections in ESG reports.

17. Corporate spending or resource commitments related to SDG 7 is drawn upon SDG 7.3. This indicator may indicate corporate inputs, such as cash flow and resources, in energy activities.

- 18.** Renewable energy consumption measures whether the company used renewable energy. This may indicate the extent to which companies contribute to renewable energy (GRI et al., 2015; Komor & Bazilian, 2005).
- 19.** Energy consumption aims to get an overview of corporate energy usage. Following GRI 302-1 and GRI 302-2, energy consumption covers: ‘
- a. Fuel consumption from non-renewable sources within the organization
 - b. Fuel consumption from renewable sources within the organization
 - c. Watt-hours, including electricity consumption, heating consumption, cooling consumption, and steam consumption within the organization
 - d. Watt-hours, including electricity sold, heating sold, cooling sold, and steam sold within the organization
 - e. Energy consumption outside of the organization, including both upstream categories and downstream categories’
- 20.** Energy intensity (efficiency) explores whether a company reduces negative environmental impacts and saves energy-related costs. This may reveal corporate energy consumption efficiency. Such information will be gathered in sections that talk about energy.

Table 4.4 Indicators for assessing SDG 7 energy disclosures

16	Does the company include SDG 7 or acknowledge activities related to energy in their reports (regardless of whether they specifically mention SDG7)?
17	Does the company disclose corporate spending or resource commitments to SDG 7?
18	Does the company report their renewable energy consumption?
19	Does the company report their energy consumption?
20	Does the company report energy efficiency?

Indicators for assessing corporate climate change reporting intend to capture corporate SDG 13 climate change disclosures. To facilitate data capture, this thesis draws upon the official SDG guideline, UNGC & GRI (2018) and GRI et al. (2015) to develop the following five indicators:

- 21. *Mentioning or acknowledging activities related to SDG 13*** is drawn upon UNGC & GRI (2018). This indicator explores whether companies acknowledge SDG 13 or climate change activities that the Chinese government identified as important between 2016 and 2020. Such may indicate corporate commitments and contributions towards climate change. Information is collected in non-performance sections in ESG reports.
- 22. *Corporate spending or resource commitments related to SDG 13*** is drawn upon SDG 13a. This indicator may indicate corporate inputs in climate change activities, such as cash flow and resources.

- 23. Direct (Scope 1) GHG emissions** refer to GHG emissions produced or controlled by the reporting entity itself (GRI 305-1, p.8). GHG emissions may cover emissions generated from electricity, heating, cooling and steam, physical or chemical processing, transportation of materials, and fugitive emissions (GRI 305-1, p.8).
- 24. Indirect (Scope 2) GHG emissions** ‘include but are not limited to CO2 emissions from the generation of purchased or acquired electricity, heating, cooling, and steam consumed by an organization’ (GRI 305-2, p.10). However, indirect (Scope 3) GHG emissions are excluded from this category.
- 25. Emission** may offer a view of corporate carbon emissions. According to KPMG China (2022, p. 1), ‘Scope 3 emission measurement is more complex and significantly less mature than Scope 1 and 2 measurements’, regarding data quality and availability in China. Many companies pointed out challenges in capturing Scope 3 data in their value chain to facilitate corporate disclosures (KPMG China, 2022). It was observed that none of the test samples disclosed their scope 3 information. However, since these companies disclose their total emissions or carbon emission information in their reports, this thesis replaced scope 3 with the company's emissions.
- 26. Corporate education and training related to climate change** seeks to understand whether companies make disclosures related to training sessions to improve corporate awareness of climate change risks (GRI, 2020; United Nations, 2015).
- 27. Risks and opportunities related to climate change** seek to understand whether climate change risks may pose threats to companies and whether the company identifies opportunities that can address or mitigate climate change risks (GRI, 2020; United Nations, 2015).

Table 4.5 Indicators for assessing SDG 13 Climate change disclosures

21	Does the company include SDG 13 or acknowledge activities related to energy in their reports (regardless of whether they specifically mention SDGs)?
22	Does the company disclose corporate spending or resource commitments to SDG 13?
23	Does the company report direct GHG emissions (Scope1)?
24	Does the company report energy indirect GHG emissions (Scope2)?
25	Does the company report GHG emission information in reports?
26	Does the company report corporate education and training to raise corporate awareness on climate change?
27	Does the company report risks and opportunities related to climate change?

Overall, this thesis develops 27 quantifiable and consistent measures for observing corporate SDG reporting in Chinese companies.

Table 4.6 SDG disclosure framework for Chinese companies

	Indicators	Reference
1	Whether companies explicitly mention SDGs in ESG reports	
2	Whether companies mention SDGs or Agenda 2030 in Chairman or CEO message?	Subramaniam et al. (2020)
3	Whether the report acknowledges Agenda 2030 or SDGs in 'About the report' or 'Preparation Basis' section?	Clarkson et al. (2008)
4	Does the company include SDG 1 or acknowledge activities related to poverty alleviation in their reports (regardless of whether they specifically mention SDGs)?	UNGC & GRI (2018)
5	Does the company disclose corporate spending or resource commitments to SDG 1?	Indicator developed based on SDG 1a, FS7, and G4-EC8
6	Does the company report any information related to government-related poverty alleviation activities?	
7	Does the company report their industrial development activities to support poverty alleviation in rural areas?	State Council of People's Republic of China (2016)
8	Does the company report information related to poverty alleviation through health improvement?	
9	Does the company report information related to poverty alleviation through bottom-up protection?	
10	Does the company report information related to poverty alleviation through social poverty alleviation?	
11	Does the company report information related to poverty alleviation through education?	
12	Does the company report information related to poverty alleviation through employment?	
13	Does the company report their contributions to poverty alleviation through ecological improvement?	
14	Does the company report information related to poverty alleviation through regional development?	
15	Does the company report information related to poverty alleviation through relocating for people?	
16	Does the company include SDG 7 or acknowledge activities related to energy in their reports (regardless of whether they specifically mention SDGs)?	UNGC & GRI (2018)
17	Does the company disclose corporate spending or resource commitments to SDG 7?	SDG7.3
18	Does the company report their renewable energy consumption?	GRI et al. (2015)
19	Does the company report their energy consumption?	'Linking the SDGs and GRI' guideline
20	Does the company report energy efficiency?	
21	Does the company include SDG 13 or acknowledge activities related to energy in their reports (regardless of whether they specifically mention SDGs)?	UNGC & GRI (2018)
22	Does the company disclose corporate spending or resource commitments to SDG 13?	SDG 13a
23	Does the company report direct GHG emission (Scope1)?	GRI et al. (2015)
24	Does the company report energy indirect GHG emission (Scope2)?	'Linking the SDGs and GRI' guideline
25	Does the company report GHG emission information in reports?	
26	Does the company report corporate education and training to raise corporate awareness on climate change?	SDG13.3
27	Does the company report risks and opportunities related to climate change?	GRI et al. (2015) 'Linking the SDGs and GRI' guideline

Text location

Identifying the text location for each indicator is important in data capture. In manual content analysis, human coders can locate relevant text for each goal and interpret text, based on their understanding, by skimming the report or the table of contents. However, without instructions or coding schemes, the automated tool is unable to identify the correct sections to collect data for each indicator. This will lead to omitting context-specific information that requires human coders' judgements and incorrect data capture.

PwC Slovakia (n.d.) and Tenuta and Cambrea (2022) suggest how companies should prepare their sustainability reports and what should be included in their reports. These studies highlight that companies should provide several basic areas: general requirements, environmental information, social information, and governance. Similarly, the latest GRI guideline (G4) also introduced guidelines related to general standard disclosures, such as strategies, stakeholder engagement, and organisational profile, and significant economic, environmental and social impacts.

With the preliminary review of ESG reports, this thesis observed that a sustainability (ESG) report can be classified into non-performance (generic information) and ESG performance sections.

- Generic information sections emphasize corporate commitments, objectives, planning, strategies and future planning related to ESG performance. Sections may include 'Chairman message', 'President's statement', 'About the report', 'Stakeholder communication', 'Outlook', 'Index', and 'Glossary'.
- ESG performance sections offer a detailed view of corporate sustainability performance, which is typically aligned with SDGs. Sections may cover the following topics: climate change, energy, human resources, poverty alleviation, social welfare, international cooperation, and equality.

The preliminary analysis observed that very limited SDG performance information is disclosed in non-performance sections. Therefore, to facilitate data capture, this thesis collected information related to corporate explicit references to SDGs in generic information sections and SDG 1, 7, and 13 performance information from ESG performance sections. To ensure that performance information is not overlooked in generic sections, this thesis uses these indicators to collect relevant SDG performance data from generic sections:

- Whether the company mentions SDG 1 or acknowledges activities related to poverty alleviation in reports.
- Whether the company mentions SDG 7 or acknowledges energy-related activities in reports.
- Whether the company mentions SDG13 or acknowledges climate change related activities in reports.

Unit of analysis

Unit of analysis refers to the portion of content that is used to form the basis of coding in content analysis (Milne & Adler, 1999). It can be words, sentences, paragraphs, sections or the whole page. Given the importance of the broader context and data interpretability, this thesis selects paragraphs as the unit of analysis for observing the extent of SDG disclosures.

Prior studies widely select words or sentences as the unit of analysis in manual content analysis, due to their simplicity and efficiency in data collection (Guthrie & Abeysekera, 2006; Krippendorff, 2018). Coders can also directly process the surrounding context and interpret data based on their understanding (Guthrie & Abeysekera, 2006; Krippendorff, 2018). However, a word without surrounding context may lack clear meaning or be misinterpreted. These may reduce the reliability of the data collected. Similarly, sentences within one paragraph may convey similar messages and could be evidence for the same indicator. If the study codes them as separate pieces of evidence, a larger set of keywords was developed to take into account the surrounding context and to ensure that all relevant information was collected by Python. This will introduce unnecessary complexity and redundancy to data collection process. Therefore, following Shevchenko et al. (2023), this thesis selects paragraphs as the unit of analysis. Typically, a paragraph represents complete thoughts or talks about the same idea, which may reduce noise caused by word or sentence-level units of analysis in automated content analysis. The preliminary analysis also shows that a greater number of less representative keywords were used when selecting sentences as the unit of analysis. Given that this thesis will analyse 300 observations, paragraphs may be viewed as a balanced unit of analysis without sacrificing its coherence and relevance.

Keyword selection, refining, and validation

This section introduces the setup for using Python to conduct the dictionary-based automated content analysis. To replicate the manual process, keywords for each indicator are the core components of the dictionary-based automated analysis. This section discusses keyword development, refining and validation.

Keyword development

Given that no corpus list was available for SDG disclosures, this thesis developed its corpus for each indicator. Keywords are initially developed based on existing guidelines, such as the SDG framework, the Chinese government document— ‘The 'Notice on the 13th Five-Year Plan for Poverty Alleviation’, the GRI framework, and prior studies. They were refined during the preliminary analysis. This thesis also generates keywords based on data collected through NVivo (Hussain et al., 2020) and the LDA model¹⁶ (Bae et al., 2023; Jiang et al., 2024; Ranta et al., 2023).

¹⁶ Latent Dirichlet Allocation (LDA) model is an unsupervised machine learning method for topic modelling (Bae et al., 2023; Jiang et al., 2024; Ranta et al. 2023). It serves as the basis for classifying text into different topics (Jiang et al., 2024). The

Table 4.7 TF-IDF keyword results for poverty alleviation through education

keyword	score	frequency
students	100	17
program	76.33819	16
stations	73.22191	7
cnpc	72.21839	13
service	68.898	6
rmb	63.19388	7
sichuan	62.00884	7
support	58.51859	7
henan	57.03769	8
school	56.21065	10
families	52.88876	6
high	52.39499	9
complete	49.91059	5
studies	48.35177	6
xuhang	45.1678	6
public	44.3053	7
poor	43.74637	4
provided	43.36904	3
training	40.55157	7
education	40.3344	7
poverty	38.84655	6
million	38.6559	4
welfare	38.08491	6
college	38.0743	4
china	36.67087	7

Overall, drawing on these approaches, a dictionary has been developed for each indicator using the test sample.

Keyword refining

Keywords are initially developed from the source, such as, GRI, prior studies, and government documents, which have been widely used and tested. However, Potter and Levine-Donnerstein (1999, p. 258) argue that content analysis should not be limited to existing coding rules or standards established

by prior studies or experts. Therefore, the keywords are refined during the preliminary set-up, refining and validation stages. During this process, some issues were identified and needed to be addressed.

Refining the boundary of indicators and keywords

- ***The definition of climate change.*** Some reports disclosed water, waste, and emission information under sections heavily on SDG 13 or climate change. For example, CNPC (2018) incorporated climate change as a sub-section in the ‘Responsible Operation’ section. This section also includes information on safe operation, environmental protection, products and services, and supply chain management. However, the SDG framework classifies energy, water, consumption and production, climate change, and biodiversity into different goals, with SDG 13 specifically focusing on emissions. Prior studies widely treat corporate carbon emissions as climate change disclosures (Cotter & Najah, 2012; Giannarakis et al., 2017). Therefore, this thesis observes only corporate emission information as evidence of climate change disclosures.
- ***The definition of ‘government-related institutions’*** for ‘whether companies involve in government-related poverty alleviation activities’. Consistent with Chang et al. (2020), this thesis includes all government-related institutions. During the manual data collection, this thesis found that companies cooperate with Chinese philanthropic foundations to implement poverty alleviation activities. According to Chang et al. (2020), foundations can be classified into NGO foundations and foundations with strong government connections. Therefore, this thesis expands the keyword list to include China Foundation for Poverty Alleviation, the Red Cross Foundation, the local poverty alleviation office, and the China Environmental Protection Foundation in the keyword list.
- ***The definition of ‘renewable energy’.*** It was observed that clean energy, unconventional energy, and new energy are used in some reports. For example, CNPC, an oil and gas company, considers natural gas as clean energy in reports; however, it does not clearly define clean energy in its reports. To establish a clear definition for renewable energy, this thesis reviewed GRI 302 (2016) to see how renewable energy is defined (see Table 4.8) and followed GRI 302 to define renewable energy.

Table 4.8 Examples of renewable vs non-renewable energy extracted from GRI 302 (2016, p.14)

Non-renewable energy source	Non-renewable energy sources can include fuel distilled from petroleum or crude oil, such as gasoline, diesel fuel, jet fuel, and heating oil; natural gas, such as compressed natural gas (CNG), and liquefied natural gas (LNG); fuels extracted from natural gas processing and petroleum refining, such as butane, propane, and liquefied petroleum gas (LPG); coal; and nuclear power.
Renewable energy source	geothermal, wind, solar, hydro, and biomass

- **The definition of bottom-up protection.** As discussed in Section 4.6.1, the bottom-up protection targets at the ‘three left-behind’ group and the disabled people. The ‘three left-behind’ group refers to the elderly, left-behind children, and orphanages. Observations show that companies also disclosed information related to women in rural areas, children in rural areas, and assistance towards people suffering from earthquakes or other natural disasters. This thesis returned to the Chinese government document to see how the government defines ‘three-left-behind’ group. Therefore, information concerning women and children in rural areas is considered evidence of bottom-up protection. However, assistance for those affected by earthquakes or other natural disasters is not included as they are not within the boundary of the ‘three left-behind’ group.
- **Energy consumption.** It has been revised to include both numerical and qualitative data.

Removal of irrelevant keywords from the dictionary

- **Scope 1.** A review of GRI 302 guideline assists in refining the definitions of Scope 1 and Scope 2, as some manual data was missing in Python results. For example: ‘COD emissions (metric tons) 31,300 28,800 28,700 28,500 *’ was misinterpreted as evidence for Scope 2 (CNPC, 2018, p. 70).
- **Scope 2.** Some irrelevant information was collected due to the keyword ‘electricity’. For example, CNPC (2018) stated ‘In 2018, we launched construction of intelligent gas stations to upgrade part of existing stations to multi-energy stations that provide gasoline, natural gas and **electricity**’ (CNPC, 2018, p. 44). However, this thesis intends to collect emissions from electricity as evidence for Scope 2, rather than electricity consumption or other electricity activities. This thesis returned to the GRI guideline to see how it defines the boundary of Scope 2 emissions. GRI 305-2 (2016) states that Scope 2 emissions ‘include, but are not limited to, the CO₂ emissions from the generation of purchased or acquired electricity, heating...’. Therefore, this thesis refines the keyword list of Scope 2, by removing electricity from the

current keyword list and acknowledges that the removal of keyword ‘electricity’ may lead to some relevant evidence being overlooked.

- **Poverty alleviation through social impacts.** It was observed that some keywords for poverty alleviation through social impacts are less representative and even not appropriate, such as loan, cooperative, earthquake, and infrastructure development. For example:
 - Providing loans to rural areas can be a way for the company to conduct paired-up assistance and can also be evidence of corporate spending or resource commitment related to SDG 1.
 - Regarding infrastructure development, whether the company reports poverty alleviation related to regional development also includes infrastructure development. Therefore, it will be removed from the keyword list for social impacts.

Therefore, this thesis again reviewed the ‘Notice on the 13th Five-Year Plan for Poverty Alleviation’ to refine the boundary for poverty alleviation through social impacts. The keywords mentioned above were removed from the list as they add complexity to the automated process. Additionally, this thesis also includes the following keywords to collect poverty alleviation through social impacts, including paired-up, designated, fixed-point poverty alleviation, and cadres. These are highly relevant to paired-up assistance in rural areas.

- **Poverty alleviation through employment.** Table 4.9 displays that Air China (2016, p. 55) argues ‘... carry out village work’, ‘pursue a green office in daily work’, ‘ensuring energy supply for all assignments’, and ‘.. cultural and musical work’ are collected by Python, considered as evidence for poverty alleviation through employment. However, the meaning of ‘work’ was interpreted as daily work rather than referring to job vacancy or employment opportunities provided for people in rural areas. Therefore, ‘work’ was removed from the keyword list for poverty alleviation through employment.

Table 4.9 irrelevant data for poverty alleviation through employment

Air China (2016, p. 55)	<p>Case study: Air China successfully guaranteed G20 Summit Representatives from our Zhejiang Branch swear to guarantee the ...targeted poverty alleviation work, making full use of its advantages to develop targeted poverty alleviation projects and implementation plans. By dispatching delegated poverty alleviation cadres to Mading Village to carry out village work and “poverty alleviation day” activities and organizing employees to subscribe for local agricultural and Deepening the Targeted Poverty Relief</p> <p>..... An employee from Air China is presented the award...</p> <p>Case study: An employee from Air China won the “Chinese – Hungarian People’s Friendship Award”</p>
China Life Insurance (2016, p. 16)	<p>Moreover, we have vigorously pushed ahead short-term health insurances for individuals and groups. In 2016,11... (II) Green office We encourage employees to practice green and environmental</p>

	concepts and pursue a green office in daily work . We also advocate all -staff engagement into the promotion of energy saving and environmental protection in order to ...
--	---

- In reviewing the results of **renewable energy consumption**, Table 4.10 shows that some companies do not use the word ‘consumption’. Alternative keywords like ‘utilisation’ and ‘use’ have been employed. However, data typically quantitatively discussed their actions and were context specific. Therefore, this thesis does not expand the keyword list for renewable energy consumption.

Table 4.10 irrelevant data for renewable energy consumption

CNPC (2018, p. 28)	With an eye on the future, we continued to promote the development and utilization of renewable energy including geothermal energy and biofuels. In 2018, CNPC continued to expand its new energy business, such as geothermal energy, solar energy, biofuels, and charging and battery swap stations.
--------------------	---

- Reviewing **energy efficiency** data collected by Python, this thesis found that some irrelevant data was collected based on the keywords ‘efficiency’ and ‘efficient’. The data collected is related to management efficiency and efficiency of leased assets, rather than energy efficiency.
- **Energy spending or resource commitment**. This thesis observed that ‘RMB’, ‘CNY’, and ‘Yuan’ are monetary symbols. Information collected based on these keywords is related to green credit, annual income, and cost reduction rather than energy spending or resource commitment. For example, CNPC (2016, p. 2) stated that ‘Technological innovation created more than 100 billion Yuan during ...’, which shows corporate gains from technological innovation, rather than energy spending or resource commitment. Another example is ‘corporate contribution of RMB 116.96 million and RMB 431.07 million to the enterprise annuity’ from Air China (2016, p. 48). Again, they are neither relevant to energy nor relevant to investment. Therefore, less representative keywords will be removed from the keyword list for energy spending or resource commitment.
- **Climate change spending or resource commitment**. Consistent with energy spending or resource commitment, RMB, Yuan, and CNY are less representative as keywords for spending or investment. For example, ‘the belly hold revenue increased by more than RMB 300...’ from Air China, and ‘saving about 60% of energy costs or about RMB 900,000 Yuan per year’ from Air China focused on revenue or cost saving, rather than corporate energy spending or resource commitment (See Appendix 4). Therefore, this thesis removes these keywords from the keyword list for climate-related spending or resource commitment.

Keyword validation

This section moves to keyword validation to evaluate the reliability and effectiveness of the automated tool. According to Krippendorff (1980) and Bolognesi et al. (2017b, p. 1988), it includes the following three areas: stability, replicability, and accuracy.

Part 1 Accuracy

Accuracy refers to ‘the degree to which a process functionally conforms to a known standard, or yields what it is designed to yield’ (Bolognesi et al., 2017a; Krippendorff, 1980, p. 131). Accuracy plays a vital role in enhancing the likelihood that correct information can be coded during data collection. It is advised that a third-party human coder is required to mitigate the core concerns related to content validity, inter-coder reliability and accuracy, especially utilising dictionary-based methods, (Bochkay et al., 2023; Neuendorf, 2010). To validate the results and ensure the accuracy of data, prior studies widely incorporate human validation in automated content analysis during the initial development and post measurement stages (Song et al., 2020). During the validation in the initial development stage, human coders are able to conduct ‘problem-specific’ validation, aiming to ensure the accuracy and reliability of developing dictionary and the automated tool. This could be done through micro-sampling for manual validation (Lappeman et al., 2020). Specifically, despite being an automated tool, it still requires humans to develop, refine, and validate the whole process until the model can replicate the manual process and manually collected data (Grimmer & Stewart, 2013). In this thesis, accuracy heavily relies on the quality of dictionary used by the automated tool and training provided to the coder.

To improve the accuracy of coding schemes in the manual process, this thesis involves another expert coder to check coding consistency and accuracy of data interpretation. My supervisor was involved in a consistency check by randomly selecting several reports to see whether any information was not collected based on my coding schemes or whether inconsistency in data interpretation existed. If inconsistencies occur in the content coded, I will review them, followed by refinements in coding schemes and the boundaries of each indicator. Based on the updated coding schemes, I will then reiterate the manual data collection process. This process will continue until there are no significant differences in the interpretation of coding schemes between me and my supervisor. This thesis then moves to keyword development, refining and validation, aiming to improve the reliability of the automated tool.

Part 2 Replicability

Replicability refers to the extent to which annotation will remain the same or high similarity, when different researchers analyse the data, based on the given coding scheme and process (Bolognesi et al., 2017b). To achieve replicability, multiple coders are normally employed to use the same coding scheme, data sources, and instruments, to test the replicability of the study. This process is conducted on a back-to-back condition and coders will be trained before data collection. However, this thesis relies on Python to conduct automated content analysis. Replicability is 're-defined', as this thesis relies on Python to

replicate the results of the human coder. Once the reliability of the coding schemes and keywords has been assured, the entire process should be replicable. In this thesis, replicability is conducted at content and score two different levels, by testing the similarity between manual and automated data collected.

Part 2.1: Content-level similarity tests

Content-level similarity tests are designed to evaluate how effectively the automated tool can replicate the manually collected data. Similarity measurements can be classified into string-based, corpus-based, semantic text matching, and graph-structure-based methods. Given that this thesis conducted keyword searching during automated content analysis and outputs paragraphs that contain the keyword, this thesis is limited to exact word matching rather than looking at the semantics of the text. Therefore, this thesis focuses on the string-based method and corpus-based method to test the similarity between texts.

- The string-based method evaluates similarity by examining ‘string sequences and character composition that measures similarity or dissimilarity between two text strings for approximate string matching’ (Wang & Dong, 2020, p. 421). This method is straightforward in calculating the similarity of text, as it explores exact matches between documents, which makes the order of words important. However, it is primarily constrained to short strings.
- The corpus-based method ‘uses the information obtained from the corpus to calculate the text similarity’ (Wang & Dong, 2020, p. 421). This method includes techniques, such as Bag-of-Words and term frequency–inverse document frequency (TF-IDF). Bag-of-Words simply looks at the frequency of the word, lacking the capability to capture the semantic meaning of words or to take into account the order of words. However, TF-IDF shows the result of the importance of a word is inversely related to its frequency across documents. Unlike Bag-of-Words, this method can weigh down the frequent words and increase the weight of rare words. Additionally, unlike the string-based method, TF-IDF can perform well with a larger document. A similarity score will be generated, ranging between 0 and 1. A score of 1 refers to the maximum similarity. The following equation shows how the TF-IDF score is calculated. $Freq(w,d)$ indicates the frequency of a term used in a document, while $|D|$ indicates the number of documents analysed. $N(w)$ refers to the number of documents that contain the word.

$$TF-IDF(w, d, D) = Freq(w, d) * \log \frac{|D|}{N(w)}$$

This thesis has implemented TF-IDF to see text similarity between manual and automated results. Initially, a single company was selected as a test sample. The similarity scores were less than 30%. It was observed that all manually collected data was captured by Python, while Python results also included some irrelevant or overlooked data. To address this issue, this thesis removed some irrelevant data in Python results and re-ran the similarity test, with the score increasing to 44%. Therefore, TF-IDF is heavily affected by all content captured.

To resolve these issues, a simple proportion method is also implemented in this thesis to explore whether data manually collected can be captured by Python, and to determine the percentage of manually collected data that can be captured by Python. It was conducted at two levels: the word and the paragraph level (Kenter & De Rijke, 2015; Prakoso et al., 2021).

- At the word level, this thesis calculates the proportion of words from the manual results that occur in the automated results relative to the total number of words in manual results. To implement the test, this thesis, again relies on data pre-processing techniques to remove stop words and lower the case sensitivity to reduce the noise in the comparison. However, it does not take the word order and semantics of sentences into consideration.
- This thesis also calculates text similarity at the cell (paragraph) level. This method considers each cell in the manual data collection spreadsheet as one unit of analysis. This thesis matches it with Python results to see whether the manually collected content in the cell occurs in any of the cells in Python results. A score of one is assigned to the cells that contain overlapping content with the Python. This thesis subsequently calculates the number of overlapping cells in manual results, relative to the total number of cells in manual results. This may indicate the percentage of manual results that can be replicated by Python programming.

$$\text{Similarity score} = \frac{\text{data collected manually that overlaps with data collected automatically}}{\text{data collected manually}}$$

- This thesis also tests the reliability of automated data collection itself by testing the proportion of overlapping data out of automatically collected data. Keyword searching does not take semantics into consideration and as such, paragraphs that contain keywords are directly extracted. This leads to an issue that some irrelevant paragraphs are captured, due to the occurrence of keywords in those sentences. Therefore, this thesis also calculates the number of cells in Python results that overlap with manual results out of the total number of cells in Python results, aiming to identify the content not collected by the manual process. If the proportion is low, this thesis will find out the reasons for the lower-than-expected similarity scores and refine the coding schemes.

$$\text{Similarity score} = \frac{\text{data collected automatically that overlaps with data collected manually}}{\text{automatically collected data}}$$

The scores range from zero to one, with one indicating a perfect replication of manual data collection and zero indicating no match occurring.

Table 4.11 displays the results of the similarity between manual and automated data, based on the simple proportion method. Out of the 27 SDG performance indicators, 21 indicators received a score exceeding

0.8, indicating a substantial consistency between manual and automated results and effective automation (Anastasei & Georgescu, 2020; Cavnar & Trenkle, 1994). Among all indicators, corporate explicit reference to SDG indicators, whether companies mention SDG 1, 7, or 13, whether companies report any energy investment and whether companies involve in education and training related to climate change received the highest similarity scores, above 0.95. These are followed by health care activities to reduce poverty alleviation, renewable energy consumption, energy consumption, and relocating activities to reduce poverty alleviation, with scores all above 0.9. These indicate that Python programming can effectively utilize the existing coding scheme, keyword lists, and other settings to validate and automate the manual process efficiently. For these indicators, automated data collection can replicate the results obtained through manual data collection.

Indicators like relocating, employment, energy efficiency, climate change investment, poverty alleviation spending, industrial development, government partner, education, regional development social impacts, have slightly lower, but reasonable scores, ranging from 0.89 to 0.80. This indicates reasonable consistency between manual and Python data results.

Table 4.11 Similarity test based on the simple proportion method

	manual data collection	After refinement
energy investment	0.96	
education & training	0.95	
health care	0.91	
renewable energy consumption	0.90	0.91
energy consumption	0.90	
relocating	0.90	
employment	0.89	
energy efficiency	0.89	0.97
climate change investment	0.88	
poverty alleviation spending	0.86	
industrial development	0.86	
govt partner	0.85	
Education	0.84	

regional development	0.82	
social impact	0.82	0.97
Bottom-up protection	0.80	0.82
GHG emission	0.77	
risk and opportunities	0.66	0.71
Scope 2	0.44	0.81
Scope 1	0.33	0.76
Acknowledge SDG 1	1	
Acknowledge SDG 7	1	
Acknowledge SDG 13	1	
Mention SDG in reports	1	
Mention SDG in Chairman message	1	
Acknowledge Agenda 2030 or SDGs in 'About the report' or 'Preparation Basis' section	1	

As shown in Table 4.11, some indicators have lower similarity scores, such as Scope 1, Scope 2, risks and opportunities related to climate change, GHG emissions, and bottom-up protection. This thesis manually checks the differences between manual and automated data collected. This thesis repeats the process starting from keyword refining to the end of data validation. The refinements are from the following three areas: refining the boundary of indicators and keywords, removing less representative keywords, and refining coding scheme (see Section ‘Keyword Refining’).

- **Scope 1 and Scope 2.** This thesis has removed irrelevant keywords for these indicators—electricity and COD emissions.
- **Poverty alleviation through social impacts.** This thesis reviewed the government policy document and included ‘pair-ed up, designated, fix-point poverty alleviation, and cadres’ in the keyword list and removed other less representative keywords (see Section ‘Keyword Refining’). The content-level similarity score has improved from 0.82 to 0.97.

Throughout the comparison, this thesis observed that some overlapped content was not captured by Python programming. Table 4.12 shows an example:

Table 4.12 An example of inconsistent results

Manual results	Energy structure keeps changing. Despite the accelerated energy structural transition to clean and low-carbon development, and is expected to reach 8.3~10% of primary energy consumption mix by 2020' (CNPC, 2017, p. 21).
Python results	'Energy structure keeps changing. Despite the accelerated energy structural transition to clean and low-carbon development, and is expected to reach 8.3~10% of primary energy consumption mix by 2020' (CNPC, 2017, p. 21).

This thesis has tried removing all asterisks (*) from cells, which were added during automated data collection to highlight relevant keywords. This thesis also removed line breaks to ensure consistent content formatting. However, Python programming still could not recognise them as the same content. After numerous trials, this thesis found that the Python result contained extra spaces, which made them different. Therefore, this thesis modifies Python codes to normalise the spaces in both results. Relevant codes are as follows:

```
# Remove line breaks from each cell in column 'A' and 'B'
```

```
df['A'] = df['A'].str.replace('\n', '')
```

```
df['B'] = df['B'].str.replace('\n', '')
```

```
# Remove "*" from each cell in column 'A' and 'B'
```

```
df['A'] = df['A'].str.replace('*', '')
```

```
df['B'] = df['B'].str.replace('*', '')
```

```
# Normalize the spaces
```

```
df['A'] = df['A'].str.replace(' +', '')
```

```
df['B'] = df['B'].str.replace(' +', '')
```

After making multiple-round adjustments, the similarity scores for Scope 1 and Scope 2 indicators increased to 0.7667 and 0.81, respectively.

Overall, with these refinements, the similarity scores for all indicators are over 0.75. With the highest score at 0.97 and the lowest score at 0.76, the content-level similarity between manual and Python results has reached an acceptable level (Cavnar & Trenkle, 1994).

Part 2.2: Score-level consistency tests

Given that this thesis aims to observe what companies have disclosed over time, it moves to assess score-level text consistency between manually and automatically collected data. This involves verifying the presence of evidence for every indicator. To test it, if one piece of evidence has been found for an indicator, a score of one will be assigned to that company regarding that indicator, otherwise zero assigned.

Table 4.13 displays the score-level consistency between manual and automated results for all pre-testing observations regarding SDG 1, 7 and 13, respectively. When inconsistent results are found, ‘False’ will be assigned to that observation. Again, the refining process was conducted with keyword refining, the removal of less representative/ irrelevant keywords, and coding refining three aspects.

Poverty alleviation. Of all these indicators, a 100% consistency rate between manual and automated data collection has been reached for SDG1 and related activities, SDG 1 spending or resource commitment, health care, and regional development. This suggests a robust capability of the automated tool to replicate manual data collection effectively. Social impact and relocating received 94% consistency between manual and automated data collection, indicating Python programming can effectively utilize the existing coding scheme, keyword lists, and other settings to validate and automate the manual process efficiently. This is followed by industrial development, employment, bottom-up approach, and education, with an acceptable consistency score of 0.89. However, with consistency scores of 0.83 and 0.5, the coding schemes and keyword lists for government-related activities, and ecological impacts were further refined.

Refinements:

Poverty alleviation through employment. As discussed in Section ‘Keyword refining’, this thesis removes ‘work’ from the keyword list for poverty alleviation through employment. The overall consistency score increased from 0.89 to 0.94.

Poverty alleviation through ecological impact. With the consistency score at 0.5, it indicates that a significant discrepancy occurred between manual and automated results. As shown in Table 4.13, collected evidence is limited to ecological protection and environmental protection, which are not relevant to poverty alleviation. To resolve this, this thesis adopts a two-step approach, by identifying ecological protection information in the first step searching and then identifying whether it is relevant to poverty alleviation in the second step searching. The consistency score for this indicator dramatically increased from 0.5 to 0.89 (See Part 2.3: second round score-level consistency test). An example is shown below (See Appendix 5):

Table 4.13 An example of poverty alleviation through ecological impacts

CNPC (2019, p. 73)	<ul style="list-style-type: none"> • We strictly follow safety and environmental protection regulations, formulate management systems on environmental protection that involve participation of ... <hr/> <p>Our employees picked up 225 kg of solid waste in 15 bags, supporting local environmental protection. 77</p> <p>Local environmental protection.</p>
--------------------	---

Table 4.14 shows the results of score-level consistency between manual and automated data collection regarding energy activities. Of all these indicators, acknowledging energy activities in the report

receives the highest consistency rate of 100%, while the lowest rate is only at 56% related to energy investment. With a score of 89%, the reliability of the coding scheme and keywords can be viewed as reliable for energy consumption. However, criteria for collecting energy spending or resource commitment, renewable energy consumption and energy efficiency were refined as follows:

1) The refinements of the boundary of indicators and keywords

Energy efficiency. Given that the consistency score is at 0.83, which is at an acceptable level, this thesis does not make changes to this indicator. In Part 3 Stability tests, this thesis selected 10 new observations to test the text similarity between manual and automated content analysis. It is observed that the similarity score for energy efficiency is at 100%, higher than 89% for the preliminary analysis.

2) The refining of coding schemes

Energy spending or resource commitment. *Irrelevant* data was misclassified as evidence for energy spending or resource commitment. It was observed that ‘.....increase investment of charity foundation in poverty alleviation...’ in China Life Insurance 2016 ESG report, ‘.....encourage disclosure and reporting of investment impact...’ in ICBC 2016 ESG report, and ‘...the bank made an investment of RMB 1 billion in a power battery producer through an industrial fund..’ from ICBC 2017 ESG report are some examples (See Appendix 5). Evidence did discuss corporate funding or investments; however, they focused on poverty alleviation, battery production, and employees, rather than energy spending or resource commitment. This can be explained by the fact that some reports discuss their poverty alleviation, energy, and climate change information in the same section (See Section 4.6.1--Text location). Given that the same keyword list is used for collecting spending or resource commitments for these three dimensions, current Python programming was unable to understand the thematic context and automatically classify them into poverty alleviation, energy, or climate change information if information was discussed in the same section in reports. This results in potentially accumulating irrelevant or duplicate data based on the existing coding schemes. To address this issue, the two-step method will also be used to collect energy spending or resource commitment data

Climate change. Table 4.14 also displays the consistency results for climate change indicators. Among all the indicators, acknowledging SDG 13 activities receives a 100% similarity rate, indicating that Python programming can perfectly replicate manual collection results regarding whether discussions have been provided for each indicator. Scope 1 and Scope 2 received rates at 94%, indicating that Python programming can effectively utilize the existing coding scheme, keyword lists, and other settings to validate and automate the manual process efficiently. For these indicators, the automated data collection can closely replicate the results obtained through manual data collection. Additionally, the indicator related to risks and opportunities associated with climate change achieved a commendable similarity

score of 83%. Conversely, indicators concerning climate change investments, education and training, and emissions reflected limited consistency, with scores at 66%, 22%, and 78% respectively.

- ***Training and education related to climate change.*** With a low consistency score at 0.222, this thesis reviewed the results to identify the differences between manual and automated results regarding training and education related to climate change. It was observed that Python results were heavily affected by irrelevant data, as some employment training and energy information were misclassified as evidence for training and education related to climate change. This thesis applies the two-step approach to this indicator.

Overall, this thesis **1.** removes less representative keywords from keyword lists and **2.** applies the two-step approach to the following indicators: poverty alleviation spending and resource commitment, energy spending and resource commitment, climate change spending and resource commitment, training and education related to climate change, risks and opportunities related to climate change, and poverty alleviation through ecological protection (Details see Section ‘keyword refining’ and Part 2.2: Score-level consistency tests).

Table 4.14 Score-level consistency test

Indicators	SDG 1 & acknowledge activates	spending or resource commitment	govt partner	industrial development	employment	social impact	health care	bottom-up approach	education	ecological impact	regional development	relocating	acknowledge energy activities (SDG7)
CNPC 2016	1	1	1	1	1	1	1	1	1	FALSE	1	1	1
CNPC 2017	1	1	1	1	1	1	1	1	1	FALSE	1	1	1
CNPC 2018	1	1	1	1	1	1	1	1	1	FALSE	1	1	1
CNPC 2019	1	1	1	1	1	1	1	1	1	FALSE	1	FALSE	1
Air China2016	1	1	1	1	FALSE	1	1	1	1	1	1	1	1
Air China2017	1	1	FALSE	1	1	1	1	1	1	FALSE	1	1	1
Air China2018	1	1	1	1	1	1	1	1	1	1	1	1	1
Air China2019	1	1	1	1	1	1	1	1	1	1	1	1	1
China Life Insurance2016	1	1	1	1	1	1	1	1	1	1	1	1	1
China Life Insurance2017	1	1	1	1	1	1	1	1	1	1	1	1	1
China Life Insurance2018	1	1	1	FALSE	1	1	1	1	1	1	1	1	1
China Life Insurance2019	1	1	FALSE	1	1	1	1	FALSE	1	FALSE	1	1	1
ICBC 2016	1	1	1	1	1	1	1	1	1	FALSE	1	1	1
ICBC 2017	1	1	1	1	1	1	1	1	FALSE	1	1	1	1
ICBC 2018	1	1	FALSE	1	1	FALSE	1	1	1	1	1	1	1
ICBC 2019	1	1	1	1	1	1	1	1	1	FALSE	1	1	1
Yili 2017	1	1	1	1	FALSE	1	1	1	1	FALSE	1	1	1
Yili 2018	1	1	1	FALSE	1	1	1	1	1	1	1	1	1
	18	18	15	16	16	17	18	17	17	9	18	17	18
	1	1	0.83	0.89	0.89	0.94	1	0.94	0.94	0.5	1	0.94	1

Indicators		energy investment	renewable energy consumption	energy consumption	energy efficiency	acknowledge climate change activities (SDG 13)	climate change investment	corporate education or training	total GHG emission	scope 1	scope 2	risks and opportunities
CNPC 2016		FALSE	FALSE	1	1	1	1	FALSE	1	1	1	1
CNPC 2017		FALSE	FALSE	1	1	1	1	FALSE	1	1	1	1
CNPC 2018		1	FALSE	1	1	1	1	FALSE	1	1	1	1
CNPC 2019		1	1	1	1	1	1	FALSE	1	1	1	1
Air China2016		1	1	1	1	1	1	FALSE	FALSE	1	1	1
Air China2017		1	1	1	1	1	1	FALSE	FALSE	1	1	1
Air China2018		1	1	FALSE	1	1	1	1	1	1	1	1
Air China2019		1	1	1	1	1	FALSE	FALSE	1	1	1	1
China Life Insurance2016		False	1	1	1	1	1	1	FALSE	1	1	1
China Life Insurance2017		FALSE	1	1	FALSE	1	FALSE	FALSE	1	1	FALSE	1
China Life Insurance2018		1	1	FALSE	1	1	1	FALSE	1	1	1	1
China Life Insurance2019		1	1	1	FALSE	1	1	FALSE	FALSE	FALSE	1	1
ICBC 2016		FALSE	1	1	1	1	FALSE	FALSE	1	1	1	FALSE
ICBC 2017		FALSE	FALSE	1	1	1	FALSE	1	1	1	1	FALSE
ICBC 2018		FALSE	1	1	FALSE	1	FALSE	FALSE	1	1	1	FALSE
ICBC 2019		1	1	1	1	1	1	FALSE	1	1	1	1
Yili 2017		1	1	1	1	1	FALSE	1	1	1	1	1
Yili 2018		FALSE	1	1	1	1	1	FALSE	1	1	1	1
		10	14	16	15	18	12	4	14	17	17	15
		0.56	0.78	0.89	0.83	1	0.66	0.2222	0.78	0.944444	0.944444	0.833333

Part 2.3: the second-round score-level consistency tests (after refinement)

Following these adjustments, this thesis re-ran the similarity tests to observe the level of consistency related to SDG 1, 7, and 13.

Table 4.15 displays the consistency levels of poverty alleviation data collected. The consistency scores for ‘acknowledging SDG1 activities’, ‘SDG1 spending or resource commitment’, and ‘health care’ remain the same, with scores at 100%. Scores for all the other indicators also increased. For example, scores for industrial development, social impact, education, and relocating increased to 100%. Corporate participation in government-related poverty alleviation activities increased from 0.83 to 0.94, while the score for ‘employment’ also increased to 0.94. The consistency scores for ‘regional development’ and ‘ecological impact’ are at 0.89. Therefore, the coding scheme and set up are relatively reliable to conduct automated data collection.

Table 4.15 Consistency test-- poverty alleviation

Indicators	SDG 1 & acknowledge activities	spending or resource commitment	govt partner	industrial development	employment	health care	bottom-up approach	social impact	education	regional development	relocating	ecological impact
first round	1	1	0.83	0.89	0.89	1	0.94	0.94	0.94	1	0.94	0.5
After adjustments	1	1	0.94	1	0.94	1	0.89	1	1	0.89	1	0.89

Table 4.16 displays the consistency levels of energy data collected. The consistency scores for all indicators have improved, with the highest score at 1 and the lowest at 0.83. Specifically, the score for ‘acknowledging energy activities’ remains 100%, while the score for ‘energy consumption’ also increased to 100%. The similarity scores for ‘renewable energy consumption’ and ‘energy efficiency’ increased to 0.83 and 0.89 respectively. Interestingly, after employing a two-step analysis for ‘spending or resource commitment’ indicators, the score for observing corporate energy spending or resource commitment dramatically increased from 0.56 to 0.89. Therefore, current coding schemes and setup are relatively reliable for replicating manual data processes for collecting SDG 7 information.

Table 4.16 Consistency test-- Energy

Indicators	acknowledge energy activities	Energy spending or resource commitment	renewable energy consumption	energy consumption	energy efficiency
first round	1	0.56	0.78	0.89	0.83
After adjustments	1	0.89	0.83	1	0.89

Table 4.17 displays the consistency levels of climate change data collected. The consistency score for the indicator ‘acknowledging climate change activities’ remains 100%, indicating that both manual and

automated processes have identified at least one piece of evidence for this indicator. Consistency scores for ‘climate change spending or resource commitments’, ‘corporate education or training’, ‘risks and opportunities’ and ‘total GHG emissions’ have improved after adjustments, such as the implementation of two-step analysis and the modification of keyword lists. The consistency between manual and automated data collected for climate change spending or resource commitments increased from 0.66 to 0.89. The consistency between manual and automated data collected for corporate education or training increased from 0.22 to 0.67. The consistency scores of ‘risks and opportunities’ and ‘total GHG emissions’ reached 100%. However, the score of ‘Scope1’ slightly decreased to 0.89. Given that all scores are over 0.8, except for the indicator ‘corporate training or education’, this indicates that the coding scheme and setup are relatively reliable to replicate the manual data collection process for collecting climate change information.

Table 4.17 Consistency test-- Climate change

Indicators	acknowledge climate change activities	climate change investment	corporate education or training	risks and opportunities	total GHG emission	scope 1	scope 2
first round	1	0.66	0.2222	0.78	0.944444	0.944444	0.833333
After adjustments	100%	0.89	0.67	100%	100%	0.89	0.89

Overall, after conducting several rounds of data collecting and refinement, current coding scheme, keyword lists, and set up are relatively reliable to conduct data collection.

Part 3: Stability tests

Stability refers to the ability to collect the same text over time, assuming that the same coder follows the same coding scheme in a given data source (Bolognesi et al., 2017). Given that this thesis adopts automated content analysis, stability refers to the extent to which existing coding schemes and keyword lists can be generalised to data collection for other reports. This thesis repeats the entire process by manually collecting data for these new observations based on the updated coding schemes to identify relevant pieces of evidence, executing the updated Python coding to collect data, and comparing the similarity between manual and automated results based on the updated coding schemes to see the stability of the automated tool, coding schemes, and the dictionary (keywords). Following the previous test sample selection criteria, Bank of China and Railway Group were selected. As stated before, this thesis selected the top-ranking companies as test samples due to their higher possibility of making SDG disclosures.

Table 4.18 compares the score-level similarity of poverty alleviation indicators between the initial 5 companies (23 observations, 2 observations without English reports) and the 2 new companies (10 new observations). The similarity scores of indicators for observing ‘SDG 1 acknowledgement activities’, ‘spending or resource commitment’, ‘industrial development’, ‘social impact’, ‘education’, and

‘relocating’ remain 100%. Interestingly, the similarity scores for the following indicators improved to 100%: ‘government activities’, ‘employment’, ‘bottom-up approach’, and ‘ecological impact’. However, the similarity score for healthcare decreased from 100% to 90%. As these scores are still over 80%, the criteria can be considered stable. Overall, the average score of poverty alleviation indicators improved from 96% to 98%, indicating that the coding schemes and setup are sufficiently stable to be used for automated data collection.

Table 4.18 Consistency test-- Poverty alleviation

Poverty alleviation	SDG 1 & acknowledge activities	spending or resource commitment	govt activities	industrial development	employment	social impact	health care	bottom-up approach	education	ecological impact	regional development	relocating	average score
10 new observations	100%	100%	100%	100%	100%	100%	90%	100%	100%	100%	90%	100%	98%
23 old observations	100%	100%	94%	100%	94%	100%	100%	89%	100%	89%	89%	100%	96%

Table 4.19 shows the consistency of test results of energy indicators. Out of the five indicators, the similarity score for ‘acknowledging energy activities’ decreased from 100% to 90%, while the similarity score for energy consumption remains the same, at a score of 100%. The similarity scores for ‘energy investment’, ‘renewable energy consumption’, and ‘energy efficiency’ all improved.

Table 4.19 Consistency test--Energy

Energy	SDG7 & acknowledging related activities	energy investment	renewable energy consumption	energy consumption	energy efficiency	average score
Energy: 10 new observations	90%	90%	100%	100%	100%	98%
Energy: 23 old observations	100%	89%	83%	100%	89%	92%

Table 4.20 compares the score-level similarity of climate change indicators between the initial 5 companies (23 observations) and the 2 new companies (10 new observations). Again, the similarity scores for observing ‘climate change activities’, ‘total emission’, and ‘risks and opportunities’ remain the same, at 100%. The scores for ‘Scope 1’, and ‘Scope2’ increased from 89% to 100%. Regarding the ‘training and education’, the score also reached 90%. However, the scores for observing corporate

‘climate change spending and resource commitment’ decreased from 89% to 70%¹⁷. Overall, the tests suggest that the automated tool is reliable for collecting climate change information.

Table 4.20 Consistency test-- climate change

<i>Climate change</i>	climate change	cc spending	total emission	scope 1	scope 2	training and education	Risk and opportunities	Average score
Climate change: 10 new observations	100%	70%	100%	100%	100%	90%	100%	93%
Climate change: 23 old observations	100%	89%	100%	89%	89%	67%	100%	91%

Limitations of automated content analysis

With multiple rounds of data collection and refinement, the automated tool is ready for collecting SDG disclosure data. This thesis needs to acknowledge some limitations that may influence the consistency between manual and automated results; however, they do not significantly affect the overall results. Potential limitations are shown as follows:

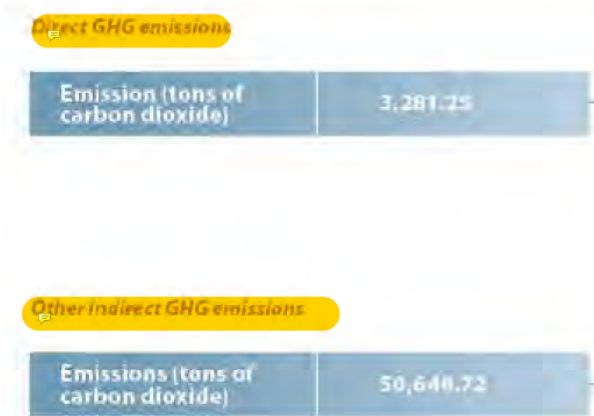
The semantics of keywords. The semantic meaning of some keywords may be interpreted differently and is irrelevant to this thesis. For example, this thesis adopts ‘industrial’ as a keyword for capturing corporate industrial development. It was observed that two pieces of evidence have been collected, where ‘industrial’ was part of a company’s name, specifically ‘2018 Inner Mongolia Yili industrial Group Co., Ltd’. This is obviously a piece of irrelevant data to the indicator.

Table 4.21 An example of the semantics of keywords

World Integrally Sharing Health 2018 Inner Mongolia Yili **industrial** Group Co., Ltd. Innovation Social Health Our Vision and Goals Win-Win About Yili Caring for the Youth Yili Ark Children’s safety education class In December 2018, on the (Yili, 2018, p.1).
Yili Ark’s Seven-Color Sails helps children learn to protect themselves Yili Ark released the first Company-wide Public Charity Initiative of China’s Dairy Industry The sixth anniversary of the “Yili Ark” Children Safety Public World Integrally Sharing Health 2018 Inner Mongolia Yili **industrial** Group Co., Ltd. Innovation Social Health Our Vision and Goals Win-Win About Yili Yili Future Park “Yili Future Park: Connecting ... (Yili, 2018, p. 48).

Data presentation challenge. Total GHG emissions in the 2018 ICBC CSR report and 2019 ICBC CSR report are in the image format, shown in the picture. Although the document was converted to an OCR-friendly version to facilitate Python readability, the Python coding faced challenges recognizing characters and figures within the images. This led to a discrepancy between manual and automated data collection consistency.

¹⁷ The consistency score was decreased to 70%, as irrelevant data was collected. Python has collected ‘... it attracted the subscriptions of many well-known green ******investment****** agencies’ as one piece of evidence for Bank of China (2017) and (2018). However, it is not related to climate change investment.



Formatting limitation. When extracting paragraphs from PDF files, some space may be removed, leading to several words combining into one. Since this thesis sets the boundary of collecting words that exactly match the keywords in the keyword lists, some keywords may be ignored. For example, Python cannot identify China Foundation for Poverty Alleviation in ‘JiangxiChina Foundation for Poverty AlleviationIn’. To address this issue, this thesis conducts an additional analysis, by replacing `rf\b{keyword}(?![a-zA-Z])` with `rf {keyword}` to capture all relevant words. However, irrelevant data will be collected, as some words may be part of another word. Based on the preliminary analysis, more irrelevant data is collected, as the keyword is part of another word, such as, grant-migrant, and yuan (currency)-songyuan (the name of a place), which affects the overall quality of data collected. Therefore, this thesis does not refine the coding.

The surrounding context. As shown in Table 4.22, Python programming extracted it as evidence for poverty alleviation through ‘participating in government-related activities’ using the keyword ‘government’. However, although it was extracted from the poverty alleviation section, it is related to the Belt and Road Initiative. Therefore, it is considered a limitation in this thesis.

Table 4.22 An example of the surrounding context

China Life Insurance (2019, p. 27)	For government officials and corporate legal persons participating in the Belt and Road initiative, China Life provides security for death and disability, medical expenses due to illness, accidents and emergency rescue services overseas, and overseas risk protection.
---------------------------------------	---

Keyword selection. Reviewing energy efficiency data collected by Python, this thesis found that some irrelevant data was collected based on the keywords ‘efficiency’ and ‘efficient’. Data collected is related to, such as management efficiency and efficiency of leased assets, rather than energy efficiency.

However, given that the consistency score is at 0.83, which is at an acceptable level, this thesis does not make changes to this indicator and will test it based on more samples.

Overall, this thesis acknowledges the semantics of keywords or surrounding context as the limitation of this thesis. However, the consistency scores indicate that the stability and reliability of the automated tool are at an acceptable level, which may not significantly affect the extent of SDG disclosures.

4.6.2 Text cleaning

Text normalisation and cleaning

Text normalisation is ‘the process of transforming unstructured text into a structured format’, aiming to enhance both the performance of the analysis and the accuracy of the measurements (Hickman et al., 2022; Mun et al., 2021; Vel, 2021, p. 879). It involves several stages, such as removing irrelevant data (i.e. stop words), and data cleaning (i.e. lemmatisation, and lowercase conversion) (Hickman et al., 2022). Given that this thesis adopts keyword searching in automated data collection, Python programming is designed to retrieve exactly the same word input by the researcher. To achieve this, this thesis conducted data pre-processing, including case sensitivity to enhance the precision of our measurements.

Extracting and cleaning text from reports

This thesis starts with converting reports into a machine-readable format for further data processing. Some reports are in the picture format, which Python cannot read. Therefore, this thesis adopts optical character recognition (OCR) to transform reports into a suitable format.

PyPDF2¹⁸ is adopted in this thesis to perform tasks on PDF documents, such as extracting texts from PDFs, merging files, and splitting pages of a PDF document. For example, this thesis initially attempted to use ‘pdfplumber’ to read PDF files, as this library can be used to extract both text and tables. However, some documents have a double-column layout on one page and pdfplumber cannot handle this layout in visual order. Therefore, this thesis selects PyPDF2 to extract basic text and uses **tabula-py** to extract tables from files and convert the table into a data frame (table formatting).

This thesis provides instructions on how to remove irrelevant elements from the text, such as the cover page, table of contents, headers and footers. To streamline the whole process, this thesis also modifies the code to automatically use the filename from the file path as part of the file name for the CSV file. This will reduce the complexity of file naming, particularly useful when multiple files are output simultaneously, and the names need to be changed each time a document is read.

¹⁸ PyPDF2 is a PDF library in Python, which is capable of ‘splitting, merging, cropping, and transforming the pages of PDF files’ (GitHub, 2024).

Text normalisation and cleaning

Text normalisation and cleaning involve several stages, such as removing irrelevant data (i.e. stop words) and data cleaning (i.e. lemmatisation, and lowercase conversion) (Hickman et al., 2022). As Python code is limited to keywords that exactly match the words in the keyword list, the tense and form of words will also affect the keyword searching results. To address this, the study sought to include all relevant forms of the word, and the process of lemmatisation and stemming were also considered. While stemming aims to transform words into their stems, it does not consider the grammar of the language, which may change the meaning of the word. For example, ‘caring’ is converted to ‘car’, while US and us were converted into ‘U’ and ‘u’. Therefore, stemming is not adopted, as it changed the original meaning of the context. This thesis then used lemmatisation, which ‘removes the inflectional endings and returns the base or dictionary form of the word to reduce variations of the word’ (Gupta et al., 2012; Pramana et al., 2022, p. 1). Unlike stemming, lemmatisation will output actual words and try to ‘identify the information of the word’. It will analyse whether it is a verb or a noun and provide the subsequent results. If any keyword occurs in the paragraph, the original paragraph without lemmatisation will be output.

Formatting

The formatting of sentences lacks consistency. Sentences retain their formatting or structure from PDF documents, leading to redundant spaces (see Table 4.23). To resolve this, this thesis refined codes to replace ‘link breaks’ with a space in instances where there is no full stop before the line break.

Table 4.23 An example of the formatting issue

In 2016, we distributed bracelets to 700 elderly people without family and autistic children in Songyuan City, and set up emergency teams to provide timely help for those wearing the bracelets
Helping Migrant Workers
Return Home
TThe Spring Festival in 2017 marks the sixth consecutive year for us to provide free food and gasoline for migrant workers at the Company’ s 65 service stations along national and provincial highways in Fujian, Jiangxi, Hunan and Hubei provinces

Coding development and improvement

As discussed in Section 4.6.2, this thesis develops Python code to enable the automated tool to replicate the manual process at its largest capacity and capture data in the most relevant sections. To achieve this, the process includes multiple steps, including splitting the whole document into different small documents based on table of contents, word frequency tests to identify documents that focus on SDG 1, 7, and 13 respectively, and keyword generating for SDG 1, 7, and 13.

- **Splitting the whole PDF document into small ones based on its table of contents.** It begins with using PDF-related libraries to extract the table of contents of the PDF and utilises the TOC () function to automatically split the whole documents. However, the TOC () function was ineffective in this situation, as some PDF documents are unstructured, and the table of contents is an outline page without linking to any respective pages. Another issue is the complex structure of the table of contents, including both sections and subsections. This thesis aims to split the entire document based on sections, rather than sub-sections. However, the layout of the report, such as sections or subsections, irregular formatting, and variations in content presentation, may complicate the task of dividing the file into smaller units. Additional pre-processing steps may be required to handle these complexities effectively. Therefore, this thesis manually splits PDF documents into multiple files based on page numbers.
- **Identifying focused sections for each goal.** This thesis also conducted a word frequency test based on keyword lists for energy, climate change, and poverty alleviation to identify the main section for that goal. Table 4.24 displays an example of using a word frequency test to identify the main section for SDG 1, 7, and 13. The first column shows a list of sections in reports and the following three columns show the word frequency related to poverty alleviation, energy and climate change. With the word frequency of poverty alleviation keywords at 77, the poverty alleviation section will be used to collect SDG 1 information.

Table 4.24 An example of word frequency test based on Air China 2019 ESG report

File	Poverty alleviation	Energy	Climate change
ABOUT THE REPORT	4	3	3
CHAIRMAN MESSAGE	4	1	2
NON-CSR TOPICS1	4	1	1
POVERTY ALLEVIATION	77	0	0
SAFETY STICK TO AND BUILD ON THE FOUNDATION FOR DEVELOPMENT	0	0	0
A HEARWARMING JOURNAL	0	0	0
BUILDING THE WORLD LEADING ENTERPRISE	0	1	0
BUILDING BEAUTIFUL HOME TOGETHER	0	46	50
BUILDING A BETTER LIFE TOGETHER	3	0	0
OTHER SECTIONS	9	12	3

This thesis develops a loop to identify paragraphs with keywords, save all relevant paragraphs into one column in a CSV file and duplicate rows from the output for each category are subsequently dropped. Instructions are given to Python to highlight keywords in red to visualise results. However, since a CSV file does not support text formatting, such as adding colour, bold or underlining, this thesis adds the following code: `cell_str = re.sub(rf' {keyword}'), f'**{keyword}**', cell_str, flags=re.IGNORECASE)`, to highlight keywords by surrounding them with '**'. Therefore, an Excel file has been created,

containing evidence for each indicator in one column and keywords highlighted by surrounding them with ‘**’.

This thesis then converts the CSV file into an Excel file to facilitate highlighting keywords in red. The ‘Panda’ and ‘openpyxl’ libraries are imported to apply red colour to keywords. However, irrelevant data was collected (i.e. grant—migrant; Yuan—Changjiangyuan), as some keywords that are part of another word are also highlighted. To resolve this, the code was improved to only highlight a keyword when it appears as a standalone word, using `rf\b{keyword\b}`. However, this method may overlook some keywords that are immediately followed by full stop or other punctuation (i.e. funds.) Therefore, `rf\b{keyword}(?![a-zA-Z])` is used to explore the keyword if it appears as a standalone word, even if it is followed immediately by a punctuation mark.

Current codes are designed to search keywords corresponding to one indicator in one PDF file. To streamline the whole process, this thesis developed a loop over the keyword file paths, to highlight keywords for each piece of evidence, and then save outputs in excel files. Then, it will loop over the output files to consolidate all results into one sheet with separate columns for each keyword file.

Based on the coding schemes and codes, an automated tool is ready to collect SDG disclosure information in Chinese ESG reports.

4.7 Dependent variable—SDG disclosures

Given that this thesis aims to observe what is disclosed by companies over time—a cross-sectional analysis, a dummy coding system is adopted to provide ‘an aggregated measure of the quantity of disclosure within reports’ (Guthrie & Abeysekera, 2006, p. 11). The disclosure scores range from 0 to 27 (See Appendix 7) to make allowance for observing the variation in SDG disclosures and subsequently investigating factors that may predict the level of SDG disclosures. The higher the score, the higher the level of SDG disclosures.

The analysis employs the disclosure scores as the dependent variable, rather than converting it into a dummy variable. As reviewed in Section 2.7, existing SDG disclosure studies widely employ a dummy variable to proxy SDG disclosures, such as the adoption of SDGs or mentioning these goals in reports (García-Meca & Martínez-Ferrero, 2021; van der Waal & Thijssens, 2020). The dummy method is due to data availability as SDGs were introduced in 2015 (García-Meca & Martínez-Ferrero, 2021; van der Waal & Thijssens, 2020). However, since 2015, an increasing number of SDG disclosure guidelines have been introduced by professional bodies (GRI et al., 2018; PwC, 2018) and academic studies (Michalczuk & Konarzewska, 2018; E Nechita et al., 2020). Despite offering more forward-looking disclosure measures, these studies focus on the overall level of SDG disclosures rather than measuring SDG performance. Therefore, this thesis develops a framework, containing 27 indicators, designed to better observe corporate SDG disclosures in Chinese companies.

Another reason is that prior studies widely adopt ESG scores to proxy sustainability disclosures, with data extracted from Bloomberg and MSCR KLD (Khan et al., 2016). As discussed in Khan et al. (2016, p. 1701), ESG scores offer ‘performance on a specific issue in a standardized format rather than the presence or absence of disclosures’. It provides more insightful information in understanding corporate ESG disclosure practices. Therefore, this thesis will employ the scoring system to provide a richer understanding of corporate SDG disclosure practices.

Adopting a scoring system may increase the richness of the prediction analysis and make better predictions. This thesis employs 159 predictor variables, which are used to analyse the variability of SDG disclosures of each observation. As the quantity of information on influential variables is quite high for each observation, a dummy variable may not effectively assist in making a prediction that may reflect a real-life situation. Therefore, this thesis adopts disclosure scores to measure the level of SDG disclosures without converting them into a dummy format.

However, SDG disclosure scores have the following two limitations: (i) a very limited range of values, and (ii) the number of indicators in each dimension is not equal, which may lead to the skewness of data.

- (i) A continuous variable with a limited range of values: SDG disclosure scores range from 0 to 27, indicating a very limited range of values. However, the framework is informative in observing SDG disclosure variation. Given that this thesis employs the gradient boosting machine, its n-fold cross-validation method can conduct multiple-round iterations to acquire prediction results, which may effectively use each data point. Consequently, this may mitigate the limitation of using a limited range of values.
- (ii) The skewness of data: 3 indicators have been adopted to observe the overall level of SDG disclosures, 12 indicators are used to proxy poverty alleviation disclosures, 5 indicators are used to observe corporate energy disclosures, and 7 indicators are used to observe climate change disclosures. Despite the number of indicators for observing corporate poverty alleviation disclosures being the highest among all dimensions, it only accounts for 44% out of all indicators. Furthermore, the gradient boosting machine is less sensitive to the skewness of data, as it, being a non-parametric model, does not assume normal distribution of data (Jones, 2017). That is, the gradient boosting machine adopts decision trees as base learners, which splits data based on feature values rather than statistical inference. Therefore, this should not be a problem in this thesis.

4.8 Chapter conclusion

To overcome the challenge of inter-coder reliability, such as their divergence of understanding, subjectiveness, and human errors, this chapter developed a dictionary-based automated content analysis tool to capture SDG disclosure data.

This thesis constructed 27 indicators (covering SDG1, SDG7 and SDG13) to capture the variation and to observe what companies disclose over time. In the framework, poverty alleviation is assessed based on the nature and extent of disclosures that focus on government-related poverty alleviation activities and TPA programs, including industrial development, employment, health care, social impact, bottom-up approach, education, regional development, relocating impoverished, and ecological improvement. Energy disclosures are evaluated based on the nature and extent of disclosures focused on energy efficiency, energy consumption, and renewable energy consumption, while climate change disclosures are based on Scope 1, Scope 2, emissions, corporate education and training related to climate change, and corporate risks and opportunities related to climate change. This thesis also includes explicit references to SDGs in the framework.

To observe corporate SDG disclosures, a sample of 60 companies listed on SSE and SZSE were examined during the period of the 13th National FYP (2016-2020) being effective, providing a preliminary analysis of Chinese corporate SDG reporting. Information is collected from English ESG reports, due to the uniqueness of the Chinese context, global accessibility and investor interests.

Therefore, a map of SDG disclosures in Chinese companies has been created. As this thesis facilitates a cross-sectional analysis to observe the level of SDG disclosures across Chinese companies, disclosure scores were constructed, ranging from 0 to 27, to make allowance for observing the variation of SDG disclosures and subsequently investigating what may motivate Chinese companies to voluntarily disclose SDG information.

Chapter 5 Research Design

5.1 Chapter introduction

Chapter 3 has sought to review factors that predict the level of SDG disclosures and identified gaps in existing studies.

Chapter 5 moves onto the research design of RQ2 what factors may predict the level of SDG disclosures. Specifically, Section 5.2 introduces the empirical framework, Gradient Boosting Machine, that this thesis employed to conduct analysis. Section 5.3 focuses on variable selection for RQ2, including government control, financial performance, corporate governance, ESG performance, and other control variables.

5.2 Empirical framework—Gradient Boosting Machine

This thesis employs the gradient boosting machine, a machine learning technique, to investigate factors that may predict the level of SDG disclosures in Chinese companies. As discussed in Section 3.4, existing studies rely on traditional statistical methods, which are limited to the in-sample analysis, do not take into account the complex non-linear associations or interaction effects among predictors, and are limited to accommodate a limited number of variables in the model. To address these challenges, this thesis employs the gradient boosting machine to **1.** identify the predictive power of a wide range of predictors from government control, financial performance, corporate governance, and ESG performance; **2.** offer a more reliable out-of-sample analysis.

5.2.1 An introduction of the gradient boosting machine

This thesis adopts a commercial version of the gradient boosting machine, TreeNet®¹⁹. It involves from exploring parameter space to applying numerical optimisation in function estimation (Chen et al., 2022). It attempts to reach an optimal function ($F^*(x)$) with the lowest value of the loss function ($L[y, F(x)]$). According to Friedman (2001), several loss functions have been widely employed, such as mean squared error, absolute error, and negative binomial log-likelihood.

$$F^*(x) = \underset{F}{\operatorname{argmin}} L[y, F(x)] \quad \text{Function (1)}$$

It is also known as an ensemble approach, comprising a set of trees to solve regression and classification problems (Friedman, 2001). Specifically, the gradient boosting machine uses decision trees as base learners, where each single tree has relatively weak predictive power (Carmona et al., 2022; Xia et al., 2017). By adding the prediction of all learners, weak learners are combined into stronger ones to develop a strong ensemble model (Friedman, 2001; Natekin & Knoll, 2013).

To achieve this, it combines both gradient boosting and steepest descent algorithms. Gradient boosting is employed to guide how the model combines weak learners together; it sequentially adjusts the model when each additional weak learner is added to create additive functions (See Function (2)). \hat{f}_n refers to the additive model at the n^{th} iteration. $h(x, \theta_n)$ is the new-added weak learner, while ρ_n refers to the size of the gradient descent step (the adjustments).

The steepest descent algorithm is adopted to minimise errors based on a ‘line-search’ approach. It is a crucial component of the gradient boosting algorithm, as it improves the model by ‘minimising the error at each step’. A weak learner is trained on the errors of the entire ensemble predictions by correcting the error in an ‘additive and sequential way’; it seeks to yield maximum improvements in a relatively straight-line direction (Friedman, 2001; Natekin & Knoll, 2013). Function (3) shows the gradient

¹⁹ This thesis also conducted the analyses based on the gradient boosting machine and XGBoost in Python to see whether similar results can be broadly found. The advanced gradient boosting model (XGBoost) also adds a regularisation term to avoid the possibility of overfitting.

descent step-size for each iteration (ρ). Function (4) shows how the parameter θ_n is calculated, which treats $-g_n(x_i)$ as a response variable and obtains the linear projection of $h(x, \theta_n)$.

$$\widehat{f}_n \leftarrow \widehat{f}_{n-1} + \rho_n h(x, \theta_n) \quad \text{Function (2)}$$

$$\rho_n = \underset{\rho}{\operatorname{argmin}} \sum_{i=1}^N \varphi [y_i, \widehat{f}_{n-1}(x_i) + \rho h(x_i, \theta_n)] \quad \text{Function (3)}$$

$$\theta_n = \underset{\theta, \beta}{\operatorname{argmin}} \sum_{i=1}^m [-g_n(x_i) - \beta h(x, \theta_n)]^2 \quad \text{Function (4)}$$

The gradient boosting machine facilitates a high-dimensional analysis, where the number of predictors can be greater than the sample size (Jones, 2017). It addresses the limitation of statistical inference analysis of being unable to accommodate a large number of predictors. The regularisation term is added to the model to penalise model complexity to avoid the possibility of model overfitting, when each single tree is added.

5.2.2 The limitation of the model

The gradient boosting machine is often criticised as a black box model due to its complexity and lack of interpretability. The model may include hundreds or thousands of decision trees, where each tree sequentially corrects the previous trees' errors. We are unable to trace how it works behind the scenes. This thesis employs relative variable importance scores, partial dependence plots and the interaction effect tool to visualise the results. However, they may often face criticism. For example, relative variable importance scores are often criticised for inflating or diminishing due to interaction effects or multicollinearity problems (I will discuss this in the Result interpretation tool section).

5.2.3 Why gradient boosting machine is adopted in my research?

The nature of data: The gradient boosting machine may fit the unique setting of my research, such as a greater number of predictor variables ($p > N$), a small sample size, and mixed-type data. Unlike statistical inference analyses, the gradient boosting machine is a tree-based method that does not make assumptions about data²⁰. It can conduct a high-dimensional analysis, by accommodating a greater number of variables in the model, even if the number of variables is greater than the sample size.

1. In my research, the sample consists of 60 Chinese companies; however, 159 variables and all possible interaction effects will be considered. The gradient boosting machine conducts a high-dimensional analysis, which well handles the situation where p is greater than N .
2. This thesis includes both continuous and categorical variables. As the gradient boosting machine works on tree splits, the nature of the data will not affect the overall prediction outcome.

²⁰ It does not need to specify the degree of freedom to calculate significance levels of variables or calculate an estimator of the variance of OLS, which both require $N > k$.

Thereby, we can obtain accurate results of how each variable contributes to the level of SDG disclosures, regardless of the sample size and data types.

3. As discussed in Section 3.3.2.1, recent studies argue the value relevance of traditional accounting information is deteriorating, particularly for stakeholders (Lev, 2018). However, Sun et al. (2021) employ the gradient boosting machine and argue that the prediction with raw accounting indicators outperforms the one with financial ratios. Therefore, this thesis makes the gradient boosting machine a unique statistical tool to observe the power of raw accounting numbers, financial ratios, and market financial indicators to predict the level of SDG disclosures.

Prediction vs statistical significance: The gradient boosting machine focuses on prediction rather than statistical significance testing. Unlike traditional statistical analyses relying on p-values, t-tests, and confidence intervals, the gradient boosting machine is an ensemble method that combines weaker learners into a stronger one for prediction and pattern discovery (Carmona et al., 2022; Xia et al., 2017). The gradient boosting machine adopts relative variable importance and partial dependency plots to visualise results. It also includes error metrics to indirectly demonstrate its model performance. Therefore, it can avoid the issues, such as p-hacking, snooping, and violating statistical assumptions (Beck et al., 2018; Jones, 2017). Unlike the statistical models, the gradient boosting machine is less sensitive to missing values, outliers, and multicollinearity problems. In addition, the nature of hypothesis development is slightly different. While traditional statistical analysis is theory-driven and attempts to accept or reject the hypotheses based on pre-defined statistical models, gradient boosting machine offers data-driven results in the black box to understand data patterns (Bao et al., 2020; Chen et al., 2022). This enables the data to tell what factors may predict the level of SDG disclosures, especially when we accommodate 159 predictors in the model and the sample size is 60 Chinese public companies.

Validity and generalisability: One of the unique characteristics of machine learning methods is their out-of-sample prediction accuracy. Data will be divided into training and test sets. The gradient boosting machine initially adopts the training set to learn and optimise the hyperparameters to improve the model performance. Once the training process is complete, the model's validity and generalisability will be ensured based on the test set. In addition, models can also be validated based on the k-fold cross-validation test. Specifically, the model will use training, and validation sets to optimise the hyperparameters. This is followed by the final validation based on the test sets.

The gradient boosting machine outperforms traditional methods in bias-variance trade-offs to acquire better prediction accuracy. It includes regularisation hyperparameters to avoid overfitting issues (Chen & Xie, 2022). Therefore, the gradient boosting machine offers greater accuracy, validity, and generalisability.

Interaction & complexity: The gradient boosting machine can also consider the complex interactions in the model to provide better predictive accuracy (Geertsema & Lu, 2023). According to Friedman (2001), gradient boosting machine can build ‘an ensemble of decisions with high interaction depth’. Compared with parametric models, it can automatically consider all interaction effects and rank them to see the predictive power of these effects on the level of SDG disclosures.

Non-linear vs linear relationship: The gradient boosting machine can deal with non-linear associations, as it is a non-parametric model (Van Der Heijden, 2022). This method will help test nonlinear ‘relationships’ between the level of SDG disclosures and factors that may predict this in my research. As discussed above, prior SDG disclosure studies have employed regression or logistic analysis to explore factors that may influence SDG disclosures; the overall model performance in prior research is relatively low. However, the gradient boosting machine, a non-parametric model, offers out-of-sample predictions to help better understand factors that may predict the level of SDG disclosures.

The comparison of machine learning The gradient boosting machine outperforms other machine learning techniques, such as Random Forests, AdaBoost, and deep learning, regarding prediction accuracy (Bahad & Saxena, 2020; Touzani et al., 2018), complexity (Bahad & Saxena, 2020; Gu et al., 2020), and flexibility (Bahad & Saxena, 2020). Prediction accuracy refers to the extent to which the model can mitigate bias and variance (Bahad & Saxena, 2020; Touzani et al., 2018). According to Touzani et al. (2018) and Gu et al. (2020), handling complex may include small sample size, data type, and non-linear predictions. Flexibility may discuss more about loss function selection and hyperparameter tuning (Bahad & Saxena, 2020).

Gradient boosting machine outperforms Random Forests regarding their prediction accuracy, flexibility, and complexity handling (Bahad & Saxena, 2020; Touzani et al., 2018). On the one hand, the gradient boosting machine offers better prediction accuracy, such as mitigating bias and variance (Bahad & Saxena, 2020; Touzani et al., 2018). This could be explained by different algorithms these methods adopted. Adopting bagging, Random Forests create a number of de-correlated trees, allowing ‘the averaging process will be less variable and more reliable’ (Jones, 2023, p. 94). This allows different fresh sets of m predictors to be considered during the tree-building process (Breiman, 2001). However, as suggested by Touzani et al. (2018), the averaging process leads to the same contributions to the prediction results. While it may reduce variance, this may not assist in ‘achieving bias reduction’, especially when individual trees underfit (Touzani et al., 2018, p. 1535). In contrast, the gradient boosting machine can minimise variance as well as bias through boosting (Bahad & Saxena, 2020). It employs the steepest descent algorithm, which minimises errors based on a ‘line-search’ approach (Friedman, 2001). This ‘constructive iterative strategy’ allows the greater ability to correct errors of the previous ensemble (Touzani et al., 2018). On the other hand, the gradient boosting machine has superior performance in offering flexibility. Unlike Random Forests, the gradient boosting machine has more

hyper-parameters for tuning, such as the learning rate, regularisation parameters and the number of estimates (boosting iterations). These parameters enable the model to be more flexible and adaptable. Unlike Random Forests that are limited to bagging, the gradient boosting machine also adopts a stage-wise additive method, which can effectively handle complex and nonlinear predictions. These make gradient boosting machine particularly suitable for my research setting, such as complex and high-dimensional data. As discussed above, this thesis has mixed-type data, including both categorical and continuous predictors, while the number of predictors is greater than the sample size. Additionally, the gradient boosting machine can accommodate and effectively handle complex and non-linear prediction, as it adopts a stagewise additive modelling.

While AdaBoost and gradient boosting machine both apply the boosting method, capable of improving prediction accuracy and handling complexity, the gradient boosting machine still offers better flexibility than AdaBoost in focusing on the selection of loss function, achieving more accurate prediction results (Bahad & Saxena, 2020). Like the gradient boosting machine, AdaBoost is an ensemble method by implementing boosting, where trees grow sequentially (Freund & Schapire, 1996; Shanmugasundar et al., 2021). At each iteration, greater weights will be assigned to observations that were incorrectly predicted in the previous round, allowing the subsequent weak learner to focus on these difficult cases (Freund & Schapire, 1996; Jones, 2023). This iterative re-weighting method allows the AdaBoost model to reduce bias. These sequential adjustments also enable AdaBoost to reduce variance to produce a stronger classifier (Freund & Schapire, 1996).

$$err_m = \frac{\sum_{i=1}^n w_i I[y_i \neq h_m(x_i)]}{\sum_{i=1}^n w_i}$$

$$w_i \leftarrow w_i \cdot \exp\{\alpha_i \cdot I[y_i \neq h_m(x_i)]\} = \begin{cases} w_i \cdot \left(\frac{1 - err_m}{err_m}\right), & y_i \neq h_m(x_i) \\ w_i, & y_i = h_m(x_i) \end{cases}$$

However, AdaBoost, in its classic form, is a ‘forward stagewise additive modelling’ using the exponential loss function, which is a ‘special case of the gradient boosting machine’ (Chen, 2021; Jones, 2023). It may be affected by noisy data and outliers (Jones, 2023). That is, this method may be less robust to outliers, as the exponential weighting may overly penalise outliers. In contrast, the gradient boosting machine is a generalised version of AdaBoost and can accommodate different types of loss functions. This offers greater flexibility in loss functions and enables the model to be less sensitive to outliers and noise, leading to more accurate prediction. In this thesis, raw data has been used for data analysis, which may include missing values, extreme values, unbalanced classes or other noisy data. The gradient boosting machine allows the selection of the proper loss function and combines the steepest descent and stagewise approaches to capture the underlying patterns.

While deep learning and gradient boosting machine are both capable of improving prediction accuracy, being flexible in loss function selection and handling complexity, gradient boosting machine, however, outperforms deep learning, when a relatively small amount of data is used (Gu et al., 2020) and when the data is structured (Gilpin et al., 2018). Deep learning utilises multiple processing layers to ‘learn representations of data with multiple levels of abstraction’ (LeCun et al., 2015, p. 436). It may require a large number of data to learn the complex representation. However, Gu et al. (2020) found that when a small number of data is used, neural network performance peaks at three layers and the performance declines when more layers are added. Therefore, ‘shallow learning’ methods, such as Gradient boosting machine and Random Forests, appear to outperform deeper learning when there is a relatively small dataset (Gu et al., 2020). Additionally, the commercial version of the gradient boosting machine may more effectively capture structured/ cross-sectional data. It argued that the deep learning method has better performance on image, text and speech data and provides accurate prediction results; however, it is hard to interpret the highly complex results, even if existing tools increase the interpretability (Gilpin et al., 2018). Therefore, given that deep learning may be less effective in small and structured datasets, this thesis employs the gradient boosting machine to conduct a cross-sectional analysis sampling 60 companies for more effectively capturing underlying patterns.

Given that gradient boosting machine tends to have superior performance in terms of prediction accuracy, flexibility in loss function selection and hyperparameter tuning, and complexity handling, this thesis adopts this model to explore factors that may predict the level of SDG disclosures.

Commercial version vs Python-based gradient boosting machine. This thesis adopts both commercial and Python-based versions of the gradient boosting machine for data analysis. The results and the R-squared scores are similar for both versions, indicating that the prediction accuracy appears to be broadly similar. However, the commercial-version gradient boosting machine outperforms the Python one regarding speed, handling missing values, and predictive power of different variables. A commercial-version model can output prediction results more quickly and effectively. This thesis compared the predictive power of different variables in the RVI table between commercial and Python-version models. It is observed that the overall ranking of predictors is broadly the same; however, the RVI scores in the commercial model tend to be higher than others in the Python model.

5.2.4 Model setup and validation

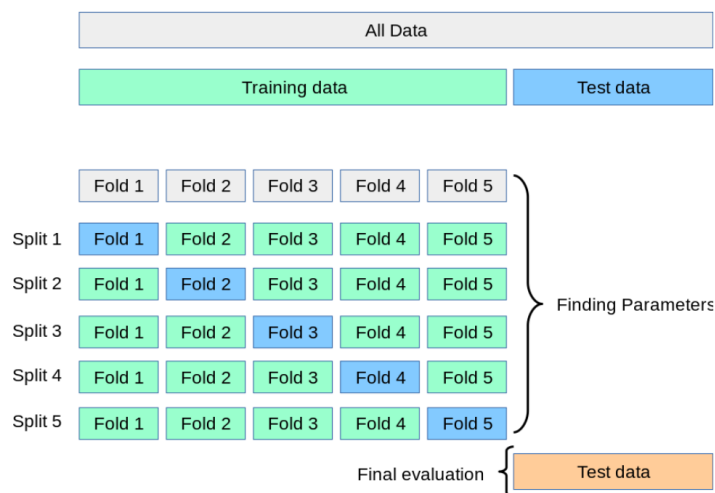
This section introduces model validation methods and the approaches to how this thesis trains and validates the model.

Traditional training and test split methods attempt to train and validate the model based on a single and random split (Gunasegaran & Cheah, 2017; Xu & Goodacre, 2018). The entire dataset will be split into training and test samples. A training set is used to train the model to reach an acceptable level, while

the test set is used to validate the model and ensure that it will perform equally. As discussed in Nguyen et al. (2021), prior studies have selected different ratios (i.e. 10/90, 20/80, 30/70, 40/60, and 50/50) to split the entire dataset into training and test sets to evaluate model performance. However, the predictive capability of the model is greatly affected by the training/test split ratios, due to bias. The random selection of the test set may also lead to some variability in the prediction effectiveness (Varshney & Singh, 2024).

Unlike the traditional training-test sample split, K-fold cross-validation is a widely used technique for evaluating model performance and selecting optimal hyperparameters in machine learning methods (Xu & Goodacre, 2018). Commonly, a test set will be left out for final model validation. The rest dataset will be classified into K folds. (K-1) folds are used for model training and the remaining part is adopted in each iteration as the validation set (Xu & Goodacre, 2018). The parameters will be tweaked continuously until the model can perform optimally in each loop. After k iterations, the parameters will be the average of the values generated in the loops. This method can help find the optimal parameters and avoid overfitting.

Figure 5-1 K-fold cross-validation, extracted from Scikit-Learn (2019)



This thesis adopts a 10-fold cross-validation. The whole dataset was randomly classified into 10 subsets of equal size. In each iteration, one subset is used to validate the model, while the remaining 9 subsets are used to train the model. The whole process repeats 10 times, allowing each subset to be used for validation for one time. The final results are the average of 10 results acquired. The primary advantage of this method is that the entire dataset has been used to train and validate the model to find the optimal prediction model and overcome the challenge of the traditional training and test sample split method, which 'may provide an erroneous estimation of model performance' (Xu & Goodacre, 2018).

5.2.5 Result interpretation tools

Given that machine learning techniques are hard to interpret results, this thesis employs relative variable importance and partial dependence plots to visualise results to enable a better interpretability.

Variable importance

Relative variable importance visualises a variable's relative contribution to predicting the variability of SDG disclosures. The most influential variable is assigned a score of 100, while all the other variables are assigned a score that is scaled relative to the most influential variable (Jones, 2017).

In a single tree A, containing n splits, the variable importance will be assessed based on all non-terminal nodes (n-1). The variable importance is determined based on the number of splits and the weighted influence of the variable that is measured by empirical squared improvements (Friedman, 2001; Natekin & Knoll, 2013). The following function represents the model for calculating the influence of a specified variable in a single tree A and I^2 represents the empirical squared improvement of the variable.

$$Influence_x(A) = \sum_{i=1}^{n-1} I_i^2 L(\text{splitting variable} = x)$$

Then, the relative variable importance is determined by taking the average of the variable importance of the variable over all trees. This method only visualises the importance of a variable relative to all the other variables; therefore, it does not show the actual impact of the influential variable in the response variable.

$$Influence_x = \frac{1}{N} \sum_{i=1}^N Influence_x(A)$$

However, the relative variable importance scores may be inflated or diminished due to interactions or multicollinearity. For example, multicollinearity problems may occur when two predictors are strongly correlated. In a tree-based model, if two predictors offer similar information, the tree may randomly select one to conduct analysis, making one more important than it is. The one selected may receive an inflated importance score, while the other one may be diminished. However, as suggested by Jones (2017), the gradient boosting machine is less sensitive to multicollinearity issues.

Partial Dependency Plot

Partial dependency plots have been primarily employed in sophisticated machine learning analysis, to illustrate the directional contributions of a feature to the output prediction (Frost et al., 2023; Jones, 2017). The plots visualise the marginal effect of a given predictor on the variation in the level of SDG disclosures. To obtain the univariate plot regarding a particular predictor, the remaining predictors are fixed 'at the joint set of values from the dataset' (Jones, 2017, p.1382). x_s refers to the predictor plotted based on the partial dependence model, while X_C refers to the other predictors used in the gradient boosting model. The partial dependence function visualises the relationship between the predictor in the set S and the dependence variable. By marginalising over the other predictors in set C, the plot will display a plot that relies on only predictors in S.

$$\hat{f}_s(x_s) = E_{X_C}[\hat{f}(x_s, X_C)] = \int \hat{f}(x_s, X_C) dP(X_C)$$

5.3 Potential factors that may predict the level of SDG disclosures

Based on the gradient boosting machine, this thesis investigates potential factors that may predict SDG disclosures from 1. contextual factors, 2. corporate factors, and 3. Other control variables.

5.3.1 Contextual factors

As theorized in Section 3.3, this thesis intends to explore whether government control, as a contextual factor, can predict the level of SDG disclosures in the gradient boosting machine. Specifically, this thesis will include 1. Regulatory influence, 2. Shareholding influence, 3. Political influence, and 4. Incentivising influence (See Section 3.3.1).

5.3.1.1 Regulating influence

To explore whether regulatory influence can predict the level of SDG disclosures based on the gradient boosting machine, this thesis employs national regulatory guidelines as the proxies for regulatory influence, including the SDG national project and the Belt and Road Initiative. Unlike corporate-level initiatives and guidelines, SDGs are an inter-governmental initiative, and they are incorporated into national-level strategies in China. These two strategies have highlighted the greater role of Chinese companies in SDG and SDG disclosures. However, this thesis is limited to IDZ location data and does not have data directly on whether the company participated in IDZ or acquired any subsidies for this program.

SDG regulatory influence is measured by whether companies are in cities that participated in the IDZ project. The location of the company is collected from the CSMAR database. A list of cities that have been approved to participate in the IDZ project can be found online. Since 2017, 11 cities have received approval from the State Council to participate in the IDZ project, including Taiyuan, Guilin, Shenzhen, Chenzhou, Lincang, Chengde, Ordos, Xuzhou, Huzhou, Zaozhuang, and Hainan Tibetan (Xinhua, 2022). This thesis manually checks whether sampled companies are from cities that have participated in the IDZ program. A company will receive a score of one if it is located in one of these cities; otherwise, it will receive a score of zero.

Belt and Road Initiative (BRI) As discussed in Section 3.3.1.1, inconsistent results have been found between CSR disclosures and corporate BRI engagement, due to different measurements of disclosures and research contexts. While companies are encouraged to participate in BRI to jointly realise the goals in agenda 2030, this thesis will explore the predictive power of BRI on SDG disclosures.

It is measured by whether companies have BRI deals. A company will receive a score of one if it has BRI deals; otherwise, it will receive a score of zero. BRI information is collected from the Thomson Reuters database.

5.3.1.2 Shareholding influence

To explore whether shareholding influence can predict the level of SDG disclosures based on the gradient boosting machine, this thesis explores shareholding influence based on different types of SOEs. They are selected as prior studies suggest inconsistent results between government shareholding influence and corporate ESG disclosures, due to different levels of governments (Parsa et al., 2021; Situ & Tilt, 2018). These studies are limited to classifying shareholding influence into regional and central levels (Chu et al., 2013; Zeng et al., 2012). However, the nature of SOEs may also affect how the company operates.

This thesis follows ‘Opinions of the State Council on Development of An Economy of Mixed-Ownership of State-Run Enterprise’ and classifies SOEs into commercial and public-interest SOEs, supervised and overseen by different authorities. Specifically, commercial SOEs include ‘perfect competitive sectors and strategic sectors (i.e. key industries related to national security and national economic lifelines)’ (Lin et al., 2020, p. 39). Public-interest SOEs target the quality and efficiency of public goods and services, to ensure people’s livelihood and serve society. They are encouraged to disclose their information publicly and accept the oversight from the public. Therefore, two proxies have been developed—commercial vs non-commercial SOEs, and public-interest vs non-public interest SOEs. A company will receive a score of one if it is a commercial SOE (public-interest SOE); otherwise, it will receive a score of zero. Data for classifying companies as SOEs and non-SOEs is obtained from the CSMAR database and then this thesis manually classified SOEs into commercial-SOE and public-interest SOEs based on the ‘Opinions of the State Council on Development of An Economy of Mixed-Ownership of State-Run Enterprise’ guideline.

5.3.1.3 Political influence (political connection)

To explore whether political influence can predict the level of SDG disclosures based on the gradient boosting machine, this thesis explores political influence from political connections at personal levels and party committees embedding two levels. Unlike regulatory influence, political influence can be considered a more indirect tool through connections to influence corporate performance and disclosure practices. Disclosure may indicate that companies are strategically building a stable relationship with the government, to acquire government support and maintain corporate political legitimacy (Li et al., 2021, p. 81). Political influence data is collected from the CSMAR database.

Political connection at personal levels

As discussed in Section 3.3.1.3, there are no consistent measurements on corporate political connections, due to different types of political connections (government officials vs members of NPC or CPPCC) and different levels of political connections (i.e. CEO, chairman, management, or directors). Their impacts on corporate performance and corporate disclosures also vary widely depending on how political connections are measured. This thesis observed that existing studies often lack clear definitions

of government-related jobs, and government official positions, which are crucial criteria in exploring corporate political connections. Therefore, this thesis provides a more comprehensive definition of political connections and explores the connections based on their corporate positions.

Once data has been classified and collected by the automated tool, this thesis will turn them into the following several proxies:

Counts	Political connection overall counts
	Political connection CEO/executives counts
	Political connection director counts
	Political connection senior management team counts
The presence of political connection	Political connection overall binary
	Political connection CEO/ executives binary
	Political connection director binary
	Political connection senior management team binary

Variable definition

This thesis follows Wang et al. (2018b) and L. Wang et al. (2021) and argues that a company is qualified as having a political connection, if any of its CEO, directors, supervisors, or managers is 1) a former government official; 2) a current or former member of NPC or CPPCC, or 3) a current or former member of the military party. Specifically, this thesis defines government officials as members who have working experience in Chinese governments.

1. Senior management team: 1) a former government official, 2) a current or former member of NPC or CPPCC, or 3) a current or former member of military party.
2. Chairman or CEO: 1) a former government official, 2) a current or former member of NPC or CPPCC, or 3) a current or former member of the military party.
3. Directors: 1) a former government official, 2) a current or former member of NPC or CPPCC, or 3) a current or former member of the military party.

I will discuss this now:

1. ***The definition of the senior management team.*** Prior studies are limited to one type of the senior management team, such as the CEO, board of directors, or chairman, to observe their political influence on corporate disclosures. However, the CSMAR database states that the senior management team includes chairman, board members (i.e. director, non-executive director, independent director, non-independent director), supervisors (i.e. supervisor general, external supervisor, and employee supervisor), managers, and secretaries. Therefore, following this, this thesis will observe corporate political connections from several levels:

- *Senior management team.* This thesis will acquire an overview of the whole management team to see whether any senior management team member is politically connected. This method has been adopted by Wang et al. (2018b).
 - *Chairman and CEO.* They are the major decision-makers in the top management team. Based on the Chinese Corporate Law, chairman is the legal representative of the company, who has the highest authority in the company and is responsible for corporate operations (Wu, Wu, Zhou, et al., 2012). The CEO is the general manager in a company, elected by the board (Wu, Wu, Zhou, et al., 2012). Chairman, the first-in-command, has the highest control of corporate operations and decision making. He/She is also the chair of the board of directors, who has the power in supervising and encouraging corporate activities (B. Y. Wang et al., 2021). Prior studies have suggested significant impacts of politically connected CEOs and chairman on corporate disclosures and performance (B. Y. Wang et al., 2021). Therefore, this thesis also investigates whether top leader of the company may predict corporate SDG disclosures in Chinese companies.
 - *Board of directors.* Board of directors are expected to serve for the interests of stakeholders, as they are accountable for supervising corporate connections with stakeholders, developing social agenda for companies, and formulating strategies and plans (M. K. Khan et al., 2021). Recent studies also suggest that politically connected board of directors may predict corporate performance and disclosures (de Andres et al., 2022; L. Wang et al., 2021)²¹.
2. ***The definition of government officials.*** Consistent with Wang et al. (2018b) and L. Wang et al. (2021), this thesis focuses on their working experience in Chinese governments. Prior studies do not have a clear description of government officials. Concerns include what can be viewed as government departments and what should be considered as official positions (See Section 3.3). It is observed that government institutions may include governments (i.e. local, municipal, provincial, and national levels), and other regulatory bodies, such as China Securities Regulatory Commission and institutions governed by the State Council. However, official positions in government-affiliated institutions, such as Road Safety bureau, and Animal bureau,

²¹ For example, de Andres et al. (2022) argue that corporate political connections negatively influence corporate financial performance, and this relationship becomes even stronger due to directors' greater financial expertise. This is explained by the fact that politicians often 'lack management, professionalism, and know-how' (EL Ammari, 2023, p. 1073). These members pay more attention to their own benefits and political career, rather than corporate performance. Some also develop personal links with politicians in order to acquire personal gains, rather than protect minority interests, which is called a grabbing-hand approach.

Hiring directors with political background may also provide companies with easier access to government resources, such as loans, government subsidies, and other scarce resource. These members tend to be more familiar with the government and may affect corporate capability to 'reap government favours' (EL Ammari, 2023, p. 1075). Finally, this may increase corporate competitive advantage over their rivals. However, it is unclear whether this may affect corporate SDG disclosures.

may have limited power in assisting companies in acquiring government support or in facilitating corporate political goals.

According to Shen (2023), officials can also be classified into five levels: national, provincial, department, division, and section levels. Officials at the division level or higher levels hold leadership roles, whereas those at the section level do not have leadership roles. However, this will not be considered in this thesis, as different leadership roles in the government affect the management team's behaviour, rather than their disclosure transparency.

This thesis measures political connection at the personal level based on the automated content analysis, developed in Chapter 4. It involves four main steps: 1.) data source and data cleaning; 2.) developing a comprehensive keyword list for automated content analysis; 3.) matching keywords with cleaned text and keyword refining; 4.) data validation to check the reliability of the automated content analysis tool.

Data source

Corporate political connection data is collected from resumes of senior management teams, where resumes were downloaded from the CSMAR database. This database is used as it offers a comprehensive view of corporate political connection descriptions by gathering resume data from corporate annual, semi-annual and quarterly reports and interim announcements from the board of directors, board of supervisors and shareholders' meetings.

Some studies collected political connection data from annual reports, as required by the CSRC, which mandates companies to report directors' personal background information in annual reports. This thesis did manually check senior management information presented in annual reports. However, limited information has been provided across companies. For example, some companies (i.e. PingAn Bank) provide brief descriptions for each senior management member, while some companies only listed the names and roles of their senior management, without offering additional details. Therefore, this thesis will use resumes downloaded from the CSMAR database. Among 300 observations, 8135 resumes were downloaded from the database.

Unit of analysis

Since data was directly downloaded from the CSMAR database into an Excel file, the unit of analysis and content remained identical for both manual and automated processes. This thesis selected paragraph as the unit of analysis

Keyword development and refining

This thesis adopts automated content analysis to analyse whether corporate senior management has any political connections with the government. This method was established in Chapter 4 and applied here to collect political influence data.

Keyword development

During manual data collection, this thesis creates a list of keywords for automated data collection and sets the variable's boundary (See Table 5.1).

- Some members of the senior management team are members from other societies, rather than NPC or CPPCC. During manual data collection, this thesis observed that senior management team members tend to be members of other societies, such as *a member of the Expert Advisory Committee on Industrial Economic Operation of the Ministry of Industry and Information Technology*, and *a member of the Think Tank Committee of the All-China Federation of Industry and Commerce*. These societies cannot offer suggestions to CPC and government's judicial activities and thereby do not have strong power to involve in national decisions. Therefore, this thesis only considers members of NPC and CPPCC as political connections with the government, consistent with Wang et al. (2018b) and L. Wang et al. (2021).
- Following Yu and Chi (2021), this thesis will also view senior management team members with military experience or background as evidence of corporate political connections.

Table 5.1 Political connection keyword

Keywords
Government
Army
Military
People's congress
National People's Congress
Chinese People's Political Consultative Conference
consultative conference

Keyword refining

The automated tool collected 55 pieces of evidence showing that senior management members are politically connected. Out of the 55 pieces of evidence, five are irrelevant data, which are collected based on the keyword 'government'. As they are related to government awards of senior management rather than their previous government-related working experience, these will not be considered evidence of senior management's political connection in this thesis (see Table 5.2).

Table 5.2 Examples of irrelevant political connection data based on the keyword 'government'

Company name	Ending Date of Statistics	Name	Resume of Senior Management	Data source	Specific Position
ZTE Corporation	2016-12-31	Xie Daxiong	'He is a national candidate of the Million Talents Project and enjoys the special government allowance issued by the State Council ...'	CSMAR	Chairman of the Board of Supervisors

ZTE Corporation	2017-12-31	Xie Daxiong	'He is a national candidate of the Million Talents Project and enjoys the special government allowance issued by the State Council ...'	CSMAR	Chairman of the Board of Supervisors
ZTE Corporation	2018-12-31	Xie Daxiong	'He is a national candidate of the Million Talents Project and enjoys the special government allowance issued by the State Council ...'	CSMAR	Chairman of the Board of Supervisors
ZTE Corporation	2019-12-31	Xie Daxiong	'He is a national candidate of the Million Talents Project and enjoys the special government allowance issued by the State Council ...'	CSMAR	Chairman of the Board of Supervisors
ZTE Corporation	2020-12-31	Xie Daxiong	'He is a national candidate of the Million Talents Project and enjoys the special government allowance issued by the State Council ...'	CSMAR	Chairman of the Board of Supervisors

Keyword validation

This thesis identified 50 pieces of evidence of political connections through manual data collection and 55 pieces of evidence through automated data collection. Since data was directly downloaded from the CSMAR database into an Excel file, the unit of analysis and content remained identical for both manual and automated processes. Therefore, rather than relying on complex or advanced statistical methods, this thesis employs simple statistical techniques to evaluate the similarity between data collected manually and automatically collected, by calculating Ratio1 No. of overlapping evidence out of the total number of manual evidence and Ratio2 No. of overlapping evidence out of the total number of automated evidence. 'Number of overlapping evidence' refers to instances where the pieces of manually collected evidence match those collected automatically.

Ratio1:
$$\frac{\text{number of overlapping evidence}}{\text{the total number of manually collected evidence}}$$

Ratio 2:
$$\frac{\text{number of overlapping evidence}}{\text{the total number of automatically collected evidence}}$$

1. In Ratio1, the total number of manual evidence refers to the number of manually collected evidence that shows the senior management team has political connections with the government. The ratio may indicate the extent to which manually collected evidence is captured by the automated tool. Table 5.3 shows that Ratio1 is 100%, indicating that the automated tool has captured all manually collected data based on the given keyword list. This may indicate the effectiveness of the automated tool and the political connection keyword list.

2. Ratio2 may indicate the extent to which automatically collected data is ‘correct’. In the first-round test, Ratio 2 is 82.76%, shown in Table 5.3. Out of 55 pieces of evidence, five pieces are irrelevant (shown in Table 5.3). Evidence is collected based on the keyword— ‘government’. As they are related to government awards of senior management rather than their previous government-related working experience, these will not be considered evidence of senior management’s political connection in this thesis.
3. Regarding CEO or chairman political connections, this thesis identified 6 pieces of evidence of political connection through manual data collection and 11 pieces of evidence through automated data collection. Table 5.4 shows that Ratio1 is 100%, indicating that the Python code has captured all manually collected data, based on the given keyword list. The result for Ratio2 is 0%, which is affected by irrelevant data.

Table 5.3 Similarity test for the first-round data collection at senior management level

Senior management	CPPCC	military	govt	govt without irrelevant data
Ratio1	100% (18)	100% (8)	100%	100%
Ratio2			82.76%	100%
	18	8	24vs29	24vs 24

Table 5.4 Similarity test for the first-round data collection at CEO & chairman level

CEO or Chairman	CPPCC	military	govt	govt without irrelevant data
Ratio1	100%	100%		100%
Ratio2	100%	100%	0%	100%
	5	1	0vs5	0vs0

The second-round preliminary testing

This thesis randomly selected additional five companies, encompassing 624 observations, to test the validity and reliability of keyword lists, criteria, and the automated tool. Again, it includes three parts: manual data collection, automated collection, and similarity tests.

During the manual data collection, this thesis observed that some senior management team members have working experience in bureaucratic institutions, such as audit bureau, finance bureau, administration bureau, and taxation bureau, and institutions governed by the State Council, such as the Ministry of Commerce. Since these institutions are also government-related, they fall outside the boundaries established in this thesis. As discussed in Section ‘Variable definition’, expanding the scope of government departments may introduce unnecessary complexity, providing limited value to this thesis. Therefore, this thesis will not expand the definition of political connection.

Similarity test:

Senior management team: This thesis identified 31 pieces of evidence through manual data collection and 32 pieces of evidence through automated data collection, which can be considered as having political connections. Ratio 1 is 100%, indicating that all manually collected data has been collected by Python code. Ratio2 for government officials is only 95.65% (see Table 5.5).

CEO or chairman: This thesis identified 7 pieces of evidence through both manual and automated data collection. Ratio1 and Ratio 2 are 100%, shown in Table 5.6. It indicates that the automated tool well replicates the manual data collection process, and the criteria and procedure are reliable.

Table 5.5 Similarity test for the second-round data collection at senior management level

Senior management	CPPCC	military	govt	govt without irrelevant data
Ratio1	100%	100%	100%	100%
Ratio2	100%	100%	95.65%	96%
	4vs4	5vs5	22v 23	22vs 23

Table 5.6 Similarity test for the second-round data collection at CEO & chairman level

CEO and chairman	CPPCC	military	govt	govt without irrelevant data
Ratio1	100%	100%	100%	100%
Ratio2	100%	100%	100%	100%
	2vs2	0vs0	5vs5	5vs5

Third round preliminary testing

In the third round of preliminary testing involving 572 selected observations, this thesis identified 46 pieces of evidence of political connections through manual data collection and 53 pieces of evidence through automated data collection that can serve as evidence for political connections. Out of these pieces of evidence, this thesis also identified 16 pieces of evidence of political connections through manual data collection and 17 pieces of evidence through automated data collection, showing that the CEO or chairman can be qualified as being politically connected. Irrelevant data is shown in Table 5.7.

Table 5.7 Irrelevant data based on the keyword 'government'

Date	Name	Role	Evidence	Source
2016-12-31	Zhou Yunzhi	Vice President	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR
2017-12-31	Zhou Yunzhi	Vice President	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR

2018-12-31	Zhou Yunzhi	Vice President	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR
2019-12-31	Zhou Yunzhi	Vice President	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR
2020-12-31	Wu Gang	Chairman of the Board	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR
2020-12-31	Zhou Yunzhi	Vice President	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR
2020-12-31	Hu Ruimin	Independent Director	'Mr. Zhou Yunzhi, is a researcher-level senior engineer. He enjoys the special government allowance of the State Council,'	CSMAR

Based on first round and third round preliminary tests, this thesis observed that irrelevant data is all related to special government allowances, based on the keyword 'government'. Therefore, this thesis refines the coding scheme to remove 'government allowance' related evidence from the results.

Table 5.8 compares the results using the original criteria and the new criteria, which exclude irrelevant data. The table demonstrates that the results for all instances of 'ratio2' improved when irrelevant data was excluded. However, further tests are needed to demonstrate whether the reliability of the new method is significantly improved.

Table 5.8 Comparison between data collection methods

		govt	govt without irrelevant data
Part 1	Ratio1	100%	100%
	Ratio2	83.76%	100%
		24vs29	24vs 24
Part2	Ratio1	100%	100%
	Ratio2	95.65%	96%
		22v 23	22vs 23
Part3	Ratio1	100%	100%
	Ratio2	68.18%	100%
		15vs22	15vs 15

Additional tests

This thesis also explores whether the ranking of each company based on different criteria has changed due to irrelevant data and whether the new method can reduce the possibility of collecting irrelevant data. For the government official indicator, this thesis calculates the number of government officials for

each company in each sampled year and ranks them accordingly. Due to the nature of the data, this thesis employs non-parametric tests, Spearman’s rank correlation test and Kendall’s Tau method, to test whether different criteria significantly affect the ranking order of the number of government officials for each company.

Spearman's rank correlation: this can be used to evaluate the relationship between ranked variables based on different criteria, employing a monotonic function or ‘the direction of the arbitrary monotonic association between two ranked variables or one ranked variable and one measurement’ (Xiao et al., 2016, p. 3869). Given that it is a non-parametric ranking tool, it does not require any distribution assumptions or assume linear relationships between ranked variables. As it calculates the squared error, this method, thereby, can be used to test sensitivity to see the discrepancy of extreme values. Xiao et al. (2016, p. 3869) argue that it measures a monotonic relationship, making it an alternative when Pearson’s r are not desirable in handling in-pair variables.

Kendall's Tau can also be used to test ranking differences, by measuring the statistical dependence between rank-transformed quantities. It is an alternative to Spearman’s rank correlation test. Xiao et al. (2016, p. 3869) state that the results for these two methods are quite close in most cases.

Compared with Spearman’s rank correlation, Kendall’s tau provides a more intuitive interpretation, by comparing the proportion of concordant pairs and the proportion of discordant pairs. It is more applicable to a small sample size to better estimate the corresponding population parameter. This method is more conservative than Spearman’s rank correlation.

Table 5.9 shows the results of Spearman’s rank correlation and Kendall’s Tau to investigate whether the ranking order of Python results based on the measurement without irrelevant data has higher similarity with the ranking order of manually collected data. The coefficient range is between -1 and 1 for Spearman’s rank correlation test. With a coefficient of 0.8, a stronger positive correlation between these two methods has been identified. The value increased to 0.97 when utilising Method 2—without irrelevant data.

Kendall’s Tau value also ranges from -1 to 1. The value for method1 is 0.76, indicating a strong positive association between these two rankings. This implies a high level of agreement in ranking order. The value of Kendall’s Tau increased to 0.98 based on Method 2.

Consequently, both sets of results suggest that method 2 without irrelevant data shows a high degree of agreement in the ranking orders between manually and automatically collected data.

Table 5.9 results of Spearman’s rank correlation and Kendall’s Tau

	Spearman's Ranking:	Kendall's Tau:
Method1: with irrelevant data	0.80	0.76
Method2: without irrelevant data	0.98	0.98

This thesis also manually checked whether any piece of evidence has been misclassified. If so, this thesis manually deletes that piece of evidence.

Party committee embedding

Party committee embedding is another form of political influence, which is measured based on whether the company has party committee embedding and the number of the senior management team members in the party committee.

Data is collected from resumes of senior management teams, where resumes were downloaded from the CSMAR database. Unlike political connections that are at individual levels, a party committee is a formal group, strongly exercised by the CPC, that is integrated into organisations to bridge the gap between the enforcement of external regulations and the development of the internal culture of social responsibility (Yin et al., 2019). This variable is selected as prior studies suggest inconsistent results between party committee and CSR disclosures, due to the measurements of party committee embedding.

Method:

Consistent with the method for collecting data of political connections at personal levels, this thesis measures party committee embedding based on the automated content analysis, involving several main steps: 1.) developing a comprehensive keyword list for automated content analysis; 2.) matching keywords with cleaned text and keyword refining; 3.) data validation to check the reliability of the automated content analysis tool. This method was established in Chapter 4 and applied here to collect political influence data.

Keyword development and refining

Based on the initial 602 observations, this thesis found that corporate party committees may include the party committee, the Propaganda Office of the party Office, the deputy manager of the party branch, or a member of the Political Office of the branch. Keywords are in Table 5.10:

Table 5.10 Keywords for party committees

keyword	party branch
	party committee
	party office
	party secretary
	director of the party

Keyword validation

After reviewing 602 resumes, this thesis observed that 108 pieces of resumes are manually identified as evidence of senior management's party committee embedding, while 124 pieces of resumes are automatically collected as evidence of their party committee embedding. This thesis also adopts Ratio 1 and 2 to see the similarity between manual and automated data. Results are shown in Table 5.11.

First-round test:

Ratio 1 No. of overlapping evidence out of the total number of manual evidence shows the extent to which manually collected resumes that may indicate senior management's party committee embedding, have been identified by the automated tool. The test suggests that 100% of manually collected evidence has been captured by the automated tool.

Ratio 2 No. of overlapping evidence out of the total number of automated evidence may show the percentage of automatically collected party committee embedding data that is 'correct'. Table 5.11 suggests that 87% pieces of automatically collected evidence overlap with manual evidence. Among the 16 pieces of irrelevant evidence, they were collected using the keyword 'party'. For example, some pieces of evidence show individuals being members of the CPC, reflecting their individual political identity rather than the embedding of party committees. This issue is considered noise in the study. Therefore, this thesis tried addressing it by eliminating evidence containing the keyword 'a member of the CPC'. However, this method removed some 'correct' evidence, affecting the overall reliability of the automated data collection process. Consequently, this thesis intends to manually review this issue to ensure accuracy.

Second and third round tests

Table 5.11 summarises the number of data collected during manual and automated data collection, along with the results of similarity tests. 87 pieces of evidence related to party committee embedding were manually collected, while 94 pieces of evidence were automatically collected. Out of the 87 manually collected pieces of evidence, this thesis observed that 97.7% were also captured during the automated data collection. Approximately 90% of the automatically collected data were relevant data.

The third-round preliminary test involves 572 observations. 14 pieces of data were manually collected as evidence of corporate party committee embedding, while the same evidence was automatically collected by Python. It is worth noting that there was a complete overlap, with all 14 manually collected data points also being identified by the automated process. This indicates that Python code used for automated data collection demonstrates a 100% reliability rate in identifying manually collected evidence.

Overall, the percentage of manually collected evidence collected by Python code is 100%, 97.7%, and 100% respectively in three round tests, indicating that the keyword list, criteria and Python code are reliable and valid for collecting the rest of the data.

Table 5.11 Consistency test results

	Manually collected data	Automatically collected data	Percentage of manual evidence that is identified by Python code	Percentage of ‘correct’ evidence out of automatically collected data
First round	108	124	100%	87%
Second round	87	94	97.7%	90.4%
Third round	14	14	100%	100%

Section 5.3.1.4 Incentivising influence

To explore whether incentivising influence can predict the level of SDG disclosures based on the gradient boosting machine, this thesis measures whether sampled companies issue green bonds during the sampled period. A score of one will be assigned if the company has engaged in the green bond market; otherwise, it will receive a score of zero.

Green bond information will be collected from the Thomson Reuters database, which includes green bonds issued in China (Day) and green bonds issued overseas by domestic entities (Day). This proxy is selected, as the Chinese government and official authorities have detailed requirements for the issuance of green bonds, rather than disclosure requirements, while only green investors and institutional bodies require companies to report their allocation of funds in their projects (Zhou & Cui, 2019). Prior studies suggest significant results between green bonds and CSR disclosures; however, mixed results have been found due to the measurement of disclosures and green bonds. This provides the opportunity to investigate the impact of green bond on corporate SDG disclosures.

5.3.2 Corporate characteristics

Financial performance

As theorized in Section 3.3, this thesis explores whether financial performance may predict the level of their corporate SDG disclosures from 1. Accounting-based variables, and 2. Market-based variables.

Accounting-based financial performance

This thesis employs a number of accounting-based variables. As discussed in Section 3.3.2.1, the results widely vary regarding variable selection, different jurisdictions, and disclosure measurements. Therefore, this thesis will explore: *1.* whether variable selection may influence the prediction results; *2.* Given that existing Chinese literature focuses on traditional financial numbers and financial ratios, this thesis also includes more forward-looking financial performance indicators to see the predictive power

of different proxies; 3. the value relevance of traditional accounting numbers and ratios in predicting the level of SDG disclosures.

Data is collected from the CSMAR database, which is the most widely used and comprehensive database for collecting financial data in China. This thesis also employs Capital IQ and Thomson Reuters as supplemental databases to fill in missing values if the measurement for a particular variable is the same.

Accounting-based indicators. While financial ratios are widely used financial performance proxies in ESG studies, this thesis includes some unscaled financial indicators, following Jones (2017), Frost et al. (2023), Tangngisalu (2020) and L. Y. Lu et al. (2017). It is argued that the impact of financial performance on ESG disclosures or the relationships between financial measures can be non-linear (Jones, 2017). Unlike financial ratios, unscaled indicators do not face ratio distortion as a result of ‘the strict linearity between numerator and denominator’ (Jones, 2017, p. 1370). Despite unscaled indicators leading to heteroscedasticity within the database, gradient boosting machine in this thesis is less sensitive to this issue (Jones, 2017). Therefore, 102 accounting-based financial performance indicators have been selected.

- ***Traditional and unscaled accounting indicators*** are collected from the CSMAR database, with indicators from financial statements and income statements. Indicators have been used in Jones (2017), Frost et al. (2023), Tangngisalu (2020) and L. Y. Lu et al. (2017) and appear to be useful indicators in Chinese studies. Indicators include total cash, cash and short-term investment, total current assets, total assets, total current liabilities, total liabilities, total receivables, total equity, total debts, total revenue, net income after tax, intangible assets, gross profit, net accounts receivable, net inventories, short-term loan, total long-term liabilities, total non-current liabilities, total operating revenue, total operating costs, R&D expense, total profit, and net profit. This thesis will explore the value relevance of these traditional accounting numbers in predicting the level of SDG disclosures and whether useful accounting numbers in Chinese studies can also predict SDG disclosures.
- ***Cash flow performance indicators*** are collected from the CSMAR database, with indicators from cash flow statements. Consistent with Tangngisalu (2020), indicators include cash flow from operating activities, cash flow from investing activities, and cash flow from financing activities. These indicators are selected, because prior studies suggest that cash flow is a useful indicator in Chinese studies, such as corporate bankruptcy (Jiang & Jones, 2018). This leads us to think about whether some useful raw financial numbers in Chinese studies may also affect the level of SDG disclosures.

- **Earning capacity** indicators focus on corporate earnings, which are financial ratios, collected from the CSMAR database. Indicators include ROA, EBIT, EBITDA, Return on Equity, ratio of net profit to total profit, ratio of total profit to EBIT, ratio of EBIT to total assets, return on long-term capital, Gross Operating Margin, total leverage, and EV to EBITDA (Frost et al., 2023; Jiang & Jones, 2018; Uyar et al., 2024). These indicators are widely used in prior studies; however, mixed results have been found, which may be due to variable selection, contexts, or model selected.
- **Solvency and capital structure indicators** are also financial ratios and collected from the CSMAR database and Capital IQ. Consistent with Jiang and Jones (2018), Matuszak and Róžańska (2017), and Ting (2021), indicators include net cash flow from operating activities/current liabilities, working capital, cash flow-based interest coverage ratio, equity multiplier, liabilities to assets ratio, working capital to loans, net cash flow from operating activities to current liabilities, long-term loans to total assets, liabilities to tangible assets, equity to debt, equity multiplier, debt to equity, debt to long-term capital, long-term debt to working capital, and EBITDA to total liabilities.
- **Liquidity indicators** are collected from the CSMAR database and Capital IQ. Indicators include current ratio, quick ratio, cash ratio, working capital to loans, and working capital (Frost et al., 2023; Jiang & Jones, 2018; Uyar et al., 2024). These are widely used liquidity ratios in prior studies and this thesis aims to explore whether proxy selection will affect the prediction results.
- **Other financial ratios** include ratio of current assets, ratio of cash assets, ratio of receivable assets, ratio of working capital to current assets, ratio of working capital, ratio of working capital to net assets, ratio of non-current assets, ratio of fixed assets, ratio of intangible assets, tangible assets ratio, ratio of owners' equity, ratio of retained earnings to total assets, ratio of shareholders equity to fixed assets, ratio of current liabilities, ratio of operating liabilities, and ratio of accounts receivable to income. Prior studies also widely employed financial ratios as proxies for financial performance (Frost et al., 2023; Jiang & Jones, 2018; Uyar et al., 2024); however, mixed results have been found in prior studies, due to variable selection, heterogeneity of disclosures, and model selected.

Market-based financial performance

This thesis also employs several market-based predictors to see whether variable selection and proxy selection for the same variable may differently predict the level of SDG disclosures. Following Davern et al. (2019), this thesis will also investigate some non-traditional and forward-looking market-based indicators that affect CSR disclosures or carbon disclosures, to see whether these indicators may also affect SDG disclosures, particularly in the Chinese context.

Book-to-Market ratio, which is widely explored in existing studies, will be used in this thesis (Naheed et al., 2021). It is defined as total assets divided by market value. Data is collected from the accounting information quality database in the CSMAR database.

Earnings per share has been used in Jones (2017) and de Villiers et al. (2024). Given that prior studies have adopted different measurements for this variable, this thesis will include the following proxies to see their power in predicting the level of SDG disclosures. Proxies include earnings per share 1, earnings per share TTM1, earnings per share 2, earnings per share TTM2, earnings per share 3, earnings per share TTM3, earnings per share 4, and earnings per share TTM4. These are useful market indicators in predicting corporate bankruptcy and ESG disclosures, however, it is still unclear whether they may predict SDG disclosures.

Tobin's Q and **enterprise value** are widely used in prior studies to evaluate firm value. Data is collected from the CSMAR database and Capital IQ. However, mixed results have been found due to the heterogeneity of disclosures.

Beta has been used to evaluate stock price volatility (Jones, 2017). This thesis explored it from Beta 1, Beta 2, beta for 5 years and Yearly Beta Value of Composite Market several proxies.

Alpha proxies are collected from the Thomson Reuters database, including alpha based on regional ranking, alpha based on country ranking, alpha based on sector ranking, alpha based on industry ranking. This indicator is selected, as it is a useful variable in Chinese accounting studies.

Corporate governance

As theorised in Section 3.3.2.2, this thesis will test whether corporate governance may predict the level of their corporate SDG disclosures based on the gradient boosting machine, from 1. Board characteristics, and 2. Ownership concentration. China is unique for its top-down and concentrated governance structure, where board characteristics tend to be a less effective governance body and institutional shareholders are a minority group (Jiang & Kim, 2024). However, whether this will predict the level of SDG disclosures is unclear. Therefore, this thesis will employ a wide range of corporate governance indicators to test their predictive power in SDG disclosures. Data is collected from the CSMAR database, Thomson Reuters database, and Capital IQ. These databases are selected as they are widely used databases for collecting Chinese data and have different types of information.

Board characteristics.

This thesis will include widely used board characteristics, including the number of executives, the number of directors, the number of non-independent directors, and the number of independent directors. These indicators have been used in Buniamin et al. (2020) and Pizzi et al. (2021c). Different levels of the board of directors may have different power in corporate decision-making (Qian & Chen, 2021),

which offers opportunities to see the power of different types of board members to predict the level of SDG disclosures.

The number of executives measures the total number of executives in a company. Data is collected from corporate governance database in the CSMAR database. The database is limited to the number of executives disclosed in the annual reports, including general managers, president, CEO, deputy general manager, vice president, and secretary to the board of director. As discussed in Section 5.3.1.3, they tend to be the major decision-makers in the top management team and in corporate governance.

The number of directors measures the total number of directors in a company. Data is collected from corporate governance database in the CSMAR database. As discussed in 3.3, board size may indicate corporate capability in overcoming skill and experience discrepancies in corporate governance (Dyduch & Krasodomska, 2017).

The number of non-independent directors measures the total number of non-independent directors in a company. Data is collected from the corporate governance database in the CSMAR database.

The number of independent directors measures the total number of independent directors in a company. Data is collected from the corporate governance database in the CSMAR database. As discussed in Khlif et al. (2021), independent directors have financial and nonfinancial expertise and may expect companies to disclose information to meet their expectations. This is particularly the case in the developing countries, where investor protection may be relatively weaker.

Board meeting. This thesis also includes the frequency of board meetings as a proxy for board characteristics, consistent with Martínez-Ferrero and García-Meca (2020) and Dalla Via and Perego (2018). It is a proxy for board diligence. Data is collected from the CSMAR database, by measuring the number of board meeting held each year.

Ownership concentration.

This thesis explores ownership concentration from institutional ownership, insider ownership, shares owned by the top 5 shareholders, and shares owned by the top 10 shareholders. Indicator selection follows Frost et al. (2023), Lu and Abeysekera (2014), and Wang et al. (2013).

- ***Institutional ownership***. Consistent with Lu and Abeysekera (2014) and Wang et al. (2013), this thesis will include institutional ownership, which is a powerful group due to ‘the capital stake they control and their level of specialization and sophistication’ (Christensen et al., 2021, p. 1193). However, in China, institutional shareholders are not major players in the company, which leads to a gap in exploring its impact on SDG disclosures. This thesis collected

institutional shareholding data from the CSRMAR database, by measuring the number of shares owned by institutional shareholders in a company.

- **Insider ownership.** Following Nurleni and Bandang (2018), this thesis also explores the impact of insider ownership on the level of SDG disclosures. Insider ownership is measured based on the number of shares owned by insider shareholders. This indicator is selected, as managerial shareholders may face greater pressure from the government and society to do ‘politically correct’ activities, by showing their compliance with ESG activities in their reports. (Shahab & Ye, 2018; Zheng et al., 2014).
- **Shares owned by top shareholders.** Following Frost et al. (2023), this thesis employs the percentage of shares owned by the top 5 shareholders and the percentage of shares owned by top 10 shareholders. Given the unique concentrated governance structure in Chinese companies, this offers opportunities for exploring whether the concentrated governance structure can also predict the level of SDG disclosures. Data is collected from the CSMAR database. Specifically, this thesis collected the proportion of shares held by the top 5 and top 10 shareholders from the shareholder database in the CSMAR database.

ESG performance

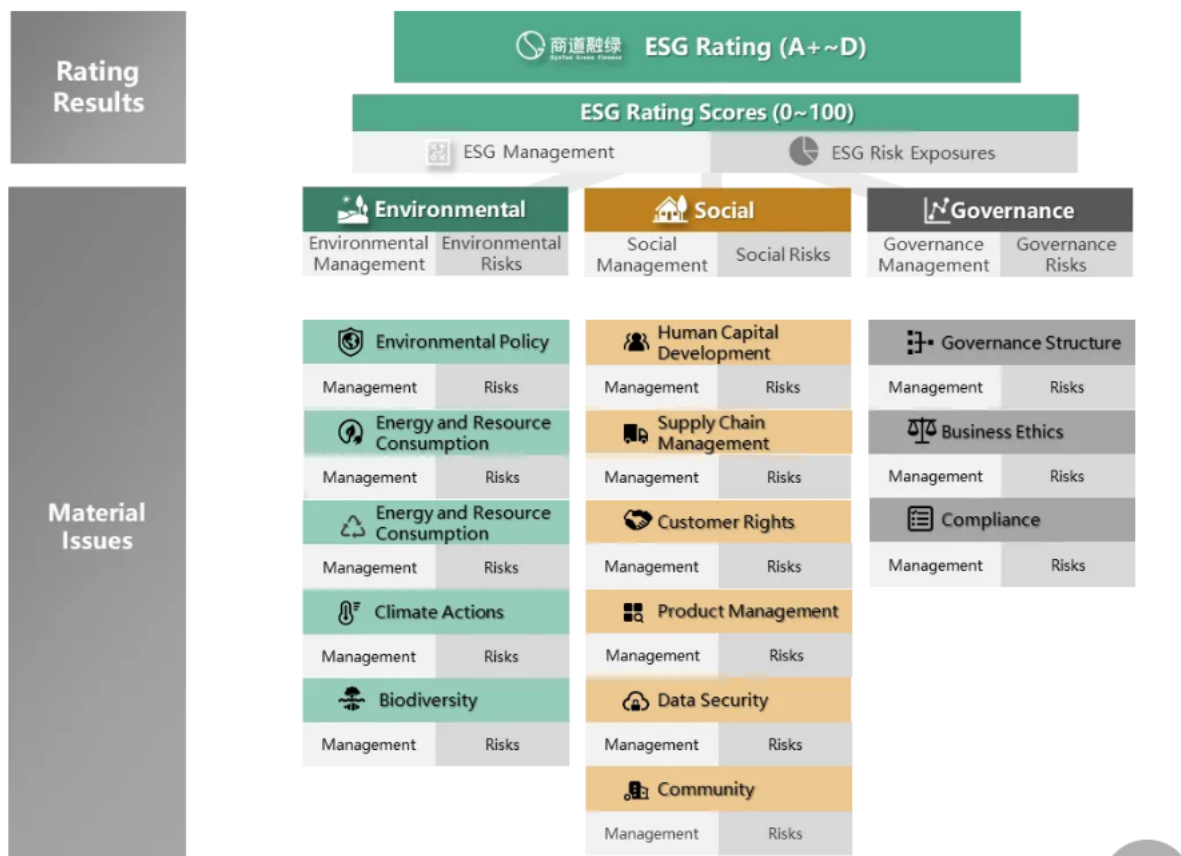
As theorized in Section 3.3.2.3, this thesis will test the impacts of different types of observable ESG performance activities on the level of SDG disclosures, from environmental, social and governance several aspects. These three areas are selected, as inconsistent results have been found between ESG performance and disclosures, due to ESG performance measurements, and research contexts. Divergence may occur between observable ESG disclosure practices and their ESG performance, as corporate response to sustainable development may be underpinned by regulatory force and institutional influence (Frost et al., 2023).

Data will be collected from SynTao Green Finance, which has been used by J. Wang et al. (2023) and Li and Xu (2024). It is widely used in Chinese studies to evaluate corporate ESG performance, due to its comprehensiveness, diversity, and accessibility. Some alternative ESG performance databases include Hexun CSR or the Wind databases; however, Hexun CSR collapsed in 2022, and the accessibility of this database is affected. I contacted the Wind database; however, they showed that their ESG performance information was collected from ESG reports. It is unclear whether it is performance or reporting data. The accessibility of the Wind database is another issue.

SynTao Green Finance developed the ESG rating system in 2015. The rating database offers ratings for all listed companies in China, including Mainland China and Hong Kong listed companies (<https://en.syntaogf.com/pages/methodology>). SynTao Green Finance offers both ESG scores (0 to 100) and ESG ratings (A+ to D). As shown in Figure 5.1, the ESG performance framework is developed to assess corporate management risk exposure to environmental, social, and governance three aspects,

which is made up of 14 key issues. Specifically, environmental issues include environmental policy, energy and resource consumption, renewable energy and resource consumption, climate actions, and biodiversity (<https://en.syntaogf.com/pages/methodology>). Social issues include human capital development, supply chain management, customer rights, product management, data security, and community (<https://en.syntaogf.com/pages/methodology>). Governance issues include governance structure, business ethics, and compliance. To capture information related to all 14 issues, SynTao Green Finance developed over 200 indicators, including core indicators and sector-specific indicators. Information is collected from reports, regulatory sources, media, macro data, geographical data, and satellite data (<https://en.syntaogf.com/pages/methodology>).

Figure 5.1 ESG rating system, extracted from <https://en.syntaogf.com/pages/methodology>



5.3.3 Control variables

Consistent with Beck et al. (2018) and Frost et al. (2023), this thesis will also include firm size, audit assurance, industry, analyst forecast, and financial risk as potential factors to investigate the determinants of SDG disclosures in Chinese companies. This thesis also includes the COVID-19 pandemic as a control variable. Data will be collected from CSMAR, S&P Global database, Capital IQ and Thomson Eikon.

Firm size is measured by market capitalisation (van der Waal & Thijssens, 2020; van Zanten & van Tulder, 2020). Data is collected from the CSMAR database. Studies widely suggest a positive relationship between firm size and SDG disclosures in various contexts (van der Waal & Thijssens, 2020; van Zanten & van Tulder, 2020). These studies suggest that larger companies may have greater access to resources and materials and have more exposure in the market, which may encourage companies to disclose SDG information. Some studies also argue that larger companies may face greater public scrutiny, which motivates companies to disclose sustainability information (Christensen et al., 2021). Therefore, this thesis includes firm size as a control variable.

Audit assurance is measured by investigating whether the company has audit assurance for sustainability information (Nichita et al., 2020; Rosati & Faria, 2019a). Data is collected from the Capital IQ database. Audit assurance of sustainability disclosures is inspired by financial assurance (Fernandez-Feijoo et al., 2018). Prior research also suggests that audit assurance is positively associated with SDG disclosure levels (Nichita et al., 2020; Rosati & Faria, 2019a). The assurance of sustainability disclosures did improve the quality of reports, thereby reducing stakeholder pressure on the companies and facilitating their survival and development (Belé et al.; Fernandez-Feijoo et al., 2018). Prior studies argue that audit assurance may increase the confidence of report users in their decision-making (Nichita, Nechita, Manea, Manea, & Irimescu, 2020; Rosati & Faria, 2019a). This thesis, thereby, also includes it as a control variable.

Industry background is proxied by classifying companies based on GICS 11 sectors. Data is collected from the CSMAR database and manually continues to be classified based on GICS 11 sectors. Prior research also suggests that the variation in SDG disclosures is affected by industries (Pizzi et al., 2021c; Rosati & Faria, 2019b). Specifically, prior studies observed that there is a higher level of SDG disclosure in environmentally vulnerable countries (García-Meca & Martínez-Ferrero, 2021; Rosati & Faria, 2019b). This thesis will explore whether similar results can be found in the Chinese context.

Analyst recommendation is proxied by analyst recommendation scores, following García-Sánchez et al. (2020) and Hummel and Rötzel (2019). Data is collected from the Capital IQ database. Scores range from 0 to 5, with a lower score referring to a stronger ‘buy’ recommendation and a higher score referring to a more negative recommendation—a ‘sell’ recommendation. This method has been used in García-Sánchez et al. (2020). It argues that analysts’ recommendations may reflect institutional pressure, where corporate reactions towards recommendation and relevant strategies would be considered legitimate (García-Sánchez et al., 2020). Disclosure serves as the basis for showing the extent to which companies have responded to analyst recommendations and meet stakeholders’ information needs (García-Sánchez et al., 2020). Therefore, this thesis intends to explore whether this relationship will remain in the Chinese context.

Financial risk is measured by Altman Z score, credit rating and going concern audit opinions, which will be collected from CSMAR, Capital IQ, and S&P Global databases. Prior studies widely adopt the Altman Z score, which is a traditional proxy for financial risks in Jones et al. (2017) and Frost et al. (2023). Given that the Altman Z score is a traditional measure of financial risks. This thesis also collected corporate credit ratings from S&P Global database.

Credit rating has attracted considerable attention from investors, researchers and regulators, indicating corporate financial performance and default risks (Bonsall et al., 2018; Cubas-Díaz & Martínez Sedano, 2018). Following Ben Saad et al. (2024), this thesis collected credit rating data from the S&P database. Given that limited data has been collected, this thesis follows Jones et al. (2015), and converts it into a dummy variable. If a company acquires a credit rating at AAA, it will be assigned a score of one, indicating a higher corporate credit rating. Otherwise, a score of zero will be assigned. The empty cells will be left blank.

This thesis also includes going concern audit opinions as a proxy for financial risks, with data collected from the CSMAR database (Tuo et al., 2023). A score of one will be assigned to the company that received going concern-related opinions; otherwise, a score of zero will be assigned. Given that sustainability disclosures may serve as the basis for indicating corporate financial health, corporate sustainability awareness and governance, it may further mitigate auditors' concerns about corporate financial risks or going concern problems (Rezaee, 2016). Sustainability reports may indicate corporate financial ability, which will further reduce auditors' concerns about corporate going concern and financial risks. This thesis explores whether this may also predict the variation in the level of SDG disclosures in Chinese companies.

COVID-19 is measured based on whether COVID-19 happened in a particular year, following Zambrano-Monserrate et al. (2020) and Alkayed et al. (2024). This thesis assigned a score of one to the year 2019 and assigned a score of 0 to all other years. Given that COVID-19 increased the uncertainty of the social and economic world, it was found that it slowed down the progress towards Agenda 2030, and removed some SDG achievements (Mukarram, 2020). However, Alkayed et al. (2024) found a positive impact of the COVID-19 on sustainability disclosures. This study explains that the COVID-19 pandemic offers an opportunity to see corporate reactions towards how 'they give back to society during a pandemic' (Alkayed et al., 2024). Disclosure may serve as the basis for companies to demonstrate corporate alternative reactions towards the pandemic and to show resilience to investors (Alkayed et al., 2024). While COVID-19 has negative impacts on SDG progress, there remains a gap in understanding whether COVID-19 may have impacts on corporate sustainability and finally, influence corporate responses through reporting.

5.4 Chapter conclusion

To investigate RQ2 factors that may predict the level of SDG disclosures, this chapter introduced the gradient boosting machine to overcome the limitations of traditional models, handle complex nonlinear relationships, make out-of-sample accurate predictions, and understand interaction effects.

This chapter also specifies data sources, proxies and measurements for government control, financial performance, corporate governance, ESG performance, and other control variables, to explore factors that may predict SDG disclosures in Chinese companies.

Chapter 6 Data analysis

6.1 Chapter introduction

The previous chapter discussed the research design for RQ2. This chapter moves on to data analysis, focusing on visualising the variation in SDG disclosures in Chinese companies in Section 6.2 and analysing whether government control, financial performance, corporate governance, and ESG performance may predict the level of SDG disclosures based on the gradient boosting machine in Section 6.3. Section 6.4 offers a discussion of the results, while Section 6.5 concludes this chapter.

6.2 Descriptive analysis of the level of SDG disclosures

This section answers RQ1 what is the extent of SDG disclosures in Chinese companies, following China's adoption of Agenda 2030, from the disclosure trend across the sampled period and across different categories. As discussed in Section 4.6, this thesis developed a framework of 27 indicators to capture corporate explicit reference of SDGs and implicit SDG performance to fill the gap that the 'quality' of existing SDG disclosure frameworks is limited.

6.2.1 Box plots for visualising the level of SDG disclosures

Boxplots are used to visualise the variation of SDG disclosures in Chinese companies across the sampled period 2016-2022. Box plots provide a good understanding and overview of a dataset's central tendency, variability and distribution shape without detailed calculations. Figure 6.1-6.4 visualise the variation in the extent of SDG disclosures, from total SDG disclosures, poverty alleviation, energy and climate change, and reflect the variation in disclosures across all 300 firm-year observations. The x-axis shows the timeframe, while the y-axis represents disclosure scores. Each dot represents an individual data point, offering insights into the spread of data and extreme values of SDG disclosures. In each box, there is a black line and a green triangle. The black line shows the median of disclosure scores for a particular year, while the green triangle represents the mean of corporate disclosure scores for that year. This thesis observed improved SDG disclosures in Chinese companies over time. This may indicate that after the introduction of SDGs in China, Chinese companies are willing to align with these goals, even if companies do not have to comply with them. It is not surprising that companies gradually made a higher level of disclosures, as they need time to react to these goals, incorporate them into business operations and finally, output a delayed response through disclosures. Now, I will discuss them in detail.

Figure 6.1 displays the variation in total SDG disclosures between 2016 and 2020. Black dots show wide-ranging disparities in disclosures across companies. The lowest value is at zero, while the highest value reaches 25. Boxes represent the interquartile range, visualising the middle 50% of total SDG disclosure scores each year. It is observed that in 2016 and 2017, interquartile ranges were taller, ranging from 0 to 18. This may suggest a greater disparity in total SDG disclosures in 2016 and 2017. One of the reasons for explaining this could be that companies were still working on how they can incorporate SDG into business practices and make SDG disclosures, especially for poor disclosers. Boxes become shorter over time and the bottom line (25% percentile) dramatically increased from 2018 onwards, indicating that companies with lower SDG disclosures were catching up with their SDG practices. Additionally, this thesis observed a reduced disparity of SDG disclosures across companies each year. Interestingly, top disclosers still make a consistent level of SDG disclosures, while poor disclosers appear to make greater efforts to achieve a higher level of SDG disclosures. Overall, the disparity in

SDG disclosures is higher in the early years following the introduction of SDGs. Companies provide a higher level of SDG disclosures over time.

The mean of corporate SDG disclosures (green triangles) consistently increased, from 10 to nearly 15. The horizontal line in each box represents the median of SDG disclosure scores each year. Despite fluctuating between 13 and 17, the median of disclosures remains relatively stable. The significance of the variation in SDG disclosures over the sampled period is analysed in Section 6.2.3.

Figure 6-1 Total SDG disclosure scores by year

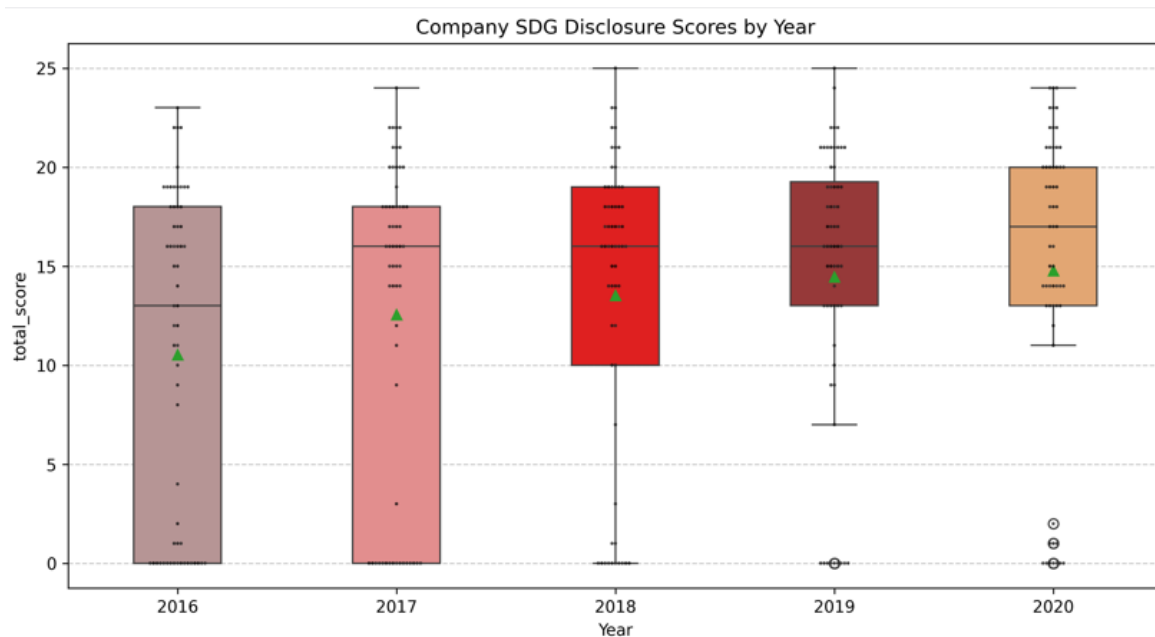


Figure 6.2 displays the variation in poverty alleviation disclosures between 2016 and 2020, where scores range from 0 to 12. Again, in 2016 and 2017, boxes covered scores ranging from 0 to 10, which are taller than those in 2018, 2019, and 2020. This indicates higher disparities in poverty alleviation disclosures in 2016 and 2017 across companies. It is observed that a number of companies had lower-level poverty alleviation disclosures when SDGs were introduced in China. However, the lowest 25% of disclosers have dramatically increased their disclosure levels since 2018, indicating a transition from being aware of SDGs to actively engaging in SDG disclosures. It may raise the questions about the rationale for the dramatic increase in poverty alleviation disclosures in 2018.

The mean score for poverty alleviation disclosures experienced a slight increase, increasing from 5.8 to approximately 7, while the median score of poverty alleviation disclosures also varies between 6 and 10. The median scores increased from 7 in 2016 to around 9 in 2017. The median of poverty alleviation scores slightly fluctuated between 2018 and 2020. However, these scores seem to vary slightly over the sampled period. The significance of the variation in poverty alleviation disclosures over the sampled period is analysed in Section 6.2.3.

Figure 6-2 Poverty alleviation disclosure scores

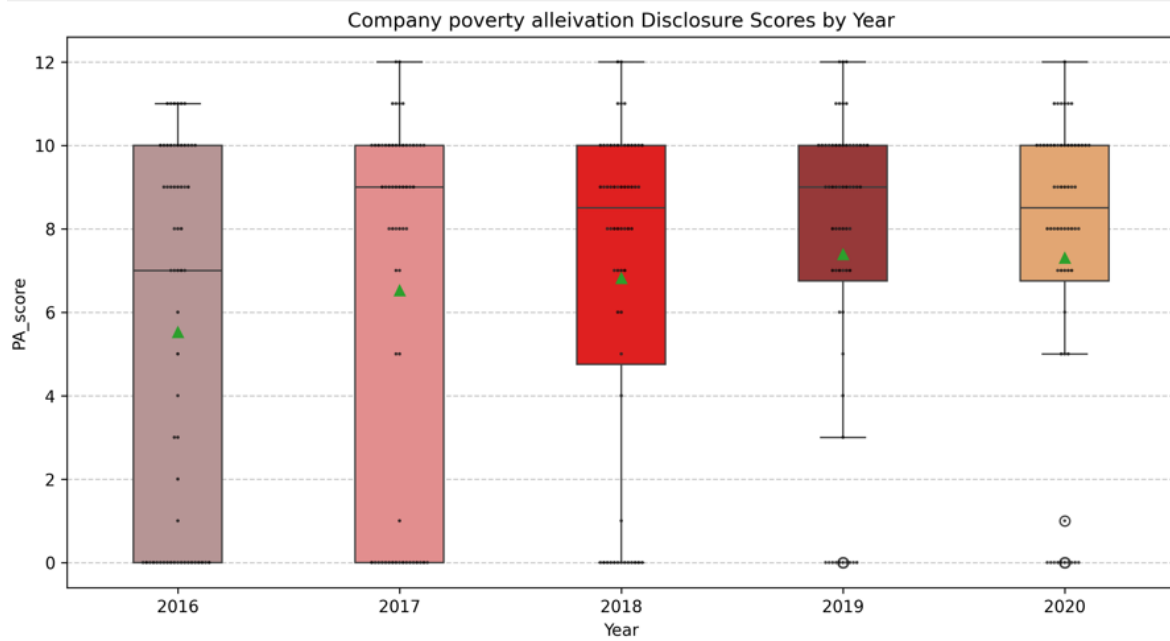


Figure 6.3 displays the variation in energy disclosures between 2016 and 2020, with the scores ranging from 0 to 5. Similar to Figures 6.1 and 6.2, more companies make energy disclosures during this reporting period. In 2016, a proportion of Chinese companies fell within the zero-disclosure category in energy disclosures. This situation has improved over time. By 2020, it is evident that most companies had achieved scores of four or five, suggesting that SDG 7 clean energy has gradually become an issue for companies to widely engage in and report on. Again, this thesis observed an increase in energy disclosures after the Chinese government signed on SDGs and incorporated them into national strategies. However, the whisker lengths across the sampled period are all from 0 to 5, indicating that there remain varying levels of SDG disclosures across sampled companies.

Figure 6-3 Energy disclosure scores

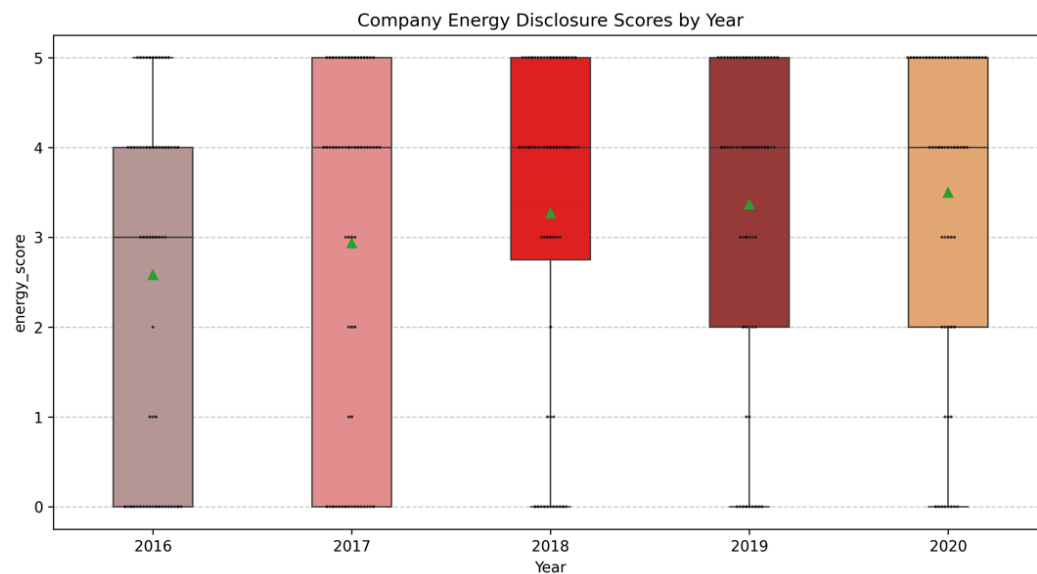
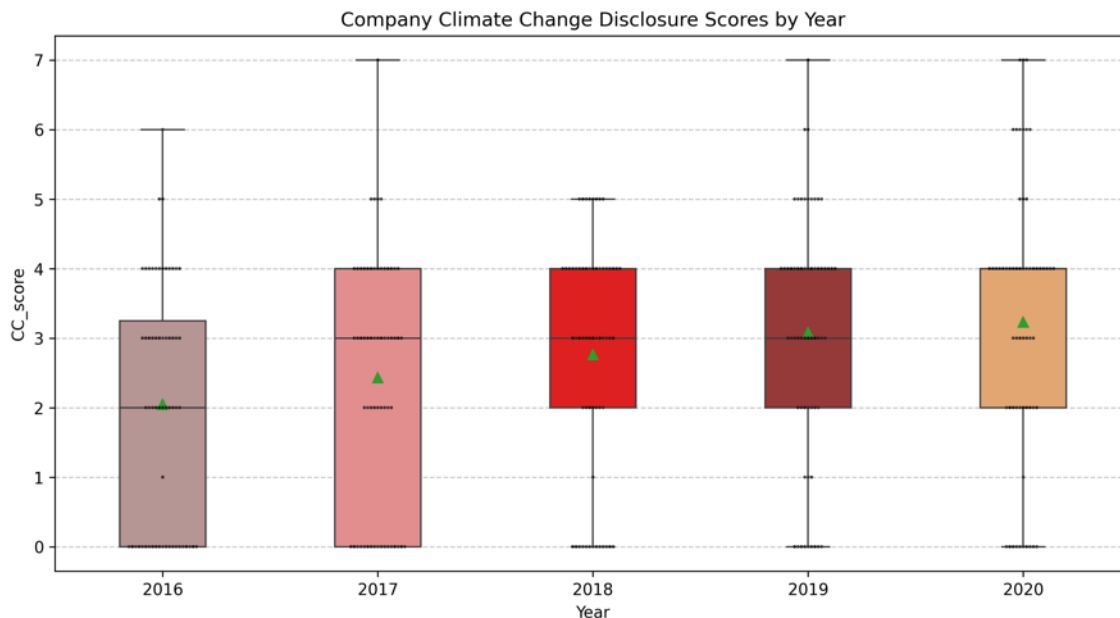


Figure 6.4 displays the variation of climate change disclosures from 2016 to 2020, with scores ranging from 0 to 7. Again, the pattern observed in climate change disclosures is broadly consistent with that of total SDG, poverty alleviation, and energy disclosures. However, unlike poverty alleviation and energy disclosures, Chinese companies tend to have lower levels of climate change disclosures, with mean and median disclosure scores between 2 and 3. It is substantially below the maximum score of 7. This raises the question of whether the lower level of climate change disclosures may be due to relatively weaker regulatory enforcement. As discussed in Section 2.3, while an increasing number of regulations and guidelines have been introduced to guide corporate ESG disclosures, limited guidance has been provided related to climate change disclosures. In 2024, climate change disclosure regulations were initially introduced and mandated. This may offer insights for whether significant changes in disclosures may occur after the introduction of mandatory climate change disclosure requirements. In 2016, a number of sampled companies fell into the zero-disclosure category, while the highest score was six. Subsequently, the number of companies with a score of zero decreased, and more companies achieved a score of seven by 2020. This indicates that an increasing number of companies disclosed climate change information over time. However, given the extended lengths of the whiskers, the disparity of climate change disclosures across companies is still substantial. This raises the question of whether relatively lower levels of climate change disclosures may be due to the weaker regulatory environment.

Figure 6-4 Climate change disclosure scores



6.2.2 Descriptive statistics of SDG disclosures

To analyse what is the extent of SDG disclosures, this thesis also visualises the variation of SDG disclosures across different reporting categories among sampled Chinese companies. This thesis observed that imbalanced disclosures have been observed across different categories. Specifically, while only a small proportion of companies explicitly reference SDGs in reports or chairman message, this

thesis did observe that a number of Chinese companies widely acknowledge poverty alleviation, energy and climate change activities in their reports. However, this thesis is limited to exploring whether the company discloses relevant information, overlooking the nature of information and the substantiveness of information. Now, I will discuss them in detail.

Table 6.1 displays an overview of the number of observations involved in each disclosure category during the sampled period. Of the 12 poverty alleviation categories, TPA of education and social impact tend to be the most widely disclosed ones, with the sum of the observations at 219 and 218 respectively during the sample period. Of the 300 observations, over 200 also disclosed industrial poverty alleviation, employment, health care, and regional development. The least disclosed categories include TPA of bottom-up approach, relocation impoverished, and ecological improvement. In China, corporate poverty alleviation disclosures focus more on industrial development, health care, and government activities, which is consistent with Wang (2017). However, Subramaniam et al. (2020, p. 26) suggest that ‘ANZ Bank focuses on financial inclusion, employment and community programs, to help customers...’. Therefore, poverty alleviation under the umbrella of SDG1 No Poverty is perceived economically different in different countries. As expected, the conservations about poverty alleviation are different between developing and developed countries. This could be explained that institutional background, regulatory enforcement, and corporate characteristics may differently affect corporate practices (See Section 3.3). Overall, different levels of assistance have been offered to the poor areas by Chinese companies. One reason for the uneven disclosures could be the limited and immature regulatory requirements on corporate poverty alleviation engagement and disclosures.

Imbalanced disclosures have been observed in energy disclosures, with ‘mentioning energy activities in reports’ as the most widely disclosed category and ‘companies claiming that they invest in energy activities’ as the least disclosed category. Of the 300 observations, 226 observations mention SDG 7 or acknowledge energy activities in their ESG reports, while 216 observations disclose their energy consumption in reports. However, less than half of the observations (140) disclose corporate spending or resource commitment towards SDG 7. Approximately 180 observations disclose their energy efficiency and renewable energy consumption information in ESG reports. Compared with poverty alleviation and climate disclosures, energy disclosures appear to be more balanced. One reason for explaining this could be the increasing number of voluntary and mandatory environmental regulations and guidelines introduced since 2008 (See Section 2.3). This leads us to think about whether the ESG regulatory system may affect corporate ESG disclosures.

Limited disclosures have been observed in climate change disclosures. Unlike energy disclosures, a number of Chinese companies engaged in corporate energy investment, while a small number of companies made Scope 1 and Scope 2 disclosures. However, the CDP questionnaire (2022) suggests that 75% Chinese companies disclosed Scope 1 emissions, while 58% disclosed Scope 2 information

in their CDP reports. The inconsistent results may indicate that Chinese companies may provide varying information across different platforms to meet various stakeholders' information needs. As suggested by Khalid et al. (2022), the Chinese government was working on the assessment mechanisms for climate change. While an increasing number of Chinese companies have adopted the CDP reporting framework as a guideline for climate change disclosures, the number of Chinese companies that responded to the CDP, was still behind developed countries.

Table 6.1 Descriptive statistics of the number of observations involving in each category over the sampled period.

Categories	Valid	Sum statistic
General SDG information	300	102
Chairman message	300	41
About the report	300	42
Mentioning SDG 1 or acknowledging activities in this space	300	210
Involving in government-related poverty alleviation activities	300	182
Industrial development	300	203
Employment	300	200
Health care	300	203
Bottom-up approach	300	57
Social impact	300	218
Education	300	219
Regional development	300	203
Relocating impoverished	300	23
SDG 1 spending or resource commitment	300	199
Ecological improvement	300	95
Mentioning SDG 7 or acknowledging activities in this space	300	226
Energy consumption	300	216
SDG7 spending or resource commitment	300	140
Energy efficiency	300	176
Renewable energy consumption	300	181
Mentioning SDG13 or acknowledging activities in this space	300	214
Emission	300	216
SDG 13 spending or resource commitment	300	168
Scope1	300	23
Scope2	300	29
Education and training related to climate change	300	121
Risks and opportunities related to climate change	300	43

6.2.3 Kruskal-Wallis test for SDG disclosures across years

Kruskal-Wallis test is widely used to analyse the significance of variations in sustainability disclosures (Bhatia & Tuli, 2014; Machado et al., 2021). Given that it is a non-parametric test, it is employed when the assumption of normality and homogeneity of variance is not satisfied across samples (Ostertag et al., 2014). It determines whether each group has the same distribution (Ostertag et al., 2014). Thereby, this thesis uses Kruskal-Wallis tests to analyse whether significant variations in SDG disclosures can be observed across sampled years.

Table 6.2 displays the results of Kruskal-Wallis tests for total SDG disclosures, poverty alleviation disclosures, energy disclosures, and climate change disclosures respectively. It is observed that with a significance level at 0.05, the distribution of overall SDG disclosures is not the same across sampled years. Similar results have been found in the distribution of energy disclosures over the sampled years, with a significance level at 0.04. With a significance level at 0.00, the distribution of climate change disclosures is also significantly different over the sampled period, particularly at 1% significance level. However, the distribution of poverty alleviation disclosures over the sampled period does not have significant variations across sampled years (p-value at 0.30).²²

Table 6.2 Kruskal-Wallis results

	Total N	Test Statistic	Degree of freedom	Asymptotic Sig. (2-sided test)
Total SDG disclosures	300	9.87	4	0.05**
Poverty alleviation disclosures	300	4.82	4	0.30
Energy disclosures	300	10.27	4	0.04**
Climate change disclosures	300	16.26	4	0.00***

6.3 Empirical analysis of factors that may predict the level of SDG disclosures

To analyse what factors may predict the level of SDG disclosures in Chinese companies, this thesis employed gradient boosting model to analyse whether government control, financial performance, corporate governance and ESG performance can predict the level of SDG disclosures. Section 6.3.1 provides an overview of the dataset and Section 6.3.2 shows the interpretation of empirical results.

6.3.1 Descriptive statistics

Appendix 8 displays descriptive statistics for all variables based on 300 observations during the period between 2016 and 2020. Specifically, this thesis includes the following factors to summarise the nature of the dataset, such as the number of observations, range, minimum, maximum, mean, standard deviation, skewness, and kurtosis. This thesis observed that the current dataset contains missing values, different types of variables, such as categorical and continuous variables, and the skewness of data.

N shows the number of observations for each predictor. With the highest value at 300 and the lowest value at 116, this suggests that some variables include missing values. As shown in Appendix 8, missing values primarily occur in financial performance variables. These include traditional financial variables, such as gross profit, total receivables, total equity, and cash and cash equivalents, financial ratios, such as current ratio, working capital to loans, alpha measures, such as combined Alpha model region rank,

²² This study also conducted post hoc tests, which are widely employed as additional tests for Kruskal Wallis tests. The post hoc test aims to identify whether significant differences can be identified in homogeneous groups. It tests whether significant variations of SDG, poverty alleviation, energy and climate change disclosures based on homogeneous year groups. However, significant differences can only be detected between 2016 and 2020 only, for SDG disclosures, poverty alleviation disclosures, energy disclosures and climate change disclosures. No significant results have been found for the rest homogeneous group (See Appendix11).

and combined Alpha mode country rank and ESP measures, such as earnings per Share TTM1, earnings per Share 2, earnings per Share TTM2, earnings per Share 3 ,earnings per Share TTM3 (See Appendix 8).

Range, minimum, and maximum may visualise the spread of data. It can be observed that this thesis includes both categorical and continuous variables. Categorical variables include BRI, CDP disclosures, SDG IDZ, the nature of companies (Commercial SOE, public interest SOEs, and other types of companies), political connection, credit rating, and industry, while other variables are continuous variables, such as financial performance, corporate governance, and sustainability performance variables.

Skewness may indicate the distribution of data, whether it is symmetric or not. As shown in Appendix 8, the range of skewness is between 9.58 and -10.04. The distribution of data widely varies across different variables. Specifically, the skewness range of financial performance variables is between 9.58 and -10.04. Examples of highly non-symmetric variables include working capital to loans (skewness value=9.58), total leverage (skewness value=8.03), ratio of working capital (skewness value=7.16), return on equity (skewness value=-10.04), ratio of net profit to total profit (skewness value=-8.29), and return on long-term capital (skewness value=-7.45). Examples of financial performance variables that are close to symmetric distributions include ratio of current asset (skewness value=0.11), working capital (skewness value=0.14), ratio of operating liabilities (skewness value=-0.09), and ratio of non-current assets (skewness value=-0.14). With the highest skewness value at 11.6 and lowest skewness value at -0.85, this suggests that the distributions of corporate governance variables widely vary. Specifically, the highest skewness value at 1.65 and lowest skewness value at -0.05 indicate that the distributions of sustainability performance variables are slightly asymmetric.

Overall, descriptive statistics summarise the nature of the dataset, such as missing values, central tendency, the spread of data, and the distribution of data. It is observed that raw data includes both categorical and continuous variables, missing values, and asymmetric distributions.

6.3.2 Empirical result interpretation

The empirical results are interpreted from model fit, relative variable importance, interaction effects, and partial dependence plots aspects.

6.3.2.1 Model error summary

Figure 6.5 displays how the mean squared error changes for learn, learn sampled, and test samples when the number of trees increases based on the gradient boosting machine. The Y-axis shows the changes in mean squared error (MSE), while the X-axis shows the number of trees used in the model, where the number of trees refers to the number of iterations. The figure visualises the optimal number of trees to achieve the smallest MSE value (that is, the MSE can be optimised at k trees). Specifically, the red curve displays the changes in MSE for the test sample, while the green one refers to the changes in MSE

based on learn sampled set. Minitab (2019, p. 111) argues that 'learn sampled error curve displays the result of a random subsample in each iteration and each iteration corresponds to fitting a tree to the generalised residuals'. The mean squared errors for both sets 'continuously and monotonically' decrease as the number of trees increases. The curves decrease dramatically for the initial 200 trees. The curves flatten out when the number of trees increases.

The optimal MSE value occurs at 2000 trees, with the lowest MSE value for the test sample at around 40. Figure 6.5 shows that the MSE value for learn sampled set is also optimised at 2000 trees, with the lowest MSE value at around 10. The test curve (red curve) is slightly above the learn sampled one (green curve), indicating that the model performs better based on learn sampled data.

Figure 6-5 MSE for learn, test and learn sampled set for gradient boosting model

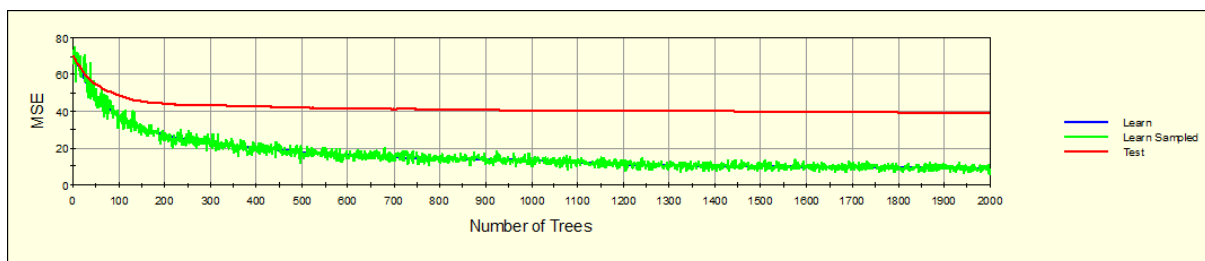


Figure 6.6 visualises how the value of R-squared changes when the number of trees increases. The blue curve shows the changes in the R-squared of the learn sample. As shown in Figure 6.6, the curve sharply increases for the first 600 trees. It flattens out when the number of trees is around 2000. The R-squared value is optimised at around 0.85 when 2000 trees are adopted. The red curve displays the changes of R-squared based on the test sample. It is observed that the optimal R-squared for the test sample occurs at 2000 trees, with the R-squared close to 0.4. Overall, the model is optimised at 2000 trees.

Figure 6-6 R-Sq for learn, test and learn sampled set for gradient boosting model

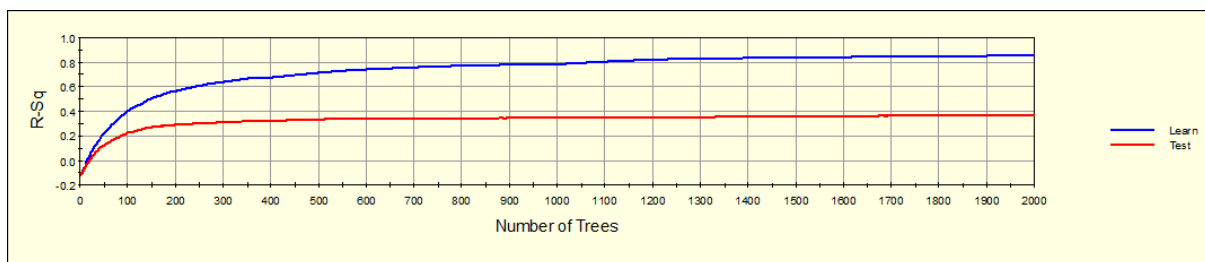


Table 6.7 summarises model error measures, based on learn, test and learn sampled sets. **The average R squared (R-Sq)** for the learn sample is 0.85, indicating that the data of independent variables in the training dataset can predict 85% of the variation in SDG disclosures. Across all test samples, the average out-of-sample R-Sq is 0.37, higher than prior studies employing statistical inference analysis. R-Sq or adjusted R-Sq in van der Waal and Thijssens (2020), Sasanti et al. (2020), and Gambetta et al. (2021) range from 0.16 to 0.22, which are relatively lower. These studies employed OLS and logit models, which are often criticised for, such as IID problems and P-hacking. In this thesis, out of sample R-Sq

and R-Sq norm are around 0.4. This indicates a reasonable level of model fit in predicting the level of SDG disclosures²³, which is consistent with Nechita et al. (2020a) and Pizzi et al. (2021c). However, these studies are limited to a small number of variables. Even if they have even higher R-squared (See Section 3.4), the gradient boosting machine still outperforms other traditional statistical methods, as a result of its superior capability of accommodating a greater number of predictors.

Table 6.3 Model error summary

Name	Learn	Test	Learn Sampled
RMSE	3.02	6.26	2.84
MSE	9.11	39.17	8.08
MAD	1.63	4.83	1.59
MAPE	0.29	0.64	0.20
SSY	18755.67	18755.67	n/a
SSE	2732.84	11750.40	1211.92
SAE	489.84	1447.52	238.81
R-Sq	0.85	0.37	n/a
R-Sq Norm	0.86	0.38	n/a
AIC	980.79	1418.36	n/a
AICc	1344.22	1781.79	n/a
BIC	1569.70	2007.26	n/a
ROC (Area Under Curve)	0.95	0.74	n/a

6.3.2.2 Data analysis of the variation in the power of predictors in predicting the level of SDG disclosures

This section analyses whether governmental control, financial performance, corporate governance, and sustainability performance may predict the level of SDG disclosures within the gradient boosting machine, by using relative variable importance and partial dependence plots²⁴.

²³ **Model accuracy** measures, mean squared error (MSE), Root Mean Squared Error (RMSE), Median Absolute Deviation (MAD), and Mean Absolute Percentage Error (MAPE), are primary tools for evaluating how actual and predicted values are different from each other, which are widely used in machine learning analysis (Noorunnahar et al., 2023; Zheng et al., 2020). MSE is a widely used metric for evaluating the difference between actual and predicted values. The average MSE for learn samples is 9.11, which increases to 40 in the test set. The RMSE is the square root of the MSE, which is scale-dependent and penalises data points that have higher residues (Kotu & Deshpande, 2018). This study observes a value of 3.02 for the learn sample and a value of 6.26 for the test sample. Similarly, MAD (mean absolute deviation) measures the absolute deviation between actual and predicted values, which is less sensitive to outliers. The values of MAD for learn, test and learn sampled sets are 1.63, 4.83, and 1.59 respectively, which are lower than the values of RMSE and MSE. MAPE (Mean absolute percentage error) evaluates errors in the percentage format. It is argued that MAPE is widely adapted when sufficient data are available (De Myttenaere et al., 2016). The MAPE value for the learn sample is 0.29, indicating that the model's predictions are 29% deviating from the actual values. MAPE for the test sample is higher, at 0.63.

Table 6.3 also includes Akaike information criterion (AIC), corrected Akaike Information Criterion (AICc), and Bayes information criterion (BIC) to evaluate model performance. AIC for learn sample is 980.79, and AIC for test sample is 1418.36, which is slightly higher than that of the learn sample. AICc made some adjustments, designed for a small sample size. The AICc for the learn sample is 1344.22, which is higher than AIC. AICc for test sample is 1781.79, slightly higher than AIC for the test sample.

²⁴ This thesis has adopted relative variable importance and partial dependence plots to visualise results, which are widely used in Jones (2017) and Frost et al. (2023). Specifically, as discussed in Section 5.2, it visualises the relative contribution of a variable to the prediction results. The strongest variable is assigned a score of 100, while all the other variables are assigned a scaled score relative to the most influential variable (Jones, 2017). As discussed in Section 5.2, partial dependency plots illustrate the directional contributions of a variable to the output prediction (Jones, 2017; Frost et al. 2023). The X-axis shows the actual value of the predictor, while the y-axis shows the marginal effect of the dependent variable (SDG disclosures). Each figure contains two curves: the blue curve displays the actual impact of a variable, and the green curve is the first-order single-

Table 6.8 displays that out of 159 variables, 135 have non-zero RVI scores. That is, all 135 variables have out-of-sample predictive power in predicting the level of SDG disclosures, with the strength of variables widely varying. Despite offering a high-dimensional analysis (the number of variables can be greater than the sample size), the gradient boosting model still may introduce random noise to the results. For example, some variables may have low or zero RVI scores. Following Jones (2017), this thesis considers variables with low/ small RVI scores as random noise in predicting the level of SDG disclosures. Possible reasons for explaining the weaker RVIs could be the measurements of variables, sample size, and disclosure measurements. The directional contributions of governmental control and financial performance to the level of SDG disclosures appear to be mixed, due to variable selection and the measurement of each variable. It is interesting to note that ESG performance indicators, institutional shareholding, and shares owned by the top 5 shareholders consistently and positively predict the level of SDG disclosures. Now, I will discuss them in detail.

Government control

As discussed in Section 3.3.1, government control is unique in the Chinese context, due to the authoritative setting and top-down governance structure. As discussed in Section 3.3.1, despite moving to a more capitalist market, the government still substantially influences businesses to achieve their political aims (Situ et al., 2018). Unlike in Western contexts, government controls in China are not limited to voluntary and mandatory guidelines. This thesis explores whether governmental control may have predictive power in predicting the level of SDG disclosures in the gradient boosting model, from regulatory influence (proxied by BRI and SDG IDZ project), government shareholding (proxied by commercial SOEs, public-interest SOEs, and other types of SOEs), political influence (party committee size, the presence of corporate political connection, the number of senior management members being politically connected, the presence of politically connected CEO/executives, the number of politically connected CEO/executives, the presence of politically connected directors, and the number of politically connected directors) and incentivising influence (the issuance of green bonds) aspects (See Section 5.3).

Relative variable importance of government control variables

Table 6.4 displays that a number of governmental control variables have predictive power on the level of SDG disclosures. However, with the highest RVI of 59.19 and the lowest RVI of 0, the predictive power of direct and indirect government control tools widely varies. Specifically, some political influence factors, such as party committee and the number of politically connected senior management team members, have the highest predictive values across all government control factors. Incentivising influence and the nature of the companies have relatively lower predictive values. This may alert the

knot spline, aiming to smooth out the direction of the prediction. The plots also visualise whether the impact is linear or non-linear.

government and regulatory bodies about what can be more effective control tools in China to encourage companies to engage in sustainable development. It is also interesting to note that this thesis did find that the predictive power of different levels of political connections on SDG disclosures widely varies. As suggested by Qian and Chen (2021), Chinese companies have unique top-down governance structures, where power distance may affect corporate decisions. However, this thesis finds that political influence predictors proxied by the presence of political connections do not have predictive power on the level of SDG disclosures. This can also be explained by the measurements of political connections (Details will be discussed below). However, given that some prior studies did find the presence of political connection affects the issuance of ESG disclosures, the inconsistent results may be due to the heterogeneity of disclosures. For example, Marquis and Qian (2014) and Wang et al. (2018) focus on the issuance of ESG disclosures, while this thesis looks at the level of SDG disclosures. This may offer some opportunities for future research to explore whether the results may remain in other types of ESG disclosures in Chinese companies, and to explore whether the definition of political connection may also affect the overall prediction results. Now, I will discuss them in detail.

Political influence. Table 6.4 displays that political connection through party committees is the strongest government control variable, with an RVI of 59.15. This indicates that a party committee serves as an effective governance body in predicting corporate SDG disclosures. As suggested by Yin et al. (2019), a party committee is integrated into organisations to bridge the gap between the enforcement of national strategies and the development of the culture of social responsibility. This is followed by the number of politically connected directors (RIV=54.58), the number of politically connected CEOs or executives (RVI=52.12), and the number of politically connected senior management members (RVI = 46.45). The variation in the predictive power of these variables may be due to different levels of corporate positions for connected members. As Qian and Chen (2021) suggested, different levels of corporate positions may affect corporate decisions. Therefore, this thesis encourages future research to explore whether these results may remain the same in other Chinese companies.

Interestingly, the presence of politically connected members, such as directors, CEOs, executives, or other senior management members, does not have power to predict the level of SDG disclosures, with RVI scores of zero. This is inconsistent with Wang et al. (2018a) and Marquis and Qian (2014), which argue that the presence of corporate political connections may significantly influence the issuance of CSR disclosures. These studies argue that disclosure may enable companies to maintain their legitimacy in the eyes of the government and they may sacrifice the 'quality' of disclosures. The heterogeneity of the disclosures could explain inconsistent results. Prior studies focus more on the issuance of CSR disclosures, while this thesis looks at the level of SDG disclosures. This raises the question of whether the presence of political connections may predict the adoption of SDG disclosures, rather than the level of SDG disclosures. Additionally, CSR performance and disclosure practices are largely promoted

through voluntary and mandatory regulations, while SDGs are incorporated into national strategies and political ideologies. From this perspective, it is unclear whether politically connected parties may respond to these sustainable development activities differently. Another possible reason for explaining the inconsistent results could be the measurements of political connections. Marquis and Qian (2014) are limited to exploring political connections related to CPPCC or government officials, while this thesis defines any of these experiences—governmental officials, NPC, CPPCC, and military background—as political connections. Another reason could be the time period, due to the changes in regulations. Specifically, the research period for Marquis and Qian (2014) is between 2006 and 2009, when Regulation 18 was not introduced. China's Central Organization Department introduced Regulation 18 in 2013, regulating that government officials have to resign from businesses (<http://renshi.people.com.cn/n/2013/1030/c139617-23380219.html>). It is unclear whether regulatory changes may also affect their political motivations in engaging in sustainable development activities and making relevant disclosures to maintain legitimacy in the eyes of the government. Additionally, given that some predictors have low or zero predictive values, this may be due to the small sample size, the definition of political connections, and the measurements of political connections. Future research could explore whether similar results could be found using a larger sample size, and different types of sustainable development disclosures. It is also unclear whether Western political factors may affect their SDG disclosures.

Government shareholding is another direct government control tool, proxied by commercial SOEs, public-interest SOEs, and other companies. This thesis observed that the RVI for Public interest SOEs is 21.6, while the RVI for commercial SOEs is 14.43. The nature of SOEs may affect how companies respond to SDG disclosures differently. One of the possible explanations for this could be that public-interest companies may be concerned more about sustainable development, while commercial SOEs tend to focus more on financial benefits. However, the proxies for the nature of SOEs have relatively weaker power in predicting the level of SDG disclosures, which could be due to measurements of the nature of SOEs. The guide provides a relatively general guideline for this thesis to define what can be classified as commercial and public-interest SOEs. This may lead to a gap in understanding whether the results may be generalisable to a larger sample size and whether the measurements of the nature of SOEs may affect the overall results.

Regulatory influence is a direct government control tool, proxied by the BRI and SDG IDZ project. Unlike direct corporate-level government control tools, indirect national strategies and initiatives have relatively lower power in predicting the variation of corporate SDG disclosures. Specifically, the RVI for the Belt and Road Initiative is 34.31, indicating that corporate engagement in BRI has some predictive power on the level of SDG disclosures. The RVI of the SDG IDZ project is relatively lower, with a score of 10.93. This may indicate that national-level and centralised regulatory tools may have greater power in predicting the level of SDG disclosures. Consistent with Situ et al. (2018), Chinese

companies have formed the habit of following the latest national ideologies and making disclosures to show their support to the government. The variation in the predictive values for these proxies could be due to the enforcement and attitude of national and regional governments towards sustainable development, which may subsequently affect how companies respond to it. As discussed in Section 2.2.1 and Section 3.3.2.1, BRI is a national strategy promoted by the central government, while SDG IDZ projects are implemented at regional levels. Their expectations and incentives for companies to engage in it may be different, which may be due to the differences between national and regional governments. This offers some opportunities to further differentiate these strategies based on different levels of the Chinese governments to better understand the relationships between regulatory influence and the level of SDG disclosures.

Incentivising influence is proxied by whether the company issued green bonds; however, green bonds merely have weak power in predicting the level of SDG disclosures, with an RVI of 6.3. As discussed above, an RVI of 6.3 could be considered the random noise in this thesis. A lower predictive value may be due to the measurement of green bonds, as this thesis collected green bond data from Thomson Reuters, where limited data is available. Another reason for explaining this could be that the Chinese green bond market focus on general ESG issues, rather than SDG issues. Therefore, there remains scope for understanding whether different measurements of green bonds may affect SDG disclosures in the Chinese context, and whether similar results can be found in Western contexts.

Table 6.4 Relative Variable Importance Scores for government control predictors

	Variable	RVI Score	
Political influence	POLITICALCONNECTION_PARTY_COMMITTEE	59.15	
	POLITICALCONNECTION_DIRECTOR_COUNTS	54.58	
	POLITICALCONNECTION_CEO_EXECUTIVES_COUNTS	52.12	
	POLITICALCONNECTION_OVERALL_COUNTS	46.45	
	POLITICALCONNECTION_DIRECTOR_BINARY	0.00	
	POLITICALCONNECTION_CEO_EXECUTIVES_BINARY	0.00	
	POLITICALCONNECTION_OVERALL_BINARY	0.00	
The nature of companies	OTHER_TYPES_OF_COMPANIES	41.88	
	PUBLIC_INTEREST_SOES	21.60	
	COMMERCIAL_SOES	14.43	
Regulatory influence	BELT_AND_ROAD_INITIATIVE	34.94	
	SDG_INNOVATION_DEMONSTRATION_ZONE	10.93	
Incentivising influence	GREEN_BOND_ISSUANCE_BINARY	6.30	

Partial dependence plots for government control predictors

This section displays the directional contributions of government control predictors that have strong predictive power on the level of SDG disclosures. This thesis observed that mixed results have been found between political influence and the level of SDG disclosures, which may be due to different levels

of political connections. As discussed in Section 3.3.1.3, political connection individual embedding can be classified into CEO/executives, directors, and the senior management team, to capture the comprehensiveness of political connections that a company may own. As suggested by Qian and Chen (2021), different corporate positions of members that are politically connected may have different degrees of power in corporate decision-making and disclosures, a unique governance structure in China and its companies. Additionally, this thesis observed that the impacts of CEO/executive political connections on SDG disclosures are inconsistent with prior studies (Marquis & Qian, 2014). One of the reasons for explaining this could be the measurements of political connections and how we define political connection. Although this thesis considers all types of political connections, such as government officials, being a member of NPC or CPPCC, or having military backgrounds, this thesis does not further investigate the impact of the degree of the government that the person is connected or differences in different types of political connections. As suggested by Qian and Chen (2021), the enforcement and focus of different types of political connections may be different. This offers opportunities to explore the impacts of political connections at different levels on corporate SDG disclosures. Now, I will discuss them in detail.

Figure 6.7 displays a partial dependency plot for the number of politically connected CEOs and executives. The regression spline plot shows that politically connected CEOs and executives negatively predict the level of SDG disclosures. However, the predictive power of this variable flattens out, when the number of politically connected CEOs and executives continues to increase. That is, the presence rather than the number of politically connected CEOs and executives, negatively affects the level of corporate SDG disclosures. However, this is inconsistent with RVI results, where we found that the presence of politically connected CEOs and executives has zero RVI. RVI is a good global measure of a feature's contribution to model predictions, while partial dependence plots show localised, marginal effects of a feature on the target variable. A variable with zero or low RVI can still have a strong partial dependence indicating that the feature, while not globally important, may still have a significant local effect, often due to interaction effects with other features or other factors.

The negative prediction result is consistent with Li et al. (2021), which also suggests a negative relationship between top executives' political connections and the level of environmental disclosures, suggesting that disclosures may demonstrate their compliance with relevant policies to acquire government support and enjoy a relaxed regulatory environment. A greater level of disclosures may increase the risks of scrutiny of their environmental issues and companies may not be willing to disclose more or negative information. However, inconsistent results have been found in Marquis and Qian (2014) and Qian and Chen (2021), possibly due to the measurement of political connections and the research contexts. Marquis and Qian (2014) explore the impact of CEO CPPCC membership on the issuance of CSR disclosures and explore whether the CEO as a government official may affect the substantiveness of CSR disclosures. This thesis finds that the CEO's political connection positively

predicts the issuance of CSR reports rather than the substantiveness of the reports, as the CEO is willing to maintain legitimacy in the eyes of the government. Marquis and Qian (2014) define government official and CPPCC connections as different types of connections with different degrees of power in facilitating political tasks, while this thesis categorises political connections based on corporate positions. Another reason could be the regulation change. The research period for Marquis and Qian (2014) is between 2006 and 2009; however, China's Central Organization Department introduced Regulation 18 in 2013, regulating government officials to resign from businesses (Xinhua, 2013a). It is unclear whether regulatory changes may also affect their political motivations to engage in sustainable development activities and in making relevant disclosures to maintain legitimacy in the eyes of the government. Therefore, measuring CEO and executive political connections may affect their SDG disclosures.

Figure 6-7 Partial dependency plot for the number of politically connected CEO & executives

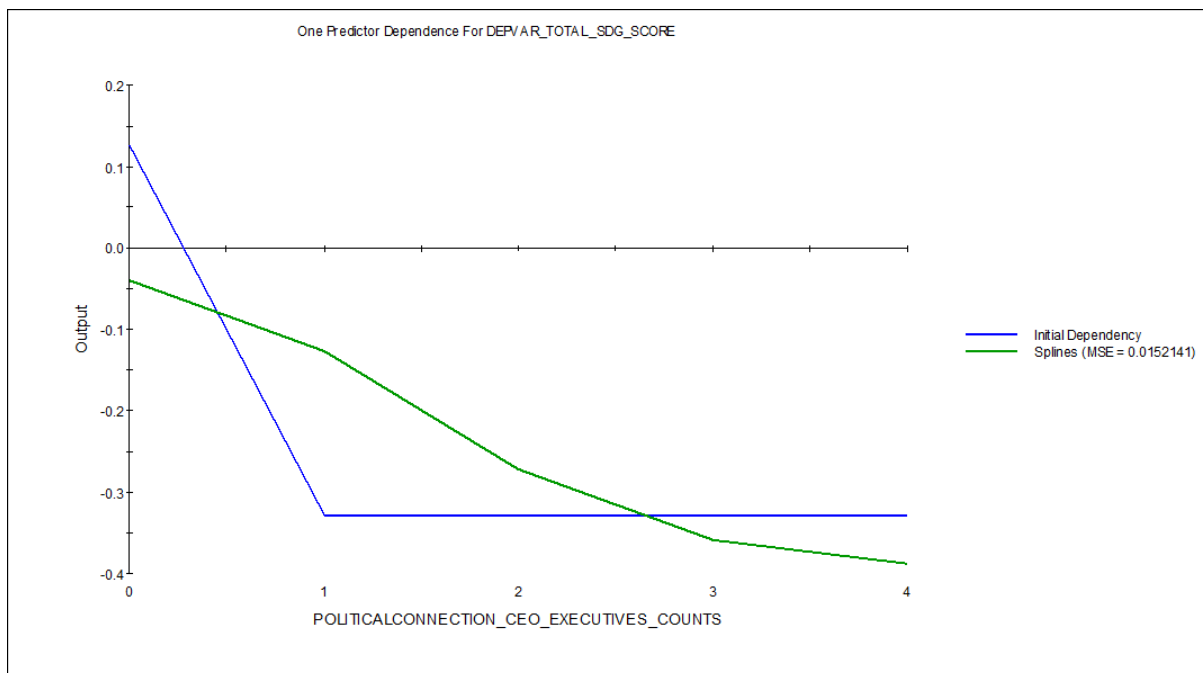


Figure 6.8 displays a partial dependency plot for the number of politically connected directors. The regression spline plot shows the positive marginal effects of this variable on the level of SDG disclosures. Specifically, this variable has strong positive marginal effects, when the number of politically connected directors is between 1 and 3. When the number of politically connected directors continues to increase, its power to predict the level of SDG disclosures flattens out. The result is broadly consistent with Ting and Lee (2024), and Dicko et al. (2020), which focus on the political connection of independent directors in Chinese companies and the board of directors in Canadian companies. As suggested by Ting and Lee (2024) and Qian and Chen (2021), stronger connections may indicate their stronger legitimising motives to fulfilling their political obligations and thereby having a higher level

of disclosures beyond merely complying with the strategies and regulations. At the same time, a greater level of disclosures may also demonstrate that the company does not over-invest in sustainable development activities to fulfil their political tasks (Ting & Lee, 2024). Therefore, the impact of political connections of directors on corporate disclosures appears consistent across different research contexts, and measurements of directors' political connections.

Figure 6-8 Partial dependency plot for the number of politically connected directors

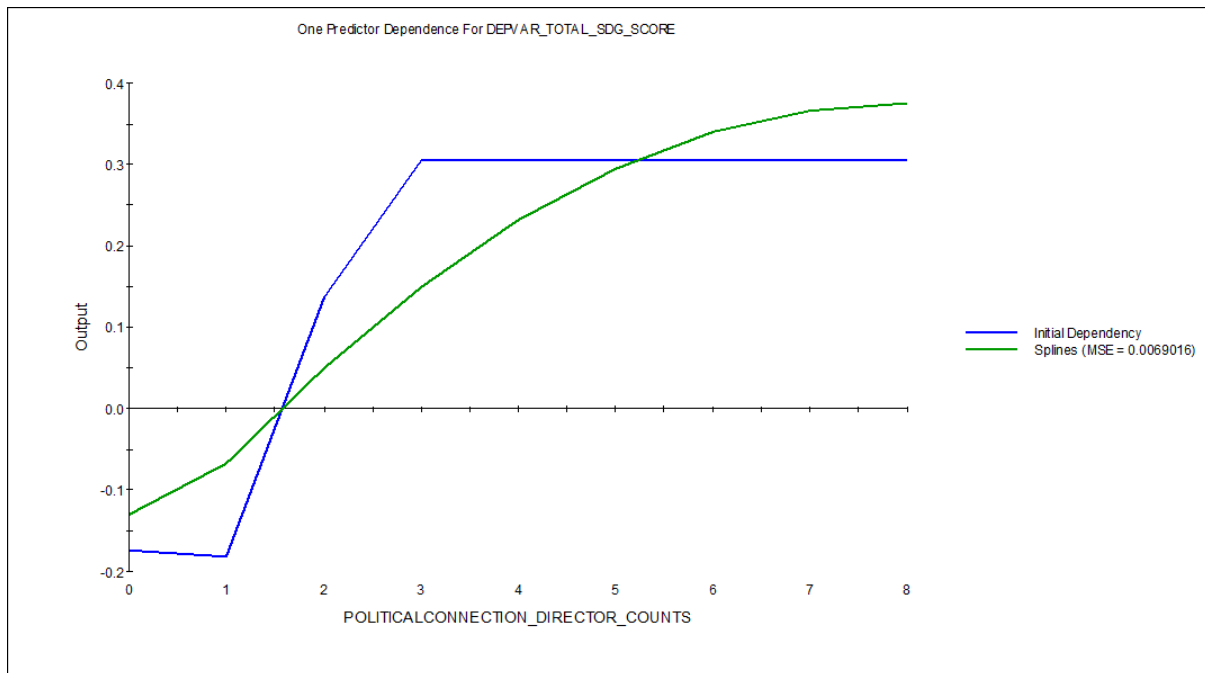
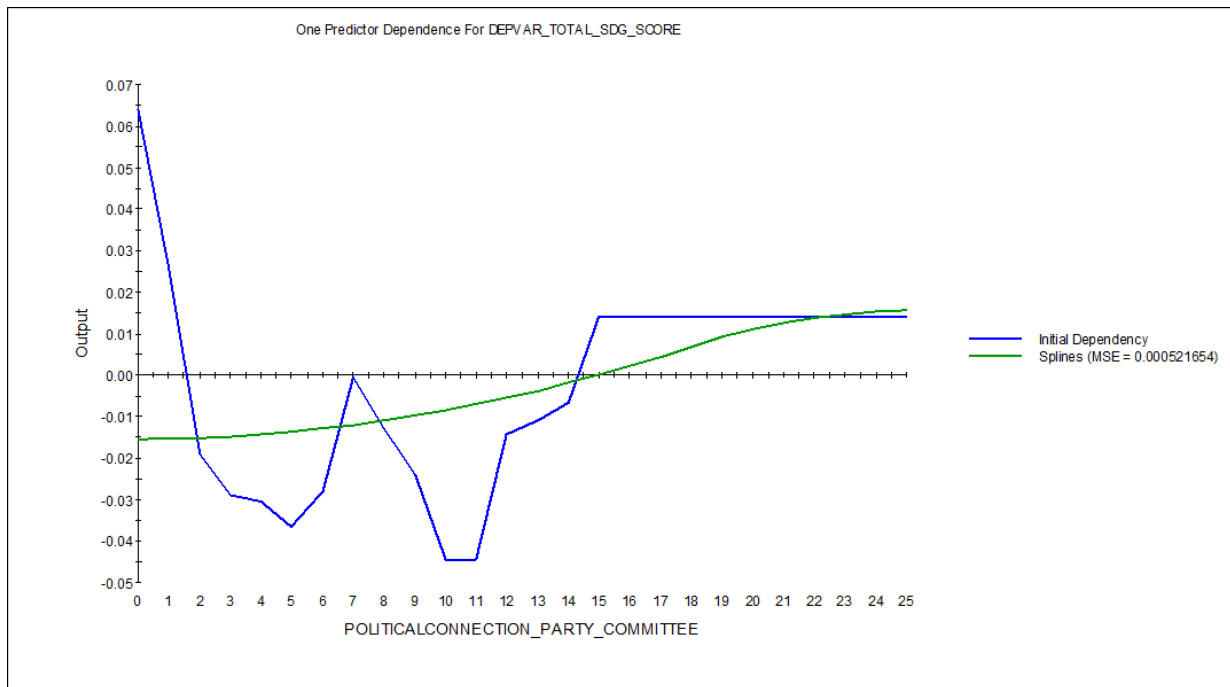


Figure 6.9 displays a partial dependency plot for the party committee. A company can only establish their party organisations when the government is satisfied with the company's performance. The first-order single-knot spline shows generally positive impacts of a party committee on SDG disclosures. This is consistent with Cheng et al. (2017), which also tests the number of party members on the executive board in order to measure the degree of political connections. Cheng et al. (2017) focus on individual-level party connections, rather than the party committee. However, the relationship in this thesis is non-linear, which may be due to the measurement of the party committee and the model selected. This provides some opportunities for future research to investigate whether the presence of a party committee or corporate positions of the people in the party committee in a company may offer different insights. The relationship seems to be negative when the party organisation is relatively small. This relationship turns positive when the size of the party committee is greater than 10 members.

Figure 6-9 Partial dependency plot for party committee



Financial performance

As discussed in Section 3.3.2.1, financial performance may indicate corporate financial capability in participating in sustainable development activities. This thesis includes a number of financial performance predictors, such as raw accounting numbers, accounting-based financial ratios, and market-based financial performance predictors (See Section 5.4.1).

The predictive power of financial performance predictors widely varies due to variable selection. Specifically, as a number of raw accounting numbers have strong predictive values, the value relevance of traditional accounting variables is not deteriorating, which is inconsistent with Lev (2018). This could be because the gradient boosting model is less sensitive to unscaled data and accommodates a greater number of predictors to provide more accurate out-of-sample results. Additionally, consistent with Sun et al. (2021), gradient boosting machine prediction with some traditional financial numbers outperforms the one with financial ratios and market-based financial predictors, regarding their predictive power. Now, I will discuss them in detail.

Raw accounting numbers. With the highest RVI of 100 and the lowest RVI of 21.92, all raw accounting numbers have predictive power on the level of SDG disclosures. Predictive values widely vary due to the variables selected. Specifically, with an RVI of 100, total debt is the strongest variable overall. According to Jones (2017), total debts are a widely used indicator for evaluating the effects of firm size. This may indicate that firm size effects have strong power in predicting the variation in the level of SDG disclosures. This thesis also includes total assets (RVI=39.78), total revenue (RVI=84.1) and

market capitalisation (RVI= 82.47) as indicators for evaluating firm size effects; however, their predictive values widely vary, indicating that variable selection will affect the overall prediction results. This is followed by cash flow from financing (RVI= 89.47), cash flow from investing (RVI=87.04), and cash flow from operations (RVI=85.24). Consistent with Marquis and Qian (2014), this thesis finds that cash flow predictors have strong predictive values on the level of SDG disclosures. This may indicate that corporate financial resources may be important for predicting disclosure variations. Besides these, short-term loans (RVI=78.93), cash and cash equivalents (RVI=78.7), net-intangible assets (RVI=76.16), and net accounts receivables (RVI=68.53) also have strong power in predicting the level of SDG disclosures. Some raw numbers also have relatively lower power in predicting the level of SDG disclosures. For example, EBIT appears to have the lowest predictive power, with a RVI of 25.92. Total assets are another widely used indicator for firm size and have reasonable power in predicting the level of SDG disclosures. This may raise the question of the useful proxies for measuring firm size.

Accounting-based financial ratios. With the highest RVI of 98.98 and the lowest RVI of 0, the predictive values of accounting-based financial ratios widely vary due to variable selection. It is observed that the ratio of net profit to total profit (RVI=98.98), and the ratio of fixed assets (RVI=96.91) are among the top 5 strongest variables, indicating that financial ratios have strong power in predicting the level of SDG disclosures. Some accounting ratios also have strong predictive power, among the top 20 strongest variables in the analysis: return on equity (RVI=86.94) and return on assets (RVI=83.40). These ratios are widely used to measure corporate profitability. This is inconsistent with Rosati and Faria (2019b) and Martínez-Ferrero and García-Meca (2020), which find weak or no impacts of these variables on the adoption of corporate SDG disclosures. Therefore, the measurement of SDG disclosures may affect the overall prediction results. However, it is interesting to note that all earnings per share indicators have no power to predict the level of SDG disclosures. As suggested by Gul et al. (2011), earnings per share has also been used to measure profitability. Given the varying power of profitability proxies such as net profit to total profit ratio, ROE and ROA and earnings per share, this may raise the question of what are the useful proxies for profitability and whether similar results may remain in other Chinese companies or other contexts²⁵.

²⁵ Some financial ratios have strong predictive power, with RVIs ranging from 82 to 70. These ratios include debt to long-term capital (RVI=82.2), ratio of cash assets (RVI=81.65), liabilities to tangible assets (RVI=81.61), ratio of shareholder equity to fixed assets (RVI=81.38), ratio of retained earnings to total assets (RVI=80.63), short-term loans (RVI=78.93), cash and cash equivalents (RVI=78.7), ratio of intangible assets (RVI=76.36), tangible assets ratio (RVI=76.18), net intangible assets (RVI=76.16), ratio of receivable assets (RVI=73.63), ratio of accounts receivable to income (RVI=72.95), long-term loans to total assets (RVI=72.77), ratio of total profit to EBIT (RVI=70.7), net cashflow from operating activities to total liabilities (RVI=69.9), and net accounts receivables (RVI=68.53).

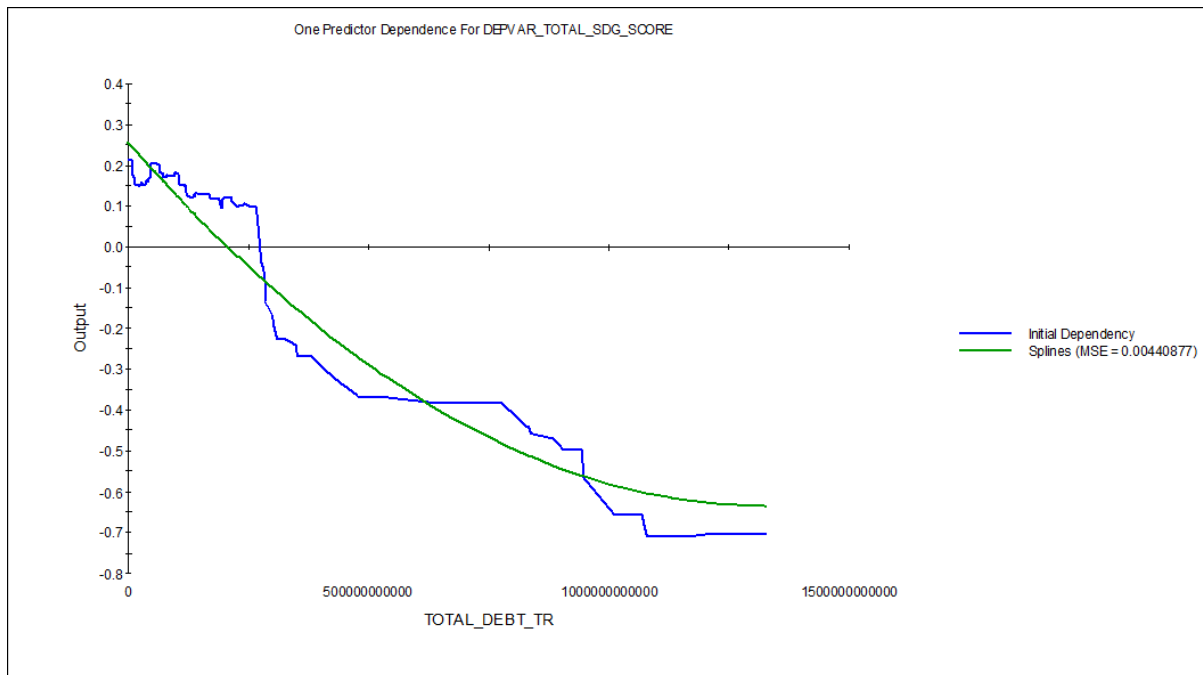
BOOK_TO_MARKET_RATIO	60.07	
RATIO_OF_CURRENT_LIABILITIES	59.71	
TOTAL_LONG_TERM_LIABILITIES	59.47	
BASIC_EARNINGS_PER_SHARE	58.65	
EARNINGS_PER_SHARE_1	58.14	
LIABILITIES_TO_EQUITY_MARKETCAP	57.69	
TOTAL_PROFIT	57.11	
COMBINED_ALPHA_MODEL_SECTOR_RANK	56.70	
CASH_RATIO_Y	56.65	
LONG_TERM_DEBT_TO_WORKING_CAPITAL	52.03	
EARNINGS_PER_SHARE	51.86	
EV_TO_EBITDA	51.77	
TOTAL_OPERATING_REVENUE	51.36	
LIABILITIES_TO_ASSETS_RATIO	51.12	
EPS_ESTIMATE_GAAP	51.05	
TOTAL_LIABILITIES	50.74	
TOTAL_LEVERAGE	50.13	
NET_CASHFLOW_FROM_OPERATING_ACTIVITIES_TO_INTERESTBEARING_LIABILITIES	50.00	
NET_CASHFLOW_FROM_OPERATING_ACTIVITIES_TO_CURRENT_LIABILITIES	49.89	
RATIO_OF_CURRENT_ASSETS	49.81	
TOTAL_EQUITY	49.68	
TOTAL_CURRENT_ASSETS	49.65	
TOTAL_RECEIVABLES	49.55	
NET_PROFIT	49.36	
EPS_NORMALISED_ACTUAL	48.64	
NET_INVENTORIES	47.12	
TOTAL_EQUITY_TO_PARENT_COMPANY	46.63	
POLITICALCONNECTION_OVERALL_COUNTS	46.45	
RATIO_OF_WORKING_CAPITAL_TO_NET_ASSETS	46.14	
TOTAL_CURRENT_LIABILITIES	45.90	
R_D_EXPENSES	45.34	
RATIO_OF_WORKING_CAPITAL	43.28	
EPS_NORMALISED_ESTIMATE	43.08	
RETURN_ON_LONG_TERM_CAPITAL	41.98	
EBITDA_TO_TOTAL_LIABILITIES	40.70	
RATIO_OF_NON_CURRENT_ASSETS	40.14	
TOTAL_ASSETS	39.78	
EARNINGS_PER_SHARE_2	39.63	
EARNINGS_PER_SHARE_4	39.58	
MARKET_VALUE	39.46	
TOTAL_LIABILITIES_TO_NET_TANGIBLE_ASSETS	39.43	
GROSS_PROFIT	39.20	
EBITDA_CSMAR	38.80	
WORKING_CAPITAL_TO_LOANS	38.73	
TOTAL_NON_CURRENT_LIABILITIES	38.05	
TOTAL_SHAREHOLDERS_EQUITY	36.90	
RATIO_OF_EBIT_TO_TOTAL_ASSETS	34.31	
EQUITY_MULTIPLIER	33.80	
GICS_CONSUMER_DISCRETIONARY	33.56	
EVALUATION_OF_ACCOUNTING_INFORMATION_QUALITY	33.35	
ACTUAL_EPS_GAAP	32.53	
DILUTED_EARNINGS_PER_SHARE	28.73	
EBIT_CSMAR	25.92	
EARNINGS_PER_SHARE_TTM1	0.00	
EARNINGS_PER_SHARE_TTM2	0.00	
EARNINGS_PER_SHARE_TTM3	0.00	
EARNINGS_PER_SHARE_TTM4	0.00	
RATIO_OF_OWNERS_EQUITY	0.00	
DEBT_TO_EQUITY	0.00	
EQUITY_TO_DEBT	0.00	
RATIO_OF_WORKING_CAPITAL_TO_CURRENT_ASSETS	0.00	
TOTAL_LIABILITIES_AND_SHAREHOLDER_EQUITY	0.00	

Partial dependence plots for financial performance predictors

The directions of the contribution of financial performance variables on SDG disclosures widely vary due to the variables selected. Traditional accounting numbers, such as cash flow from investing, cash flow from operating activities, and cash flow from financing still have strong power and have positive marginal effects on SDG disclosures, while total debts and financial ratios, such as ROA and ROE, broadly negatively influence corporate SDG disclosures. Now, I will discuss them in detail.

Figure 6.10 displays a partial dependency plot for total debt, which is the strongest predictor in this thesis, with an RVI of 100. The first-order single-knot spline shows that the marginal effect of total debt is slightly non-linear and negatively affects SDG disclosures, consistent with Singhania and Chadha (2023) and Andrikopoulos and Krikilani (2013). This may indicate that indebted companies may not be willing to disclose a higher level of SDG disclosures. Figure 6.10 shows that when a company's debt is lower than 250 billion Chinese Yuan, the impact of total debts on SDG disclosures is slightly negative. When total debts increase, a steep decline in SDG disclosures is observed. This indicates that the predictive power of total debts on disclosure levels becomes stronger, when total debts become higher. However, the predictive power of this variable changed, after it reached 500 billion. This suggests that an initial increase in total debts negatively predicts the level of SDG disclosures. Singhania and Chadha (2023) argue that increased debts may limit corporate capability to participate in SDG disclosures, as sustainability practice is costly. It is also worth noting that beyond this debt level, the decrease in SDG disclosures continues; however, it is less dramatic. This may indicate that once the debt reaches a certain level, the negative impact of total debts on SDG disclosures may become weaker. When a company's total debts are extremely high, its influence on SDG disclosures may flatten out. Overall, indebted companies are less likely to disclose their SDG information.

Figure 6-10 Partial dependency plot for total debts



Some traditional financial variables, such as cash flow from investing, operating activities, and financing, also have strong and broadly positive marginal effects on SDG disclosures, shown in the first-order single knot splines. This thesis extends Naheed et al. (2021) to explore the impact of different types of cash flow on corporate disclosures and broadly consistent results have been found. As suggested by Marquis and Qian (2014), cash flow proxies may indicate corporate financial resources. The positive result may suggest that disclosure may demonstrate that companies have sufficient financial resources to participate in SDG activities. While some studies find no relationships between cash flow from operating activities and disclosures, they focus on sustainability disclosures, suggesting that the heterogeneity in different types of observable sustainability disclosures may affect the direction of the prediction result (Dissanayake et al., 2016). The research setting may also be another possible reason for the inconsistent results. Dissanayake et al. (2016) are limited to a single-year analysis, which may cause lags in the prediction results. **Figure 6.11** displays a partial dependency plot for cash flow from financing. The relationship between cash flow from financing and SDG disclosures is non-linear. When cash flow from financing is between -10 million and 10 million, its impact on SDG disclosures rapidly increases. This may indicate that companies may improve their SDG disclosures when they have more cash flow from financing. When cash flow from financing exceeds the 10 million level, the relationship turns negative and finally flattens out. **Figure 6.12** displays a partial dependency plot for cash flow from operating activities. The marginal effect of this variable on SDG disclosures is non-linear. When cash flow from operating activities is negative, up to -1 million, there is little or no power in predicting the level of SDG disclosures. This is followed by a steep decrease in SDG disclosures.

This may suggest that companies with extremely negative or extremely low positive cash flow from operating activities tend to face constrained financial ability. This may subsequently affect corporate investment in SDG disclosures. However, this relationship turns positive when cash flow from operating activities exceeds 5000. The predictive power of this variable is even stronger when cash flow from operating activities is between 20 million and 25 million. Greater cash flow from operating activities may imply corporate financial ability to facilitate corporate operating activities, such as corporate SDG engagement (Putri, 2021). This suggests that companies with greater cash flow from operating activities tend to improve their SDG disclosures. When cash flow from operating activities exceeds a certain level, its marginal effects on SDG disclosures turn negative. However, the result is inconsistent with Ching et al. (2017), which find no relationship with the quality of sustainability disclosures. This thesis looks at all companies listed on Corporate Sustainability Index, which does not take into account regulatory and institutional circumstances and pressure in a country.

Figure 6.13 displays a partial dependency plot for cash flow from investing. Broadly positive marginal effects on SDG disclosures have been observed based on the first-order single-knot spline. Specifically, when negative cash flow from investing is extremely high, there is little or no power to predict SDG disclosures. When negative cash flow from investing is low, particularly between -30 million and 0, its impact on SDG disclosures turns positive. This indicates that companies with a lower negative cash flow from investing offer greater transparency of their SDG activities. One explanation for the negative prediction result could be showing corporate financial capacity to invest in SDG areas.

Figure 6-11 partial dependency plot for cash flow from financing

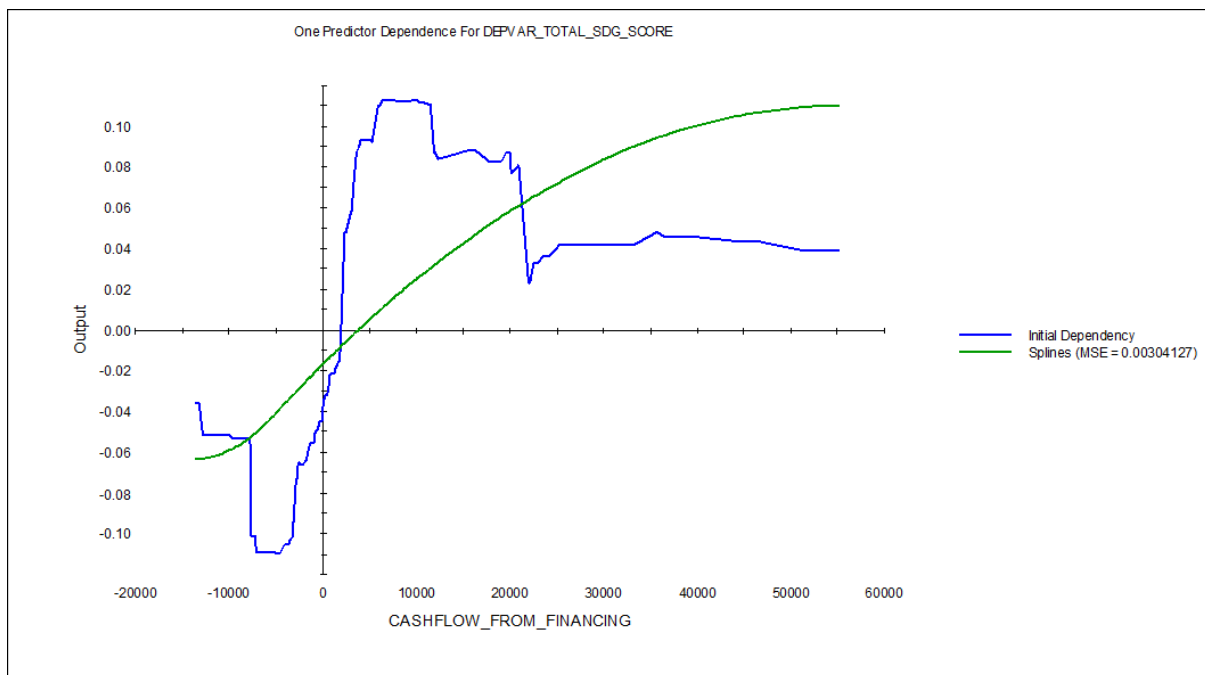


Figure 6-12 partial dependency plot for cash flow from operations

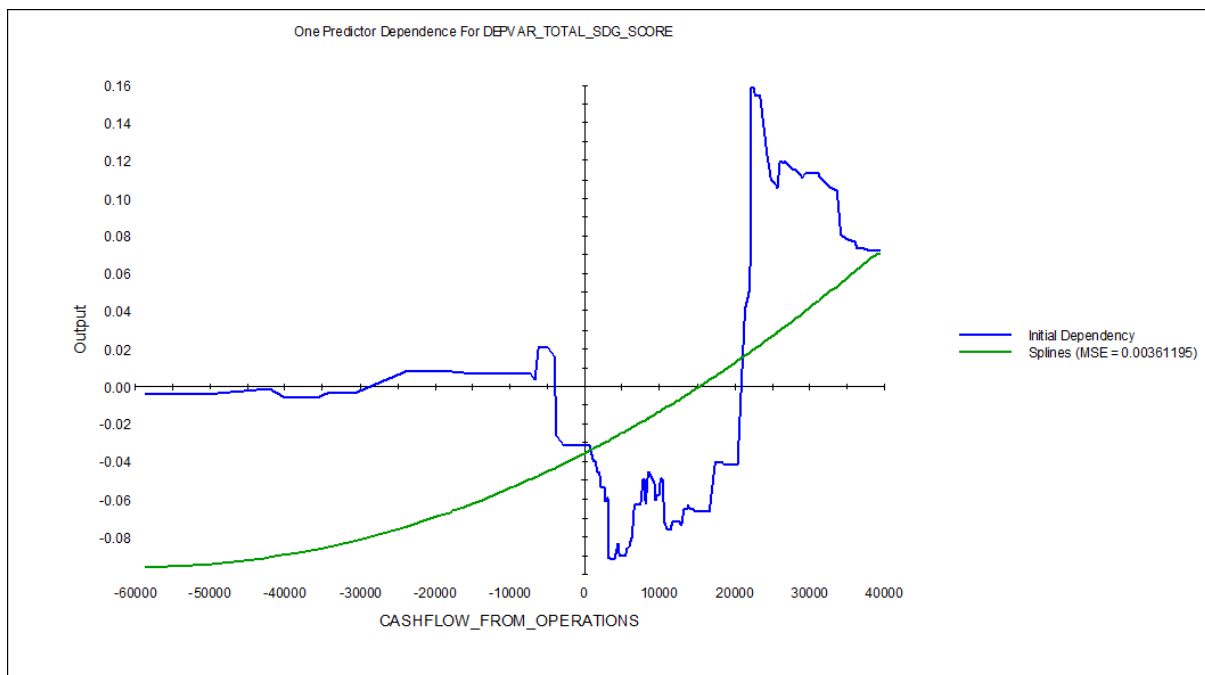
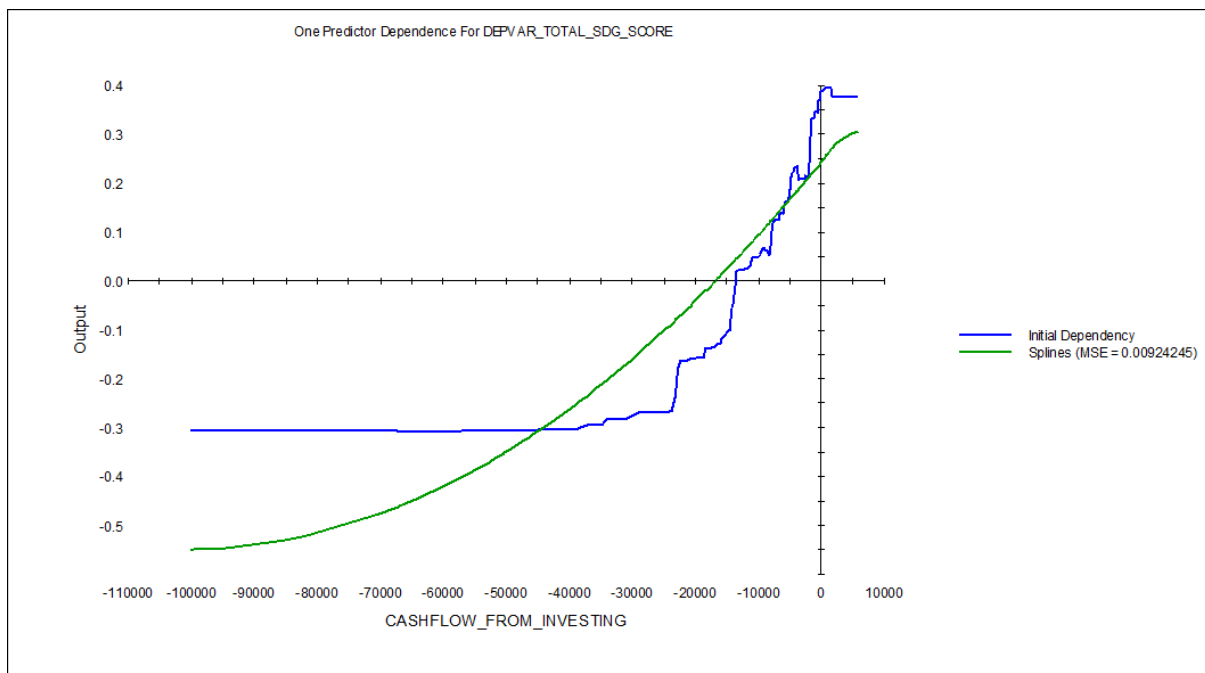


Figure 6-13 partial dependency plot for cash flow from investing activities



Figures 6.14 and 6.15 display partial dependency plots for ROE and ROA, which are widely used proxies for corporate profitability. The first-order single-knot splines show inconsistent results between profitability and the level of SDG disclosures. However, the impacts of these variables on SDG disclosures are broadly consistent and non-linear, when the value is below 0.1. These variables are observed to have negative marginal effects on SDG disclosures, when the values of ROA and ROE are lower than 0.1. However, the marginal effects of these variables turn positive, when ROA and ROE are

greater than 0.1. Companies with greater profitability have a greater capacity to undertake SDG activities in Chinese companies, consistent with Lu and Abeysekera (2014). Disclosure may serve as the basis for showing investors and external stakeholders’ corporate capacity in participating in SDGs (X. H. Meng et al., 2014; Rosati & Faria, 2019b). However, as discussed in Section 3.3.2.1, ROA and ROE do not appear to influence SDG disclosures in Rosati and Faria (2019b) and Martínez-Ferrero and García-Meca (2020). One of the reasons for explaining this is that the measurements of SDG disclosures are different. Rosati and Faria (2019b) and Martínez-Ferrero and García-Meca (2020) explored the adoption of SDG disclosures. Pizzi et al. (2021c) evaluate corporate compliance with the SDG Compass by evaluating the number of SDG indicators that the company has disclosed out of the total number of indicators developed by the SDG Compass. These studies focus on corporate compliance with SDGs or SDG Compass, which may limit the scope of disclosure analysis. The ‘quality’ of disclosures in these studies is limited, as these studies did not reflect actual corporate SDG performance, and the framework does not take into account specific contexts. Another reason for explaining the difference is the context. Prior studies widely investigate SDG disclosures in European companies, where SDG disclosures are mandated. The institutional background, governance structure, and SDG enforcement are different in China.

Figure 6-14 partial dependency plot for ROE

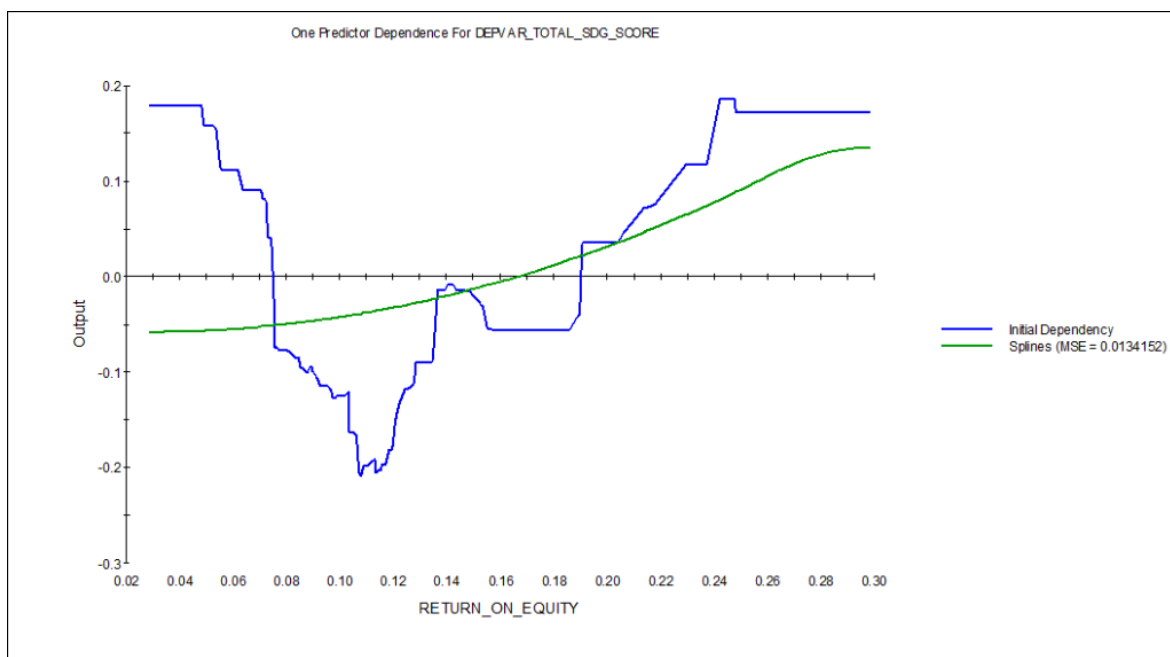
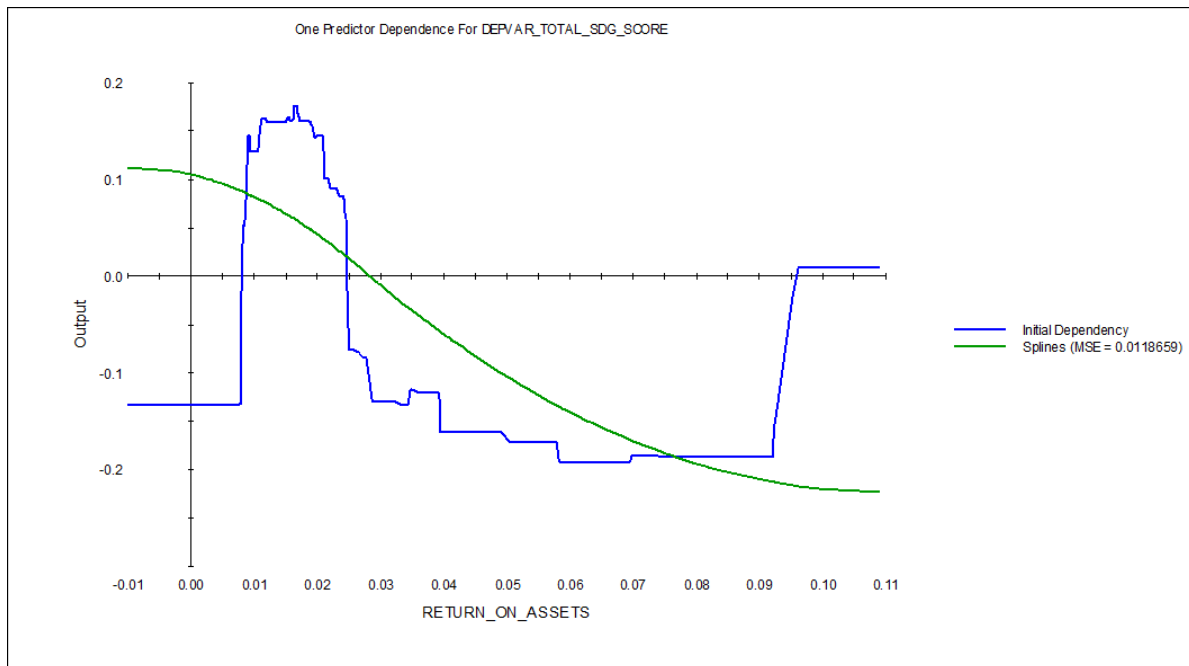


Figure 6-15 partial dependency plot for ROA



This thesis also tests whether different proxies for firm value may affect the level of SDG disclosures, with firm value proxied by Tobin's Q and enterprise value. This thesis suggests that similar non-linear patterns have been found for these two proxies, indicating that mixed results in prior studies may be due to the research contexts, different types of ESG disclosures, and empirical models employed. **Figure 6.16** displays a partial dependency plot for Tobin's Q. The first-order single knot spline shows that Tobin's Q has non-linear impacts on SDG disclosures. The regression Spline plot shows that a higher level of Tobin's Q negatively predicts corporate SDG disclosures. When the value of Tobin's Q is lower than 1, there is a sharp decrease in the level of SDG disclosures, consistent with Dhaliwal et al. (2011). Companies with a lower Tobin's Q may be perceived as underperforming and have limited financial flexibility, affecting their capability and willingness to participate in SDG disclosures. When Tobin's is between 1 and 2, the impacts on SDG disclosures turn positive. This may indicate that higher firm value may have power in predicting better corporate SDG disclosures. However, extremely high Tobin's Q may have limited or no power in predicting corporate SDG disclosures. As discussed in Section 3.3.1.1, prior studies have found mixed results between Tobin's Q and SDG disclosures, which may be due to different research contexts. Specifically, Girón et al. (2020) focused on developing countries, while García-Meca and Martínez-Ferrero (2021) sampled European companies. **Figure 6.17** displays a partial dependency plot for enterprise value. The regression spline plot suggests a positive marginal impact on SDG disclosures, which is consistent with García-Meca and Martínez-Ferrero (2021). This may indicate that companies with greater enterprise value tend to disclose more SDG information. Disclosure may send messages to stakeholders about corporate capability to engage in

SDGs. Specifically, this variable tends to have stronger predictive power when enterprise value is lower. This relationship flattens out, when enterprise value continues to increase. It is worth noting that different proxies for firm value, such as Tobin's Q and enterprise value, have similar patterns in predicting the level of SDG disclosures in Chinese companies.

Figure 6-16 partial dependency plot for Tobin's Q

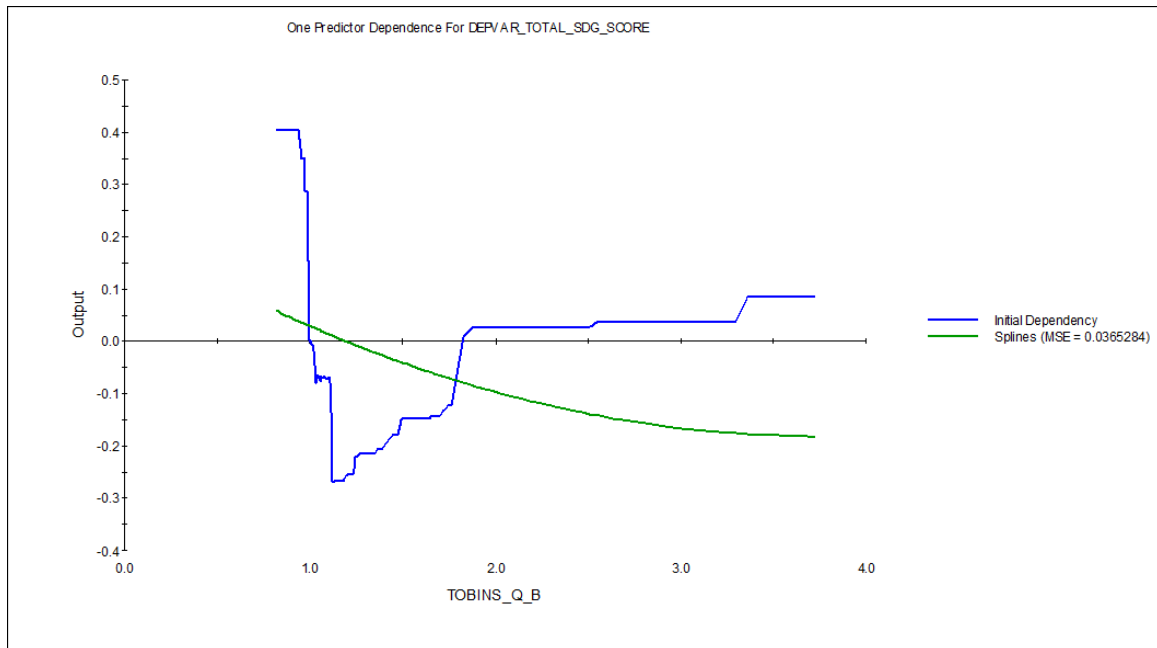
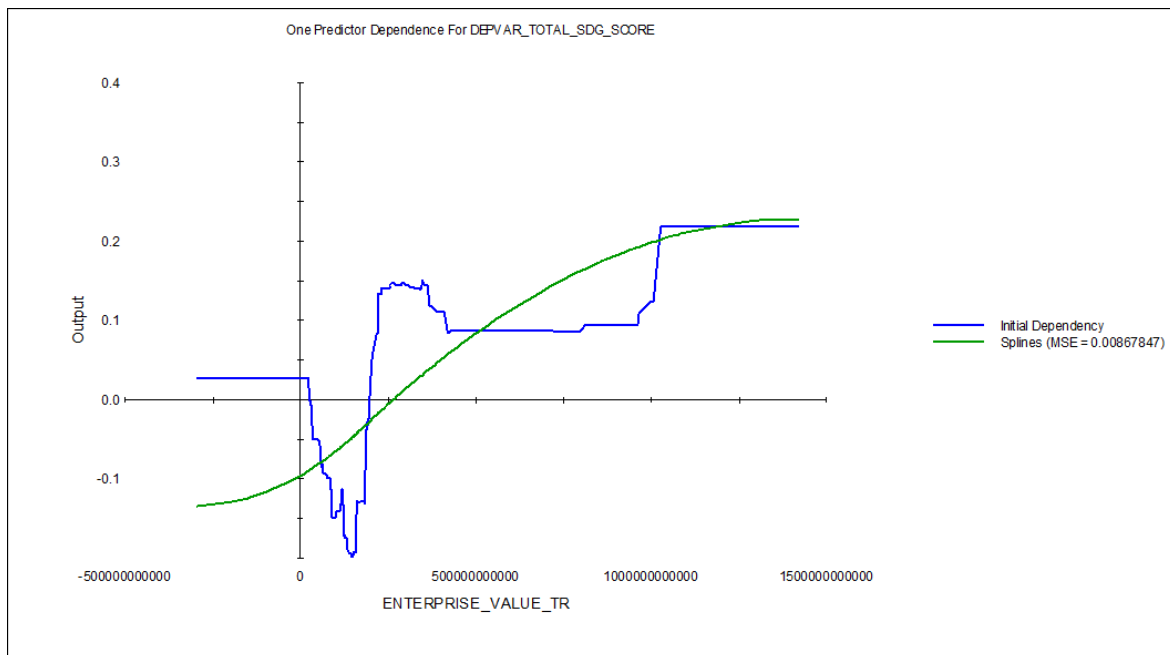


Figure 6-17 partial dependency plot for Enterprise value



Corporate governance

As discussed in Section 3.3.2.2, corporate governance is complex and broad, this thesis explores corporate governance in Chinese companies from ownership concentration and board characteristics aspects.

Table 6.10 displays that all corporate governance variables have non-zero RVIs, indicating that corporate governance has power to predict the level of SDG disclosures. With the highest RVI of (RVI=85.07) and the lowest RVI of 34.03, this thesis found that ownership concentration variables have stronger predictive values compared with board characteristics. This is consistent with the governance structure in Chinese companies—a single or multiple large shareholder governance structure (Jiang & Kim, 2018). Board characteristics still have strong power in predicting the level of SDG disclosures, reflecting that they are not just a ‘formality’. The predictive values widely vary across different variables. This thesis offers insights into the unique corporate governance model in China in understanding the variation in the level of SDG disclosures and offers insights into corporate governance in other countries. It also offers opportunities for Western countries to understand the corporate governance system in China, to cooperate with Chinese companies and ‘tie with China’.

Ownership concentration is measured based on the percentage of shares owned by the top 5 shareholders, shares owned by the top 10 shareholders, institutional ownership, and insider ownership. The percentage of shares owned by the top five shareholders is the strongest corporate governance predictor, with an RVI of 85.07. This is consistent with Jiang and Kim (2020) and Jiang and Kim (2024) that the concentrated governance structure has stronger power to predict corporate practices in China. Shares owned by the top 10 shareholders have a relatively lower RVI at 76.14. The percentage of shares owned by the top 10 shareholders appears to have a relatively lower predictive value, indicating that ownership dispersion may prevent larger shareholders from over-investing or under-investing in sustainable development activities and focus on joint values (Jiang & Kim, 2024). Some other predictors include institutional shareholding (RVI=84.27) and inside shareholding (RVI=80.58). It is also interesting to note that even if institutional investors in Chinese companies are minority stakeholders, they still have strong predictive power in predicting the level of SDG disclosures, which may be due to their rights in active trading (Jiang et al., 2018). Additionally, large shareholders in SOEs and non-SOEs may also be fundamentally different (Jiang & Kim, 2024). As suggested by Jiang and Kim (2024), larger shareholders in SOEs are the agents of the governments, rather than owners. They may act on their interests rather than pursuing the interests of nations. Therefore, this offers opportunities for future research to explore whether ownership concentration in SOEs and non-SOEs may be different.

Board composition predictors have strong predictive power in the analysis, including the number of non-independent directors (RVI=72.37), the number of board meetings (RVI=68.4), the number of

directors (RVI=63.12), the number of non-executive directors (RVI=59.72), the number of independent directors (RVI=48.45), and the number of executive directors (RVI=34.3). Overall, with the RVIs ranging from 72.34 to 34.30, the impacts of board characteristics on the level of SDG disclosures are consistently strong. These predictors suggest board characteristics have strong power in predicting the level of SDG disclosures, indicating that the board of directors may offer expertise and capability in allowing companies to engage in SDG disclosures (Buniamin et al., 2020). A relatively ‘lower’ predictive value for the number of independent directors could be explained by Jiang and Kim (2024), which suggest that independent directors are minority shareholders in Chinese companies, and they do not act as engaged owners.

Table 6.6 Relative Variable Importance Scores for corporate governance predictors

Variable	RVI Score	
TOP5_HOLDING_PERCENTAGE_CSMAR	85.07	
INSITUTIONAL_SHAREHOLDING_CSMAR	84.27	
INSIDER_SHAREHOLDING	80.58	
TOP10_HOLDING_PERCENTAGE_CSMAR	76.14	
NO_NON_INDEPENDENT_DIRECTORS	72.37	
THE_NUMBER_OF_BOARD_MEETING	68.40	
NO_DIRECTORS	63.12	
NO_INDEPENDENT_DIRECTORS	48.45	
NO_EXECUTIVE_DIRECTORS	34.03	

Partial dependence plots for corporate governance predictors

The directions of the contributions of corporate governance variables, such as institutional shareholders, and shares owned by top shareholders on SDG disclosures are broadly positive. However, prior studies find mixed results between ownership concentration and corporate disclosures, due to variable selection. For example, Li and Zhang (2010) find a positive relationship between shares owned by top shareholders and CSR practices in SOEs, and an inverse relationship for non-SOEs. One of the reasons for explaining this could be that China’s concentrated governance structure may be different in these two types of companies. In SOEs, large shareholders could be the government, while non-SOEs may need to face the diverse interests of multiple large stakeholders—short-term financial benefits and long-term sustainable development. However, this thesis is limited to broadly exploring the impact of shares owned by top shareholders. This offers opportunities for future research to investigate whether the marginal effects of ownership concentration on the level of SDG disclosures may be different in SOEs and non-SOEs. Now, I will discuss them in detail.

Figure 6.18 displays a partial dependency plot for institutional shareholding. The first-order single-knot spline shows that the concentration of institutional shareholding positively predicts the level of SDG disclosures. However, it is observed that when companies have more institutional shareholding, particularly over 20 billion, it does not further influence the variation in the level of SDG disclosures. The result is broadly consistent with Khelif et al. (2017), which suggests that institutional shareholders have financial and nonfinancial know-how and expect the company to provide more transparency about

their ESG practices. In developing countries, investor protection may be weaker, offering more motives for institutional shareholders to request greater SDG disclosures. However, this is inconsistent with Acar et al. (2021), which argue that institutional shareholders focus more on short-term financial benefits rather than long-term sustainable development. Acar et al. (2021) explore environmental disclosures from biodiversity impact reduction, climate change risks and opportunities, environmental investment initiatives and spill impact reduction dimensions, while this thesis looks at SDG disclosures. Therefore, the expectations of institutional shareholders may widely vary in terms of different types of sustainable development disclosure practices. Another potential explanation for this could be the research context. As suggested by Jiang and Kim (2020), Chinese companies have a concentrated ownership structure, either one controlling or multiple large shareholders. Institutional shareholders may be a minority group and therefore they may have stronger motives to request a greater level of disclosures for their information needs.

Figure 6-18 partial dependency plots for institutional shareholding

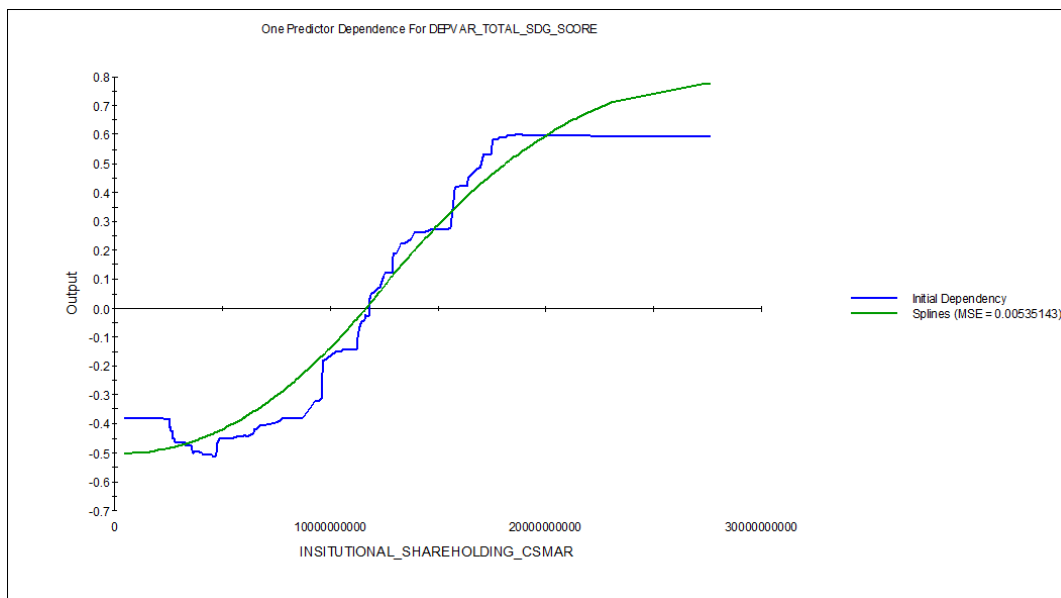
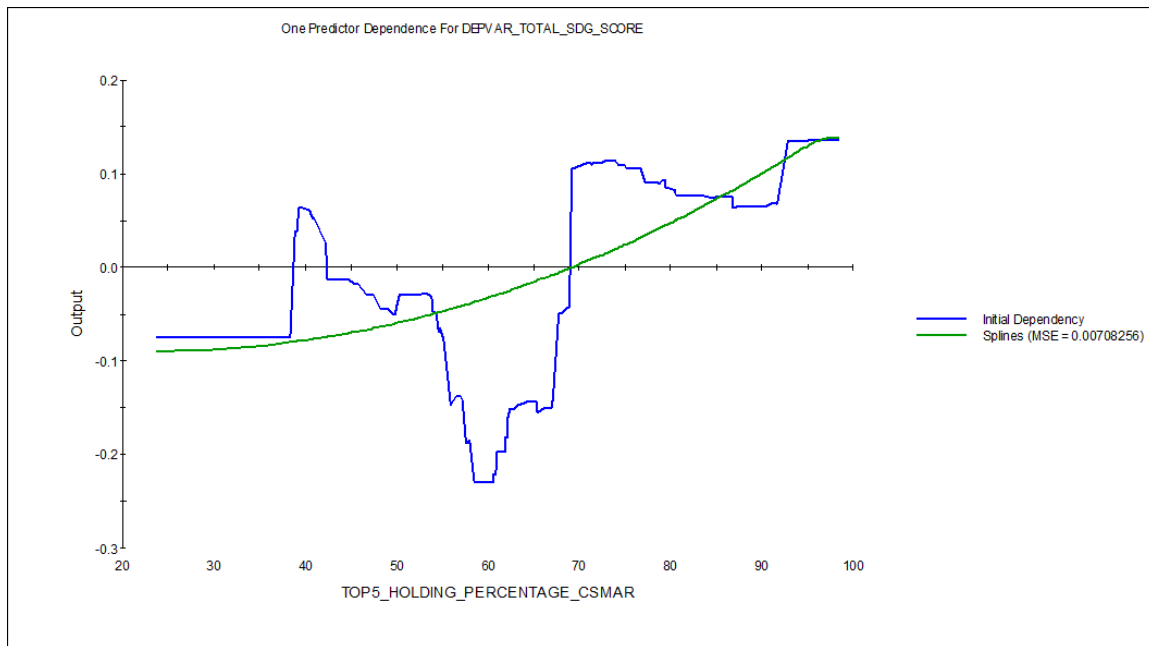


Figure 6.19 displays a partial dependency plot of shares owned by the top 5 shareholders. The regression spline plot smooths the prediction, indicating that companies with more shares owned by the top 5 shareholders tend to disclose more SDG information. However, when the top five shareholders hold less than 60% shares, companies with greater shares among the top five shareholders tend to have lower levels of corporate SDG disclosures. This relationship turns positive, when the shares owned by the top 5 shareholders are more than 60%. The non-linear result is inconsistent with Khlif et al. (2017), which argue a negative relationship between ownership concentration and voluntary disclosures. The inconsistent result may be due to the measurement of ownership concentration. This thesis focuses on shares owned by the top 5 shareholders, while Khlif et al. (2017) provide a meta-analysis and suggest a negative relationship between ownership concentration and voluntary disclosures. Therefore, variable selection may affect the prediction result.

Figure 6-19 partial dependency plots for shares owned by top 5 shareholders






ESG performance

As discussed in Section 3.3.2.3, corporate ESG performance may represent the increasing concern for stakeholders and regulators and is important to corporate financial performance. While prior studies widely explored the relationship between ESG performance and disclosures, ESG performance can be delineated into different ESG activities, such as environmental, social and governance activities, which may affect how companies respond to different ESG activities. Therefore, this thesis explored ESG performance from ESG risk scores, ESG management scores, environmental performance scores, social performance scores, and governance scores.

All sustainability performance variables have strong power in predicting the level of SDG disclosures, with RIVs ranging from 86.40 to 64.13. Consistent with Frost et al. (2023) and X. Meng et al. (2014), observable sustainability performance activities may have different power in predicting the level of SDG disclosure, which may be due to how companies respond differently to stakeholder pressure and regulatory demands. With RIVs of 86.40 and 83.84, ESG and governance scores have the strongest predictive values among all sustainability performance proxies. It is followed by ESG management score (RVI= 82.26), social scores (RVI=76.24), environmental scores (RVI=68.85), and ESG risk score (RVI=64.13).

Table 6.7 Relative Variable Importance Scores for ESG performance predictors

Variable	RVI Score	
SYNTAOGREENFINANCE_ESGSCORE	86.40	
SYNTAOGREENFINANCE_GSCORE	83.84	

SYNTAOGREENFINANCE_ESGMANAGEMENTSCORE	82.26	
SYNTAOGREENFINANCE_SSCORE	76.24	
SYNTAOGREENFINANCE_ESCORE	68.85	
SYNTAOGREENFINANCE_ESGRISKSCORING	64.13	

Partial dependence plots for ESG performance predictors

Figure 6.20 displays a partial dependency plot for ESG performance, measured by ESG scores from SynTao GreenFinance. This thesis observes a non-linear impact of ESG performance on SDG disclosures. When ESG scores are lower than 50, its marginal effects on SDG disclosures are negative, indicating that companies with lower ESG scores are more likely to disclose SDG information. This is consistent with X. Meng et al. (2014) and Li et al. (2017). However, when ESG performance scores are at a fair level, between 50 and 62, this thesis observed broadly positive impacts of ESG performance on SDG disclosures. They seem motivated to involve in SDG disclosures to show their capability in SDG activities, and the relationship between ESG performance and SDG disclosures flattens out, when corporate ESG performance scores are extremely high. This is inconsistent with Li et al. (2017). One of the reasons for explaining this is that Li et al. (2017) focus on environmental performance and disclosures in Chinese companies. Environmental disclosures have been mandated since 2008, and several corporate-level guidelines have also included environmental dimension. Unlike environmental disclosures, the Chinese government is still working on assessment mechanisms to guide corporate SDG disclosures. Social, regulatory, institutional and stakeholder pressure may affect how companies make relevant disclosures. Another explanation could be different types of observable activities, such as ESG performance and environmental performance, may affect corporate disclosures differently.

Figure 6-20 partial dependency plot of ESG performance

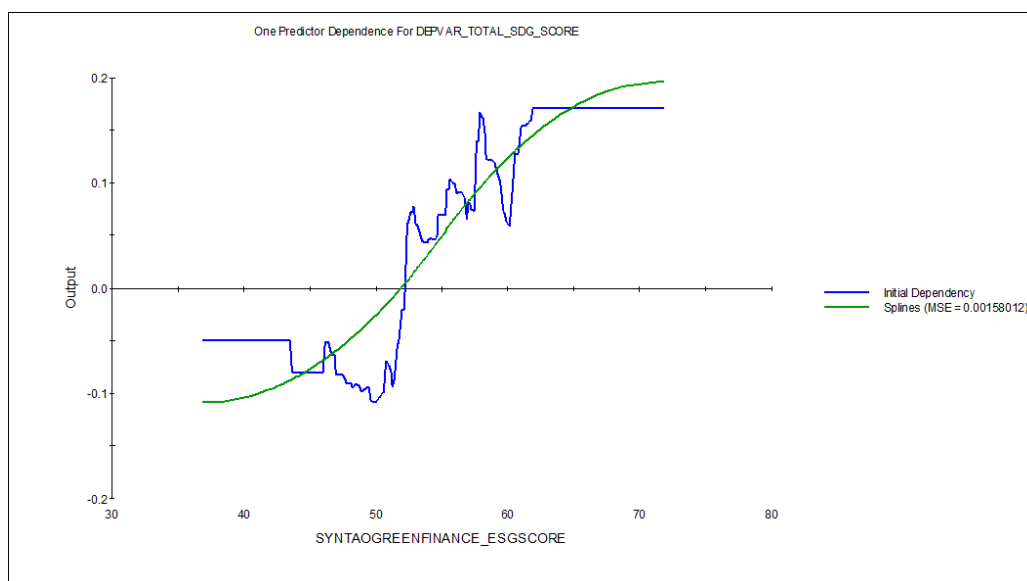


Figure 6.21 displays a partial dependency plot for the governance pillar of ESG performance²⁶. A broadly linear relationship between governance performance and SDG disclosures has been observed. The positive impact of governance performance on the level of SDG disclosures sharply increases, when the governance score is between 10 and 20. Disclosure may serve as a tool for companies to show their capacity to participate in SDG activities. Again, the predictive power of governance performance flattens out when the scores increase. This is inconsistent with X. Meng et al. (2014) and Li et al. (2017), indicating that different types of observable ESG activities may impact corporate disclosures differently.

Figure 6-21 partial dependency plot of governance score

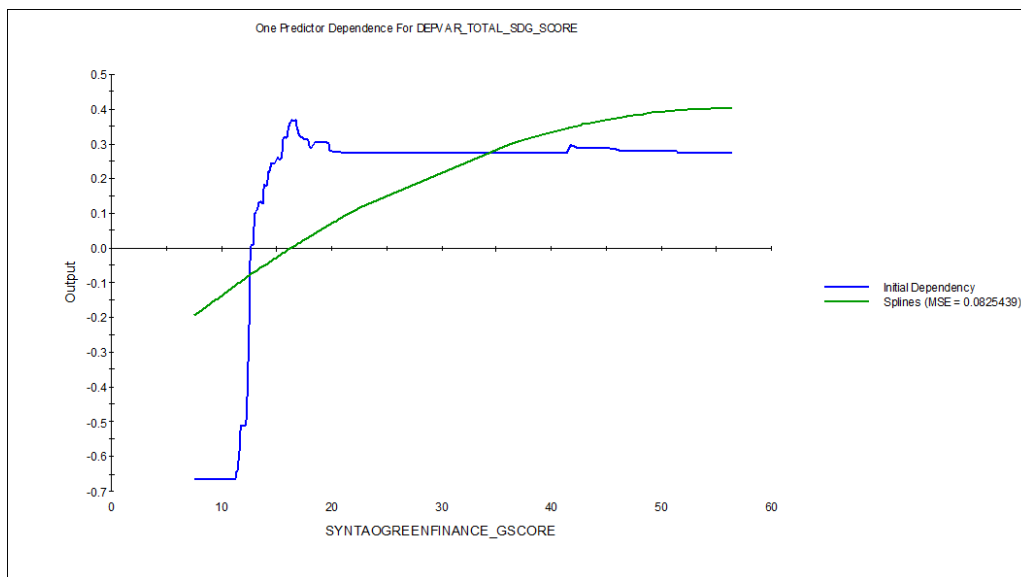
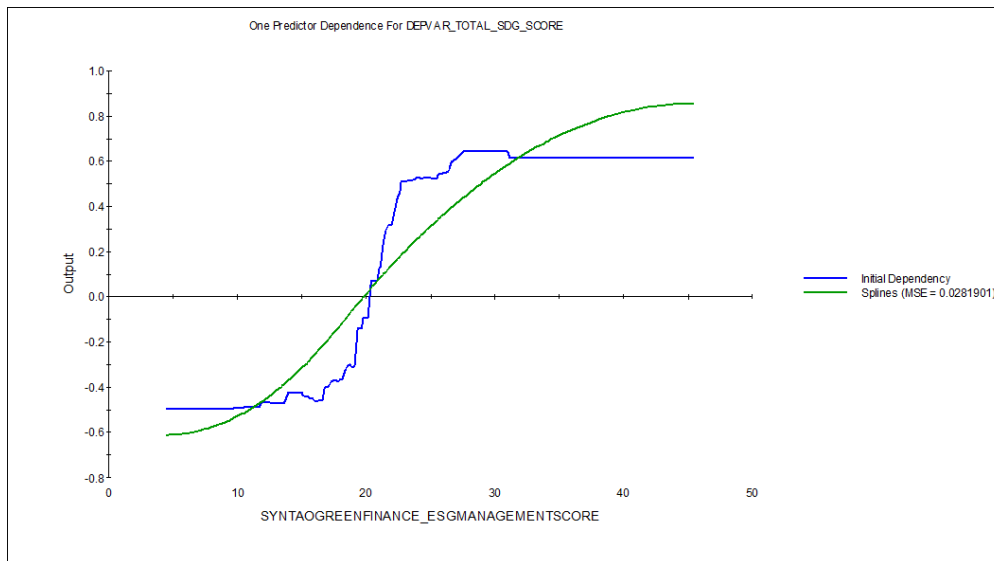


Figure 6.22 displays a partial dependency plot for ESG management scores. Consistent with ESG scores and ESG management risk scores, a positive marginal impact of ESG management scores on SDG disclosures has been found. The result is consistent with Hummel and Schlick (2016). Disclosure may indicate corporate better performance in ESG management risks. The predictive power of ESG management scores is stronger when scores are between 10 and 30. It is also observed that when the ESG performance scores are lower than 10 or higher than 30, the relationship between ESG management and SDG disclosures flattens out.

²⁶ ESG governance pillar is fundamentally different from corporate governance proxies in this thesis. ESG governance pillar focuses on business ethics (i.e. information disclosure, audit independence), governance structure (i.e. anti-corruption and bribery, whistleblowing systems, and tax transparency) and compliance (i.e. compliance management, risk management etc.) (<https://en.syntaogf.com/pages/methodology>).

Figure 6-22 partial dependency plot for ESG management scores



Control variables

As discussed in Section 5.3.3, this thesis employs several control variables in the gradient boosting machine, which influence the level of SDG disclosures in prior studies. Control variables include firm size, analyst recommendation, audit assurance, industry effects, COVID-19, and financial risks. Table 6.12 displays the predictive values widely across control variables, with RVI scores ranging from 90.91 to 0. This indicates that some control variables, such as firm size, audit assurance, industry effects, and financial risks can predict the level of SDG disclosures in the gradient boosting model. The lowest or zero RVIs may be due to the measurements of control variables. For example, financial risks have been proxied by Altman Z Score, credit ratings, and audit opinions, which are collected from multiple databases, such as Thomson Reuters, Capital IQ, and CSMAR. However, merely Altman Z Scores and credit ratings have non-zero RVIs. This may offer opportunities for future research to consider variable selection and proxy measurements. It also offers opportunities to explore whether the results may be generalisable to other Chinese companies and the Western contexts. Now, I will discuss them in detail.

Analyst recommendation is proxied by analyst recommendation scores, ranging from 0 to 5. As suggested by García-Sánchez et al. (2020), it acts as an ‘information intermediary in the market’ to provide useful information for investors’ decision-making. With an RVI of 90.91, analyst recommendation has strong power to predict the variation in the level of SDG disclosures. It indicates that Chinese companies have used SDG disclosures as a communication strategy to respond to analysts’ recommendations. It offers opportunities to investigate whether the results can be generalisable to other Chinese companies and other types of ESG disclosures, and whether SDG disclosures could be a new communication tool in Western countries to change the perception of analysts and recommendation results.

Firm size is proxied by market capitalisation. With an RVI of 83.04, it is suggested that firm size has strong power in predicting the level of SDG disclosures, broadly consistent with Frost et al. (2023). Firm size may indicate a company’s access to resources, financial capability, and exposure in the market.

Audit by Big 4 measures the impact of audit assurance on the level of SDG disclosures. With an RVI of 39.36, it has some power in predicting the variation in the level of SDG disclosures. Consistent with Beck et al. (2018) and Frost et al. (2023), the assurance of the quality of sustainability reports may influence the level of SDG disclosures.

Financial risks are proxied by Altman Z score, crediting rating indicators, and different types of audit opinions. With the predictive values varying from 41.81 to 0, the predictive power of financial risks widely varies due to variable selection. Altman Z score appears to be a useful predictor of the variation in SDG disclosures. This is followed by crediting rating_higher (RVI=24.25). However, audit opinions do not have power to predict the level of SDG disclosures, which may be due to its measurements. This thesis is limited to whether the company issued the opinions related to ‘going concern’, ‘unqualified opinion with emphasis of matter’, ‘standard unqualified opinion’, and ‘qualified opinion with emphasis of matter’. Limited data has been collected for audit opinions. Therefore, it offers insights into the proxy selection for financial risks and whether results may be different if we use a different method to measure audit opinion and if we increase the sample size.

Industry effects are measured based on GICS 11 classification criteria. With the highest RVI of 33.56 and lowest RVI of 0, this thesis observed that the impacts of industry effects on the level of SDG disclosures in the gradient boosting model widely vary. With a RVI of 33.56, GICS Consumer Discretionary appears to have reasonable power in predicting the level of SDG disclosures. Unlike prior studies suggest that environmentally sensitive industries tend to have stronger impacts on the level of SDG disclosures (Pizzi et al., 2021c; Rosati & Faria, 2019b), this thesis observed that consumer discretionary industry have impacts on the level of SDG disclosures. According to GICS, consumer discretionary sectors, automobiles, apparel, hotels, leisure, media and retailers, which are highly sensitive to consumer preferences. Other types of industry effects appear to have low or no predictive power in predicting the level of SDG disclosures. CSMAR database also has industry data based on the Chinese classification rule, and therefore, future studies could explore the impacts of industry effects on SDG disclosures based on a different classification method.

Table 6.8 Relative Variable Importance Scores for control predictors

ANALYST_RECOMMENDATION_CAPITALIQ	90.91	
FIRM_SIZE_MARKET_CAPITALISATION	82.47	
ALTMAN_ZSCORE	41.81	
AUDIT_BY_BIG4	39.36	
GICS_CONSUMER_DISCRETIONARY	33.56	
CREDITRATING_HIGHER_LT	24.25	
CREDITRATING_HIGHER_ST	18.53	
CDP_REPORTING_LOWRANKING	17.52	
GICS_ENERGY	16.43	

COVID_19	11.62	
CDP_REPORTING_BINARY	11.44	
GICS_INDUSTRIALS	9.79	
CREDITRATING_LOWER_ST	7.73	
GICS_INFORMATION_TECHNOLOGY	7.60	
GICS_CONSUMER_STAPLES	7.45	
GICS_MATERIALS	6.81	
TYPE_OF_AUDIT_OPINION_UNQUALIFIED_OPINION_WITH_EMPHASI S_OF_MATTER_PARAGRAPH	0.00	
TYPE_OF_AUDIT_OPINION_STANDARD_UNQUALIFIED_OPINION	0.00	
GICS_FINANCIALS	0.00	
GOING_CONCERN_AUDIT_OPINION_BINARY	0.00	
GICS_REAL_ESTATE	0.00	
CREDITRATING_LOWER_LT	0.00	
TYPE_OF_AUDIT_OPINION_QUALIFIED_OPINION_WITH_EMPHASIS_O F_MATTER_PARAGRAPH	0.00	
GICS_HEALTH_CARE	0.00	
TYPE_OF_AUDIT_OPINION_DISCLAIMER_OF_OPINION	0.00	
GICS_COMMUNICATION_SERVICE	0.00	
CDP_REPORTING_HIGHRANKING	0.00	

Partial dependence plots for control predictors

Figure 6.23 displays a partial dependency plot for analyst recommendation. The first-order single-knot spline shows that the impact of analyst recommendations on SDG disclosures is highly non-linear, in a concavity shape, which is inconsistent with prior studies. The non-linear prediction result can be explained by the gradient boosting model, which provides richer and non-linear interpretive insights into the predictive power of analyst recommendations on SDG disclosures, particularly when analysts offer strong ‘buy’ and strong ‘sell’ recommendations. As discussed in Section 5.3.3, the scores of analyst recommendations range from 0 to 5, with higher scores indicating a stronger ‘sell’ recommendation by the analyst. When the analyst provides a strong ‘buy’ or ‘sell’ recommendation, the line is completely flat, indicating that analysts’ strong ‘buy’ or ‘sell’ perceptions do not seem to have power to predict the level of SDG disclosures. However, when recommendation scores are between 1.5 and 1.8, ‘buy-side’ analyst recommendation tends to have negative marginal effects on SDG disclosures. When the value of analyst recommendation is between 2.8 and 3.2, the marginal effect of analyst recommendation on SDG disclosures turns positive, which is consistent with García-Sánchez et al. (2020). This may indicate that companies are more likely to disclose their SDG information when they receive a negative recommendation (sell recommendation). Disclosure may serve as a basis for showing corporate capability in participating in SDG activities. Providing incremental information may assist in reducing information asymmetries and thereby mitigate analysts’ negative perceptions (Hinze & Sump, 2019).

Figure 6-23 partial dependency plot for shares owned by analyst recommendation

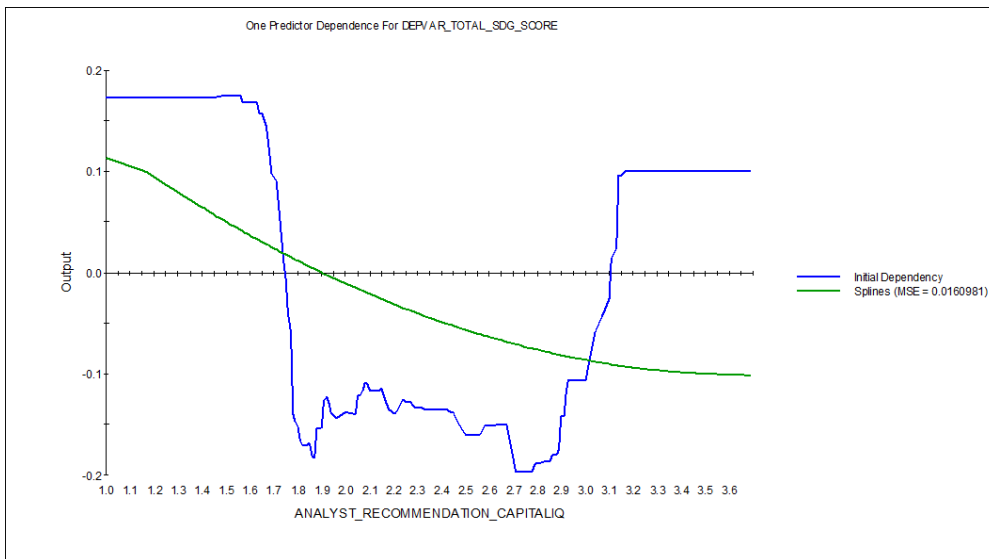
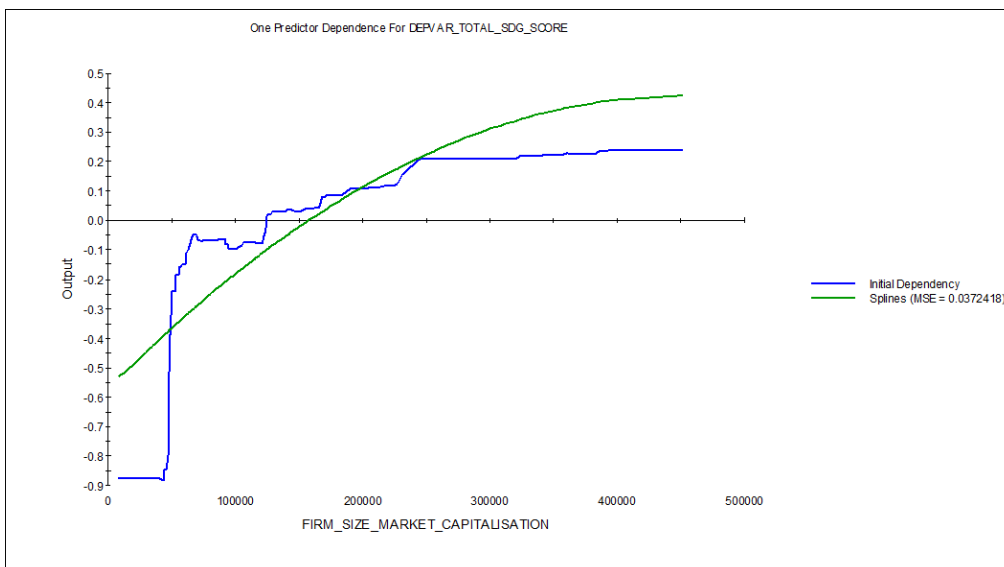


Figure 6.24 displays a partial dependency plot for firm size. The first-order single-knot spline shows that the marginal effect of this variable on SDG disclosures is positive. Consistent with Western studies, larger companies disclose more SDG information (García-Meca & Martínez-Ferrero, 2021). Disclosure can serve as a tool for companies to respond to public and stakeholder pressure to show their engagement in SDGs and report on SDGs. Specifically, the positive impact on SDG disclosures sharply increases, when firm size is lower than 10000. Its impact on disclosures continues to increase, when firm size continues to increase; however, the predictive power of this variable becomes weaker. When corporate market capitalisation is greater than 0.3 million, its impacts on SDG disclosures flattens out.

Figure 6-24 partial dependency plot for firm size



6.3.2.3 Interaction effects of variables

Interaction effects in gradient boosting models reveal how features work together to influence the target variable beyond their individual contributions. They improve prediction accuracy, enhance interpretability, and help detect non-linear dependencies, making them important for understanding complex relationships in data. It can be classified into whole-variable interaction effects and two-variable interactions. In the gradient boosting machine, variables can only interact with other variables within the same individual tree (Minitab, 2019). To observe interaction effects of predictors, this thesis provides whole variable interaction scores and 2-way (paired) interaction scores.

Whole-variable interaction scores

Appendix 9 shows whole-variable interaction scores, ranked by their overall interaction strength. These interactions show ‘the effects that has the strongest impacts’ on the overall prediction (Minitab, 2019). The higher the score, the greater the predictor's interaction effects on the overall prediction (Minitab, 2019). The values are in % scale. For example, it is observed that the ratio of fixed assets has the greatest impact on the model’s overall prediction, with a whole-variable interaction score of 5.02%, indicating that interactions between the ratio of fixed assets and other predictors explain 5.02% of the overall variation in the level of SDG disclosures. The predictor beta for 5 years has a whole-variable interaction score of 4.65%, indicating that interactions between beta for 5 years and other predictors explain 4.65% of the overall variation in the level of SDG disclosures. All other predictors have relatively lower whole-variable interaction scores in influencing the model’s overall prediction.

Interaction effects of strong pairs

Table 6.9 shows the strongest pairs of interacting predictors. The ‘score global’, on a percentage scale, shows the ‘contribution of an interacting pair after the normalisation’ to the total variability of the dependent variable (Minitab, 2019). For example, a score of 0.97% reflects that the interaction between ESG management scores and working capital contributes to 0.97% of the variation in the level of SDG disclosures. Similarly, the interaction between governance scores and working capital contributes to 0.96% of the variation in the level of SDG disclosures.

The ‘score local’ reflects the contribution of interaction pairs after normalisation ‘to the two-variable partial dependency function of the pair of variables under consideration’ (Minitab, 2019, p. 49). For example, a score of 16.55% means that 16.55% of the variation in the partial dependencies of working capital and ratio of shareholders’ equity to fixed assets can be explained by the interaction effects between these two predictors. With a score of 12.61%, this suggests that 12.61% of the variation in two-variable partial dependency between ratio of working capital and ratio of fixed assets is accounted by the interaction between these variables. Additionally, with a score of 12.23%, 12.23% of the variation in the two-variable partial dependency quick ratio and working capital is accounted by the interaction between these variables.

Table 6.9 shows the top 5 pairs that have the strongest interaction effects on the level of SDG disclosures, including ESG management scores and working capital, governance scores and working capital, working capital and ratio of shareholders' equity to fixed asset, quick ratio and working capital, as well as ratio of working capital and ratio of fixed assets.

Table 6.9 Top Variable interaction pairings

A. Top Variable Interaction Pairings by Score Global			
Score Global	Score Local	Predictor 1	Predictor 2
0.97	10.42	ESG management scores	Working capital
0.96	11.85	Governance scores	Working capital
0.83	16.55	Working capital	Ratio of shareholders equity to fixed assets
0.70	12.23	Quick ratio	Working capital
0.62	12.61	Ratio of working capital	Ratio of fixed assets
Pairs are sorted by Score Global and have either Score Global or Score Local greater than or equal to 10.000.			
B. Top Variable Interaction Pairings by Score Local			
Score Global	Score Local	Predictor 1	Predictor 2
0.83	16.55	Working capital	Ratio of shareholders equity to fixed assets
0.62	12.61	Ratio of working capital	Ratio of fixed assets
0.70	12.23	Quick ratio	Working capital
0.96	11.85	Governance scores	Working capital
0.97	10.42	ESG management scores	Working capital
Pairs are sorted by Score Local and have either Score Global or Score Local greater than or equal to 10.000.			

6.4 Discussion of the results

To answer what is the extent of SDG disclosures in Chinese companies, following China's adoption of Agenda 2030 (RQ1), this thesis developed a framework of 27 indicators, focusing on corporate explicit references to SDGs and implicit SDG performance related to SDG 1, SDG 7, and SDG 13.

This thesis observed the extent of SDG disclosures from *1)* the variation of SDG disclosures across the sampled period and *2)* the variation of SDG disclosures across different indicators/categories.

- 1) The variation of SDG disclosures across the sampled period.* As discussed in Section 6.2.1, boxplots show that an increasing number of Chinese companies voluntarily disclose SDG information, poverty alleviation, energy, and climate change information across the sampled period 2016-2020. In 2016 and 2017, the taller boxes indicate greater variability in disclosure scores across companies. This may indicate a disparity between better SDG disclosers and relatively lower SDG disclosers. The boxes become shorter in the rest of the sampled period, especially from 2018 onwards, with the lower quartile (Q1/25th percentile of the data) line rising

steadily across sampled years. That is, an improvement in SDG disclosures in the lowest-performing 25% of the sampled companies was observed. This may indicate that Chinese companies need time to be aware of SDGs and subsequently make SDG disclosures. Additionally, unlike SDG, poverty alleviation or energy disclosures, climate change disclosure levels appear to be relatively lower. One of the reasons for explaining this could be that the government was still at an early stage of working on the assessment mechanisms of climate change regulations during the sampled period; however, in 2024, a mandatory regulation was introduced in China to achieve the 3060 goal—carbon peak and neutral goals.

2) *The variation of SDG disclosures across different indicators/categories.* This thesis observed imbalanced disclosures in SDG disclosures across different indicators. While a number of observations widely acknowledged their poverty alleviation, energy and climate change activities in reports, imbalanced disclosures were also observed in implicit SDG performance categories. This may raise concerns about the quality of their disclosures—whether they are merely committed to SDG disclosures, rather than substantively making SDG disclosures. This may offer some insights into the limitations of existing corporate guidelines and requirements on corporate engagement and disclosures related to poverty alleviation, energy and climate change.

Chinese companies have offered different levels of assistance to the poor areas. A number of Chinese companies claimed that they have widely engaged in different types of poverty alleviation projects, such as industrial development, employment, healthcare, social impacts, and education. However, a limited number of Chinese companies have participated in relocating and improved ecological protection in the poor areas. This thesis observed more balanced disclosures in energy activities, while companies provided limited disclosures in climate change activities. One of the reasons for explaining this could be that while an increasing number of environmental disclosure regulations and guidelines have been introduced to standardise energy and carbon disclosures, limited climate change guidelines have been provided. This thesis observed that a number of companies widely disclosed total carbon emissions and acknowledged SDG 13 climate activities in reports, while less than 50 observations disclosed Scope 1 and Scope 2 information. However, the CDP questionnaire (2022) suggests companies widely disclosed their Scope 1 and Scope 2 information in their CDP reports. This may alert regulatory bodies about how they can improve their ESG disclosure guidelines and requirements to enable more balanced disclosures in Chinese companies.

To answer what factors may predict the level of SDG disclosures in Chinese companies (RQ2), this thesis analysed whether government control, financial performance, corporate governance, and ESG performance may predict the level of SDG disclosures in Chinese companies based on the gradient boosting machine. Out of the 159 predictors, 135 have non-zero power in predicting the level of SDG

disclosures, indicating that a number of government control, financial performance, corporate governance, and ESG performance variables can predict the level of SDG disclosures. The variation in the predictive values may be due to variable selection.

Government control. This thesis developed more novel methods to evaluate whether regulatory influence, political influence, government shareholding, and incentivising influence may predict the level of SDG disclosures.

With the highest RVI of 59.15 and lowest RVI of 0, the predictive power of government control predictors widely varies due to different types of government control tools and variable selection. A cluster of political influence predictors have the strongest power in predicting the level of SDG disclosures, such as party committees, the number of politically connected CEO/executives, the number of politically connected directors, and the number of politically connected senior management members, have stronger predictive values across all government control tools. This indicates that political influence is a more effective government control tool to influence corporate disclosure practices.

This is followed by the nature of SOE proxies, which are also direct government control tools. As discussed in Section 6.3.2.2, this thesis finds that the nature of SOEs differently predicts the level of SDG disclosures. This can be explained that public-interest SOEs may have stronger motivations to engage in long-term sustainable development activities, while commercial SOEs may focus on financial benefits. Future research could explore different measurements of the nature of SOEs, as existing guidelines for measuring the nature of SOEs have limited clarification of the boundaries for each type of SOE.

This thesis observed that direct government control tools have stronger predictive power than indirect government control tools, such as regulatory influence predictors and incentivising predictors. This may alert the government and regulatory bodies on the most effective tools to influence corporate disclosures.

It is interesting to note that all predictors measuring the presence of political connections do not have power to predict the level of SDG disclosures. One of the possible reasons for explaining this could be the measurement of political connections and the heterogeneity of disclosures. Marquis and Qian (2014) and Wang et al. (2018) also explored the impact of the presence of political connections on corporate CSR disclosures and significant results have been found. They look at CEO political connections at government officials and CPPCC levels and they are limited to the issuance of CSR disclosures. While this thesis has a broader definition of political connections—any of the following roles can be considered political connections: government official, CPPCC, NPC or military experience, this thesis does not consider that different types of political connection may have different power in predicting the level of SDG disclosures. This offers some opportunities for future research to develop more comprehensive measurements to evaluate political connections. Now, I will discuss them in detail.

Financial performance. This thesis improves the understanding of what financial performance factors may predict the level of SDG disclosures in the gradient boosting machine. Prior studies widely adopt financial ratios, and recent studies encourage the use of more forward-looking market-based indicators, as the value relevance of traditional accounting numbers is deteriorating (Lev, 2018). To overcome these challenges, this thesis takes into account raw accounting numbers that are widely used in Chinese studies, financial ratios and market-based indicators.

With the highest RVI of 100 and lowest RVI of 0, the predictive power of financial performance predictors widely varies due to variable selection. This thesis observed that traditional accounting numbers, such as total debts (RVI=100) and cash flow from financing (RVI=89.49), outperform some financial ratios and market-based ratios, indicating that the value relevance of raw accounting numbers is not deteriorating in predicting SDG disclosures, which is inconsistent with Lev (2018). Additionally, as observed in Section 6.3.2.2, all raw accounting numbers have non-zero predictive values.

Mixed results have been found between financial performance and the level of SDG disclosures, due to variable selection and model selection. As discussed in Section 3.3.2, this thesis observed that the marginal effect of total debt is slightly non-linear and negatively affects SDG disclosures. Consistent with Singhania and Chadha (2023) and Andrikopoulos and Kriklani (2013), this may indicate that indebted companies may not be willing to disclose a higher level of SDG disclosures, as increased debts may limit corporate capability in participating in SDG disclosures. However, broadly positive impacts have been found between cash flow indicators and the level of SDG disclosures. Consistent with Naheed et al. (2021) and Marquis and Qian (2014), companies with more financial resources tend to disclose more SDG information to show their superior financial performance in participating in sustainable development activities. Therefore, this thesis observed that a consistent direction has been found between the proxies for the same variable and the level of disclosures, while mixed results may be due to different types of financial performance variables selected.

Corporate governance. This thesis improves the understanding of what types of corporate governance structure may predict the level of SDG disclosures in Chinese companies in the gradient boosting machine.

With the highest RVI of 85.07 and lowest RVI of 34.03, the predictive power of corporate governance predictors widely varies due to variable selection. As discussed in Section 6.3.2, this thesis observed that all ownership concentration predictors, such as shares owned by the top 5 shareholders (RVI=85.07) and institutional ownership (RVI=84.27), have stronger predictive power in the level of SDG disclosures, compared with board characteristics. This is consistent with the concentrated governance structure, which appears to be more effective corporate governance tool in predicting the level of SDG disclosures. It is interesting to note that institutional shareholders also have strong power to predict the level of SDG disclosures. As suggested by Jiang and Kin (2018), institutional shareholders are the

minority group in Chinese companies. The result could be explained that they have rights in active trading to pursue their interests, rather than direct corporate engagement (Jiang et al., 2018). Additionally, this thesis also observed that board characteristics have a reasonable level of power in predicting the level of SDG disclosures. The predictive values widely vary due to their corporate positions, as they may have different power in corporate decision-making (Qian & Chen, 2021).

This thesis observed that institutional shareholding has positive impacts on the level of SDG disclosures, which is consistent with Khelif et al. (2017) and García-Sánchez et al. (2020). Institutional investors have financial and nonfinancial know-how and different information needs to evaluate corporate performance and make investment decisions, especially in developing countries, where capital markets are less mature and investor protection is weaker (Khelif et al., 2017). Disclosure may serve as a tool to meet their information needs. Overall, this thesis offers a better understanding of the corporate governance structure in Chinese companies and what works in China. This may also offer some implications for foreign investors who are willing to enter the Chinese capital market to see how they can tie with China and Chinese companies. This may also offer insights for other countries, regarding whether the Chinese governance structure may apply to other countries and other companies.

ESG performance. This thesis improves the understanding of different types of observable ESG performance on the level of SDG disclosures in the gradient boosting machine. Different types of observable ESG activities are measured from ESG performance, ESG management, ESG risks, environmental performance, social performance, and governance performance. With the highest RVI of 86.40 and lowest RVI of 64.13, all sustainability performance variables have strong power in predicting the level of SDG disclosures. Consistent with Frost et al. (2023), observable sustainability performance activities may have different power in predicting the level of SDG disclosures, which may be due to the way how companies respond differently to stakeholder pressure and regulatory demands.

ESG performance predictors, such as ESG performance, governance, and ESG management, have consistently positive impacts on the level of SDG disclosures. Disclosure may demonstrate corporate capacity in participating in SDG activities and their superior performance (X. Meng et al., 2014). However, the result is inconsistent with Li et al. (2017). One of the reasons for explaining this could be that Li et al. (2017) focus on environmental performance and disclosures in Chinese companies. Environmental disclosures have been mandated since 2008, and several corporate-level guidelines have also included environmental dimension. Unlike environmental disclosures, the Chinese government is still working on assessment mechanisms to guide corporate SDG disclosures. Different levels of social, regulatory, institutional and stakeholder pressure may affect how companies make relevant disclosures. Another reason could be that different types of ESG activities may predict the level of disclosures differently.

6.5 Chapter conclusion

This chapter answers RQ1 and RQ2, by analysing the variation of SDG disclosures during the period 2016- 2020, and factors that may predict the level of SDG disclosures in Chinese companies. Box plots visualise the variation of SDG disclosures from 2016 to 2020, including total SDG, poverty alleviation, energy, and climate change disclosures.

Employing the gradient boosting model, this thesis observed that financial performance, corporate governance, ESG performance, government control, and other influential variables have power in predicting the level of SDG disclosures. However, the predictive power of variables and the direction of their contributions on the level of SDG disclosures widely vary due to variable selection, different research contexts, and heterogeneity of sustainable development disclosure practices. Specifically:

- **Government control.** Among all government control variables, political connection variables appear to have strong power in predicting the level of SDG disclosures, with the highest RVI of 59.15. This may offer insights into the most direct and effective government tools to influence corporate disclosure practices. However, the direction of their impacts on SDG disclosures widely varies due to the variables selected. Therefore, this thesis can conclude that contextual focus in developing countries can offer new insights into the practice of understanding and forming their SDG disclosures, consistent with Qian et al. (2021).
- **Financial performance.** Top-ranked financial performance variables include total debts, beta for 5 years, yearly beta, fixed asset ratio, net profit ratio, cash flow from investing, cash flow from financing, cash flow from operating activities, total revenue, and Tobin's Q, with RVI ranging from 100 to 84. This suggests that traditional financial numbers, financial ratios, and other market variables that are widely adopted in finance areas all have power in predicting the level of SDG disclosures. It is also interesting to note that inconsistent with Lev (2018), traditional financial numbers are still variable for overall prediction. Consistent with Sun et al. (2021), gradient boosting machine prediction with traditional financial numbers outperforms the one with financial ratios regarding its interpretability and predictive power.
- **Corporate governance.** With the RVIs ranging from 85.07 to 34.30, corporate governance variables have strong power in predicting the level of SDG disclosures. Ownership concentration variables have stronger predictive power than board composition variables. As suggested by Jiang and Kim (2020), Chinese companies have a concentrated ownership structure, either one controlling or multiple large shareholders. Compared with concentrated shareholders, the board of directors are a formality rather than an effective governance body (Jiang et al. 2018). Partial dependency plots suggest that companies with higher levels of ownership concentration tend to disclose more SDG information.

- ESG performance. With the RVIs ranging from 86.4 to 64.13, ESG performance variables have strong power in predicting the level of SDG disclosures. These variables consistently suggest positive impacts on SDG disclosures, consistent with Hummel (2019).

Chapter 7 Conclusion

7.1 Chapter introduction

Section 7.2 starts with the contributions of this thesis. This is followed by the implications in Section 7.3. Section 7.4 discusses the limitations. Section 7.5 offers insights into future research. Section 7.6 concludes the chapter.

7.2 Contributions

This thesis contributes to several areas: 1. Overcoming methodological challenges; 2. The nature and extent of SDG disclosures in the Chinese context; 3. Identifying predictors that may predict the variation in the level of SDG disclosures; 4. Applying the gradient boosting machine in accounting research.

The *first* contribution of this thesis is overcoming the methodological issues in prior studies, from data capture for SDG disclosures and technology embracing aspects.

This thesis is a novel study that explores SDG disclosures in Chinese companies, following China's adoption of Agenda 2030. Prior studies are limited to existing international SDG disclosure guidelines or boilerplate disclosure frameworks without specificity. In contrast, this thesis developed a novel SDG disclosure framework, by constructing 27 indicators to proxy SDG disclosure quantity and quality in the Chinese context. The framework considers China's unique circumstances, focusing on SDG 1, SDG 7, and SDG 13. This may also offer some insights into what is happening in the real world after the introduction of SDGs as part of China's regulatory system. It is not surprising that Chinese companies need time to react to the voluntary regulations, gradually implement them and make disclosures. However, this may raise questions about enforcing voluntary regulations on Chinese companies' disclosures and the nature of disclosed information, such as boilerplate language or details.

This thesis also offers insights into technologies embraced in accounting studies. This thesis developed an automated content analysis tool to capture SDG data from ESG reports to overcome inter-coder reliability issues, such as their divergence of understanding, subjectiveness, and human errors. Prior studies focus on manual content analysis, existing ESG databases, or software. However, manual content analysis is often criticised for human coders' subjective understanding and data collection efficiency (Kasperuniene et al., 2020). Existing databases provide limited transparency and comparability across data vendors (Lin et al., 2024), while machine learning techniques have different focus, such as sentiment analysis or topic modelling (Huang et al., 2023). In contrast, this thesis measures SDG disclosures based on the automated content analysis tool. Specifically, this thesis develops a dictionary for data capture, as the automated tool is unable to process text as the human brain to classify them automatically. This raises the question of whether this tool can be applied to other data capture analyses.

The *second* contribution is the nature and extent of SDG disclosures in the Chinese context. Prior studies are limited to exploring the adoption of SDG disclosures and the frequency of SDG terms in reports. Lacking strategic information may undermine corporate SDG performance, which may not reflect corporate SDG efforts and their impacts. Some studies focus on score-based SDG disclosures, from commitments, prioritisation, strategic integration, and KPIs. However, these studies did not offer details for each goal or outline actions that the company undertakes. Instead, they focus more on whether the

company has stated that it has related KPIs, without mentioning the substance of its actions. This thesis extends existing studies by exploring the extent of SDG disclosures in Chinese companies, from explicit references to SDGs and implicit SDG performance aspects. This thesis offers insights into disclosure patterns based on different types of disclosures (See Section 6.3). Companies appear to have greater levels of poverty alleviation and energy disclosures, while limited data has been provided on climate change disclosures. As discussed in Section 5.3, Khalid et al. (2022) suggest that Chinese companies have adopted the CDP framework as their climate change disclosure guideline; however, a small number of Chinese companies responded to the CDP. Until 2024, the government introduced mandatory requirements for climate change disclosures in China. However, it is still unclear whether the variation of SDG disclosures may be due to the enforcement of different types of ESG regulations.

The *third* contribution is a multifaceted framework for testing government control. Specifically, this thesis develops a framework for a better understanding of government control in China. Prior studies are limited to focusing on one type of government control due to model selection and research design. Specifically, prior studies heavily focus on the impact of government shareholding on corporate ESG disclosures; however, they are often limited to classifying companies into SOE and non-SOEs. In contrast, going beyond government shareholding, this thesis takes into account regulatory, political, government shareholding and incentivising influence to see whether direct corporate-level or indirect national-level government controls may have stronger predictive power on corporate SDG disclosures. Specifically, this thesis extends prior studies to define political connections and measure political connections based on corporate positions of members that are connected. This thesis also extends existing literature by exploring the nature of SOEs, following China's government policy documents. Additionally, this thesis employs the automated content analysis tool, allowing me to collect data for different government control tools.

The *fourth* contribution is the factors that may predict the level of SDG disclosures in Chinese companies. Among all government control tools, this thesis finds that indirect control through connections, has the strongest predictive power. Additionally, this thesis also quantifies and tests a number of corporate-level predictors to identify relevant predictive variables that may predict the variation in the level of SDG disclosures. Prior studies are limited to a small number of variables and mixed results have been found due to variable selection and empirical models. This thesis extends existing studies by accommodating 159 predictors in one single model, to provide out-of-sample accurate prediction results. This thesis observed that variable selection may affect the power and directional contributions of SDG disclosures. Traditional accounting numbers still have the power to predict the level of SDG disclosures based on the gradient boosting machine.

The *fifth* contribution is the application of machine learning techniques in accounting research. This thesis extends existing studies about how the machine learning technique can overcome challenges in

traditional statistical analysis tools to articulate the predictive power of different variables on SDG disclosures. Prior studies widely employ statistical analysis tools, such as regression or logit models, to build an explanatory model, which can accommodate a limited number of variables and may be subject to p-hacking or multicollinearity problems. In contrast, the gradient boosting machine offers a high-dimensional analysis, which can accommodate 159 predictors and sample 60 Chinese public companies. It can provide more accurate out-of-sample prediction results. To some extent, the gradient boosting machine is often criticised for its interpretability challenges. However, as discussed in Section 5.2, this thesis relies on RVIs, partial dependence plots, and interaction scores, which improve the interpretability of the results in this thesis.

7.3 Implications

This thesis offers implications from several areas: regulators, researchers and potential investors.

Regulators: This thesis may alert regulators about the qualifications of government control tools for predicting the variation in SDG disclosures. This thesis found that political connections through party committees appear to be the most effective indirect government control tool to predict corporate behaviours. Direct tools, such as national guidelines and regional projects, appear to have relatively weaker power to predict the level of SDG disclosures. This might align with the argument by Situ et al. (2021) and Situ and Tilt (2018), suggesting that disclosures may be a ‘politically embedded phenomenon’. This may offer insights to regulators on how they may encourage companies to engage in national strategies and turn words into practices.

This thesis also offers insights to regulators on how Chinese companies interpret SDG 1, 7, and 13 and how companies responded to national priorities, following China’s adoption of Agenda 2030. However, the template for SDG disclosures is customised, which may not be easier for regulators to evaluate corporate SDG practices. Regulators may introduce additional regulations if they require more information. However, the prioritised goals in the Five-Year National Plans will change every five years, which leads to the question of what may happen after five years.

This thesis may also offer insights for regulators and foreign investors on what type of corporate governance structure works in China. Given that this thesis identifies that ownership concentration has stronger predictive values than board characteristics, this may improve foreign investors’ understanding of Chinese companies if they are willing to work with the Chinese government and Chinese companies.

Researchers. This thesis offers some implications to researchers and investors regarding factors that may predict the level of SDG disclosures in Chinese companies from contextual and corporate characteristics. Given that different types of observable ESG activities have strong and positive predictive power on the level of SDG disclosures. This suggests that disclosures may reflect corporate reactions to different types of ESG issues. This thesis identifies a number of financial and corporate governance variables that have strong predictive power on the variation in the level of SDG disclosures.

Mixed results have been found in Chinese companies. This may indicate that variable selection and proxy selection for each variable may affect the power and the direction of the impact on the level of SDG disclosures.

This thesis also provides methodological implications to researchers. Prior studies widely hypothesized the relationships between SDG disclosures and different factors; however, they may be subjective to a small number of variables. This could be due to their theoretical framework or model employed. The gradient boosting machine conducts a high-dimensional analysis about the role and the impacts of predictors and prediction directions. Given that the gradient boosting machine focuses on prediction based on a data-driven model, the reasons/explanations behind these findings are still unclear. This calls for further research on the explanatory power of these indicators in explaining the level of SDG disclosures and whether similar results can be found in other contexts.

This thesis may also offer some methodological implications regarding the gradient boosting machine. As discussed in Section 6.3.2.2, predictors with lower or zero RVI scores may be seen as random noise in the gradient boosting machine (Jones, 2017). This may offer some insights for researchers to re-run the model to see whether similar prediction results can be found. This may offer researchers opportunities to further compare the predictive accuracy across different types of machine learning methods.

7.4 Limitations

This section summarises the limitations of this thesis from methodological and machine learning aspects.

Methodological limitations

The methodological limitations will be summarized in several aspects, including the sample size, research period, reporting framework, measurements of indicators, gradient boosting machine, and the automated content analysis tool several aspects.

Sample size. This thesis is limited to 60 Chinese companies listed on the SSE and SZSE Stock Exchanges, which may potentially raise concerns about sample bias and generalisability. This thesis selected top publicly listed companies, which may not fully represent the actions of the broader community. For example, the prediction results may not be generalisable to small and medium companies in China or other Western companies. Additionally, this thesis is limited to selecting companies that have English ESG reports. This may raise the concern that SDG information by companies that only have ESG reports in Chinese may be overlooked, leading to language-based sample bias. However, China is at an early stage of recognising and interpreting SDGs between 2016 and 2020. Chinese companies need time to incorporate SDGs into business strategies and operations and make SDG disclosures. As it is observed that limited information has been disclosed by Chinese companies, we do not expect more SDG disclosures from other companies, particularly small and medium

companies. Therefore, this thesis expects larger companies to be more motivated to participate in and disclose SDGs (See Section 4.2). Given that Zhang and Ye (2022) suggest that English reports for Chinese companies can be acquired from the Hong Kong Stock Exchange, future research could increase the number of observations to conduct a more comprehensive analysis.

Research period. This thesis is limited to the research period between 2016 and 2020, when the 13th National FYP was effective. It does not take into account the variation in SDG disclosures in recent years, where the Chinese government is still actively promoting SDGs. Since 2016, the Chinese government and regulatory bodies continued to promote SDGs and Agenda 2030. Specifically, during the period when the 13th National FYP was effective, the national strategies and regulatory bodies laid the foundation for various parties to be aware of and get involved in SDGs and the 14th National FYP continues to promote SDGs, by focusing on 'high-quality' development. These national plans appear to put forward the emphasis on SDGs and laid the foundation for SDG disclosure analysis in China. During the period between 2016 and 2020, data availability is limited due to the initial stage of the adoption of SDGs in China. It is challenging for companies, as they were just beginning to recognise SDGs and to involve in relevant disclosure guidelines. Extending the research time period to include the 14th National FYP may allow a more comprehensive analysis and improve the availability and quality of SDG disclosure data. However, as discussed in Section 4.4, the Chinese government introduced the 14th National FYP in 2021. While SDGs are still prioritised in the latest national plan, the focus of the 14th FYP has changed to 'Building a Modern Socialist Country', which would add complexity and noise to this thesis. Given that this thesis focuses on the starting period of the adoption of SDGs, future research could extend the research period to explore how corporate SDG disclosures may shift in response to regulatory changes and to explore the potential influence of such regulatory changes on corporate SDG disclosures.

Data source. This thesis is limited to ESG reports. As suggested by Lin et al. (2024), annual and ESG reports appear to complement each other rather than substituting one another. This thesis also conducted data collection in annual reports during the preliminary analysis; however, annual reports provide limited information related to SDG 1, 7, and 13. Given that ESG reports appear to be focused reports of corporate SDG disclosures, offer standardised data and enable cross-sectional analyses, this thesis follows Dhaliwal et al. (2011) to collect data from ESG reports. Future research could use other methods and data sources to capture SDG disclosure data.

Reporting framework (The scope of SDGs and the number of categories). Different jurisdictions may have alternative means of implementing SDGs and attitudes to promote SDGs. This thesis is limited to SDG 1, 7, and 13, as the Chinese government used SDGs to operationalise the 13th National FYP and these three goals are prioritized in Chinese companies, which provides a unique setting in exploring SDGs. However, the authoritative setting and the imprinting process of political ideologies in Chinese

companies may not apply to other jurisdictions. It is unclear whether the reporting framework can apply to other contexts.

This thesis develops an SDG reporting framework, comprising 27 indicators for SDG 1, 7 and 13. However, the number of indicators for each goal is not balanced. This may increase the risk of data skewness and developing a continuous variable with a very limited range of values. However, this thesis developed these indicators based on multiple guidelines, such as the Chinese government policy document, the GRI indicators, prior studies, and the official SDG guideline. As discussed in Section 4.7, despite the number of indicators for observing poverty alleviation disclosures being the highest among all dimensions, it only accounts for 44% of all indicators. The gradient boosting machine is less sensitive to data skewness, as it adopts a decision tree as base learners. Additionally, even if the reporting framework has a limited range of values, it is still informative to observe SDG disclosure levels. Given that this thesis adopts the gradient boosting machine, the K-fold cross-validation method enables multiple-round iterations to effectively use each data point and subsequently obtain accurate prediction results. This will mitigate this limitation.

Automated content analysis. This thesis acknowledges some limitations of the automated content analysis in Section 4.6. For example, the semantic meaning of some keywords may be interpreted differently. Given that this thesis adopts a dictionary-based content analysis, it is unable to tell whether the paragraph collected can be the evidence in terms of its surrounding context. To address this issue, this thesis splits the whole report into several documents based on the table of contents, as each ‘chapter’ may have its topic or focus. This thesis also employed a two-stage data collection method to improve the accuracy of data collection. Some space may be removed when extracting paragraphs from PDF files, leading to several words combining into one. To address this issue, this thesis conducts an additional analysis, by replacing `rf\b{keyword}(?![a-zA-Z])` with `rf {keyword}` to capture all relevant words. Irrelevant data will be collected, as some words may be part of another word. However, the preliminary analysis of 50 observations suggests that these issues are individual cases that will not significantly affect the overall results.

Machine learning limitations:

Given that this thesis employs the gradient boosting machine, the model identifies predictors that may predict the level of SDG disclosures. However, machine learning is a data-driven model; it focuses on the predictive power of the model rather than the explanations behind it. Therefore, there remains scope for exploring the best explanatory model for SDG disclosures in Chinese companies and drawing causal inferences. We are still unclear about how we can explain these indicators in the variation of SDG disclosures and their explanatory power in influencing the variation in the level of SDG disclosures.

This thesis employs the gradient boosting machine, known for its accurate out-of-sample analysis. A common criticism of the advanced machine learning techniques is their interpretability, due to their black box effects. The gradient boosting machine involves a number of non-linear associations and complex interaction effects between predictors that lack transparency. This thesis employs RVI and partial dependence plots enabling some-level interpretation of results. Another limitation could be its slow training speed and high memory requirements; however, the commercial TreeNet® significantly improves the speed. Additionally, as suggested by Jones (2017), the gradient boosting machine may introduce some random noise in the results, such as predictors with lower and zero RVIs. Future research could re-run the model without these low-power predictors to see the overall prediction accuracy.

7.5 Future research

Several suggestions are offered for future research:

Future research could conduct the discourse analysis to explore what the company claims to have done in ESG reports. While this thesis developed 27 indicators to evaluate corporate SDG disclosures, it is limited to identifying whether at least one piece of evidence can be found for each indicator for a particular company. SDG disclosure data was turned into a continuous variable with a limited range of values. However, this thesis does not analyse the semantic information of SDG disclosures. Specifically, it is unclear whether the company claims that they are committed to SDGs or provides evidence that they actually implement SDGs. This offers opportunities for further research to explore whether SDG disclosures in Chinese companies are symbolic, greenwashing, or substantive and what the company claims it has done in reports.

Future research could use statistical inference analyses to develop an explanatory model for identifying factors that may influence the level of SDG disclosures. While this thesis finds that 139 predictors have non-zero power to predict the level of SDG disclosures, there remains a gap in identifying the useful indicators for explaining the level of SDG disclosures and the best explanatory model for the variation in the level of SDG disclosures.

Future research could focus on the predictive accuracy of the gradient boosting machine. The gradient boosting machine suggests that the predictive values of predictors widely vary in predicting the level of SDG disclosures, with RVI scores for some predictors lower than 10. As suggested by Jones et al. (2017), these indicators may result from random noise. Therefore, future research could re-estimate the prediction model, excluding predictors with low RVIs. On the other hand, this thesis suggests that the zero RVI scores may result from the measurements of the variables, sample size, and the heterogeneity of disclosure practices. Despite the gradient boosting machine being a widely used machine learning technique in accounting research, this thesis suggests that future research can compare the performance of the gradient boosting machine with other machine learning techniques, such as Random Forests and deep learning.

Future research could focus on developing a theoretical framework for explaining the variation in SDG disclosures. This thesis collected data for financial performance, corporate governance, and ESG performance data from existing databases, which are second-hand data. Even if I manually coded political connections, government shareholding, and voluntary regulatory influence proxies, their data are still from existing databases. Therefore, this thesis focuses on more quantifiable factors. This provides opportunities for future research to find reasons for their disclosures by collecting primary data. For example, future research could focus more on case studies to better understand why top-performing companies have SDG disclosures. This could also enable the development of a theoretical framework for explaining corporate SDG disclosure practices in Chinese companies.

Future research could explore alternative measures for predictors. This thesis developed novel methods for different government controls. As discussed above, the definitions and boundaries of these indicators may need to be further standardised.

1. This thesis develops a method for evaluating corporate political connections based on their corporate positions, such as politically connected CEOs and executives, directors, and senior management team members. As suggested by Qian and Chen (2021), the strength of the government may also affect the strength of their political connections and their corporate decision-making. Future research could develop a more comprehensive method to evaluate corporate political connections. Given that this thesis does not find the impact of the presence of political connections on the level of SDG disclosures. This thesis encourages future research to explore whether similar results may be found in other Chinese companies, or whether different measurements of the presence of political connection may affect SDG disclosures in Chinese companies.
2. This thesis develops a novel method for evaluating government shareholding on the level of SDG disclosures, by measuring the nature of SOEs. However, the Opinions of the State Council on Development of An Economy of Mixed-Ownership of State-Run Enterprise' provide limited guidance and blurry boundaries for each type of SOEs. Future research could investigate all types of SOEs based on a larger sample size and develop a more standardised way to measure the nature of SOEs.
3. This thesis developed a novel method for evaluating corporate engagement in regional SDG projects, by evaluating corporate engagement in the SDG IDZ projects. However, due to data availability, this thesis is limited to identifying whether the company is located in the cities that have been approved to participate in the SDG IDZ project. Future research could explore an approach to evaluate corporate direct engagement or investments in this project.

The motivations behind corporate SDG disclosures in Chinese companies may change due to the time period, national strategies, and stakeholder pressure. As discussed above, the 14th National FYP has

changed its focus on ‘Building a Modern Socialist Country’, targeting digital transformation and innovation. Therefore, SDG priorities may also be changed from SDG 1 to innovation-related goals, such as SDG 9. This provides some opportunities for future research to develop a new reporting framework to accommodate the changes in Chinese national circumstances. Unlike SDG 1 No Poverty which is a more Chinese-oriented goal, innovation and digital transformation are widely discussed worldwide, which may offer more opportunities to compare how companies in different countries may interpret these ideas, how they are turning their words into practices, and whether the reporting framework on innovation may be more generalisable to different contexts and jurisdictions.

This thesis is limited to 60 Chinese public companies and the SDG disclosure framework also takes into account the complex Chinese national circumstances. This leads us to consider whether the results and the reporting framework may be generalisable to other developing contexts or whether this may provide insights into future reporting guideline development.

7.6 Chapter conclusion

This thesis investigates the nature and extent of SDG disclosures in Chinese companies, following China’s adoption of Agenda 2030 and identifies factors that may predict voluntary SDG disclosures in Chinese companies based on the gradient boosting machine. This thesis developed an SDG disclosure framework and an automated content analysis tool to capture disclosure data. This thesis observed the variation of SDG disclosures in Chinese companies and imbalanced disclosures across different categories.

This thesis also employs an advanced machine learning technique, the gradient boosting machine, to identify potential factors that may predict the level of SDG disclosures. This thesis does find that financial performance, corporate governance, sustainability performance, and government control have power to predict the variation in the level of SDG disclosures, with predictive values widely varying across proxies selected. The results are consistent with the Chinese unique and complex hierarchical structure and concentrated nature of corporate governance set up in Chinese companies.

It offers insights to regulators on how SDGs are interpreted and implemented in Chinese companies and how they can be more effective in encouraging corporate engagement in SDG disclosures. This thesis also introduced the gradient boosting machine to SDG disclosure literature to output more reliable out-of-sample results. This thesis also provides a new reporting framework and automated content analysis tool for observing SDG disclosures, which may offer insights for researchers, regulators, and professional bodies in reporting framework development and data capture. This thesis also extends current literature related to factors that may predict voluntary SDG disclosures.

Reference

- Abeysekera, I., Li, F., & Lu, Y. (2021). Financial disclosure quality and sustainability disclosure quality. A case in China. *Plos One*, *16*(5), e0250884.
- Aboagye-Otchere, F. K., Simpson, S. N. Y., & Kusi, J. A. J. B. S. (2019). The influence of environmental performance on environmental disclosures: An empirical study in Ghana. *Business Strategy & Development*.
- Acar, E., Tunca Çalıyurt, K., & Zengin-Karaibrahimoglu, Y. (2021). Does ownership type affect environmental disclosure? *International Journal of Climate Change Strategies and Management*, *13*(2), 120-141.
- Adams, C. A. (2017). *The Sustainable Development Goals, integrated thinking and the integrated report*.
- Adams, C. A., Druckman, P. B., & Picot, R. C. (2020). Sustainable development goal disclosure (SDGD) recommendations. *ACCA: London, UK*.
- Air China. (2016). *2016 Corporate social responsibility report Air China Limited*.
- Air China. (2017). *2017 Corporate social responsibility report Alir China Limited*.
- Air China. (2019). *2019 Corporate social responsibility report Air China Limited*.
- Al-Tuwajjri, S. A., Christensen, T. E., & Hughes Ii, K. (2004). The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach. *Accounting, Organizations and Society*, *29*(5-6), 447-471.
- Alkayed, H., Yousef, I., Hussainey, K., & Shehadeh, E. (2024). The impact of COVID-19 on sustainability reporting: A perspective from the US financial institutions. *Journal of Applied Accounting Research*, *25*(2), 279-297.
- Alon, I., Wang, H., Shen, J., & Zhang, W. (2014). Chinese state-owned enterprises go global. *The Journal of Business Strategy*, *35*(6), 3-18.
- Anastasiu, I.-D., & Georgescu, M. R. (2020). Automated vs manual content analysis—A retrospective look. *Scientific Annals of Economics and Business*, *67*, 57-67.
- Andrikopoulos, A., & Krikiani, N. (2013). Environmental disclosure and financial characteristics of the firm: The case of Denmark. *Corporate Social Responsibility Environmental Management*, *20*(1), 55-64.
- Avrampou, A., Skouloudis, A., Iliopoulos, G., & Khan, N. (2019). Advancing the Sustainable Development Goals: Evidence from leading European banks. *Sustainable Development (Bradford, West Yorkshire, England)*, *27*(4), 743-757.
- Bae, J., Yu Hung, C., & van Lent, L. (2023). Mobilizing text as data. *European Accounting Review*, *32*(5), 1085-1106.

- Bahad, P., & Saxena, P. (2020). Study of adaboost and gradient boosting algorithms for predictive analytics. International Conference on Intelligent Computing and Smart Communication 2019: Proceedings of ICSC 2019
- Bannier, C., Pauls, T., & Walter, A. J. J. o. B. E. (2019). Content analysis of business communication: introducing a German dictionary. *Journal of Business Economics*, 89, 79-123.
- Bao, R., & Wei, L. (2024). The impact of companies disclosing ESG reports in multiple languages on the enthusiasm of foreign investors for holding shares. *Plos One*, 19(3), e0299643.
- Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research*, 58(1), 199-235.
- Bebbington, J., & Unerman, J. (2018). Achieving the United Nations Sustainable Development Goals: An enabling role for accounting research. *Accounting, Auditing & Accountability Journal*, 31(1), 2-24.
- Beck, C., Frost, G., & Jones, S. (2018). CSR disclosure and financial performance revisited: A cross-country analysis. *Australian Journal of Management*, 43(4), 517-537.
- Belé, N. A., Ferná, n., Feijó, n., Sou, o., Romero, S., & Blanco, S. R. (2012). Measuring quality of sustainability reports and assurance statements: characteristics of the high quality reporting companies. *International Journal of Society Systems Science*, 4(1), 5-27.
- Ben Saad, S., Laouti, M., & Ajina, A. (2024). Corporate social responsibility and credit rating: evidence from French companies. *Review of Accounting and Finance*, 23(3), 330-352.
- Bhatia, A., & Tuli, S. (2014). An empirical analysis of sustainability disclosure practices: Evidence from India and China. *IIM Kozhikode Society & Management Review*, 3(2), 135-148.
- Bi, X. (2021). The Communist Party Committee and corporate takeovers. *International review of Financial Analysis*, 78, 101948.
- Bin-Feng, C., Zhe, L., Mirza, S. S., & Safdar, R. (2023). Targeted poverty alleviation initiatives and corporate sustainable value creation: Determining the role of corporate internal governance control in China. *Heliyon*, 9(7), e17686.
- Blagov, Y. E., & Petrova-Savchenko, A. A. (2020). The transformation of corporate sustainability model in the context of achieving the UN SDGs: evidence from the leading Russian companies. *Corporate governance (Bradford)*, 21(2), 307-321. <https://doi.org/10.1108/CG-01-2020-0047>
- Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2023). Textual analysis in accounting: What's next? *Contemporary Accounting Research*, 40(2), 765-805.
- Bohnett, E., Coulibaly, A., Hulse, D., Hoctor, T., Ahmad, B., An, L., & Lewison, R. (2022). Corporate responsibility and biodiversity conservation: challenges and opportunities for companies participating in China's Belt and Road Initiative. *Environmental Conservation*, 49(1), 42-52.
- Bolognesi, M., Pilgram, R., & van den Heerik, R. (2017). Reliability in content analysis: The case of semantic feature norms classification. *Behavior Research Methods*, 49, 1984-2001.

- Bonsall, S. B., Green, J. R., & Muller, K. A. (2018). Are credit ratings more rigorous for widely covered firms? *The Accounting Review*, 93(6), 61-94.
- Borghei, Z., Leung, P., & Guthrie, J. (2018). Does voluntary greenhouse gas emissions disclosure reduce information asymmetry? Australian evidence. *Afro-Asian Journal of Finance and Accounting*, 8(2), 123-147.
- Bose, S., & Khan, H. Z. (2022). Sustainable development goals (SDGs) reporting and the role of country-level institutional factors: An international evidence. *Journal of Cleaner Production*, 335, 130290.
- Brammer, S., & Pavelin, S. (2006). Voluntary environmental disclosures by large UK companies. *Journal of Business Finance & Accounting*, 33(7-8), 1168-1188.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- Buniamin, S., Jaffar, R., Ahmad, N., & Johari, N. H. (2020). Exploring the involvement of Malaysian companies in Sustainable Development Goals (SDGs). *Global Business & Management Research*, 12(4).
- Büyükköçkan, G., & Karabulut, Y. (2018). Sustainability performance evaluation: Literature review and future directions. *Journal of Environmental Management*, 217, 253-267.
- Calabrese, A., Costa, R., Gastaldi, M., Levaldi Ghiron, N., & Villazon Montalvan, R. A. (2021). Implications for Sustainable Development Goals: A framework to assess company disclosure in sustainability reporting. *Journal of Cleaner Production*, 319, 128624.
- Calabrese, A., Costa, R., Ghiron, N. L., Tiburzi, L., & Montalvan, R. A. V. (2022). Is the private sector becoming cleaner? Assessing the firms' contribution to the 2030 Agenda. *Journal of Cleaner Production*, 132324.
- Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business Finance*, 61, 101649.
- Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization. *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*.
- CDSB, & WBCSD. (2017). The Reporting Exchange: An overview of sustainability and corporate reporting in China. *The Reporting Exchange*.
- Chakraborty, B., & Bhattacharjee, T. (2020). A review on textual analysis of corporate disclosure according to the evolution of different automated methods. *Journal of Financial Reporting Accounting*, 18(4), 757-777.
- Chang, Y., He, W., & Wang, J. (2020). Government Initiated Corporate Social Responsibility Activities: Evidence from a Poverty Alleviation Campaign in China. *Journal of Business Ethics*, 173, 1-25.

- Chauvey, J.-N., Giordano-Spring, S., Cho, C. H., & Patten, D. M. (2015). The Normativity and Legitimacy of CSR Disclosure: Evidence from France. *Journal of Business Ethics*, 130(4), 789-803.
- Chen, Q. (2021). *机器学习及 Python 应用*.
- Chen, X., Cho, Y. H., Dou, Y., & Lev, B. (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research*, 60(2), 467-515.
- Chen, Z., & Xie, G. (2022). ESG disclosure and financial performance: Moderating role of ESG investors. *International Review of Financial Analysis*, 83, 102291.
- Cheng, Z., Wang, F., Keung, C., & Bai, Y. (2017). Will corporate political connection influence the environmental information disclosure level? Based on the panel data of A-shares from Listed companies in Shanghai stock market. *Journal of Business Ethics*, 143(1), 209-221.
- China Life Insurance. (2016). *2016 Corporate social responsibility report*.
- China Life Insurance. (2017). *2017 Environmental, social and governance report*.
- China Life Insurance. (2019). *2019 Environmental, social and governance report*.
- Ching, H. Y., Gerab, F., & Toste, T. H. (2017). The quality of sustainability reports and corporate financial performance: Evidence from Brazilian listed companies. *SAGE Open*, 7(2), 2158244017712027.
- Chiu, C. L., Zhang, J., Li, M., Wei, S., Xu, S., & Chai, X. (2020). A study of environmental disclosures practices in Chinese energy industry. *Asian Journal of Sustainability and Social Responsibility*, 5(1), 1-21.
- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176-1248.
- Christopher, T., & Filipovic, M. (2008). The extent and determinants of disclosure of Global Reporting Initiative Guidelines: Australian evidence. *Journal of Contemporary Issues in Business and Government*, 14(2), 17.
- Chu, C. I., Chatterjee, B., & Brown, A. (2013). The current status of greenhouse gas reporting by Chinese companies: A test of legitimacy theory. *Managerial Auditing Journal*, 28(2), 114-139.
- Chua, W. F., Dai, N. T., Tan, Z. S., & Yu, L. (2024). Researching the accounting-state-market dynamic in China: A literature review and research agenda. *The British Accounting Review*, 101444.
- Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society*, 33(4-5), 303-327.
- CNPC. (2016). *2016 Corporate social responsibility report*.
- CNPC. (2017). *2017 Corporate social responsibility report*.
- CNPC. (2018). *2018 Corporate social responsibility report*.
- CNPC. (2019). *2019 Corporate social responsibility report*.

- Coenen, J., Bager, S., Meyfroidt, P., Newig, J., & Challies, E. (2021). Environmental governance of china's belt and road initiative. *Environmental Policy and Governance*, 31(1), 3-17.
- Communist Party of China. (2016a). *The 13th Five-year Plan for Economic and Social Development of the People's Republic of China (2016-2020)*. Central Compilation & Translation Press.
- Communist Party of China. (2016b). China's National Plan on Implementation of the 2030 Agenda for Sustainable Development.
- Communist Party of China. (2017). China's Progress Report on Implementation of the 2030 Agenda for Sustainable Development.
- Consolandi, C., Phadke, H., Hawley, J., & Eccles, R. G. (2020). Material ESG Outcomes and SDG Externalities: Evaluating the Health Care Sector's Contribution to the SDGs. *Organization & Environment*, 1086026619899795.
- Cosma, S., Venturelli, A., Schwizer, P., & Boscia, V. (2020). Sustainable development and European Banks: A non-financial disclosure analysis. *Sustainability (Basel, Switzerland)*, 12(15), 6146.
- Cotter, J., & Najah, M. M. (2012). Institutional investor influence on global climate change disclosure practices. *Australian Journal of Management*, 37(2), 169-187.
- Cubas-Díaz, M., & Martínez Sedano, M. (2018). Do credit ratings take into account the sustainability performance of companies? *Sustainability (Basel, Switzerland)*, 10(11), 4272.
- Dalla Via, N., & Perego, P. (2018). Determinants of conflict minerals disclosure under the Dodd–Frank Act. *Business Strategy and the Environment*, 27(6), 773-788.
- Dang, H.-A. H., & Serajuddin, U. (2020). Tracking the sustainable development goals: Emerging measurement challenges and further reflections. *World Development*, 127, 104570.
- Dang, V. Q., Otchere, I., & So, E. P. (2022). Does the nature of political connection matter for corporate social responsibility engagement? Evidence from China. *Emerging Markets Review*, 52, 100907.
- Datta, S., & Goyal, S. (2022). Determinants of SDG reporting by businesses: A literature analysis and conceptual model. *Vision*, 09722629221096047.
- Davern, M., Gyles, N., Hanlon, D., & Pinnuck, M. (2019). Is financial reporting still useful? Australian evidence. *Abacus*, 55(1), 237-272.
- de Andres, P., Garcia-Rodriguez, I., Romero-Merino, M. E., & Santamaria-Mariscal, M. (2022). Political directors and corporate social responsibility: Are political ideology and regional identity relevant? *Review of Managerial Science*, 1-35.
- De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, 192, 38-48.
- de Villiers, C., Ma, D., & Marques, A. (2024). Corporate social responsibility disclosure, dividend payments and firm value–relations and mediating effects. *Accounting & Finance*, 64(1), 185-219.
- Deloitte. (2019). *Sustainability Risk Management: Powering performance for responsible growth*.

- Deswanto, R. B., & Siregar, S. V. (2018). The associations between environmental disclosures with financial performance, environmental performance, and firm value. *Social Responsibility Journal*, 14(1), 180-193.
- Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, 86(1), 59-100.
- Dicko, S., Khemakhem, H., & Zogning, F. (2020). Political connections and voluntary disclosure: the case of Canadian listed companies. *Journal of Management and Governance*, 24, 481-506.
- Dissanayake, D., Tilt, C., & Xydias-Lobo, M. (2016). Sustainability reporting by publicly listed companies in Sri Lanka. *Journal of Cleaner Production*, 129, 169-182.
- Drobetz, W., Merikas, A., Merika, A., & Tsionas, M. G. (2014). Corporate social responsibility disclosure: The case of international shipping. *Transportation Research. Part E, Logistics and Transportation Review*, 71, 18-44.
- Du, X., & Zhou, Z. (2010). Political connections and information transparency of listed companies: Empirical evidence based on the Shenzhen Stock Exchange information disclosure evaluation. *Journal Of Zhongnan University of Economics and Law*, 178(1), 126-131.
- Dyduch, J., & Krasodomska, J. (2017). Determinants of corporate social responsibility disclosure: An empirical study of Polish listed companies. *Sustainability*, 9(11), 1934.
- EL Ammari, A. (2023). Does political connection affect corporate financial performance? The moderating role of directors' financial expertise. *Journal of Financial Reporting and Accounting*, 21(5), 1073-1099.
- Elalfy, A., Weber, O., & Geobey, S. (2021). The Sustainable Development Goals (SDGs): a rising tide lifts all boats? Global reporting implications in a post SDGs world. *Journal of Applied Accounting Research*, 22(3), 557-575.
- Erin, O. A., Bamigboye, O. A., & Oyewo, B. (2022). Sustainable development goals (SDG) reporting: an analysis of disclosure. *Journal of Accounting in Emerging Economies*.
- Ervits, I. (2021). CSR reporting by Chinese and Western MNEs: patterns combining formal homogenization and substantive differences. *International Journal of Corporate Social Responsibility*, 6(1), 1-24.
- Ervits, I. (2021a). CSR reporting in China's private and state-owned enterprises: A mixed methods comparative analysis. *Asian Business & Management*, 1-29.
- EY Global. (2017). *Why Sustainable Development Goals should be in your business plan*.
- Fernandez-Feijoo, B., Romero, S., & Ruiz, S. (2018). Financial auditor and sustainability reporting: Does it matter? *Corporate Social-Responsibility and Environmental Management*, 25(3), 209-224.

- Faye, A., Bob, M., Fall, A., & Fall, C. (2012). Primary health care and the millennium development goals. *Medecine et Sante Tropicales*, 22(1), 6-8.
- Florini, A., & Pauli, M. (2018). Collaborative governance for the sustainable development goals. *Asia & the Pacific Policy Studies*, 5(3), 583-598.
- Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm.
- Friedman, H. L., & Ormazabal, G. (2024). The role of information in building a more sustainable economy: A supply and demand perspective. *Journal of Accounting Research*.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.
- Frost, G., Jones, S., & Yu, M. (2023). Voluntary Carbon Reporting Prediction: A Machine Learning Approach. *Abacus*.
- Gambetta, N., Azcárate-Llanes, F., Sierra-García, L., & García-Benau, M. A. (2021). Financial institutions' risk profile and contribution to the sustainable development goals. *Sustainability*, 13(14), 7738.
- Gamerschlag, R., Möller, K., & Verbeeten, F. (2011). Determinants of voluntary CSR disclosure: empirical evidence from Germany. *Review of Managerial Science*, 5(2), 233-262.
- Gao, Y., & Feng, J. (2022). Communist party participation and corporate disclosure quality: Governance effect or reputation effect? *Journal of the Asia Pacific Economy*, 1-33.
- García-Meca, E., & Martínez-Ferrero, J. (2021). Is SDG reporting substantial or symbolic? An examination of controversial and environmentally sensitive industries. *Journal of Cleaner Production*, 298, 126781.
- García-Sánchez, I.-M., Aibar-Guzmán, B., Aibar-Guzmán, C., & Rodríguez-Ariza, L. (2020). "Sell" recommendations by analysts in response to business communication strategies concerning the Sustainable Development Goals and the SDG compass. *Journal of Cleaner Production*, 255, 120194.
- García-Sánchez, I. M., Rodríguez-Ariza, L., Aibar-Guzmán, B., & Aibar-Guzmán, C. (2020). Do institutional investors drive corporate transparency regarding business contribution to the sustainable development goals? *Business Strategy and the Environment*, 29(5), 2019-2036.
- Geertsema, P., & Lu, H. (2023). Relative valuation with machine learning. *Journal of Accounting Research*, 61(1), 329-376.
- Ghosh, S., & Rajan, J. (2019). The business case for SDGs: an analysis of inclusive business models in emerging economies. *International Journal of Sustainable Development and World Ecology*, 26(4), 344-353.
- Giannarakis, G. (2014). Corporate governance and financial characteristic effects on the extent of corporate social responsibility disclosure. *Social Responsibility Journal*, 10(4), 569-590.
- Giannarakis, G., Zafeiriou, E., & Sariannidis, N. (2017). The impact of carbon performance on climate change disclosure. *Business Strategy and the Environment*, 26(8), 1078-1094.

- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*.
- Girón, A., Kazemikhasragh, A., Cicchiello, A. F., & Panetti, E. (2020). Sustainability reporting and firms' economic performance: evidence from Asia and Africa. *Journal of the Knowledge Economy*, 1-19.
- Github. (2023). *Why Python keeps growing, explained*.
- Goldman, E., Rocholl, J., & So, J. (2013). Politically connected boards of directors and the allocation of procurement contracts. *Review of Finance*, 17(5), 1617-1648.
- Grainger-Brown, J., & Malekpour, S. (2019). Implementing the Sustainable Development Goals: A Review of Strategic Tools and Frameworks Available to Organisations. *Sustainability (Basel, Switzerland)*, 11(5), 1381.
- GRI. (2017). *The Power of Sustainability Reporting*. <https://www.globalreporting.org/information/about-gri/Pages/default.aspx>.
- GRI. (2020). *Linking the SDGs and the GRI Standards*.
- GRI. (2024). *A short Introduction to the GRI standards*. <https://www.globalreporting.org/media/wtafl4tw/a-short-introduction-to-the-gri-standards.pdf>
- GRI 302. (2016). *GRI 302: Energy 2016* <https://www.globalreporting.org/pdf.aspx?id=12467&page=8>
- GRI 305. (2016). *GRI 305: Emissions 2016*. <https://www.globalreporting.org/publications/documents/english/gri-305-emissions-2016/>
- GRI, PRI, & United Nations Global Compact. (2018). *UN in focus: addressing investor needs—Address reporting on the SDGs*.
- GRI, UN Global Compact, & WBCSD. (2015). *Linking the SDGs and GRI* <https://sdgcompass.org/wp-content/uploads/2015/09/SDG-Compass-Linking-the-SDGs-and-GRI.pdf>
- GRI, & United Nations Global Compact. (2018). *Integrating the Sustainable Development Goals into corporate reporting: A practical guide*.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.
- Gunasegaran, T., & Cheah, Y.-N. (2017). Evolutionary cross validation. *2017 8th International Conference on Information Technology (ICIT)*.
- Gul, F. A., Cheng, L. T., & Leung, T. (2011). Perks and the informativeness of stock prices in the Chinese market. *Journal of Corporate Finance*, 17(5), 1410-1429.
- Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, 93(2), 332-359.

- Guping, C., Safdar Sial, M., Wan, P., Badulescu, A., Badulescu, D., & Vianna Brugni, T. (2020). Do board gender diversity and non-executive directors affect CSR reporting? Insight from agency theory perspective. *Sustainability*, *12*(20), 8597.
- Gupta, D., Kumar, Y., & Sajjan, N. J. I. J. o. C. A. (2012). Improving unsupervised stemming by using partial lemmatization coupled with data-based heuristics for Hindi. *International Journal of Computer Applications*, *38*(8), 1-8.
- Guthrie, J., & Abeysekera, I. (2006). Content analysis of social, environmental reporting: What is new? *Journal of Human Resource Costing & Accounting*.
- Gutterman, A. (2020). SDG-Related Reporting. Available at SSRN 3823833.
- Hansen, B. (2010). *Econometrics*. University of Wisconsin.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- Heichl, V., & Hirsch, S. (2023). Sustainable fingerprint—Using textual analysis to detect how listed EU firms report about ESG topics. *Journal of Cleaner Production*, *426*, 138960.
- Heras-Saizarbitoria, I., Urbieto, L., & Boiral, O. (2022). Organizations' engagement with sustainable development goals: From cherry-picking to SDG-washing? *Corporate Social Responsibility and Environmental Management*, *29*(2), 316-328.
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, *25*(1), 114-146.
- Hinze, A.-K., & Sump, F. (2019). Corporate social responsibility and financial analysts: A review of the literature. *Sustainability Accounting, Management Policy Journal*, *10*(1), 183-207.
- Hoang, T.-H.-V., BERROU, R., & Pham, L. (2022). The impact of green bond issuance on firms' financial and ESG performance: Does the proportion of green bonds matter? *SSRN*.
- Horne, J., Recker, M., Michelfelder, I., Jay, J., & Kratzer, J. (2020). Exploring entrepreneurship related to the sustainable development goals - mapping new venture activities with semi-automated content analysis. *Journal of Cleaner Production*, *242*, 118052.
- Huan, Y., Li, H., & Liang, T. (2019). A new method for the quantitative assessment of Sustainable Development Goals (SDGs) and a case study on Central Asia. *Sustainability*, *11*(13), 3504.
- Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806-841.
- Huber, B. M., Comstock, M., Simith, H., Polk, D., & LLP, W. (2018). *UN Sustainable Development Goals — The Leading ESG Framework for Large Companies*. <https://corp.gov.law.harvard.edu/2018/10/04/un-sustainable-development-goals-the-leading-esg-framework-for-large-companies/>

- Hummel, K. (2019). Reporting on the Sustainable Development Goals—Early Evidence from Europe. *SSRN*.
- Hummel, K., & Rötzel, P. (2019). Mandating the sustainability disclosure in annual reports—evidence from the United Kingdom. *Schmalenbach Business Review*, 71(2), 205-247.
- Hummel, K., & Schlick, C. (2016). The relationship between sustainability performance and sustainability disclosure—Reconciling voluntary disclosure theory and legitimacy theory. *Journal of Accounting and Public Policy*, 35(5), 455-476.
- Hussain, S., Liu, L. Y. J., & Miller, A. D. (2020). Accounting as a dichotomised discipline: An analysis of the source materials used in the construction of accounting articles. *Critical Perspectives on Accounting*, 66, 102086.
- ICBC. (2016). *2016 Corporate social responsibility report*.
- ICBC. (2017). *2017 Corporate social responsibility report*.
- Ionaşcu, E., Mironiuc, M., Anghel, I., & Huian, M. C. (2020). The Involvement of Real Estate Companies in Sustainable Development—An Analysis from the SDGs Reporting Perspective. *Sustainability*, 12(3), 798.
- Izzo, M. F., Ciaburri, M., & Tiscini, R. (2020a). The Challenge of Sustainable Development Goal Reporting: The First Evidence from Italian Listed Companies. *Sustainability*, 12(8), 3494.
- Izzo, M. F., Dello Strologo, A., & Granà, F. (2020). Learning from the Best: New Challenges and Trends in IR Reporters' Disclosure and the Role of SDGs. *Sustainability (Basel, Switzerland)*, 12(14), 5545.
- Jeanjean, T., Lesage, C., & Stolowy, H. (2010). Why do you speak English (in your annual report)? *The International Journal of Accounting*, 45(2), 200-223.
- Jiang, F., Cai, W., Wang, X., & Zhu, B. (2018). Multiple large shareholders and corporate investment: Evidence from China. *Journal of Corporate Finance*, 50, 66-83.
- Jiang, F., & Kim, K. A. (2020). Corporate governance in China: A survey. *Review of Finance*, 24(4), 733-772.
- Jiang, F., & Kim, K. A. (2024). Understanding corporate governance in China. *The British Accounting Review*, 56(5), 101459.
- Jiang, J. X., Kong, J., & Shen, M. (2024). The information content of rating action reports: A topic modeling approach. *The British Accounting Review*, 101435.
- Jiang, Y., & Jones, S. (2018). Corporate distress prediction in China: A machine learning approach. *Accounting & Finance*, 58(4), 1063-1109.
- Jones, S. (2017). Corporate bankruptcy prediction: a high dimensional analysis. *Review of Accounting Studies*, 22(3), 1366-1422.
- Jones, S. (2023). The Rise of the Machines. In *Distress Risk and Corporate Failure Modelling: The State of the Art* (pp. 242). Routledge.

- Jones, S., Johnstone, D., & Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance* 56, 72-85.
- Jones, S., & Wright, C. (2018). Fashion or future: does creating shared value pay? *Accounting & Finance*, 58(4), 1111-1139.
- Kartal, M. T., Taşkın, D., Shahbaz, M., Depren, S. K., & Pata, U. K. (2024). Effects of Environment, Social, and Governance (ESG) disclosures on ESGscores: Investigating the role of corporate governance for publicly traded Turkish companies. *Journal of Environmental Management*, 368, 122205.
- Kasperuniene, J., Briedienė, M., & Zydziunaite, V. (2020). Automatic content analysis of social media short texts: scoping review of methods and tools. *Computer Supported Qualitative Research: New Trends on Qualitative Research (WCQR2019)* 4, 89-101.
- Kenter, T., & De Rijke, M. (2015). Short text similarity with word embeddings. *Proceedings of the 24th ACM international on conference on information and knowledge management*.
- Khalid, F., Ye, Z., Voinea, C. L., & Naveed, K. (2022). Carbon disclosure project: Chinese chief executive officer background and corporate voluntary climate change reporting. *Carbon Management*, 13(1), 321-336.
- Khan, H.-U.-Z. (2010). The effect of corporate governance elements on corporate social responsibility (CSR) reporting. *International journal of law and management*, 52(2), 82-109.
- Khan, M., Serafeim, G., & Yoon, A. J. T. a. r. (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6), 1697-1724.
- Khan, M. K., Zahid, R., Saleem, A., & Sági, J. (2021). Board composition and social & environmental accountability: A dynamic model analysis of Chinese firms. *Sustainability*, 13(19), 10662.
- Khan, P. A., Johl, S. K., & Akhtar, S. (2021). Firm Sustainable Development Goals and firm financial performance through the lens of green innovation practices and reporting: A proactive approach. *Journal of Risk and Financial Management*, 14(12), 605.
- Khlif, H., Ahmed, K., & Souissi, M. (2017). Ownership structure and voluntary disclosure: A synthesis of empirical studies. *Australian Journal of Management*, 42(3), 376-403.
- Khurram, M. U., Xie, W., Mirza, S. S., & Tong, H. (2023). Green bonds issuance, innovation performance, and corporate value: Empirical evidence from China. *Heliyon*, 9(4).
- Komor, P., & Bazilian, M. J. E. P. (2005). Renewable energy policy goals, programs, and technologies. 33(14), 1873-1881.
- Kotu, V., & Deshpande, B. (2018). *Data science: concepts and practice*. Morgan Kaufmann.
- KPMG. (2017). *Sustainable Development Goals (SDGs): leveraging CSR to achieve SDGs*. https://assets.kpmg/content/dam/kpmg/in/pdf/2017/12/SDG_New_Final_Web.pdf
- KPMG. (2018). *How to Report on the SDGs: What Good Looks Like and Why It Matters*.
- KPMG. (2020). *The Time Has Come: The KPMG Survey of Sustainability Reporting 2020*.
- KPMG China. (2022). *Scope 3 emissions: Disclosing data about the value chain*.

- Krippendorff, K. (1980). Validity in content analysis.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
- Lagasio, V., & Cucari, N. (2019). Corporate governance and environmental social governance disclosure: A meta-analytical review. *Corporate Social Responsibility and Environmental Management*, 26(4), 701-711.
- Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of accounting and economics*, 60(2-3), 110-135.
- Lappeman, J., Clark, R., Evans, J., Sierra-Rubia, L., & Gordon, P. (2020). Studying social media sentiment using human validated analysis. *MethodsX*, 7, 100867.
- Lassala, C., Orero-Blat, M., & Ribeiro-Navarrete, S. (2021). The financial performance of listed companies in pursuit of the Sustainable Development Goals (SDG). *Economic Research-Ekonomska Istraživanja*, 34(1), 427-449.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lev, B. (2018). The deteriorating usefulness of financial report information and how to reverse it. *Accounting and Business Research*, 48(5), 465-493.
- Lewis, D. J., Yang, X., Moise, D., & Roddy, S. J. (2021). Dynamic synergies between China's Belt and Road Initiative and the UN's Sustainable Development Goals. *Journal of International Business Policy*, 4(1), 58-79.
- Li, D., Zhao, Y., Sun, Y., & Yin, D. (2017). Corporate environmental performance, environmental information disclosure, and financial performance: Evidence from China. *Human and Ecological Risk Assessment: An International Journal*, 23(2), 323-339.
- Li, J., & Xu, X. (2024). Can ESG rating reduce corporate carbon emissions?—An empirical study from Chinese listed companies. *Journal of Cleaner Production*, 434, 140226.
- Li, M., & Cheng, H. (2021). Participation of communist party organizations in corporate governance and innovation: Empirical evidence from listed state-owned enterprises. *Journal of Systems & Management*, 30(3), 401.
- Li, Q., & Guo, M. (2022). Do the resignations of politically connected independent directors affect corporate social responsibility? Evidence from China. *Journal of Corporate Finance*, 73, 102174.
- Li, W., & Zhang, R. (2010). Corporate social responsibility, ownership structure, and political interference: Evidence from China. *Journal of Business Ethics*, 96(4), 631-645.
- Li, X., Chan, K. C., & Ma, H. (2020). Communist party direct control and corporate investment efficiency: evidence from China. *Asia-Pacific Journal of Accounting & Economics*, 27(2), 195-217.
- Li, Y., Wong, C. W., & Miao, X. (2021). Political connection, political promotion and corporate environmental information disclosure in China. *Chinese Management Studies*.

- Li, Z., Jia, J., & Chapple, L. J. (2022). Textual characteristics of corporate sustainability disclosure and corporate sustainability performance: evidence from Australia. *Meditari Accountancy Research*.
- Limkriangkrai, M., Koh, S., & Durand, R. B. (2017). Environmental, Social, and Governance (ESG) profiles, stock returns, and financial policy: Australian evidence. *International Review of Finance*, 17(3), 461-471.
- Lin, H.-Y., & Hsu, B.-W. (2023). Empirical study of ESG score prediction through machine learning—A case of non-financial companies in Taiwan. *Sustainability*, 15(19), 14106.
- Lin, K. J., Lu, X., Zhang, J., & Zheng, Y. (2020). State-owned enterprises in China: A review of 40 years of research and practice. *China Journal of Accounting Research*, 13(1), 31-55.
- Lin, N. (2011). Capitalism in China: A centrally managed capitalism (CMC) and its future. *Management and Organization Review*, 7(1), 63-96.
- Lin, Y., Shen, R., Wang, J., & Julia Yu, Y. (2024). Global evolution of environmental and social disclosure in annual reports. *Journal of Accounting Research*.
- Liu, X., & Anbumozhi, V. (2009). Determinant factors of corporate environmental information disclosure: an empirical study of Chinese listed companies. *Journal of cleaner production*, 17(6), 593-600.
- Lu, F., Kozak, R., Toppinen, A., D'Amato, D., & Wen, Z. (2017). Factors influencing levels of CSR disclosure by forestry companies in China. *Sustainability*, 9(10), 1800.
- Lu, J., Liang, M., Zhang, C., Rong, D., Guan, H., Mazeikaite, K., & Streimikis, J. (2021). Assessment of corporate social responsibility by addressing sustainable development goals. *Corporate Social-Rresponsibility and Environmental Management*, 28(2), 686-703.
- Lu, L. Y., Shailer, G., & Yu, Y. (2017). Corporate social responsibility disclosure and the value of cash holdings. *European Accounting Review*, 26(4), 729-753.
- Lu, Y., & Abeysekera, I. (2014). Stakeholders' power, corporate characteristics, and social and environmental disclosure: evidence from China. *Journal of cleaner production*, 64, 426-436.
- Luo, X. R., Danqing, W., & Jianjun, Z. (2017). Whose Call to Answer: Institutional Complexity and Firms' CSR Reporting. *Academy of Management Journal*, 60(1), 321-344.
- Luo, X., & Liu, W. (2019). Political connection and fir's environmental violation punishment: Sheltering or monitoring. *Journal of Shanxi University of Finance and Economics*, 41(10), 85-99.
- Lyulyov, O., Kwilinski, A., Pimonenko, T., Us, Y., & Korobets, O. (2023). The role of environmental risk management in company profitability within sustainable development goals. *E3S Web of conferences: 3rd international interdisciplinary scientific conference "Digitalization and sustainability for development management: economic, social, and environmental aspects"* 2023, 14-15 October 2023, London.,
- Machado, B. A. A., Dias, L. C. P., & Fonseca, A. (2021). Transparency of materiality analysis in GRI-based sustainability reports. *Corporate Social Responsibility Environmental Management*. 28(2), 570-580.

- Mahmood, Z., Ahmad, Z., Ali, W., & Ejaz, A. (2017). Does environmental disclosure relate to environmental performance? Reconciling legitimacy theory and voluntary disclosure theory. *Journal of Commerce and Social Sciences (PJCSS)*, 11(3), 1134-1152.
- Manes-Rossi, F., & Nicolo', G. (2022). Exploring sustainable development goals reporting practices: From symbolic to substantive approaches—Evidence from the energy sector. *Corporate Social Responsibility and Environmental Management*, 29(5), 1799-1815.
- Marquis, C., & Qian, C. (2014). Corporate social responsibility reporting in China: Symbol or substance? *Organization Science*, 25(1), 127-148.
- Martínez-Ferrero, J., & García-Meca, E. (2020). Internal corporate governance strength as a mechanism for achieving sustainable development goals. *Sustainable Development*, 28(5), 1189-1198.
- Matuszak, Ł., & Róžańska, E. (2017). An examination of the relationship between CSR disclosure and financial performance: The case of Polish banks. *Accounting and Management Information Systems*, 16(4), 522-533.
- Meah, M. R., & Chaudhory, N. U. (2019). Corporate governance and firm's profitability: An emerging economy-based investigation. *Indian Journal of Corporate Governance*, 12(1), 71-93.
- Meng, X., Zeng, S., Shi, J. J., Qi, G., & Zhang, Z. (2014). The relationship between corporate environmental performance and environmental disclosure: An empirical study in China. *Journal of Environmental Management*, 145, 357-367.
- Meng, X. H., Zeng, S. X., & Tam, C. M. (2013). From voluntarism to regulation: A study on ownership, economic performance and corporate environmental information disclosure in china. *Journal of Business Ethics*, 116(1), 217-232.
- Michalczuk, G., & Konarzewska, U. (2018). The use of GRI standards in reporting on actions being taken by companies for sustainable development. *Optimum. Economic Studies*, 4 (94), 72-86.
- Milne, M. J., & Adler, R. W. (1999). Exploring the reliability of social and environmental disclosures content analysis. *Accounting, Auditing & Accountability Journal*, 12(2), 237-256.
- Ministry of Ecology and Environment the People's Republic of China. (2017). *The Belt and Road Ecological and Environmental Cooperation Plan*. Retrieved from https://english.mee.gov.cn/Resources/Policies/policies/Frameworkp1/201706/t20170628_416869.shtml
- Ministry of Foreign Affairs of The People's Republic of China. (2017). *China's Progress Report on Implementation of the 2030 Agenda for Sustainable Development*. https://www.fmprc.gov.cn/mfa_eng/topics_665678/2030kcxzfzyc/201708/P020210525474841044350.pdf
- Minitab. (2019). *Introducing TreeNet® Gradient Boosting Machine*.
- Moneva, J. M., & Llana, F. (2000). Environmental disclosures in the annual reports of large companies in Spain. *European Accounting Review*, 9(1), 7-29.

- Muhmad, S. N., & Muhamad, R. (2020). Sustainable business practices and financial performance during pre-and post-SDG adoption periods: a systematic review. *Journal of Sustainable Finance & Investment*, 1-19.
- Mukarram, M. (2020). Impact of COVID-19 on the UN sustainable development goals (SDGs). *Strategic Analysis*, 44(3), 253-258.
- Mun, J., Lee, S., Choi, J., Choi, J., & Bae, K. (2021). Dataset retrieval system based on automation of data preparation with dataset description model. *Concurrency and Computation: Practice and Experience*, 33(2), e5288.
- Naheed, R., AlHares, A., Shahab, Y., & Naheed, R. (2021). Board's financial expertise and corporate social responsibility disclosure in China. *Corporate governance (Bradford)*, 21(4), 716-736. <https://doi.org/10.1108/CG-08-2020-0329>
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7, 21.
- National Congress of the Chinese Communist Party. (2022). *Resolution of the 20th National Congress of the Communist Party of China on the Revised Constitution of the Communist Party of China*. Retrieved from <https://www.12371.cn/special/zggcdzc/zggcdzcqw/>
- National Development and Reform Commission. (2016). *The 13th Five-Year Plan for Economic and Social Development of the People's Republic of China (2016–2020)*. Translated by the Compilation and Translation Bureau, Central Committee of the Communist Party of China.
- National Development and Reform Commission, People's Bank of China, & China Securities Regulatory Commission. (2021). *Catalogue of Green Bond Supported Projects*. Retrieved from http://www.gov.cn/zhengce/zhengceku/2021-04/22/content_5601284.htm
- Nechita, E., Manea, C., Irimescu, A., & Nichita, E. (2020). The content analysis of reporting on sustainable development goals. *Audit Finance*, 18, 160.
- Nechita, E., Manea, C. L., Nichita, E.-M., Irimescu, A.-M., & Manea, D. (2020). Is Financial Information Influencing the Reporting on SDGs? Empirical Evidence from Central and Eastern European Chemical Companies. *Sustainability (Basel, Switzerland)*, 12(9251), 9251.
- Neuendorf, K. A. (2010). Reliability for content analysis. In *Media Messages and Public Health* (pp. 85-105). Routledge.
- Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021(1), 4832864.
- Nichita, E.-M., Nechita, E., Manea, C.-L., Manea, D., & Irimescu, A.-M. (2020). Reporting on Sustainable Development Goals. A score based-approach with company-level evidence from Central-Eastern Europe economies. *Accounting and Management Information Systems*, 19(3), 501-541.

- Nicolo', G., Zampone, G., De Iorio, S., & Sannino, G. (2024). Does SDG disclosure reflect corporate underlying sustainability performance? Evidence from UN Global Compact participants. *Journal of International Financial Management & Accounting*, 35(1), 214-260.
- Nilipour, A., de Silva, T.-A., & Li, X. (2020). The Readability of Sustainability Reporting in New Zealand over time. *Australasian Accounting, Business & Finance Journal*, 14(3), 86-107.
- Noorunnahar, M., Chowdhury, A. H., & Mila, F. A. (2023). A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *Plos One*, 18(3), e0283452.
- Nurleni, N., & Bandang, A. (2018). The effect of managerial and institutional ownership on corporate social responsibility disclosure. *International Journal of Law and Management*, 60(4), 979-987.
- OECD. (2021). *OECD Regulatory Policy Outlook 2021*. <https://www.oecd.org/en/topics/regulatory-reform.html>
- Orazalin, N., & Mahmood, M. (2020). Determinants of GRI-based sustainability reporting: evidence from an emerging economy. *Journal of Accounting in Emerging Economies*, 10(1), 140-164.
- Ordonez-Ponce, E., & Khare, A. (2021). GRI 300 as a measurement tool for the United Nations sustainable development goals: assessing the impact of car makers on sustainability. *Journal of Environmental Planning and Management*, 64(1), 47-75.
- Ostertag, O., Kováč, J., & Ostertagová, E. (2014). Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials*, 611(Applied Mechanics and Mechatronics), 115-120.
- Parhankangas, A., & Ehrlich, M. (2014). How entrepreneurs seduce business angels: An impression management approach. *Journal of Business Venturing*, 29(4), 543-564.
- Parsa, S., Dai, N., Belal, A., Li, T., & Tang, G. (2021). Corporate social responsibility reporting in China: political, social and corporate influences. *Accounting and Business Research*, 51(1), 36-64.
- Permanent Mission of the People's Republic of China to the UN. (2022). *Accelerate the Implementation of 2030 Agenda for Sustainable Development, Build a Shared Future for All Life on Earth — Remarks at Opening of Ministerial Segment of 2022 HLPF By His Excellency HUANG Runqiu, President, COP 15 of CBD Minister of Ecology and Environment of China*. http://un.china-mission.gov.cn/eng/hyyfy/202207/t20220714_10719654.htm
- Phan, H.-T.-P., De Luca, F., & Iaia, L. (2020). The “Walk” towards the UN Sustainable Development Goals: Does mandated “talk” through non-financial disclosure affect companies’ financial performance? *Sustainability*, 12(6), 2324.
- Pizzi, S., Caputo, A., Corvino, A., & Venturelli, A. (2020). Management research and the UN sustainable development goals (SDGs): A bibliometric investigation and systematic review. *Journal of Cleaner Production*, 124033.
- Pizzi, S., Rosati, F., & Venturelli, A. (2021c). The determinants of business contribution to the 2030 Agenda: Introducing the SDG Reporting Score. *Business Strategy and the Environment*, 30(1), 404-421.

- Porter, M. E., & Kramer, M. R. (2011). Creating shared value. In *Managing sustainable business* (pp. 323-346). Springer.
- Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis.
- Prakoso, D. W., Abdi, A., & Amrit, C. (2021). Short text similarity measurement methods: a review. *Soft Computing*, 25, 4699-4723.
- Pramana, R., Subroto, J. J., & Gunawan, A. A. S. (2022). Systematic literature review of stemming and lemmatization performance for sentence similarity. *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*.
- Putri, P. (2021). The effect of operating cash flows, sales growth, and operating capacity in predicting financial distress. *International Journal of Innovative Science Research Technology*, 6(1), 638-646.
- PwC. (2016). *Navigating the SDGs: A business guide to engaging with the UN global goals*.
- PwC. (2018). *SDG reporting challenge. from promise to reality: Does business really care about the SDGs? and what needs to happen to turn words into action*.
- PwC. (2020). *Private sector awareness of the Sustainable Development Goals: A survey report on business and sustainability in China*.
- PwC Slovakia. (n.d.). *ESG reporting and preparation of a sustainability report*. <https://www.pwc.com/sk/en/environmental-social-and-corporate-governance-esg/esg-reporting.html>
- Python. (2024). *What is Python? Executive Summary*. https://www.python.org/doc/essays/blurb/?external_link=true
- Qian, W., & Chen, X. (2021). Corporate environmental disclosure and political connection in regulatory and leadership changes: The case of China. *The British Accounting Review*, 53(1), 100935.
- Qian, W., Parker, L., & Zhu, J. (2024). Corporate environmental reporting in the China context: The interplay of stakeholder salience, socialist ideology and state power. *The British Accounting Review*, 56(1), 101198.
- Qian, W., Tilt, C., & Belal, A. (2021). Social and environmental accounting in developing countries: contextual challenges and insights. *Accounting, Auditing & Accountability Journal*, 34(5), 1021-1050.
- Qian, W., Zhu, P., & Tilt, C. (2022). Corporate environmental disclosure and local government environmental oversight in China. *Pacific Accounting Review*.
- Qiao, T., Han, L., & Liu, Y. (2021). Does targeted poverty alleviation disclosure improve stock performance? *Economics Letters*, 201, 109805.
- Ranta, M., Ylinen, M., & Järvenpää, M. (2023). Machine learning in management accounting research: Literature review and pathways for the future. *European Accounting Review*, 32(3), 607-636.

- Rauf, F., Voinea, C. L., Naveed, K., & Fratostiteanu, C. (2021). CSR Disclosure: Effects of Political Ties, Executive Turnover and Shareholder Equity. Evidence from China. *Sustainability*, 13(7), 3623.
- Rezaee, Z. (2016). Business sustainability research: A theoretical and integrated perspective. *Journal of Accounting Literature*, 36(1), 48-64.
- Rosati, F., & Faria, L. G. D. (2019a). Addressing the SDGs in sustainability reports: The relationship with institutional factors. *Journal of cleaner production*, 215, 1312-1326.
- Rosati, F., & Faria, L. G. D. (2019b). Business contribution to the Sustainable Development Agenda: Organizational factors related to early adoption of SDG reporting. *Corporate Social-Responsibility And Environmental Management*, 26(3), 588-597.
- Russell, E., Lee, J., & Clift, R. (2018). Can the SDGs provide a basis for supply chain decisions in the construction sector? *Sustainability*, 10(3), 629.
- Ryans, J. P. (2021). Textual classification of SEC comment letters. *Review of Accounting Studies*, 26(1), 37-80.
- Sasanti, E. E., Cahyaningtyas, S. R., & Soesatyo, R. S. H. B. (2020). Sustainable Development Goals disclosures and company values. *Jurnal Riset Akuntansi Aksioma*, 19(1), 29-45.
- Schönherr, N., Findler, F., & Martinuzzi, A. (2017). Exploring the Interface of CSR and the Sustainable Development Goals. *Transnational corporations*, 24(3), 33.
- Sciandra, A., Surian, A., & Finos, L. (2021). Supervised machine learning methods to disclose action and information in “UN 2030 Agenda” social media data. *Social indicators research*, 156(2), 689-699.
- Scikit-Learn. (2019). *User guide: contents — scikit-learn 0.22.1 documentation*. Scikit-Learn.org. https://scikit-learn.org/stable/user_guide.html
- SDG Compass. (2015). The guide for business action on the SDGs. *World Business Council for Sustainable Development (WBCSD): Geneva, Switzerland*.
- Shahab, Y., & Ye, C. (2018). Corporate social responsibility disclosure and corporate governance: empirical insights on neo-institutional framework from China. *International Journal of Disclosure and Governance*, 15(2), 87-103.
- Shanmugasundar, G., Vanitha, M., Čep, R., Kumar, V., Kalita, K., & Ramachandran, M. (2021). A comparative study of linear, random forest and adaboost regressions for modeling non-traditional machining. *Processes*, 9(11), 2015.
- Shen, C. (2023). Independent directors’ political connections and corporate environmental information disclosure: Empirical evidence from Chinese industrial listed companies. *Journal of SunYan-Sen University*, 63(302), 194-206.
- Shen, C. (2023). 独立董事政治联系与企业环境信息披露-- 来自中国工业上市公司的经验证据. *Journal of SunYan-Sen University*, 63(302), 194-206.

- Shen, H., Lin, H., Han, W., & Wu, H. (2023). ESG in China: A review of practice and research, and future research avenues. *China Journal of Accounting Research*, 100325.
- Shevchenko, I., Andreev, P., Dernova, M., & Poddubei, O. (2023). Paragraph-oriented methods for determining the coherence and thematic unity of scientific and technical texts. *Radioelectronic and Computer Systems*, (2), 27-35.
- Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5(1), 415-435.
- Singhania, M., & Chadha, G. (2023). Impact of debt on sustainability reporting: a meta-analysis of the moderating role of country characteristics. *Journal of Accounting Literature*.
- Situ, H., & Tilt, C. (2018). Mandatory? Voluntary? A discussion of corporate environmental disclosure requirements in China. *Social and Environmental Accountability Journal*, 38(2), 131-144.
- Situ, H., Tilt, C., & Seet, P.-S. (2021). The influence of the Chinese government's political ideology in the field of corporate environmental reporting. *Accounting, Auditing, & Accountability*, 34(9), 1-28.
- Situ, H., & Tilt, C. A. (2012). Chinese government as a determinant of corporate environmental reporting: a study of large Chinese listed companies. *Journal of Asia Pacific Centre For Environmental Accountability*, 18(4), 251-286.
- Situ, H., Tilt, C. A., & Seet, P.-S. (2018). The influence of the government on corporate environmental reporting in China: an authoritarian capitalism perspective. *Business & Society*, 59(8), 1589-1629.
- Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgaarden, H. G. (2020). In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication*, 37(4), 550-572.
- State Council of People's Republic of China. (2015). *Opinions of the State Council on Development of An Economy of Mixed-Ownership of State-Run Enterprises*. Retrieved from <http://lawinfochina.com/display.aspx?id=28120&lib=law>
- State Council of People's Republic of China. (2016). *13th FYP development plan for poverty alleviation*. www.gov.cn/zhengce/content/2016-12/02/content_5142197.htm
- State Council of the People's Republic of China. (2018). *农村一二三产业融合发展年度报告*. Retrieved from https://www.gov.cn/xinwen/2018-04/20/content_5284269.htm#1
- State Council of People's Republic of China. (2021). *White Paper on China's Practice in Poverty Reduction*. https://www.gov.cn/zhengce/2021-04/06/content_5597952.htm
- State Council Information Office of the People's Republic of China. (2017). *Guiding Opinion on Regulating Banking Services for Enterprises Heading Overseas and Strengthening Risk*

Prevention and Control. Retrieved from <http://www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/35861/36645/xgzc36651/Document/1551306/1551306.htm>

State Council Information Office of the People's Republic of China. (2019). *Promoting 'One Belt and One Road' to build a community of shared future* <http://www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/39595/40298/xgbd40305/Document/1652420/1652420.htm>

Steenkamp, N., & Northcott, D. (2007). Content analysis in accounting research: The practical challenges. *Australian Accounting Review*, 17(43), 12-25.

Subramaniam, N., Mori Junior, R., Akbar, S., Ji, H., & Situ, H. (2020). *SDG Measurement and Disclosure 2.0 A study of ASX150 companies*.

Sun, Y., Ma, Z., Zeng, X., & Guo, Y. (2021). A predicting model for accounting fraud based on ensemble learning. *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*.

Szennay, Á., Szigeti, C., Kovács, N., & Szabó, D. R. (2019). Through the blurry looking glass—SDGs in the GRI reports. *Resources*, 8(2), 101.

Tangngisalu, J. (2020). Effect of Cash Flow and Corporate Social Responsibility disclosure on Firm Value. *ATESTASI: Jurnal Ilmiah Akuntansi*, 3(2), 142-149.

Tenuta, P., & Cambrea, D. R. (2022). ESG measures and non-financial performance reporting. In *Corporate sustainability: Measurement, reporting and effects on firm performance* (pp. 27-57). Springer.

Teo, H. C., Lechner, A. M., Walton, G. W., Chan, F. K. S., Cheshmehzangi, A., Tan-Mullins, M., Chan, H. K., Sternberg, T., & Campos-Arceiz, A. (2019). Environmental impacts of infrastructure development under the Belt and Road Initiative. *Environments (Basel, Switzerland)*, 6(6), 72.

Tihanyi, L., Aguilera, R. V., Heugens, P., Van Essen, M., Sauerwald, S., Duran, P., & Turturea, R. (2019). State ownership and political connections. *Journal of Management*, 45(6), 2293-2321.

Ting, H.-I., & Lee, Y.-C. (2024). When politics meets sustainability: the effect of independent directors' political connections on corporate sustainability disclosure in China. *Managerial Finance*, 50(1), 28-49.

Ting, P.-H. (2021). Do large firms just talk corporate social responsibility?-The evidence from CSR report disclosure. *Finance Research Letters*, 38, 101476.

Topple, C., Donovan, J. D., Masli, E. K., & Borgert, T. (2017). Corporate Sustainability Assessments: MNE engagement with sustainable development and the SDGs. *Transnational Corporations*, 24(3), 61.

Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158, 1533-1543.

- Tuo, L., Han, S., Rezaee, Z., & Yu, J. (2023). What does corporate sustainability reporting imply to auditors? Evidence from going-concern opinions and discretionary accruals. *International Journal Of Accounting And Information Management*, 31(3), 478-503.
- Ullmann, A. A. (1985). Data in Search of a Theory: A Critical Examination of the Relationships among Social Performance, Social Disclosure, and Economic Performance of U. S. Firms. *The Academy of Management review*, 10(3), 540-557.
- UN Department of Economic and Social Affairs Sustainable Development. (2017). *SDGs Local Monitoring – China's Pilot Practice*. <https://sdgs.un.org/partnerships/sdgs-local-monitoring-chinas-pilot-practice>
- UNDP. (2016). *Sustainable Development Goals in motion: China's progress and the 13th Five-Year Plan*.
- UNDP (China). (2017). *Report on the Sustainable Development of Chinese enterprises overseas*.
- UNGC, & GRI. (2017). *Business reporting on the SDGs: An analysis of the goals and targets*. https://www.globalreporting.org/media/v5milwee/gri_ungc_business-reporting-on-sdgs_analysis-of-goals-and-targets.pdf
- UNGC, & GRI. (2018). *Integrating the SDGs into corporate reporting: A practical guideline*.
- United Nations. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. United Nations: New York, NY, USA. <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>
- United Nations. (2018). *Reporting on the SDGs - shape the future of corporate reporting on the SDGs*. <https://www.unglobalcompact.org/take-action/action>
- United Nations. (2021). *Tier Classification for Global SDG indicators*.
- Uyar, A., Gerged, A. M., Kuzey, C., & Karaman, A. S. (2024). Do CSR performance and reporting facilitate access to debt financing in emerging markets? The role of asset structure and firm performance. *Review of Accounting and Finance*, 23(2), 157-185.
- Van Der Heijden, H. (2022). Predicting industry sectors from financial statements: An illustration of machine learning in accounting research. *The British Accounting Review*, 54(5), 101096.
- van der Waal, J. W., & Thijssens, T. (2020). Corporate involvement in Sustainable Development Goals: Exploring the territory. *Journal of Cleaner Production*, 252, 119625.
- van der Waal, J. W., Thijssens, T., & Maas, K. (2021). The innovative contribution of multinational enterprises to the Sustainable Development Goals. *Journal of Cleaner Production*, 285, 125319.
- van Zanten, J. A., & van Tulder, R. (2020). Towards nexus-based governance: defining interactions between economic activities and Sustainable Development Goals (SDGs). *International Journal Of Sustainable Development And World Ecology, ahead-of-print*(ahead-of-print), 1-17.

- Varshney, N., & Singh, S. (2024). Enhancing Diabetes Prediction: A comparative analysis of Train-Test Split and Stratified 10-Fold Cross-Validation with SMOTE Integration. *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*.
- Vel, S. S. (2021). Pre-processing techniques of text mining using computational linguistics and python libraries. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*,
- Velte, P., Stawinoga, M., & Lueg, R. (2020). Carbon performance and disclosure: A systematic review of governance-related determinants and financial consequences. *Journal of Cleaner Production*, *254*, 120063.
- Vourvachis, P., & Woodward, T. (2015). Content analysis in social and environmental reporting research: trends and challenges. *Journal of Applied Accounting Research*.
- Wan, J., Wang, J., Li, X., Qin, C., Qiang, Y., & Su, J. (2021). Strategic goals and implementation path of building a Beautiful China in 2035. *Chinese Journal of Environmental Management*, *13*(5), 29-36.
- Wang, B. Y., Duan, M., & Liu, G. (2021). Does the power gap between a chairman and CEO matter? Evidence from corporate debt financing in China. *Pacific-Basin Finance Journal*, *65*, 101495.
- Wang, F., Yang, S., Reisner, A., & Liu, N. (2019). Does green credit policy work in China? The correlation between green credit and corporate environmental information disclosure quality. *Sustainability*, *11*(3), 733.
- Wang, J., & Dong, Y. (2020). Measurement of text similarity: a survey. *Information*, *11*(9), 421.
- Wang, J., Ma, M., Dong, T., & Zhang, Z. (2023). Do ESG ratings promote corporate green innovation? A quasi-natural experiment based on SynTao Green Finance's ESG ratings. *International Review of Financial Analysis*, *87*, 102623.
- Wang, J., Song, L., & Yao, S. (2013). The determinants of corporate social responsibility disclosure: Evidence from China. *Journal of Applied Business Research (JABR)*, *29*(6), 1833-1848.
- Wang, J., Tian, G., Fan, W., & Luo, D. (2017). The effect of mandatory regulation on corporate social responsibility reporting quality: Evidence from China. *Journal of Applied Business Research*, *33*(1), 67-86.
- Wang, L., Kang, S., & Wu, H. (2021). Do Politically Connected Firms Pay Less Toward Environmental Protection? Firm-level Evidence from Polluting Industries in China. *Abacus*, *57*(2), 362-405.
- Wang, S., Zheng, J., & Tu, Y. (2023). The Communist Party of China embedded in corporate governance and enterprise value: Evidence from state-owned enterprises. *Finance Research Letters*, *54*, 103766.
- Wang, X., & Liu, Q. (2024). Information disclosure and ESG rating disagreement: Evidence from green bond issuance in China. *Pacific-Basin Finance Journal*, *85*, 102350.
- Wang, Y., Guo, C.-h., Chen, X.-j., Jia, L.-q., Guo, X.-n., Chen, R.-s., Zhang, M.-s., Chen, Z.-y., & Wang, H.-d. (2021). Carbon peak and carbon neutrality in China: Goals, implementation path and prospects. *China Geology*, *4*(4), 720-746.

- Wang, Z., Braam, G., Reimsbach, D., & Wang, J. (2020). Political embeddedness and firms' choices of earnings management strategies in China. *Accounting & Finance*, 60(5), 4723-4755.
- Wang, Z., Reimsbach, D., & Braam, G. (2018a). Political embeddedness and the diffusion of corporate social responsibility practices in China: A trade-off between financial and CSR performance? *Journal of Cleaner Production*, 198, 1185-1197.
- Wang, Z., & Sarkis, J. (2017). Corporate social responsibility governance, outcomes, and financial performance. *Journal of Cleaner Production*, 162, 1607-1616.
- Weber, O. (2014). Environmental, Social and Governance Reporting in China. *Business strategy and the environment*, 23(5), 303-317.
- Wilkerson, J., & Casas, A. (2017). Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, 20(1), 529-544.
- Wu, Y., Tian, Y., Chen, Y., & Xu, Q. (2022). The Spillover Effect, Mechanism and Performance of Green Bond Issuance. *Manag. World*, 38, 176-193.
- Wu, W., Wu, C., & Rui, O. M. (2012). Ownership and the value of political connections: Evidence from China. *European Financial Management*, 18(4), 695-729.
- Wu, W., Wu, C., Zhou, C., & Wu, J. (2012). Political connections, tax benefits and firm performance: Evidence from China. *Journal of Accounting and Public Policy*, 31(3), 277-300.
- Wu, X., & Hąbek, P. (2021). Trends in Corporate Social Responsibility Reporting. The Case of Chinese Listed Companies. *Sustainability (Basel, Switzerland)*, 13(8640), 8640.
- Wu, Y., Tian, Y., Chen, Y., & Xu, Q. J. J. M. W. (2022). The Spillover Effect, Mechanism and Performance of Green Bond Issuance. 38, 176-193.
- Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyperparameter optimization for credit scoring. *Expert systems with applications*, 78, 225-241.
- Xiao, C., Ye, J., Esteves, R. M., & Rong, C. (2016). Using Spearman's correlation coefficients for exploratory data analysis on big dataset. *Concurrency and Computation: Practice and Experience*, 28(14), 3866-3878.
- Xie, H., Wen, J., & Choi, Y. (2021). How the SDGs are implemented in China—A comparative study based on the perspective of policy instruments. *Journal of Cleaner Production*, 291, 125937.
- Xing, C., Zhang, Y., & Tripe, D. (2021). Green credit policy and corporate access to bank loans in China: The role of environmental disclosure and green innovation. *International review of financial analysis*, 77, 101838.
- Xinhua. (2013a). *The Central Organization Department issued the 'Opinions on Further Regulating the Part-time Work (Employment) of Party and Government Leading Cadres in Enterprises'*. <http://renshi.people.com.cn/n/2013/1030/c139617-23380219.html>

- Xinhua. (2013b). *The Organization Department of the CPC Central Committee further regulates the part-time employment of leading Party and government officials in enterprises*. Retrieved from https://www.gov.cn/govweb/jrzg/2013-10/31/content_2518437.htm
- Xinhua. (2015). *Vision and proposed actions outlined on jointly building Silk Road economic belt and 21st-Century Maritime Silk Road*. http://www.xinhuanet.com/gangao/2015-06/08/c_127890670.htm
- Xinhua. (2017). *Full text of President Xi's speech at opening of Belt and Road forum*. http://www.xinhuanet.com/english/2017-05/14/c_136282982.htm
- Xinhua. (2018). *China approves three demonstration zones for sustainable development*. <https://www.chinadaily.com.cn/a/201802/24/WS5a9145aea3106e7dcc13ddcf.html>
- Xu, L., Situ, H., Ren, J. C., & Yang, Y. (2025). Decoding CSR reporting on poverty alleviation in China: a symbolic power perspective. *Meditari Accountancy Research*.
- Xu, L., Wu, L., & Dong, S. (2021). Internal Mechanism and Quality of CSR Disclosure: Evidence from China. In *Green and Social Economy Finance* (pp. 49-70). CRC Press.
- Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of analysis and testing*, 2(3), 249-262.
- Yang, C., Jia, X., & Yang, C. (2023). The influence of political association and institutional environment on M&A performance: Based on empirical research of listed companies in China. *Journal of Liaoning Normal University (Natural Science Edition)*, 46(1), 19-25.
- Yi, H., & Liu, Y. J. G. E. C. (2015). Green economy in China: Regional variations and policy drivers. 31, 11-19.
- Yili. (2017). *2017 Social responsibility report*.
- Yili. (2018). *2018 Social responsibility report*.
- Yi, Y., Xie, B., Zhou, L., & Wei, Y. (2020). Does CSR affect the cost of equity capital: Empirical evidence from the targeted poverty alleviation of listed companies in China. *Plos One*, 15(2), e0227952.
- Yin, H., Li, M., Ma, y., & Zhang, Q. (2019). The relationship between environmental information disclosure and profitability: A comparison between different disclosure styles. *International Journal of Environmental Research and Public Health*, 16(1556).
- Yin, W. (2019). Integrating Sustainable Development Goals into the Belt and Road Initiative: would it be a new model for green and sustainable investment? *Sustainability*, 11(24), 6991.
- Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57(1), 247-265.

- Yu, L., Zhang, W., & BI, Q. (2019). Party organization embedding and corporate social responsibility. *Collected Essays on Finance and Economics*, 245(4), 61.
- Yu, S., Sial, M. S., Tran, D. K., Badulescu, A., Thu, P. A., & Sehleanu, M. (2020). Adoption and Implementation of Sustainable Development Goals (SDGs) in China—Agenda 2030. *Sustainability (Basel, Switzerland)*, 12(15), 6288.
- Yu, L., Zhang, W., & Bi, Q. (2019). 党组织嵌入与企业环境信息披露. *Journal of Jiangxi University of Finance and Economics*, 122.
- Yu, Y., & Chi, J. (2021). Political embeddedness, media positioning and corporate social responsibility: Evidence from China. *Emerging Markets Review*, 47, 100799.
- Zambrano-Monserrate, M. A., Ruano, M. A., & Sanchez-Alcalde, L. (2020). Indirect effects of COVID-19 on the environment. *Science of the Total Environment*, 728, 138813.
- Zeng, S., Xu, X., Yin, H., & Tam, C. M. (2012). Factors that drive Chinese listed companies in voluntary disclosure of environmental information. *Journal of Business Ethics*, 109(3), 309-321.
- Zhang, A. Y., & Zhang, J. H. (2023). Renovation in environmental, social and governance (ESG) research: the application of machine learning. *Asian Review of Accounting*.
- Zhang, G., & Guan, Z. (2020). GEE Chinese corporate SDGs information disclosure research report. In *GBEE Chinese Corporate Social Responsibility reporting research report (2019)* (pp. 352-367). Social Sciences Academic Press (CHINA).
- Zhang, H., Li, L., Fan, C., Hang, Z., & Khan, H. u. R. (2021). How does corporate party committee governance affect charitable donations? Evidence from heavy-pollution Industries in China. *Sustainability*, 13(21), 12242.
- Zhang, N. (2023). Implement the requirements of the Party Committee's pre-research and discussion to promote the integration of the Party's leadership into corporate governance in state-owned enterprises: Taking Sinopec National Exploration Company as an example. *Journal of Sinopec Management Institute*, 25(1).
- Zhang, Y., Wang, C., & Fang, J. (2023). A study on the impact of political connections on the investment efficiency of resource-based enterprises: based on the mediating effect of corporate tax avoidance. *Journal of Henan Institute of Science and Technology*, 43(5).
- Zhao, C., Guo, Y., Yuan, J., Wu, M., Li, D., Zhou, Y., & Kang, J. J. S. (2018). ESG and corporate financial performance: Empirical evidence from China's listed power generation companies. *10(8)*, 2607.
- Zhao, N., & Patten, D. M. (2016). An exploratory analysis of managerial perceptions of social and environmental reporting in China: Evidence from state-owned enterprises in Beijing. *Sustainability Accounting, Management and Policy Journal*, 7(1), 80-98.
- Zheng, L., Balsara, N., & Huang, H. (2014). Regulatory pressure, blockholders and corporate social responsibility (CSR) disclosures in China. *Social Responsibility Journal*.

- Zheng, S., He, C., Hsu, S.-C., Sarkis, J., & Chen, J.-H. (2020). Corporate environmental performance prediction in China: An empirical study of energy service companies. *Journal of Cleaner Production*, 266, 121395.
- Zhilkina, A. N., Karp, M. V., Bodiako, A. V., Smagulova, S. M., Rogulenko, T. M., & Ponomareva, S. V. (2022). Socially-oriented approach to financial risk management as the basis of support for the SDGs in entrepreneurship. *Risks*, 10(2), 42.
- Zhou, & Cui. (2019). Green Bonds, Corporate Performance, and Corporate Social Responsibility. *Sustainability*, 11(23), 6881.
- Zhu, H., Hui, K. N.-C., & Gong, Y. (2022). Uncovering the nonmarket side of internationalization: the Belt and Road Initiative and Chinese firms' CSR reporting quality. *Asia Pacific Journal of Management*, 1-29.
- Zhu, J., Spence, C., & Ezzamel, M. (2021). Thinking like the state: Doxa and symbolic power in the accounting field in China. *Accounting, Organizations and Society*, 93, 101235.
- Zhu, Q., & Zhang, Q. (2015). Evaluating practices and drivers of corporate social responsibility: The Chinese context. *Journal of Cleaner Production*, 100, 315-324.
- Zhu, R., & Liu, M. (2022). Regional environmental governance motivation, political connection, and corporate environmental violation penalties: New evidence from environmental supervision system. *The Theory and Practice of Finance and Economics*, 43(4), 129-137.
- Zimon, D., Tyan, J., & Sroufe, R. (2020). Drivers of sustainable supply chain management: Practices to alignment with un sustainable development goals. *International Journal for Quality Research*, 14(1).

Appendix

Appendix 1 A manual comparison of GRI, IR, and SASB to evaluate how these reporting frameworks can help measure SDG disclosures

This section manually compares how existing reporting frameworks, including GRI, IR, and SASB, can assist in observing SDG disclosures (See Table 8.1). The comparisons were conducted across several aspects, including environmental, social, human capital, business model and innovation, economic, as well as leadership and governance. Existing reporting frameworks are partially aligned with SDGs.

GRI indicators are partially aligned with the environmental, human capital, social and economic dimensions of SDG disclosures. It is observed that the GRI framework focuses on tradition ESG issues. They do not provide appropriate indicators for measuring the business model and innovation or leadership and governance aspects. Specifically,

1. Regarding environmental SDG targets, GRI indicators can provide a framework for measuring help observe SDG 3.9, SDG 6 (6.3, 6.6, 6a, and 6b), SDG 7 (7.2 and 7.3), SDG 12 (12.4, 12.5 and 12.2), SDG 13.1, and SDG 14.2.
2. Regarding human capital SDG targets, GRI indicators can help measure SDG 5.1, SDG 3 (3.3, 3.4, 3.5, 3.7), SDG 4 (4.3, 4.4, and 4.5), SDG 8 (8.5 and 8.8), SDG 10.3, and SDG 16 (16.1, 16.5, 16.10, and 16.3).
3. Regarding economic SDG targets, GRI indicators can be used to observe SDG 1.2, SDG 3.8, SDG 5.4, SDG 8 (8.1, 8.2 and 8.5), SDG 9 (9.1, 9.4 and 9.5) and SDG 10.4.

The SASB framework provides a more tangible guide related to business model and innovation and human capital, though little guidance is provided for measuring social and leadership and governance. The SASB framework does not have appropriate indicators for the environmental and economic dimensions of SDG disclosures. Specifically,

1. Regarding human capital SDG targets, the SASB framework can be used to observe SDG 1, SDG 3, SDG 5, SDG 8, SDG 10.
2. Regarding social SDG targets, the SASB framework can help observe SDG 1, SDG 2, and SDG 16 (16.1 and 16.10).
3. Regarding business model and innovation aspects, the SASB framework can help observe SDG 3, SDG 6, SDG 7, SDG 8, SDG 9, SDG 11, SDG 13, SDG 16 and SDG 12.

IR provides general guiding principles for business model and innovation, and leadership and governance in SDG disclosures.

Regarding business model and innovation, IR may guide SDG 4a, SDG 5b, SDG 7b, SDG 9 (9.5, 9c, 9.1, 9.4, and 9b), SDG 2a, SDG 3b, SDG 7 b, SDG 8.2, and SDG 14a.

Regarding social SDG targets, IR can also measure SDG 1a, SDG2b, SDG3 (3.3, 3.4, and 3.5), SDG 5 (5.1 and 5.5), SDG 6 (6.4, 6.5 and 6.6), SDG 7 (7.2 and 7.3), SDG 8 (8.5 and 8.8), SDG 9.4, SDG 10.3, SDG 12 (12.2, 12.4 and 12.5) and SDG 16 (16.4, 16.5, 16.6 and 16.7).

Table 8.1 How GRI, SASB, and IR can be used to operationalise SDGs

	GRI	SASB	IR
Environmental	SDG3.9 SDG 6.3; 6.6; 6a; 6b SDG 7.2; 7.3 SDG 8.4; SDG 12.2; 12.4; 12.5 SDG13.1 SDG 14.2		
Human capital	SDG 5.1 SDG 3.3; 3.4; 3.5; 3.7 SDG4.3; 4.4; 4.5 SDG8.5; 8.8 SDG 10.3; SDG 16.1; 16.7	SDG 1 SDG 3 SDG 5 SDG 8 SDG 10	
Social	SDG 16.1; 16.3; 16.5; 16.10 SDG 8.8	SDG 1.2 SDG 16.1; 16.10	
Business model and innovation	SDG 12.8	SDG 3 SDG 6 SDG 7 SDG 8 SDG 9 SDG 11 SDG 12 SDG 13 SDG 16	SDG4a SDG5b SDG7b SDG9.5; 9c SDG2a SDG3b SDG 7b SDG8.2 SDG9.1 (develop infrastructure); 9.4; 9.5; 9b SDG 14a
Economic	SDG1.2 SDG 3.8 SDG 5.4 SDG 8.1; 8.2; 8.5 SDG 9.1; 9.4; 9.5 SDG 10.4		
Leadership and governance		SDG 16	SDG1a SDG2b SDG3.3;3.4; 3.5 SDG 5.1; 5.5 SDG6.4; 6.5; 6.6 SDG7.2;7.3; SDG 8.5; 8.8 SDG9.4 SDG 10.3 SDG 12.2; 12.4; 12.5 SDG16.4; 16.5; 16.6; 16.7

Appendix 2 A list of sampled companies

Company Name	Code
PING AN BANK CO LTD	000001 (SZSE)
CHINA VANKE CO LTD	000002 (SZSE)
ZTE CORP	000063 (SZSE)
TCL TECHNOLOGY GROUP CORP	000100 (SZSE)
SHENWAN HONGYUAN GROUP CO LTD	000166 (SZSE)
MIDEA GROUP CO LTD	000333 (SZSE)
CCOOP GROUP CO LTD	000564 (SZSE)
BOE TECHNOLOGY GROUP CO LTD	000725 (SZSE)
CITIC GUOAN INFORMATION INDUSTRY CO LTD	000839 (SZSE)
WULIANGYE YIBIN CO LTD	000858 (SZSE)
NEW HOPE LIUHE CO LTD	000876 (SZSE)
SUNING.COM CO LTD	002024 (SZSE)
GOLDWIND SCIENCE & TECHNOLOGY CO LTD	002202 (SZSE)
HANGZHOU HIKVISION DIGITAL TECHNOLOGY CO LTD	002415 (SZSE)
TIANQI LITHIUM CORP	002466 (SZSE)
LUXSHARE PRECISION INDUSTRY CO LTD	002475 (SZSE)
BYD CO LTD	002594 (SZSE)
SHANGHAI PUDONG DEVELOPMENT BANK CO LTD	600000 (SHSZ)
CHINA MINSHENG BANKING CORP LTD	600016 (SHSZ)
SHANGHAI INTERNATIONAL PORT GROUP CO LTD	600018 (SHSZ)
BAOSHAN IRON & STEEL CO LTD	600019 (SHSZ)
CHINA PETROLEUM & CHEMICAL CORP	600028 (SHSZ)
CHINA SOUTHERN AIRLINES CO LTD	600029 (SHSZ)
CHINA MERCHANTS BANK CO LTD	600036 (SHSZ)
SDIC CAPITAL CO LTD	600061 (SHSZ)
CHINA EASTERN AIRLINES CORP LTD	600115 (SHSZ)
SHANGHAI FOSUN PHARMACEUTICAL GROUP CO LTD	600196 (SHSZ)
HAINAN AIRLINES HOLDING CO LTD	600221 (SHSZ)
ANHUI CONCH CEMENT CO LTD	600585 (SHSZ)
SINOPEC SHANGHAI PETROCHEMICAL CO LTD	600688 (SHSZ)
HAIER SMART HOME CO LTD	600690 (SHSZ)
HAITONG SECURITIES CO LTD	600837 (SHSZ)
SINOPEC OILFIELD SERVICE CORP	600871 (SHSZ)
INNER MONGOLIA YILI INDUSTRIAL GROUP CO LTD	600887 (SHSZ)
CHINA MERCHANTS SECURITIES CO LTD	600999 (SHSZ)

CHINA SHENHUA ENERGY CO LTD	601088 (SHSZ)
AIR CHINA LTD	601111 (SHSZ)
INDUSTRIAL BANK CO LTD	601166 (SHSZ)
CHINA RAILWAY CONSTRUCTION CORP LTD	601186 (SHSZ)
GUOTAI JUNAN SECURITIES CO LTD	601211 (SHSZ)
AGRICULTURAL BANK OF CHINA LTD	601288 (SHSZ)
PING AN INSURANCE GROUP CO OF CHINA LTD	601318 (SHSZ)
BANK OF COMMUNICATIONS CO LTD	601328 (SHSZ)
NEW CHINA LIFE INSURANCE COMPANY LTD	601336 (SHSZ)
CHINA RAILWAY GROUP LTD	601390 (SHSZ)
INDUSTRIAL AND COMMERCIAL BANK OF CHINA LTD	601398 (SHSZ)
METALLURGICAL CORPORATION OF CHINA LTD	601618 (SHSZ)
CHINA LIFE INSURANCE CO LTD	601628 (SHSZ)
HUATAI SECURITIES CO LTD	601688 (SHSZ)
SHANGHAI ELECTRIC GROUP CO LTD	601727 (SHSZ)
EVERBRIGHT SECURITIES CO LTD	601788 (SHSZ)
CHINA COMMUNICATIONS CONSTRUCTION CO LTD	601800 (SHSZ)
PETROCHINA CO LTD	601857 (SHSZ)
CHINA COAL ENERGY CO LTD	601898 (SHSZ)
ZIJIN MINING GROUP CO LTD	601899 (SHSZ)
YONGHUI SUPERSTORES CO LTD	601933 (SHSZ)
CHINA CONSTRUCTION BANK CORP	601939 (SHSZ)
CHINA NATIONAL NUCLEAR POWER CO LTD	601985 (SHSZ)
BANK OF CHINA LTD	601988 (SHSZ)
CMOC GROUP LTD	603993 (SHSZ)

Appendix 3 Keyword lists generated based on TF-IDF

SDG 1 bottom-up approach keywords with TF_IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
children	100.00	49	charity	21.76	6	recorded	29.20	7	summer	15.80	5
women	97.27	35	death	21.62	6	project	28.84	9	service	15.64	7
china	76.10	44	rmb	21.18	5	county	28.74	10	initiative	15.64	5
health	61.66	19	leukemia	20.85	3	launched	28.23	8	camp	15.52	4
poverty	61.62	31	teenagers	20.73	5	provide	28.01	11	including	15.37	5
alleviation	53.21	25	illness	20.30	6	disabled	27.09	8	crew	15.24	5
rural	53.19	19	foundation	20.28	6	province	26.52	7	zhouqu	15.09	4
areas	48.44	21	welfare	20.27	8	knowledge	25.25	8	development	15.00	5
donated	48.08	15	fund	20.04	6	poor	24.99	9	cancers	14.79	3
families	46.77	17	set	19.67	8	accidental	24.91	7	social	14.44	5
life	44.87	16	activities	19.62	7	plan	24.47	8	regions	14.19	4
insurance	44.29	16	growth	19.25	6	medical	24.21	7	security	14.09	4
air	41.97	18	earthquake	19.00	5	target	24.19	3	hebei	14.07	1
care	39.16	14	work	18.77	7	day	24.06	5	olympic	14.06	8
orphans	37.14	12	cancer	18.71	3	federation	23.98	6	four	13.99	3
branch	36.82	9	support	18.38	6	total	23.75	7	department	13.97	5
elderly	36.49	16	society	17.89	6	yuan	23.69	6	assistance	13.94	3
nutrition	35.88	10	lost	17.61	5	impoverished	23.49	8	nutritional	13.92	3
people	34.34	16	lectures	17.42	5	also	23.11	9	supported	13.91	3
love	34.22	15	team	17.23	7	aviation	22.63	6			
cards	31.91	8	welcome	17.18	2	employment	22.38	5			
files	31.91	8	targeted	17.15	4	improve	21.93	6			
million	30.82	9	new	17.00	4	employees	16.29	4			
public	29.55	12	class	16.83	4	safety	16.19	4			
students	29.45	10	maternal	16.66	4	lonely	15.96	5			
keyword	score	frequency	counties	16.36	4	child	13.90	3			

SDG 1 ecological impact keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
insurance	100.00	13	commercial	25.12	2	poor	26.39	5
poverty	86.12	18	life	25.12	4	means	25.59	2
alleviation	68.41	14	also	24.44	4	supporting	25.57	2
environmental	66.96	6	public	24.41	3	villages	25.20	4
China	59.79	10	projects	24.36	3	branch	25.20	4
protection	56.17	5	explore	24.23	1	aspects	18.70	1
relief	52.29	4	visit	24.23	1	expanded	18.70	1
green	43.01	3	programs	24.23	1	direct	18.70	1
academy	42.19	4	research	24.23	1	cover	18.70	1
forestry	42.19	4	investigation	24.23	1	etc	18.70	1
Chinese	42.19	4	advantages	23.27	2			
ecological	40.89	4	volunteer	21.49	1			
local	40.06	6	implement	21.49	1			
conduct	38.32	2	reduce	21.49	1			
million	35.32	7	support	21.49	1			
consumption	33.68	2	carry	21.49	1			
development	30.82	4	educational	21.49	1			
office	30.51	2	cause	21.49	1			
activities	30.51	2	resource	21.49	1			
ecology	30.25	3	electronation	21.49	1			
strategy	29.17	3	cultural	21.49	1			
actively	28.54	2	professional	21.49	1			
targeted	28.54	3	areas	20.95	2			
foundation	28.36	3	including	20.95	2			
company	28.29	4	environment	19.79	2			
right	27.60	2	charges	18.90	3			
banner	27.60	2	sales	18.90	3			
sonid	27.60	2	specific	18.70	1			

SDG 1 education keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
students	100.00	17	reduce	26.84	2	chengdu	36.01	2	jiangxi	19.77	3
program	76.34	16	stricken	26.84	2	zhengzhou	36.01	2	outstanding	19.77	3
stations	73.22	7	jobs	26.84	2	donated	36.01	2	launched	19.38	3
cnpc	72.22	13	buses	26.84	2	based	36.01	2	teacher	18.29	3
service	68.90	6	capabilities	26.84	2	sold	36.01	2	equal	17.24	2
rmb	63.19	7	burden	26.84	2	areas	35.45	7	opportunities	17.24	2
sichuan	62.01	7	help	26.03	3	teachers	34.99	6	senior	17.24	2
support	58.52	7	people	24.74	3	educational	33.44	5	supporting	16.94	2
henan	57.04	8	experience	24.41	2	counties	33.02	4	scholarships	16.94	2
school	56.21	10	oil	24.41	2	including	32.66	4	special	15.72	2
families	52.89	6	impoverished	23.89	2	alleviation	31.93	5	fund	15.72	2
high	52.39	9	development	23.77	2	initiatives	31.45	4	funded	15.72	2
complete	49.91	5	provinces	23.51	3	hand	31.45	4	community	15.72	2
studies	48.35	6	Beijing	23.34	4	student	29.56	3	provide	15.65	2
xuhang	45.17	6	Guizhou	23.34	4	set	29.11	3	jointly	15.65	2
public	44.31	7	schools	22.70	2	university	28.64	4	campus	13.88	1
poor	43.75	4	foundation	22.62	4	nearly	28.18	3	consumers	13.88	1
provided	43.37	3	project	21.81	2	platform	27.59	4	china	36.67	7
training	40.55	7	young	21.22	3	invested	27.22	2	liter	36.01	2
education	40.33	7	primary	20.28	1	financial	26.84	2	gasoline	36.01	2
poverty	38.85	6	reconstruction	20.28	1	examinees	26.84	2	every	36.01	2
million	38.66	4	pv	20.28	1	shuttle	26.84	2			
welfare	38.08	6	county	20.28	1	enhance	26.84	2			
college	38.07	4	power	20.28	1	integrated	20.28	1			

SDG 1 employment keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
poverty	100.00	50	courses	22.50	5	processing	33.95	8	tourism	17.88	5
training	89.48	32	rid	22.24	5	personnel	32.94	7	service	17.60	5
people	75.83	21	life	21.97	7	income	30.32	6	four	17.26	5
alleviation	69.42	31	family	21.82	6	college	30.13	9	cultivation	17.23	5
county	63.25	28	food	21.75	5	recruitment	29.12	7	programs	16.92	4
poor	59.63	21	realize	21.64	5	work	28.81	9	technicians	16.84	2
areas	55.52	22	enterprises	21.35	6	provided	28.56	6	entrepreneurship	16.64	4
employment	53.22	23	special	21.35	6	help	28.22	7	government	16.47	6
development	51.82	22	province	20.18	6	total	28.07	7	management	16.27	4
employees	50.60	17	employee	20.16	3	targeted	27.98	9	autonomous	16.20	4
women	49.55	19	including	20.03	6	xinjiang	27.95	8	leaders	16.11	3
cooperatives	49.10	12	company	19.83	7	industrial	27.56	9	technical	15.53	1
cadres	45.39	11	increased	19.55	4	bank	27.28	9	together	15.47	2
local	44.24	13	supported	19.45	4	building	26.84	8	youth	15.44	3
trained	42.65	7	program	19.38	5	two	26.61	6	parties	15.44	2
impoverished	40.83	10	health	19.11	4	care	25.77	7	collaborative	15.44	2
rmb	39.50	14	cooperative	19.08	4	students	25.70	8	construction	15.38	4
jobs	39.43	4	get	18.86	5	plants	25.23	4	economy	15.18	3
opportunities	39.30	14	managers	18.34	2	relief	25.21	4	billion	14.84	3
counties	37.70	12	grassroots	18.31	2	cnpc	24.89	9	need	14.75	3
created	37.46	6	air	18.13	4	party	24.86	7	chemical	14.53	3
support	37.23	11	invested	18.12	5	outstanding	24.57	5	workers	14.49	3
project	36.50	13	providing	18.02	3	qinghe	24.42	8	production	14.48	3
China	35.18	12	young	17.96	4	families	24.34	7	promoting	14.47	3
households	34.66	10	find	17.90	1	helped	24.28	2	rural	14.44	5

SDG 1 Education keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	Frequency	Keyword	score	Frequency
poverty	100.00	50	courses	22.50	5	college	30.13	9	cultivation	17.23	5
training	89.48	32	rid	22.24	5	recruitment	29.12	7	programs	16.92	4
people	75.83	21	life	21.97	7	work	28.81	9	technicians	16.84	2
alleviation	69.42	31	family	21.82	6	provided	28.56	6	entrepreneurship	16.64	4
county	63.25	28	food	21.75	5	help	28.22	7	government	16.47	6
poor	59.63	21	realize	21.64	5	total	28.07	7	management	16.27	4
areas	55.52	22	enterprises	21.35	6	targeted	27.98	9	autonomous	16.20	4
employment	53.22	23	special	21.35	6	xinjiang	27.95	8	leaders	16.11	3
development	51.82	22	province	20.18	6	industrial	27.56	9	technical	15.53	1
employees	50.60	17	employee	20.16	3	bank	27.28	9	together	15.47	2
women	49.55	19	including	20.03	6	building	26.84	8	youth	15.44	3
cooperatives	49.10	12	company	19.83	7	two	26.61	6	parties	15.44	2
cadres	45.39	11	increased	19.55	4	care	25.77	7	collaborative	15.44	2
local	44.24	13	supported	19.45	4	students	25.70	8	construction	15.38	4
trained	42.65	7	program	19.38	5	plants	25.23	4	economy	15.18	3
impoverished	40.83	10	health	19.11	4	relief	25.21	4	billion	14.84	3
rmb	39.50	14	cooperative	19.08	4	cnpc	24.89	9	need	14.75	3
jobs	39.43	4	get	18.86	5	party	24.86	7	chemical	14.53	3
opportunities	39.30	14	managers	18.34	2	outstanding	24.57	5	workers	14.49	3
counties	37.70	12	grassroots	18.31	2	qinghe	24.42	8	production	14.48	3
created	37.46	6	air	18.13	4	families	24.34	7	promoting	14.47	3
support	37.23	11	invested	18.12	5	helped	24.28	2	rural	14.44	5
project	36.50	13	providing	18.02	3	built	24.26	5	services	14.32	3
China	35.18	12	young	17.96	4	job	24.16	7	members	14.31	3
households	34.66	10	find	17.90	1	skills	23.89	8	entrepreneurial	14.31	3
processing	33.95	8	tourism	17.88	5	increase	23.22	5	education	14.27	4
personnel	32.94	7	service	17.60	5	needy	23.06	8	registered	14.05	4
income	30.32	6	four	17.26	5	million	22.94	8	foundation	14.02	4

SDG 1 healthcare keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
medical	100.00	59	free	23.53	8	recorded	28.34	11
poverty	91.47	64	tibet	22.97	9	foundation	26.42	8
insurance	81.24	45	xinjiang	22.52	9	healthcare	26.39	10
alleviation	72.65	45	branch	22.30	8	assistance	25.62	11
china	71.36	45	accidental	21.72	7	project	25.34	10
women	69.98	28	carry	21.30	6	charity	24.91	7
county	60.49	39	projects	21.12	7	local	24.81	9
health	58.00	25	designated	20.95	4	training	24.72	8
people	57.96	35	workers	20.33	3	education	24.71	9
life	55.84	30	patrol	20.31	3	two	24.30	7
rmb	52.91	22	province	20.27	9	total	23.72	5
illness	51.57	22	funds	20.01	6	carried	23.71	5
million	48.44	23	teachers	20.01	2	population	18.46	4
poor	44.23	24	guizhou	19.78	7	major	17.96	6
service	41.59	11	areas	19.75	8	cancer	17.76	4
donated	40.98	16	death	19.64	7	burden	17.31	5
critical	38.86	16	security	19.47	6	stations	17.29	4
children	36.54	17	autonomous	19.44	7	benefiting	17.05	4
counties	36.14	15	hospital	19.27	8	special	17.01	6
impoverished	34.61	14	treatment	19.23	8	regions	16.89	6
files	32.46	13	level	19.23	6	receive	16.88	5
patients	32.00	10	invested	19.21	6	development	16.66	5
cards	30.63	12	including	19.19	6	programs	16.16	4
care	30.53	13	bank	19.17	7	yuan	15.97	4
rural	30.26	11	providing	19.06	8			
relief	30.14	9	provided	18.93	9			
expenses	29.92	9	improve	18.83	7			
households	28.99	12	aid	18.48	5			

SDG 1 industrial development keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
poverty	100.00	151	households	16.54	11	service	24.26	14	employees	13.48	7
alleviation	83.33	105	resources	16.29	11	air	24.21	20	aid	13.47	7
agricultural	55.54	48	sold	16.26	4	sales	24.06	14	platforms	13.34	8
products	51.90	42	supporting	16.08	10	zhaoping	24.04	18	food	13.09	4
rmb	49.86	35	plant	16.06	6	enterprises	23.96	19	platform	13.06	9
county	49.75	49	benefiting	15.96	8	production	21.96	13	land	13.03	7
development	48.84	48	park	15.75	8	rural	20.93	11	animal	13.02	8
project	47.41	35	yuan	15.75	4	supply	20.82	12	husbandry	13.02	8
million	46.34	42	investment	15.70	7	sonid	20.78	15	reporting	12.85	7
China	45.34	42	promote	15.53	9	farmers	20.54	17	made	12.83	9
areas	43.04	34	worth	15.33	7	industry	20.12	9	education	12.79	7
tourism	40.77	21	period	15.33	10	help	20.11	10	residents	12.75	6
poor	40.32	37	model	15.27	8	irrigation	19.98	11	province	12.70	7
people	39.99	27	total	14.94	10	purchase	19.91	11	photovoltaic	12.68	4
local	39.28	36	loans	14.91	6	village	19.69	13	sell	12.67	5
industrial	39.15	32	town	14.82	9	processing	19.31	10	icbc	12.59	9
counties	37.30	25	planting	14.51	10	right	19.29	14	new	12.53	6
bank	35.61	30	autonomous	14.41	9	designated	19.27	7	business	12.47	8
construction	30.54	16	thousand	14.21	3	cooperation	19.08	11	company	12.47	10
relief	28.40	19	mall	14.19	7	Xinjiang	19.05	12	provide	12.40	5
training	27.64	17	international	14.08	6	including	18.93	15	cultural	12.32	6
projects	27.36	18	Guizhou	13.97	8	banner	18.65	13	provided	12.19	9
invested	26.75	18	infrastructure	13.95	7	ten	18.51	5	launched	12.15	9
support	25.76	20	regions	13.92	8	cnpc	18.18	12	specialties	12.13	3
loan	25.67	18	nearly	13.89	7	special	18.15	13	featured	12.08	5
life	25.35	14	safflower	13.82	7	targeted	18.14	13	merchants	12.00	8
power	24.88	12	completed	13.55	8	qinghe	17.85	11	expand	11.97	4
helped	24.85	13	income	13.52	9	insurance	17.83	10	travel	11.92	5

SDG 1 regional development keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
poverty	100.00	49	reconstruction	19.44	4	qinghe	23.09	5	township	16.15	3
alleviation	91.22	41	tibet	19.44	4	qinghai	22.61	5	activities	16.11	3
county	59.81	23	henan	19.44	4	branch	21.91	7	used	15.86	4
counties	48.23	14	relief	19.36	4	company	21.83	5	areas	15.76	4
local	47.15	16	promote	19.26	5	national	21.59	5	inner	15.06	3
autonomous	42.03	12	industry	19.19	5	drinking	21.41	5	mongolia	15.06	3
million	41.85	14	support	18.98	6	water	21.41	5	helped	14.27	3
targeted	40.75	11	households	18.24	4	one	21.39	4	precise	14.24	2
projects	40.07	12	donated	17.92	5	districts	21.32	5	sold	14.13	1
xinjiang	39.77	11	cadres	17.90	3	fanxian	21.28	5	purchased	14.13	1
China	37.86	11	special	17.76	5	directly	20.64	6	worth	14.13	1
rmb	37.81	12	build	17.75	4	countryside	20.62	5	products	14.13	1
village	35.26	10	agricultural	17.74	2	provinces	20.42	5	population	14.06	3
project	34.39	9	irrigation	17.43	3	municipalities	20.42	5	seven	14.06	3
including	33.93	10	officials	17.38	2	programs	19.67	3	develop	14.04	4
people	31.91	7	carried	17.31	3	chongqing	19.44	4	work	13.94	3
regions	28.00	7	province	17.15	3	jiangxi	19.44	4	healthcare	13.84	2
infrastructure	26.44	7	demonstration	17.10	4	continued	19.44	4	collaboration	13.84	2
invested	26.44	7	beautiful	17.10	4						
life	26.22	6	completed	16.97	3						
education	25.86	7	assistance	16.70	4						
guizhou	24.98	5	icbc	16.59	5						
training	24.24	5	measures	16.54	3						
aid	24.08	5	development	16.54	5						
villages	23.66	6	carry	16.45	3						
benefiting	23.51	5	annual	16.45	3						
funds	23.42	7	public	16.43	5						
industrial	23.28	6	implemented	16.17	3						

SDG 1 relocating keywords with TF-IDF

keyword	score	frequency
helped	100	2
courtyard	100	2
households	100	2
relocated	100	2
rmb	50	1
residence	50	1
trees	50	1
relocating	50	1
herdsmen	50	1
qinghe	50	1
planting	50	1
permanent	50	1
per	50	1
jimunai	50	1
impoverished	50	1
vegetables	50	1
household	50	1
average	50	1
benefit	50	1
future	50	1
fruit	50	1
finding	50	1
facilitated	50	1
etc	50	1
economic	50	1
earthworks	50	1
development	50	1
cut	50	1
crops	50	1
changing	50	1
building	50	1
xinjiang	50	1

SDG 1 social impact keywords with TF-IDF (Due to over 1000 words for this indicators, this table only shows keywords with the word frequency more than 3 times).

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
poverty	100.00	141	rural	10.19	6	approach	6.11	3
alleviation	84.46	112	services	8.22	6	area	6.22	3
china	66.22	85	set	7.75	6	august	4.64	3
air	46.48	48	society	9.75	6	aviation	6.02	3
county	42.25	41	southwest	10.05	6	battle	4.96	3
targeted	35.66	28	special	8.45	6	beichuan	5.46	3
areas	30.16	27	staff	7.17	6	brought	3.95	3
local	34.13	27	township	11.55	6	built	5.74	3
people	29.72	27	villages	14.50	6	campaign	5.49	3
village	38.21	24	visited	9.11	6	capital	9.54	3
life	26.05	19	advantages	9.63	5	case	3.51	3
million	29.36	19	agricultural	9.88	5	charge	4.88	3
rmb	29.05	19	chinese	6.54	5	chongqing	6.73	3
counties	29.97	18	class	10.49	5	concept	4.89	3
children	23.57	17	community	6.59	5	continued	6.73	3
zhaoping	22.09	17	company	10.32	5	cooperated	7.03	3
work	24.60	15	environment	9.70	5	cooperative	8.68	3
activities	25.70	14	facilities	11.93	5	courtyard	6.79	3
banner	18.98	14	funds	7.98	5	cultural	6.91	3
cadres	26.12	14	guizhou	11.99	5	culture	4.10	3
charity	24.80	14	helped	9.75	5	danjiangkou	6.29	3
development	20.14	14	implementation	9.79	5	deep	5.41	3
designated	20.83	13	infrastructure	9.87	5	department	3.66	3
guangxi	17.41	13	international	9.35	5	develop	5.94	3
love	16.41	13	mading	8.83	5	difficulties	9.84	3
party	21.65	13	milk	9.71	5	directly	5.63	3
support	18.46	13	order	8.51	5	districts	6.73	3
assistance	17.53	12	participate	9.51	5	donations	7.49	3
foundation	22.13	12	places	10.47	5	ecological	6.82	3
group	15.80	12	plan	7.35	5	economic	5.80	3
help	17.88	12	president	8.96	5	effectiveness	6.68	3
province	19.91	12	primary	7.16	5	emergency	6.70	3
public	20.96	12	quality	6.54	5	etc	5.01	3

yili	23.36	12	responsibilities	9.04	5	every	3.44	3
projects	16.53	11	social	12.46	5	five	5.13	3
training	20.00	11	southern	8.20	5	flight	3.23	3
autonomous	14.82	10	supporting	11.86	5	free	6.01	3
branch	12.48	10	transportation	5.73	5	fulfilled	3.82	3
construction	17.22	10	used	9.11	5	gave	6.76	3
donated	19.48	10	walk	6.57	5	glory	3.72	3
education	14.73	10	world	8.35	5	good	3.02	3
inner	14.18	10	youth	8.95	5	henan	6.73	3
mongolia	14.18	10	yuan	9.82	5	high	3.09	3
regions	19.72	10	achieving	6.47	4	housing	5.98	3
right	14.65	10	adhering	8.13	4	hubei	6.29	3
sonid	14.65	10	assisted	7.84	4	implemented	5.82	3
volunteers	16.54	10	athletes	3.00	4	income	5.76	3
xinjiang	16.56	10	bank	6.80	4	increase	8.64	3
measures	15.28	9	benefiting	8.85	4	invited	6.62	3
poor	15.38	9	building	7.47	4	jiangkou	4.35	3
practical	12.58	9	cadre	11.49	4	jiangsu	4.26	3
project	13.16	9	city	9.14	4	jiangxi	6.73	3
welfare	17.86	9	collaboration	6.15	4	jinping	5.43	3
actively	12.65	8	completed	12.78	4	joined	5.23	3
basketball	9.42	8	cpc	6.81	4	june	5.24	3
committee	11.28	8	cross	9.28	4	leadership	5.28	3
games	5.99	8	end	6.70	4	learning	4.89	3
invested	13.45	8	general	4.52	4	leveraging	6.63	3
liangshan	15.46	8	get	8.23	4	little	5.28	3
living	17.54	8	go	9.13	4	longzhou	6.29	3
provide	10.49	8	groups	5.55	4	major	6.66	3
sichuan	11.10	8	hand	6.61	4	making	6.73	3
ability	14.68	7	health	10.61	4	mutual	5.38	3
activity	11.20	7	improvement	8.35	4	natural	4.86	3
also	9.60	7	increasing	8.72	4	nearly	6.65	3
carried	10.25	7	industry	8.96	4	need	6.08	3
central	9.40	7	investigation	10.70	4	newly	10.93	3
cooperation	20.22	7	launched	5.46	4	office	4.43	3
efforts	15.68	7	located	5.67	4	officials	8.95	3

employees	12.86	7	made	4.99	4	part	6.66	3
focus	12.99	7	medical	9.13	4	pass	5.10	3
households	13.50	7	mission	5.80	4	play	7.28	3
improve	13.89	7	municipalities	7.57	4	promote	6.04	3
insurance	13.37	7	new	6.85	4	promoting	5.66	3
key	10.62	7	organizing	7.32	4	provided	5.83	3
leaders	14.00	7	personnel	6.74	4	purchase	5.99	3
olympic	5.24	7	prefecture	7.84	4	reconstruction	6.73	3
outstanding	12.27	7	professional	6.66	4	region	5.04	3
products	12.27	7	programs	11.07	4	remote	4.68	3
provinces	11.35	7	qinghai	8.74	4	research	9.03	3
relief	20.63	7	red	9.43	4	resources	5.53	3
school	9.78	7	road	7.83	4	responsibility	3.21	3
service	9.42	7	schools	6.51	4	rid	5.39	3
team	6.38	7	seconded	7.43	4	rio	2.25	3
total	13.31	7	secretary	5.92	4	room	6.91	3
two	14.69	7	sent	11.60	4	safe	4.75	3
aid	11.32	6	situation	9.78	4	secretaries	14.37	3
beijing	9.70	6	sports	3.86	4	seven	6.73	3
carry	11.74	6	teenagers	10.06	4	since	5.73	3
elderly	4.81	6	thereby	6.43	4	six	10.23	3
first	13.69	6	three	7.94	4	smooth	2.94	3
full	10.23	6	tibet	7.96	4	strategy	4.89	3
including	10.40	6	together	5.48	4	stricken	5.18	3
industrial	10.40	6	villagers	6.56	4	summer	4.64	3
leading	9.02	6	warmth	4.71	4	system	5.59	3
members	9.51	6	well	5.29	4	taking	6.35	3
national	8.53	6	young	9.27	4	talents	5.53	3
needs	10.87	6	yunxi	8.30	4	tiandeng	6.29	3
organized	11.03	6	actions	4.59	3	town	6.71	3
use	6.32	3	ways	4.72	3	towns	6.66	3
view	5.68	3	within	4.25	3	trained	9.06	3
voluntary	4.76	3	worth	6.03	3	transport	8.02	3
volunteer	14.87	3	xi	5.43	3	university	6.42	3

SDG 1 spending keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
poverty	100.00	164	bank	17.24	16	local	19.57	25	yuan	20.93	14
alleviation	91.72	124	benefiting	16.95	10	support	19.14	24	agricultural	20.90	15
rmb	88.42	124	households	16.83	13	company	18.96	20	thousand	20.75	4
million	73.43	117	charity	16.75	15	xinjiang	18.90	15	impoverished	20.18	17
china	50.16	75	files	16.74	17	development	18.51	21	province	20.17	20
county	48.93	60	products	16.50	10	including	18.50	17	fund	20.03	13
total	44.77	34	zhaoping	16.33	10	service	18.40	14	air	13.92	14
insurance	41.97	52	party	16.30	11	autonomous	18.28	13	families	13.87	14
people	36.73	42	assistance	16.27	13	construction	17.76	12	compensation	13.87	15
life	35.87	47	patients	16.21	10	help	17.41	20	yili	13.58	10
investment	33.51	14	loans	15.98	9	activities	17.36	14	tibet	13.44	9.00
targeted	33.33	32	loan	15.79	13	qinghe	12.94	11	program	13.42	14.00
counties	32.81	33	illness	15.68	16	since	12.88	7	village	12.98	10
invested	31.91	33	cny	15.62	5	value	12.88	6			
donated	30.87	29	implemented	15.41	11	carry	12.74	9			
project	29.06	34	cards	15.06	15	public	12.70	13			
areas	28.25	31	amount	14.98	6	industrial	12.69	12			
foundation	28.04	30	guangxi	14.96	14	free	12.42	7			
medical	27.35	24	aid	14.88	10	women	12.39	6			
poor	27.21	34	children	14.47	13	scholarships	12.37	10			
students	25.33	23	recorded	14.31	14	provided	12.34	14			
education	23.58	25	community	14.30	3	living	12.19	7			
relief	23.35	22	provide	14.29	13	increase	12.18	5			
projects	23.21	18	critical	14.26	15	funds	20.02	17			
billion	23.18	22	carried	14.17	7	cnpc	19.97	21			
worth	22.46	16	funding	14.08	5	designated	13.05	9			

SDG 7 activities keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
energy	100.00	648	operation	17.17	49	million	19.45	69
gas	59.44	214	technology	16.62	43	tons	19.28	74
green	52.68	443	innovation	16.60	67	carbon	18.87	46
environmental	47.31	184	volume	16.52	28	cooperation	18.62	37
protection	46.33	176	national	16.28	45	per	17.97	35
reduction	46.14	120	promote	15.76	44	environment	17.76	50
consumption	45.80	96	meters	15.68	29	employees	11.74	29
oil	43.02	142	cubic	15.54	28	future	11.70	17
emission	42.62	116	industry	14.70	55	electricity	11.68	21
saving	40.99	140	efficiency	14.65	27	reporting	11.60	53
conservation	38.05	121	measures	14.34	35	efforts	11.54	33
development	37.46	201	fuel	14.26	35	reduce	11.39	28
management	34.18	129	utilization	14.20	42	projects	19.67	214
china	33.35	179	actively	14.13	55	supply	19.63	443
clean	32.08	117	year	14.10	75	increased	11.80	17
emissions	29.11	98	products	14.03	34	sustainable	11.79	30
natural	28.64	69	services	13.98	66			
bank	28.45	158	coal	13.53	37			
production	26.97	79	including	13.40	31			
company	25.84	108	investment	13.39	39			
domestic	24.76	49	equipment	13.29	32			
water	24.28	60	standard	12.98	44			
total	23.54	47	also	12.92	33			
new	22.49	74	ecological	12.86	92			
power	22.37	74	international	12.81	40			
air	21.33	67	service	12.80	31			
system	20.80	61	period	12.79	60			
project	20.63	83	construction	12.77	34			
resources	20.41	41	responsibility	12.65	30			
cnpc	19.84	52	proportion	12.44	13			
use	19.68	648	life	11.82	35			

SDG 7 spending or resource commitment keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency
investment	100.00	3	totalled	25.05	1
energy	88.02	29	power	24.46	5
green	81.21	19	outstanding	23.72	2
million	78.82	11	granted	23.72	2
rmb	76.75	8	increased	23.06	2
air	63.25	5	beginning	23.06	2
credit	53.16	10	reached	22.00	2
protection	47.55	7	balance	22.00	2
saving	47.09	10	support	21.50	6
vehicles	46.37	10	sectors	21.38	4
projects	42.58	6	China	19.69	5
invested	41.25	6	infrastructure	18.54	2
new	40.68	6	emission	18.45	2
period	40.17	5	charging	18.40	4
bank	39.78	8	hong	18.15	3
company	38.28	4	yili	18.01	3
bond	33.35	6	per	17.45	3
services	32.81	3	issued	17.26	4
reporting	32.77	4	qinghai	16.84	5
end	32.59	4	hydropower	16.84	5
environmental	32.40	4	development	16.44	4
clean	31.81	5	sustainable	16.34	3
domestic	31.28	4	management	16.34	3
resource	31.28	4	facilities	15.68	1
recycling	29.46	3	loans	15.68	1
ecological	29.46	3	public	15.68	1
reduction	28.82	2	corporate	15.68	1
industrial	28.74	6	generation	15.59	2
billion	28.25	5	environmentally friendly	14.84	1
transportation	28.15	4	came	14.84	1
vehicle	26.77	2	year	26.44	6

SDG7 energy consumption keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
consumption	100	195	investment	5.69	6	fleet	18.65	14	technology	5.4	5
energy	74.39	137	ecological	5.4	6	million	16.54	14	reporting	5.3	5
power	42.24	46	square	5.34	6	natural	13.46	14	relevant	5.01	5
gas	40.44	44	quality	5.33	6	efficiency	12.63	14	ensure	4.96	5
per	42.42	41	space	5.26	6	vehicle	23.64	13	material	4.92	5
fuel	43.2	32	projects	5.05	6	tons	12.61	13	areas	4.76	5
saving	29.89	31	rtk	12.12	5	electric	17.45	12	waste	4.75	5
office	32	28	policies	9.03	5	carbon	14.76	12	plan	4.59	5
green	20.59	27	thousand	8.44	5	reducing	11.98	12	achieved	4.59	5
china	20.05	26	cubic	8.11	5	operation	11.87	12	plant	4.39	5
emission	23.64	25	capacity	7.56	5	entire	10.66	12	financial	4.38	5
tonnes	29.69	24	paper	7.5	5	management	10.22	12	efforts	3.76	5
gasoline	25.54	23	electronic	6.87	5	ton	25.15	11	increase	12.12	4
water	18.42	22	promote	6.75	5	diesel	13.71	11	technologies	9.16	4
reduction	22.02	20	social	6.67	5	aircraft	13.01	11	mwh	7.87	4
use	17.26	19	employees	6.33	5	beijing	11.65	11	every	7.56	4
air	16.65	19	high	6.27	5	capita	10.4	11	resulting	6.35	4
oil	25.36	18	recycling	6.25	5	company	9.89	11	electricity	6.14	4
coal	19.33	18	resources	6.12	5	environment	9.84	11	introduced	6.07	4

environmental	15.72	18	awareness	6.07	5	data	9.68	11	performance	5.98	4
development	14.82	18	utilization	5.92	5	reduced	22.16	10	metres	5.93	4
total	20.34	17	production	5.89	5	new	12.05	10	proportion	5.81	4
standard	18.09	17	national	5.79	5	packaging	10.19	10	corporate	5.75	4
reduce	16.6	17	source	5.65	5	supply	10.01	10	fossil	5.74	4
bank	14.98	16	various	5.6	5	equipment	9.98	10	primary	5.74	4
kwh	17.66	15	units	5.5	5	protection	9.59	10	biogas	5.58	4
headquarters	12.02	15	lighting	5.46	5	system	9.43	10	intensity	5.58	4
head	20.92	14	time	5.45	5	group	8.58	10	vi	5.54	4
keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
mix	13.51	9	steadily	5.43	4	direct	7.49	6	cut	3.97	4
improve	10.27	9	center	5.36	4	made	6.9	6	hebei	3.53	4
vehicles	9.43	9	whole	5.17	4	work	6.72	6	area	3.45	4
improved	8.68	9	information	5.15	4	concept	6.39	6	results	3.39	4
year	8.18	9	lights	5.08	4	construction	6.35	6	factor	3.16	4
first	7.55	9	day	4.79	4	upgrading	6.32	6	belly	3.11	4
aviation	13.25	8	accounting	4.75	4	expected	6.14	6	points	3.11	4
comprehensive	11.93	8	pipeline	4.73	4	actively	6.04	6	percentage	3.11	4
equivalent	9.82	8	light	4.67	4						
yili	8.93	8	dioxide	4.55	4						

structure	8.13	8	overall	4.53	4						
measures	7.96	8	shanghai	4.49	4						
project	7.91	8	rmb	4.48	4						
materials	7.86	8	increased	4.43	4						
life	7.63	8	buildings	4.41	4						
building	7.27	8	mechanism	4.37	4						
emissions	7.1	8	optimized	4.35	4						
generation	6.5	8	thermal	4.3	4						
tce	23.34	7	friendly	4.3	4						
generated	9.39	7	people	4.25	4						
fuels	9.25	7	business	4.23	4						
according	7.53	7	end	4.19	4						
meter	7.4	7	local	4.12	4						
market	6.38	7	disclosed	4.09	4						
clean	6.04	7	services	4.07	4						
compared	5.02	7	civilization	4.06	4						
products	16.27	6	improving	4.05	4						
control	11.25	6	public	4.04	4						
unit	8.63	6	strengthened	4.03	4						
economic	7.91	6	actual	4.03	4						

SDG7 energy efficiency keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
energy	100.00	88	bank	13.43	5	technology	19.21	8	promoted	10.82	4
efficiency	62.91	49	weight	13.07	5	promote	17.36	8	innovation	10.79	4
fuel	60.04	30	water	13.04	5	vehicles	17.30	8	strengthen	10.70	4
gas	49.69	29	refining	12.89	5	operation	17.86	9	resource	12.20	4
china	41.85	25	chemical	12.80	5	environment	17.71	9	environmental	11.85	4
air	44.93	24	continuously	12.58	5	natural	17.01	9	vehicle	10.95	4
management	35.70	19	technological	12.57	5						
consumption	27.95	18	carried	12.30	5						
utilization	32.97	16	airlines	12.23	5						
green	28.80	15	structure	11.46	5						
aircraft	26.17	15	carbon	11.34	5						
improve	26.64	14	construction	11.23	5						
reduce	28.09	13	operating	11.03	5						
emission	25.94	13	data	10.74	5						
development	26.09	12	improving	10.38	5						
system	25.07	11	project	9.91	5						
conservation	21.37	11	level	9.81	5						
improved	23.88	10	special	9.56	5						
saving	23.64	10	capacity	8.36	5						
climate	23.17	10	load	7.95	5						
fleet	20.38	9	factor	7.71	5						
also	18.29	9	provide	13.85	4						

SDG7 renewable energy keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency
consumption	100.00	19	clean	16.18	5
gas	96.24	17	products	15.64	5
natural	81.39	11	utilization	14.44	4
kwh	67.84	9	actively	14.36	4
electric	64.35	8	provided	14.13	1
power	60.85	10	piles	14.13	1
energy	49.13	18	energies	13.35	4
electricity	34.20	2	renewable	13.07	4
reduce	30.14	5	solar	13.07	4
carbon	25.90	7	fuels	12.37	3
production	20.07	6	entire	12.28	2
per	19.60	5	group	12.28	2
emissions	18.22	5	areas	11.82	3
headquarters	18.20	3	cnpc	11.73	3
oilfield	17.09	4	projects	11.17	3
charging	16.61	2	year	10.87	3
geothermal	16.55	5	structure	10.52	3

SDG 13 climate change activities keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
energy	100.00	207	branch	7.43	8	investigation	8.55	3
green	58.89	148	products	7.43	8	offices	8.00	3
emissions	77.53	123	high	7.35	8	charity	7.68	3
emission	75.88	108	airlines	7.05	8	electric	5.88	3
china	42.15	83	cnpc	6.88	8	wastewater	5.54	3

environmental	58.63	82	domestic	6.74	8	philosophy	5.51	3
reduction	69.47	79	plant	6.69	8	cut	5.46	3
development	36.63	74	promoting	6.68	8	disposal	5.28	3
protection	53.47	67	mitigate	6.54	8	aircrafts	5.10	3
saving	47.83	60	process	6.37	8	saved	4.99	3
gas	39.33	59	progress	6.32	8	tion	4.98	3
bond	11.65	59	mainly	6.29	8	emisslon	4.97	3
air	33.04	58	implemented	5.66	8	ensure	4.93	3
bank	34.41	57	recognized	5.41	8	investigations	4.84	3
management	33.43	52	framework	3.69	8	serving	4.77	3
carbon	41.38	48	strengthening	3.05	8	con	4.72	3
per	32.91	48	indexconstituents	0.86	8	community	4.67	3
conservation	52.34	45	underwriter	0.86	8	innovating	4.63	3
waste	33.94	43	changes	21.94	7	reuse	4.63	3
new	28.11	39	action	14.69	7	related	4.63	3
production	26.06	38	integrated	11.35	7	consideration	4.53	3
system	26.91	37	dioxide	10.37	7	recycling	4.53	3
climate	52.30	36	materials	10.34	7	capacity	4.48	3
power	26.05	33	public	8.50	7	general	4.41	3
data	21.27	32	continued	8.45	7	forward	4.41	3
greenhouse	29.30	31	requirements	8.44	7	methods	4.38	3
total	19.56	31	head	8.33	7	regularly	4.36	3
hazardous	23.21	30	set	8.28	7	manage	4.29	3
company	20.71	30	indirect	7.99	7	training	4.27	3
consumption	25.34	28	cubic	7.84	7	constantly	4.24	3
aviation	23.35	27	meters	7.84	7	transportation	4.19	3
actively	20.08	26	prevention	7.70	7	petrochemical	4.17	3
unit	18.98	26	significant	7.57	7	characteristics	4.16	3

intensity	18.30	26	future	7.17	7	paris	4.16	3
operation	29.08	25	analysis	7.16	7	manner	4.10	3
water	23.83	25	supply	7.15	7	principle	4.07	3
project	23.25	25	business	7.12	7	organizational	4.02	3
reduce	24.86	24	paperless	6.97	7	transactions	3.99	3
environment	18.69	24	order	6.97	7	guiding	3.95	3
appropriate	17.55	24	implementation	6.87	7	participate	3.93	3
beijing	20.67	23	monitoring	6.70	7	concepts	3.91	3
measures	17.52	23	industries	6.65	7	red	3.89	3
volume	16.63	23	number	6.57	7	calculated	3.79	3
facility	16.59	23	equipment	6.40	7	roadmap	3.78	3
information	15.86	23	quality	6.25	7	packaging	3.77	3
tonnes	16.62	22	reform	5.95	7	target	3.77	3
chinese	4.86	22	industrial	5.54	7	measuring	3.73	3
employees	18.54	21	handled	5.49	7	issued	3.73	3
efforts	18.48	21	initiatives	5.43	7	energies	3.69	3
projects	17.66	21	assets	3.41	7	paper	3.69	3
fuel	17.48	21	issuance	2.69	7	detail	3.68	3
billion	9.79	20	along	2.69	7	paid	3.64	3
among	9.73	20	provide	2.61	7	factors	3.64	3
resources	24.15	19	september	2.49	7	mountains	3.58	3
ecological	14.95	19	accumulatively	1.94	7	waters	3.58	3
national	13.67	19	second	1.61	7	le	3.58	3
investment	12.75	19	continuously	11.06	6	publicity	3.56	3
office	20.13	18	rmb	10.05	6	entered	3.54	3
use	19.44	18	welfare	9.67	6	market	3.52	3
relevant	14.58	18	years	9.29	6	standardize	3.52	3
produced	12.10	18	strategy	8.97	6	team	3.48	3

international	9.59	18	carry	8.53	6	party	3.44	3
clean	15.07	17	active	7.72	6	every	3.43	3
ground	13.34	17	standard	7.49	6	ogci	3.43	3
standards	13.05	17	reduced	7.38	6	accelerating	3.41	3
tons	19.91	16	centralized	6.97	6	practice	3.41	3
promote	13.91	16	ment	6.94	6	rolled	3.40	3
life	13.65	16	discharges	6.88	6	local	3.39	3
services	12.93	16	relating	6.88	6	actions	3.37	3
included	11.36	16	strictly	6.63	6	trainings	3.37	3
first	6.76	16	promotion	6.56	6	areas	3.35	3
first	2.38	16	performance	6.15	6	transformation	3.34	3
reducing	20.78	15	develop	6.15	6	upgrading	3.34	3
vehicles	17.66	15	cooperation	5.97	6	using	3.33	3
credit	16.28	15	loan	5.85	6	environmentally	3.31	3
million	16.20	15	award	5.82	6	guangdong	3.31	3
ghg	14.05	15	alleviation	5.79	6	kept	3.26	3
regulations	13.78	15	implement	5.73	6	bay	3.25	3
based	13.48	15	disclosure	5.67	6	world	3.24	3
reporting	12.78	15	time	5.49	6	parking	3.24	3
sustainable	12.43	15	beautiful	5.19	6	optimization	3.23	3
made	12.21	15	risks	5.09	6	river	3.23	3
achieved	11.21	15	win	5.00	6	delta	3.23	3
results	10.77	15	technological	4.95	6	ability	3.23	3
technologies	19.27	14	nancial	4.85	6	targeted	3.23	3
activities	15.24	14	completed	4.84	6	battles	3.20	3
pollution	15.14	14	well	4.75	6	right	3.19	3
policies	14.79	14	main	4.48	6	integrate	3.18	3
impact	14.45	14	via	1.93	6	base	3.16	3

support	14.14	14	deal	1.78	6	verified	3.15	3
work	10.96	14	raised	1.77	6	brand	3.14	3
financial	5.19	14	certified	1.66	6	person	3.12	3
change	33.84	13	aided	0.65	6	working	3.10	3
center	16.39	13	banking	0.65	6	thereby	3.10	3
concept	13.01	13	aiding	0.65	6	friendly	3.09	3
laws	11.86	13	aised	0.65	6	advance	3.08	3
construction	11.78	13	bllion	0.65	6	hours	3.07	3
including	11.61	13	bondsand	0.65	6	asset	3.04	3
promoted	11.36	13	financigmarketdevelopment	0.65	6	optimized	3.04	3
include	11.28	13	chinesebankinginstitutions	0.65	6	regard	2.95	3
economic	10.18	13	make	9.28	5	technical	2.93	3
description	9.84	13	solid	8.96	5	outstanding	2.91	3
oil	9.11	13	adhere	7.93	5	kong	2.91	3
technology	8.25	13	treatment	7.90	5	earnestly	2.91	3
belt	4.52	13	fuels	7.64	5	strengthened	2.91	3
yili	18.50	12	establishing	7.00	5	mode	2.88	3
efficiency	17.93	12	accounting	6.78	5	took	2.87	3
industry	15.59	12	fleet	6.51	5	adheres	2.86	3
service	11.47	12	priority	6.40	5	yet	2.85	3
scope	11.03	12	effectively	6.09	5	disclosed	2.85	3
economy	11.02	12	counting	5.85	5	heating	2.85	3
social	10.35	12	gri	5.71	5	developed	2.84	3
year	9.47	12	insurance	5.66	5	supported	2.84	3
natural	9.24	12	mechanism	5.59	5	estimated	2.83	3
application	8.49	12	currently	5.46	5	online	2.82	3
report	7.45	12	save	5.43	5	renewable	2.81	3
corporate	6.84	12	administration	5.40	5	always	2.81	3

making	6.33	12	improved	5.39	5	normal	2.80	3
improve	11.25	11	end	5.39	5	space	2.79	3
generation	10.99	11	continue	5.26	5	surface	2.78	3
area	9.77	11	education	5.22	5	stable	2.78	3
enterprises	9.38	11	inspection	5.22	5	ways	2.77	3
plan	9.34	11	used	5.21	5	round	2.76	3
people	9.21	11	shall	5.20	5	terminals	2.75	3
aircraft	7.98	11	unified	5.17	5	kwh	2.74	3
issuer	7.06	11	one	5.14	5	structural	2.74	3
global	5.86	11	calculation	5.13	5	pipeline	2.74	3
balance	3.49	11	leading	5.10	5	di	2.72	3
institution	2.70	11	road	5.07	5	meet	2.69	3
best	2.34	11	healthy	5.04	5	supplied	2.64	3
bonds	1.84	11	ameco	5.02	5	initiative	2.64	3
utilization	17.58	10	whole	4.95	5	state	2.62	3
systems	10.31	10	airport	4.94	5	kunming	2.61	3
compliance	10.01	10	leadership	4.92	5	atmosphere	2.59	3
wastes	9.94	10	strong	4.71	5	focus	2.59	3
units	9.79	10	nance	4.23	5	help	2.57	3
electricity	9.41	10	important	4.19	5	rights	2.55	3
low	9.31	10	disabled	4.09	5	increased	2.54	3
discharge	8.93	10	external	4.08	5	cargo	2.53	3
special	8.77	10	demand	4.03	5	introduction	2.51	3
control	8.58	10	status	3.92	5	effective	2.46	3
responsibility	8.47	10	achievements	3.83	5	led	2.45	3
campaign	8.44	10	limited	3.77	5	platform	2.45	3
respective	8.10	10	interests	3.51	5	employment	2.44	3
procurement	7.11	10	call	1.67	5	effect	2.42	3

institutions	5.54	10	duing	1.63	5	create	2.40	3
finance	4.53	10	opinion	0.54	5	lead	2.32	3
excellent	2.19	10	themed	0.54	5	network	2.32	3
underwrote	1.74	10	ol	0.54	5	supporting	2.20	3
land	10.73	9	oversubscribed	0.54	5	restructuring	2.18	3
building	10.10	9	latest	0.54	5	exploration	2.17	3
build	9.69	9	alignment	0.54	5	accounts	2.14	3
calibre	9.40	9	ranking	0.54	5	improving	8.52	8
civil	9.11	9	shading	0.54	5	responsible	7.83	8
poverty	8.84	9	received	0.54	5	direct	7.74	8
group	8.75	9	receiving	0.54	5	key	7.48	8
sky	7.62	9	single	0.54	5			
blue	7.62	9	esg	0.54	5			
types	7.28	9	euros	0.54	5			
apu	6.83	9	climatic	10.40	5			
carrying	5.27	9	addressing	9.81	5			
research	3.89	9	substances	2.13	5			
amount	3.64	9	consumers	2.10	5			
recording	0.97	9	better	2.10	5			
comprehensive	10.20	8	real	2.08	3			
carried	10.15	8	radiation	2.03	3			
series	9.80	8	shale	1.81	3			
according	9.62	8	period	7.47	8			
also	9.40	8						
guidelines	8.62	8						
operations	8.57	8						

SDG 13 climate change investment keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
control	51.42	7	policy	10.05	1	qhse	24.00	2
company	41.29	7	yili	8.99	1	prevention	24.00	2
treatment	57.06	6	wildlife	8.99	1	risks	24.00	2
vehicles	46.05	6	working	8.99	1	invested	23.52	2
china	38.01	6	pipes	8.99	1	comprehensive	23.48	2
environmental	90.28	5	petroleum	8.99	1	pollution	20.32	2
protection	86.34	5	suspended	8.99	1	low	20.09	2
emission	36.35	5	praised	8.99	1	improve	20.09	2
investment	100.00	4	significant	8.99	1	facilities	18.66	2
exhaust	48.21	4	steel	8.99	1	th	17.99	2
mechanism	37.24	4	starlings	8.99	1	xinjiang	17.99	2
labor	24.88	4	starling	8.99	1	construction	17.99	2
female	24.88	4	spent	8.99	1	rosy	17.99	2
people	24.88	4	species	8.99	1	site	17.99	2
rmb	49.09	3	tens	8.99	1	birds	17.99	2
air	36.15	3	shining	8.99	1	ecological	17.73	2
carbon	30.14	3	prefecture	8.99	1	million	16.66	2
high	30.14	3	shelters	8.99	1	implement	15.39	2
national	24.27	3	thousands	8.99	1	resolutely	15.39	2
terms	23.14	3	scientific	8.99	1	new	15.35	2
vehicle	23.03	3	undisturbed	8.99	1	diesel	15.35	2
energy	21.45	3	putting	8.99	1	charging	15.35	2
strictly	20.69	3	protected	8.99	1	pace	15.35	2
promote	20.64	3	social	8.99	1	phasing	14.97	2
law	18.66	3	young	8.99	1	values	12.85	2
republic	18.66	3	office	8.99	1	cnpc	12.85	2
employees	18.66	3	autonomous	8.99	1	employment	12.44	2
relevant	16.99	3	green	8.99	1	opportunities	12.44	2
thousand	42.05	1	expressway	8.99	1	regulations	12.44	2
ten	42.05	1	highway	8.99	1	always	12.44	2
reconstruction	15.79	1	home	8.99	1	equal	12.44	2

sludge	15.79	1	example	8.99	1	operation	7.68	1
covering	15.79	1	engineering	8.99	1	dpf	7.68	1
plants	15.79	1	economy	8.99	1	verified	7.68	1
service	15.79	1	ecology	8.99	1	device	7.68	1
neries	15.79	1	could	8.99	1	demonstration	7.68	1
oily	15.79	1	kazak	8.99	1	airports	7.68	1
stations	15.79	1	breed	8.99	1	compliance	7.68	1
ecoenvironmental	15.79	1	bird	8.99	1	capital	7.68	1
storage	15.79	1	made	8.99	1	capability	7.68	1
governance	15.79	1	management	8.99	1	results	7.68	1
oil	15.79	1	build	8.99	1	bottlenecks	7.68	1
underground	15.79	1	migrating	8.99	1	base	7.68	1
wastewater	15.79	1	mesh	8.99	1	application	7.68	1
antiseepage	15.79	1	also	8.99	1	good	7.68	1
vocs	15.79	1	nurture	8.99	1	feasibility	7.68	1
typically	15.79	1	along	8.99	1	accelerated	7.68	1
upgrading	15.79	1	may	8.99	1	planned	7.68	1
billion	15.79	1	move	8.99	1	invest	7.68	1
standard	12.05	1	activities	8.99	1	insufficient	7.68	1
effectively	12.05	1	actively	7.68	1	install	7.68	1
followed	12.05	1	airport	7.68	1	infrastructure	7.68	1
reducing	12.05	1	special	7.68	1	pollutant	7.68	1
transforming	12.05	1	various	7.68	1	guaranteed	7.68	1
improving	12.05	1	achieved	7.68	1	achieve	7.68	1
efficiency	12.05	1	electric	7.68	1	projects	7.68	1
blue	12.00	1	fuel	7.68	1	overcame	7.68	1
program	12.00	1	full	7.68	1	maternity	6.22	1
pure	12.00	1	generally	7.68	1	child	6.22	1
battle	12.00	1	mechanisms	10.05	1	rst	6.22	1
investigation	12.00	1	gender	6.22	1	guarantee	6.22	1
audit	12.00	1	religion	6.22	1	safeguards	6.22	1
production	12.00	1	career	6.22	1	forced	6.22	1

key	12.00	1	prohibit	6.22	1	senior	6.22	1
land	12.00	1	trade	6.22	1	regardless	6.22	1
prevent	12.00	1	cally	6.22	1	fair	6.22	1
safe	12.00	1	lawful	6.22	1	executives	6.22	1
capacity	12.00	1	ts	6.22	1	race	6.22	1
fect	12.00	1	nationali	6.22	1	ethnic	6.22	1
strengthen	12.00	1	parental	6.22	1	equality	6.22	1
year	12.00	1	ty	6.22	1	enjoy	6.22	1
continue	12.00	1	aim	6.22	1	speci	6.22	1
waters	12.00	1	ment	6.22	1	host	6.22	1
system	12.00	1	leave	6.22	1	ination	6.22	1
establish	12.00	1	pregnancy	6.22	1	end	6.22	1
thereby	12.00	1	union	6.22	1	em	6.22	1
skies	12.00	1	bene	6.22	1	puts	6.22	1
evaluation	12.00	1	chinese	6.22	1	provide	6.22	1
clear	12.00	1	background	6.22	1	develop	6.22	1
contributing	12.00	1	women	6.22	1	interests	6.22	1
time	10.05	1	managers	6.22	1	cultural	6.22	1
emissions	10.05	1	approved	6.22	1	countries	6.22	1
proper	10.05	1	policies	6.22	1	international	6.22	1
introduce	10.05	1	minorities	6.22	1	conventions	6.22	1
corporate	10.05	1	ployees	6.22	1	contract	6.22	1
targets	10.05	1	laws	6.22	1	development	10.05	1
consumption	10.05	1	rights	6.22	1	strategy	10.05	1
cost	10.05	1	compensation	6.22	1	major	10.05	1
assessment	10.05	1	got	6.22	1	demand	10.05	1
product	10.05	1	government	6.22	1	water	10.05	1

SDG 13 emission keywords with TF-IDF

keyword	score	frequency	keyword	score	frequency	keyword	score	frequency
emission	100.00	36	emission	4.22	1	nations	2.21	1
emissions	85.32	37	scape	4.22	1	responsibilities	2.21	1
greenhouse	78.97	22	major	4.11	2	responded	2.21	1
gas	75.36	29	period	4.05	3	Katowice	2.21	1
reduction	69.26	29	limited	3.95	2	challenge	2.21	1
tonnes	58.68	12	units	3.89	1	commend	2.17	1
tons	48.65	24	capabilities	3.89	1	residents	2.17	1
total	48.32	16	also	3.89	1	shining	2.17	1
carbon	47.09	16	companies	3.89	1	circular	2.17	1
per	46.15	11	commitment	3.89	1	river	2.17	1
ton	44.19	5	successfully	3.89	1	teninskaya	2.17	1
ghg	36.89	6	included	3.89	1	volga	2.17	1
million	34.65	12	quota	3.89	1	advanced	2.17	1
nitrogen	29.73	8	past	3.84	2	connect	2.17	1
waste	24.27	7	production	3.84	2	started	2.17	1
China	22.87	16	accumulative	3.84	2	supported	2.17	1
reductions	22.55	5	domestic	3.83	2	Russia	2.17	1
equivalence	22.55	5	billion	3.83	2	sets	2.17	1
product	21.11	4	setting	3.83	3	structure	2.17	1
rtk	21.03	3	caron	3.81	1	commercially	2.17	1
group	20.27	5	turning	3.80	2	related	2.17	1
dioxide	20.07	5	property	3.80	2	importance	2.17	1
hazardous	19.90	4	formulate	3.80	2	generator	2.17	1
equivalent	19.81	6	work	3.80	2	equipped	2.17	1
green	19.33	16	clarify	3.80	2	pearl	2.17	1
energy	19.28	15	control	3.80	2	electric	2.17	1
entire	18.59	4	end	3.80	2	living	2.17	1
power	18.02	10	special	3.75	1	developing	2.17	1
trading	17.74	6	replacement	3.75	1	cooperation	2.17	1

office	17.25	5	vehicles	3.75	1	cut	2.17	1
air	16.78	7	traditional	3.75	1	large	2.17	1
climate	16.78	8	dongfeng	3.75	1	fueling	2.17	1
standard	16.64	9	present	3.75	1	moscow	2.17	1
kilometres	16.29	3	costs	3.75	1	people	2.17	1
coal	16.06	8	cars	3.75	1	great	2.17	1
kilograms	15.65	3	refers	3.75	1	pollutant	2.17	1
year	15.61	9	keep	3.75	1	nd	2.17	1
oxides	14.95	4	noise	3.75	1	icbc	2.17	1
data	14.69	5	currently	3.75	1	economy	2.17	1
rpk	14.68	2	level	3.75	1	operate	2.17	1
ammonia	14.50	3	fossil	3.75	1	blue	1.91	1
kept	14.18	1	db	3.75	1	serves	1.91	1
kg	14.01	2	commuter	3.75	1	equal	1.91	1
capita	13.81	2	rules	3.68	2	europe	1.91	1
emissions	13.74	1	social	3.54	3	rounds	1.91	1
including	12.92	4	balance	3.54	3	skies	1.91	1
emitted	12.92	4	enterprises	3.36	1	battle	1.91	1
indirect	12.86	4	rst	3.36	1	schedule	1.91	1
tce	12.84	4	contracts	3.36	1	solid	1.91	1
headquarters	12.51	6	lled	3.36	1	soot	1.91	1
consumption	12.14	5	ful	3.36	1	nearly	1.91	1
management	12.08	5	iv	3.23	1	producer	1.91	1
head	11.64	1	table	3.23	1	sulfur	1.91	1
institutions	11.64	1	chapter	3.23	1	foundation	1.91	1
reduce	11.48	5	ecologicalcivilization	3.23	1	four	1.91	1
oil	11.43	6	insurance	3.23	1	accounts	1.91	1
actively	11.36	7	report	3.23	1	winning	1.91	1
company	11.33	7	kpi	3.23	1	adherent	1.91	1
aircraft	10.97	3	progress	3.19	2	upgrade	1.91	1
cod	10.70	2	five	3.19	2	trillion	1.91	1
centralized	10.38	4	pollution	3.06	1	transformation	1.91	1

discharge	10.23	1	machines	3.06	1	america	1.91	1
save	10.15	5	taken	3.06	1	according	1.91	1
development	10.11	7	resources	3.06	1	took	1.91	1
saving	9.52	5	needs	3.06	1	going	1.91	1
environmental	9.36	6	secondary	3.06	1	technological	1.91	1
recycling	9.28	4	recycle	3.06	1	task	1.91	1
credit	9.10	5	intelligent	3.06	1	largest	1.91	1
institution	9.02	4	around	3.06	1	laying	1.91	1
electric	9.00	3	department	3.06	1	supplier	1.91	1
economic	8.95	5	enhance	3.06	1	supplied	1.91	1
rmb	8.85	2	awareness	3.06	1	substitution	1.91	1
scope	8.44	2	cabin	3.06	1	campaign	1.91	1
thousand	8.24	1	among	3.06	1	accomplished	1.91	1
ten	8.24	1	channels	3.06	1	particulate	1.91	1
grams	8.24	1	encourages	3.06	1	plans	1.91	1
cnpc	8.24	4	centre	3.06	1	devoted	1.91	1
employees	8.19	3	bottles	3.06	1	deployment	1.91	1
national	8.15	4	bottle	3.06	1	decades	1.91	1
beijing	8.03	3	actual	3.06	1	planning	1.91	1
performance	7.91	3	expand	3.06	1	provider	1.91	1
could	7.83	4	continue	3.06	1	imports	1.91	1
equipment	7.82	4	account	3.06	1	contributor	1.91	1
clean	7.82	5	collected	3.06	1	computer	1.90	1
years	7.69	4	strengthening	3.03	2	basic	1.90	1
revenue	7.50	1	replacing	3.02	1	generated	1.90	1
amount	7.50	1	expected	3.02	1	collect	1.90	1
density	7.50	1	renovate	3.02	1	battery	1.90	1
energies	7.49	2	kwh	3.02	1	batteries	1.90	1
driven	7.49	2	engineering	3.02	1	fluorescent	1.90	1
electricity	7.49	2	remodeled	3.02	1	collects	1.90	1
natural	7.41	4	led	3.02	1	tubes	1.90	1
installed	6.87	2	lighting	3.02	1	assets	1.90	1

reducing	6.84	2	form	3.02	1	toner	1.90	1
promote	6.72	3	metal	3.02	1	technology	1.90	1
ogci	6.63	3	cooling	3.02	1	suppliers	1.90	1
change	6.63	3	bathrooms	3.02	1	computers	1.90	1
building	6.58	3	maintenance	3.02	1	information	1.90	1
plant	6.51	3	pump	3.02	1	recover	1.90	1
local	6.51	3	hangar	3.02	1	conditions	1.90	1
list	6.49	2	ageing	3.02	1	teaching	1.90	1
use	6.46	2	heat	3.02	1	areas	1.90	1
burning	6.46	2	addition	3.02	1	ink	1.90	1
direct	6.46	2	halide	3.02	1	set	1.90	1
fuels	6.46	2	corporation	3.02	1	everyday	1.90	1
purchased	6.46	2	invested	3.02	1	procedures	1.90	1
properties	6.45	2	government	3.00	2	help	1.90	1
indicators	6.45	2	industries	3.00	2	methods	1.90	1
refer	6.45	2	reporting	3.00	2	certain	1.90	1
please	6.45	2	comprehensively	2.77	2	donation	1.90	1
disclosed	6.45	2	advance	2.77	2	schools	1.90	1
annual	6.42	3	reached	2.76	2	scrapping	1.90	1
developed	6.42	2	countries	2.38	1	wastes	1.90	1
achieve	6.42	2	pollutants	2.38	1	donate	1.90	1
activities	6.42	2	makes	2.38	1	principle	1.90	1
tianjin	6.42	2	saved	2.38	1	original	1.90	1
exchange	6.42	2	model	2.38	1	boxes	1.90	1
annually	6.42	2	participates	2.38	1	consumables	1.90	1
comprehensive	6.42	2	municipalities	2.38	1	measures	1.90	1
projects	6.42	2	consecutive	2.38	1	dispose	1.90	1
mechanisms	6.42	2	named	2.38	1	bins	1.90	1
participated	6.42	2	regions	2.38	1	manages	1.90	1
promoting	6.29	5	built	2.38	1	cartridges	1.90	1
beverage	6.12	2	road	2.38	1	manage	1.90	1
lamps	6.05	2	decline	2.38	1	higher	1.18	1

ameco	6.05	2	endeavors	2.38	1	issuance	1.18	1
meters	6.01	3	eight	2.38	1	increasing	1.18	1
cubic	6.01	3	efforts	2.38	1	toboost	1.18	1
upgrading	5.74	3	provinces	2.38	1	improve	1.18	1
quality	5.74	3	continuously	2.38	1	cultural	1.18	1
disposal	5.71	3	sink	2.38	1	acted	1.18	1
reuction	5.70	1	land	2.38	1	industry	1.18	1
od	5.70	1	thanks	2.38	1	industrial	1.18	1
process	5.64	4	average	2.38	1	points	1.18	1
protection	5.46	4	autonomous	2.38	1	political	1.18	1
first	5.35	2	investment	2.38	1	coordination	1.18	1
ecological	5.30	4	forests	2.38	1	allaspects	1.18	1
three	5.20	2	along	2.38	1	utilization	1.18	1
completed	5.20	2	undertakings	2.38	1	instrument	1.18	1
asset	5.19	2	belt	2.38	1	underwriter	1.18	1
project	5.11	4	awards	2.38	1	generation	1.18	1
corporate	5.00	3	water	2.38	1	growth	1.18	1
operation	4.87	2	improves	2.38	1	prevention	1.18	1
public	4.75	2	low	2.38	1	concepts	1.18	1
hectares	4.75	2	reducecoz	2.36	2	risks	1.18	1
welfare	4.75	2	innovation	2.36	2	thereporing	1.18	1
market	4.72	2	rated	2.36	2	branch	1.18	1
bank	4.72	4	highest	2.36	2	exploitation	1.18	1
lightweight	4.61	1	conservation	2.36	2	sharing	1.18	1
usual	4.61	1	good	2.36	2	benefits	1.18	1
fuel	4.61	1	conservationand	2.36	2	sectors	1.18	1
resulting	4.61	1	economyin	2.36	2	biomass	1.18	1
lighter	4.61	1	shift	2.36	2	milion	1.18	1
seats	4.61	1	example	2.36	2	operationin	1.18	1
far	4.61	1	supporting	2.36	2	financing	1.18	1
introduced	4.61	1	services	2.36	2	business	1.18	1
flown	4.61	1	bond	2.36	2	resource	1.18	1

life	4.59	2	loan	2.36	2	opening	1.18	1
key	4.54	2	forward	2.36	2	energy	1.18	1
continued	4.54	2	system	2.21	1	trillion	1.18	1
service	4.44	2	message	2.21	1	carrying	1.18	1
implement	4.42	2	partnership	2.21	1	extended	1.18	1
paris	4.42	2	roadmap	2.21	1	reinforced	1.18	1
global	4.42	2	established	2.21	1	the bank actively acted in line	1.18	1
agreement	4.42	2	united	2.21	1	lead	1.18	1
recovery	4.41	2	member	2.21	1	fund	1.18	1
used	4.41	2	agreed	2.21	1	pursuit	1.18	1
improving	4.34	2	investments	2.21	1	raised	1.18	1
grid	4.34	2	initiative	2.21	1	backed	1.18	1
sulfur	4.22	1	conference	2.21	1	balanced	1.18	1
no	4.22	1	meeting	2.21	1	synchronously	1.18	1
significant	4.22	1	positive	2.21	1	carried	1.18	1
direct	4.22	1	sending	2.21	1	regarding	1.18	1
sox	4.22	1	modern	2.21	1	commercial	1.18	1
osce	4.22	1	international	2.21	1	leading	1.18	1
emissions	4.22	1	response	2.21	1	percentage points	1.18	1
emissions	4.22	1	attended	2.21	1	subsidiary	1.18	1

SDG 13 scope 2 with TF-IDF

keyword	score	frequency	keyword	score	frequency			
emissions	100.00	11	branch	14.71	1	emission	15.10	1
indirect	93.10	11	power	13.12	1	beijing	14.71	1
gas	85.23	9	use	13.12	1	electricity	8.14	1
greenhouse	85.23	9	burning	13.12	1	settings	8.14	1
ghg	77.71	6	electric	13.12	1	scape	15.10	1
tons	69.75	10	purchased	13.12	1	lamps	8.14	1
emission	50.49	5	fuels	13.12	1	kwh	8.14	1
scope	34.15	3	group	13.12	1			
energy	23.49	2	entire	13.12	1			
including	22.30	2	cod	11.53	1			
emitted	22.30	2	optimizing	8.14	1			
total	20.71	1	equivalent	8.14	1			
ofc	20.71	1	emit	8.14	1			
nox	16.72	2	led	8.14	1			
per	16.27	2	reducing	8.14	1			
year	16.27	2	save	8.14	1			

SDG13 education or training with TF-IDF

keyword	score	frequency
chalks	100	1
consumption	100	1
control	100	1
damage	100	1
environment	100	1
maximally	100	1
paper	100	1
pollution	100	1
reduce	100	1
trainings	100	1

SDG13 scope1 with TF-IDF

keyword	score	frequency
emissions	100	4
tons	100	4
cod	25	1
nox	25	1

Appendix 4 Irrelevant data collected for climate change spending or resource commitments due to less representative keywords

Air China (2017, p. 46)	The bellyhold revenue increased by more than ****rmb****300 million due to the increase in bellyhold load factor Aircraft fuel saving	Not relevant to climate change
Air China (2017, p. 47)	we enabled the recovery and recycling of heat and replaced heat supply from boilers, saving about 60% of energy costs, or about ****rmb****900 ,000 ****Yuan**** per year. Maintenance Air pollution control. In terms of air pollution control, Air China strictly followed national environmental protection standard to control exhaust emission. In 2017, Air China **invested** ****rmb****8.4 million in transforming exhaust treatment facilities	Not relevant to climate change
Air China (2019, p. 58)	Increasing ****investment**** in Green Support Air China continued to strengthen its control and risk management of pollutants from the ground, ...	Not relevant to climate change
	To support the protection of new energy vehicles, we **invested** ****rmb****8.267 million to construct the 380 centralized ...	Not relevant to climate change
	In 2019, Air China **invested** ****rmb****148 million to introduce new energy vehicles and equipment, basically covering all types of vehicles. Meanwhile, Air China implemented the elimination and upgrade work of existing diesel vehicles and already **invested** ****rmb****467 million for vehicle upgrade, charging infrastructure	Not relevant to climate change
(China Life Insurance, 2017, p. 2)	☐ Increase ****investment**** of charity foundation s in poverty alleviation; Strengthen insurance education in poverty - stricken areas.	Not relevant to climate change
ICBC (2016, p. 43)	.. the Bank practiced mobility management over /five industries with serious Four policy advices on green /finance encourage /financial institutions to measure climatic, environmental and social risks in their ****investment**** portfolios4 improve institutional construction and intellectual capability through the G20 international cooperation platform3 establish green and environmental protection standards and encourage disclosure and reporting of ****investment**** impact2 Create incentive measures and reduce the cost of green ****investment**** and /f_inancing1 over-capacity, upper- and downstream industries of steel and the coal industrial chain and actively carried out differentiated credit policies with “both support and control”	Not relevant to climate change

Appendix 5 Irrelevant data for poverty alleviation through ecological protection

<p>CNPC (2019, p. 73)</p>	<ul style="list-style-type: none"> • We strictly follow safety and environmental protection regulations, formulate management systems on environmental protection that involve participation of all stakeholders and closely monitor environmental impact throughout the entire operations. We immediately restore the surface right after the completion of the operation so as to minimize the impact of production activities imposed on ecological environment. <p>In Mozambique, we organized the "Coastal Cleanup" volunteer activities at Maputo Beach. Our employees picked up 225 kg of solid waste in 15 bags, supporting local environmental protection. 77</p> <p>Local environmental protection We strongly believe in the importance of striking a balance between business development and environmental protection. Therefore, we strictly comply with the laws and international standards on environmental protection in the countries in which we operate and pay attention to protection of local ecological environments during the production operation.</p>
<p>Yili (2017, p. 50)</p>	<p>Yili prefers suppliers who lay emphasis on environmental protection; and calls on all suppliers to earnestly implement the protection measures of biodiversity.</p>
<p>Yili (2018, p. 50)</p>	<p>“....· Use environmentally friendly and energy-saving building materials Products · Reduced waste production. · Advocate separated waste recycling. · Promotion of green consumption. Employees Daily Work · Green offices · Promotion of employee awareness of environmental protection · Employee volunteer activities· Energy cascade utilization and recycling · Clean production technologies · High environmental protection requirements for supplier access. ·</p>

Appendix 6 Less representative keywords or misclassified data for energy spending or resource commitment

<p>CNPC (2016, p. 15)</p>	<p>Support education, public health and other public welfare activities. Perform social ****investment**** Enhance sustainable development capacity. Carry out disaster relief, poverty alleviation and donations to education, and support the reconstruction of disaster-stricken areas</p>	<p>Misclassified as energy spending or resource commitment evidence</p>
	<p>Technological innovation created more than 100 billion ****Yuan**** during our “12th Five-Year Plan” period.</p>	<p>‘Yuan’ not a representative keyword for spending or resource commitment</p>
<p>CNPC (2017, p. 24)</p>	<p>The application of the ASP flooding technology has enabled an accumulative increase of 20.56 million tons in oil production in Daqing Oilfield, representing a production value of ****rmb**** 67.75 billion. “ASP flooding technology for EOR and its industrial application” won the Second prize of National Science and Technology Progress ...</p>	<p>‘RMB’ not a representative keyword for spending or resource commitment</p>
<p>China Life Insurance (2016, p. 52)</p>	<p>.... Our achievement We **invested** a total of **CNY** 170 million in vocational training and trained 1.46 million person -times. We contributed **CNY** 13 million to assisted needy employees. China Life Insurance Company Limited 2016 Corporate Social Responsibility Report Part III Humanistic Care Stakeholders' expectation Satisfying remuneration and welfare, a safe and comfortable working environment and a</p>	<p>‘CNY’ not a representative keyword for spending or resource commitment</p>
	<p>housing security, medical security, daily life support and offspring security which can be cash or in non-currency forms. Six categories of welfare State-mandated welfare such as social insurance and housing ****fund****.</p>	<p>Irrelevant to SDG 7</p>
<p>ICBC (2017, p. 21)</p>	<p>.... Bearing is a basic part crucial in equipment manufacturing, and is called the "heart... period, Dalian Branch of the Bank, in cooperation with an overseas branch, tailor-made a ****loan**** product solution for</p>	<p>Not energy investment</p>

Appendix 7 SDG disclosure data for Chinese companies

Year	code	company	PA_score	energy_score	CC_score	total_score
2016	000001	Ping An Bank Co., Ltd.	0	0	0	0
2017	000001	Ping An Bank Co., Ltd.	0	0	0	0
2018	000001	Ping An Bank Co., Ltd.	0	0	0	0
2019	000001	Ping An Bank Co., Ltd.	10	2	2	14
2020	000001	Ping An Bank Co., Ltd.	9	3	2	14
2016	000002	China Vanke Co., Ltd.	10	3	3	16
2017	000002	China Vanke Co., Ltd.	10	3	3	16
2018	000002	China Vanke Co., Ltd.	10	5	4	19
2019	000002	China Vanke Co., Ltd.	10	4	4	18
2020	000002	China Vanke Co., Ltd.	10	4	4	18
2016	000063	ZTE Corporation	9	4	6	19
2017	000063	ZTE Corporation	10	4	2	18
2018	000063	ZTE Corporation	9	4	2	16
2019	000063	ZTE Corporation	10	4	4	19
2020	000063	ZTE Corporation	7	4	2	14
2016	000100	TCL Technology Group Corporation	0	0	0	0
2017	000100	TCL Technology Group Corporation	0	0	0	0
2018	000100	TCL Technology Group Corporation	7	5	4	17
2019	000100	TCL Technology Group Corporation	0	0	0	0
2020	000100	TCL Technology Group Corporation	0	0	0	0
2016	000166	Shenwan Hongyuan Group Co., Ltd.	0	0	0	0
2017	000166	Shenwan Hongyuan Group Co., Ltd.	0	0	0	0
2018	000166	Shenwan Hongyuan Group Co., Ltd.	8	5	3	16
2019	000166	Shenwan Hongyuan Group Co., Ltd.	9	2	5	16
2020	000166	Shenwan Hongyuan Group Co., Ltd.	8	3	3	14
2016	000333	Midea Group Co., Ltd.	7	2	2	11
2017	000333	Midea Group Co., Ltd.	10	4	3	17
2018	000333	Midea Group Co., Ltd.	9	5	3	17
2019	000333	Midea Group Co., Ltd.	7	4	2	13
2020	000333	Midea Group Co., Ltd.	6	5	2	14
2016	000564	Ccoop Group Co., ltd.	7	3	2	12
2017	000564	Ccoop Group Co., ltd.	0	0	0	0
2018	000564	Ccoop Group Co., ltd.	0	0	0	0
2019	000564	Ccoop Group Co., ltd.	0	0	0	0
2020	000564	Ccoop Group Co., ltd.	0	0	0	0
2016	000725	BOE Technology Group Co., Ltd.	8	5	3	16
2017	000725	BOE Technology Group Co., Ltd.	8	4	3	15
2018	000725	BOE Technology Group Co., Ltd.	5	4	3	12
2019	000725	BOE Technology Group Co., Ltd.	8	5	5	19
2020	000725	BOE Technology Group Co., Ltd.	8	5	6	20
2016	000839	CITIC Guoan Information Industry Co., Ltd.	7	4	2	13
2017	000839	CITIC Guoan Information Industry Co., Ltd.	8	5	2	15
2018	000839	CITIC Guoan Information Industry Co., Ltd.	10	3	3	16
2019	000839	CITIC Guoan Information Industry Co., Ltd.	10	5	3	18
2020	000839	CITIC Guoan Information Industry Co., Ltd.	1	0	1	2
2016	000858	Wuliangye Yibin Co., Ltd.	0	0	0	0

2017	000858	Wuliangye Yibin Co., Ltd.	0	0	0	0
2018	000858	Wuliangye Yibin Co., Ltd.	0	0	0	0
2019	000858	Wuliangye Yibin Co., Ltd.	9	3	1	13
2020	000858	Wuliangye Yibin Co., Ltd.	8	4	4	16
2016	000876	New Hope Liuhe Co., Ltd.	0	0	0	0
2017	000876	New Hope Liuhe Co., Ltd.	0	0	0	0
2018	000876	New Hope Liuhe Co., Ltd.	0	0	0	0
2019	000876	New Hope Liuhe Co., Ltd.	10	1	2	13
2020	000876	New Hope Liuhe Co., Ltd.	9	2	3	14
2016	002024	Suning.com Co., Ltd.	0	0	0	0
2017	002024	Suning.com Co., Ltd.	0	0	0	0
2018	002024	Suning.com Co., Ltd.	8	3	4	16
2019	002024	Suning.com Co., Ltd.	0	0	0	0
2020	002024	Suning.com Co., Ltd.	0	0	0	0
2016	002202	Xinjiang Goldwind Science & Technology Co., Ltd.	0	1	0	1
2017	002202	Xinjiang Goldwind Science & Technology Co., Ltd.	8	4	2	16
2018	002202	Xinjiang Goldwind Science & Technology Co., Ltd.	0	1	0	1
2019	002202	Xinjiang Goldwind Science & Technology Co., Ltd.	9	3	3	16
2020	002202	Xinjiang Goldwind Science & Technology Co., Ltd.	0	0	0	0
2016	002415	Hangzhou Hikvision Digital Technology Co., Ltd.	0	0	0	0
2017	002415	Hangzhou Hikvision Digital Technology Co., Ltd.	0	0	0	0
2018	002415	Hangzhou Hikvision Digital Technology Co., Ltd.	4	3	2	10
2019	002415	Hangzhou Hikvision Digital Technology Co., Ltd.	6	4	5	15
2020	002415	Hangzhou Hikvision Digital Technology Co., Ltd.	0	1	0	1
2016	002466	Tianqi Lithium Corporation	0	0	0	0
2017	002466	Tianqi Lithium Corporation	7	2	4	14
2018	002466	Tianqi Lithium Corporation	1	0	0	1
2019	002466	Tianqi Lithium Corporation	9	3	1	13
2020	002466	Tianqi Lithium Corporation	9	3	2	14
2016	002475	Luxshare Precision Industry Co., Ltd.	0	0	0	0
2017	002475	Luxshare Precision Industry Co., Ltd.	0	0	0	0
2018	002475	Luxshare Precision Industry Co., Ltd.	0	0	0	0
2019	002475	Luxshare Precision Industry Co., Ltd.	0	5	4	9
2020	002475	Luxshare Precision Industry Co., Ltd.	8	5	6	20
2016	002594	BYD Company Limited	4	4	2	10
2017	002594	BYD Company Limited	9	3	2	14
2018	002594	BYD Company Limited	8	4	2	14
2019	002594	BYD Company Limited	7	4	2	13
2020	002594	BYD Company Limited	5	4	4	13
2016	600000	Shanghai Pudong Development Bank Co., Ltd.	0	0	0	0
2017	600000	Shanghai Pudong Development Bank Co., Ltd.	0	0	0	0
2018	600000	Shanghai Pudong Development Bank Co., Ltd.	9	5	5	19
2019	600000	Shanghai Pudong Development Bank Co., Ltd.	9	3	3	15
2020	600000	Shanghai Pudong Development Bank Co., Ltd.	9	5	3	17

2016	600016	China Minsheng Banking Corp., Ltd.	10	5	4	19
2017	600016	China Minsheng Banking Corp., Ltd.	10	5	4	20
2018	600016	China Minsheng Banking Corp., Ltd.	9	5	4	18
2019	600016	China Minsheng Banking Corp., Ltd.	12	5	5	22
2020	600016	China Minsheng Banking Corp., Ltd.	11	4	4	19
2016	600018	Shanghai International Port (Group) Co., Ltd.	0	0	0	0
2017	600018	Shanghai International Port (Group) Co., Ltd.	0	0	0	0
2018	600018	Shanghai International Port (Group) Co., Ltd.	0	0	0	0
2019	600018	Shanghai International Port (Group) Co., Ltd.	5	2	2	9
2020	600018	Shanghai International Port (Group) Co., Ltd.	0	0	0	0
2016	600019	Baoshan Iron & Steel Co., Ltd.	11	3	4	20
2017	600019	Baoshan Iron & Steel Co., Ltd.	10	4	2	17
2018	600019	Baoshan Iron & Steel Co., Ltd.	10	4	4	19
2019	600019	Baoshan Iron & Steel Co., Ltd.	10	3	3	17
2020	600019	Baoshan Iron & Steel Co., Ltd.	8	5	7	22
2016	600028	China Petroleum & Chemical Corporation	0	1	2	4
2017	600028	China Petroleum & Chemical Corporation	8	5	3	19
2018	600028	China Petroleum & Chemical Corporation	0	1	0	3
2019	600028	China Petroleum & Chemical Corporation	8	5	5	21
2020	600028	China Petroleum & Chemical Corporation	11	5	6	24
2016	600029	China Southern Airlines Company Limited	9	5	3	17
2017	600029	China Southern Airlines Company Limited	9	4	2	15
2018	600029	China Southern Airlines Company Limited	9	4	2	16
2019	600029	China Southern Airlines Company Limited	9	4	3	17
2020	600029	China Southern Airlines Company Limited	8	5	3	16
2016	600036	China Merchants Bank Co., Ltd.	11	4	4	19
2017	600036	China Merchants Bank Co., Ltd.	12	4	5	22
2018	600036	China Merchants Bank Co., Ltd.	11	4	5	20
2019	600036	China Merchants Bank Co., Ltd.	10	5	5	21
2020	600036	China Merchants Bank Co., Ltd.	11	5	7	24
2016	600061	SDIC Capital Co., Ltd	9	5	4	19
2017	600061	SDIC Capital Co., Ltd	9	4	2	17
2018	600061	SDIC Capital Co., Ltd	9	5	3	19
2019	600061	SDIC Capital Co., Ltd	6	4	3	15
2020	600061	SDIC Capital Co., Ltd	10	5	4	21
2016	600115	China Eastern Airlines Corporation Limited	10	4	2	19
2017	600115	China Eastern Airlines Corporation Limited	10	5	4	21
2018	600115	China Eastern Airlines Corporation Limited	10	5	4	21
2019	600115	China Eastern Airlines Corporation Limited	11	5	4	22
2020	600115	China Eastern Airlines Corporation Limited	10	5	4	21
2016	600196	Shanghai Fosun Pharmaceutical (Group) Co., Ltd.	7	4	4	16

2017	600196	Shanghai Fosun Pharmaceutical (Group) Co., Ltd.	9	5	4	18
2018	600196	Shanghai Fosun Pharmaceutical (Group) Co., Ltd.	7	5	4	16
2019	600196	Shanghai Fosun Pharmaceutical (Group) Co., Ltd.	9	5	4	18
2020	600196	Shanghai Fosun Pharmaceutical (Group) Co., Ltd.	7	5	6	18
2016	600221	Hainan Airlines Holding Co., Ltd.	8	4	4	16
2017	600221	Hainan Airlines Holding Co., Ltd.	5	1	3	11
2018	600221	Hainan Airlines Holding Co., Ltd.	7	4	1	12
2019	600221	Hainan Airlines Holding Co., Ltd.	0	0	0	0
2020	600221	Hainan Airlines Holding Co., Ltd.	0	0	0	0
2016	600585	Anhui Conch Cement Company Limited	1	0	0	1
2017	600585	Anhui Conch Cement Company Limited	7	5	3	15
2018	600585	Anhui Conch Cement Company Limited	10	5	4	20
2019	600585	Anhui Conch Cement Company Limited	8	5	4	17
2020	600585	Anhui Conch Cement Company Limited	8	5	4	17
2016	600688	Sinopec Shanghai Petrochemical Company Limited	10	3	2	16
2017	600688	Sinopec Shanghai Petrochemical Company Limited	10	3	2	16
2018	600688	Sinopec Shanghai Petrochemical Company Limited	8	3	2	14
2019	600688	Sinopec Shanghai Petrochemical Company Limited	8	5	3	16
2020	600688	Sinopec Shanghai Petrochemical Company Limited	7	3	3	13
2016	600690	Haier Smart Home Co., Ltd.	0	0	0	0
2017	600690	Haier Smart Home Co., Ltd.	0	0	0	0
2018	600690	Haier Smart Home Co., Ltd.	0	0	0	0
2019	600690	Haier Smart Home Co., Ltd.	0	0	0	0
2020	600690	Haier Smart Home Co., Ltd.	5	4	3	13
2016	600837	Haitong Securities Co., Ltd.	3	3	3	9
2017	600837	Haitong Securities Co., Ltd.	1	4	4	9
2018	600837	Haitong Securities Co., Ltd.	6	4	4	14
2019	600837	Haitong Securities Co., Ltd.	7	4	2	13
2020	600837	Haitong Securities Co., Ltd.	5	4	4	13
2016	600871	Sinopec Oilfield Service Corporation	3	4	1	8
2017	600871	Sinopec Oilfield Service Corporation	5	4	3	12
2018	600871	Sinopec Oilfield Service Corporation	0	4	3	7
2019	600871	Sinopec Oilfield Service Corporation	4	4	3	11
2020	600871	Sinopec Oilfield Service Corporation	8	5	2	15
2016	600887	Inner Mongolia Yili Industrial Group Co., Ltd.	0	1	0	1
2017	600887	Inner Mongolia Yili Industrial Group Co., Ltd.	11	5	3	22
2018	600887	Inner Mongolia Yili Industrial Group Co., Ltd.	10	4	4	21
2019	600887	Inner Mongolia Yili Industrial Group Co., Ltd.	0	0	0	0
2020	600887	Inner Mongolia Yili Industrial Group Co., Ltd.	10	5	4	22
2016	600999	China Merchants Securities Co., Ltd.	9	3	3	15
2017	600999	China Merchants Securities Co., Ltd.	10	5	5	20
2018	600999	China Merchants Securities Co., Ltd.	6	1	3	10

2019	600999	China Merchants Securities Co., Ltd.	10	5	4	19
2020	600999	China Merchants Securities Co., Ltd.	0	1	0	1
2016	601088	China Shenhua Energy Company Limited	11	4	4	19
2017	601088	China Shenhua Energy Company Limited	10	5	3	18
2018	601088	China Shenhua Energy Company Limited	9	3	4	17
2019	601088	China Shenhua Energy Company Limited	8	4	4	16
2020	601088	China Shenhua Energy Company Limited	8	2	2	13
2016	601111	Air China Limited	9	5	2	18
2017	601111	Air China Limited	9	5	4	20
2018	601111	Air China Limited	11	4	4	22
2019	601111	Air China Limited	9	5	4	21
2020	601111	Air China Limited	8	5	7	23
2016	601166	Industrial Bank Co., Ltd.	6	3	2	11
2017	601166	Industrial Bank Co., Ltd.	11	2	4	17
2018	601166	Industrial Bank Co., Ltd.	11	5	5	23
2019	601166	Industrial Bank Co., Ltd.	10	5	4	21
2020	601166	Industrial Bank Co., Ltd.	11	5	4	22
2016	601186	China Railway Construction Corporation Limited	9	3	3	15
2017	601186	China Railway Construction Corporation Limited	9	4	3	16
2018	601186	China Railway Construction Corporation Limited	10	2	4	16
2019	601186	China Railway Construction Corporation Limited	10	4	4	18
2020	601186	China Railway Construction Corporation Limited	10	5	4	19
2016	601211	Guotai Junan Securities Co., Ltd.	0	0	0	0
2017	601211	Guotai Junan Securities Co., Ltd.	8	4	4	16
2018	601211	Guotai Junan Securities Co., Ltd.	8	3	3	14
2019	601211	Guotai Junan Securities Co., Ltd.	0	0	0	0
2020	601211	Guotai Junan Securities Co., Ltd.	10	4	4	18
2016	601288	Agricultural Bank of China Limited	10	5	4	19
2017	601288	Agricultural Bank of China Limited	10	4	7	22
2018	601288	Agricultural Bank of China Limited	9	5	4	18
2019	601288	Agricultural Bank of China Limited	3	1	2	7
2020	601288	Agricultural Bank of China Limited	9	3	4	17
2016	601318	Ping An Insurance (Group) Company of China, Ltd.	2	0	0	2
2017	601318	Ping An Insurance (Group) Company of China, Ltd.	0	0	0	0
2018	601318	Ping An Insurance (Group) Company of China, Ltd.	10	4	4	18
2019	601318	Ping An Insurance (Group) Company of China, Ltd.	9	4	5	19
2020	601318	Ping An Insurance (Group) Company of China, Ltd.	10	4	4	20
2016	601328	Bank of Communications Co., Ltd.	10	4	3	17
2017	601328	Bank of Communications Co., Ltd.	0	1	0	3
2018	601328	Bank of Communications Co., Ltd.	9	3	3	16
2019	601328	Bank of Communications Co., Ltd.	10	5	4	20
2020	601328	Bank of Communications Co., Ltd.	7	2	2	12
2016	601336	New China Life Insurance Company Limited	0	0	0	0

2017	601336	New China Life Insurance Company Limited	8	4	2	14
2018	601336	New China Life Insurance Company Limited	9	4	5	18
2019	601336	New China Life Insurance Company Limited	10	5	6	21
2020	601336	New China Life Insurance Company Limited	10	5	6	21
2016	601390	China Railway Group Limited	9	5	3	18
2017	601390	China Railway Group Limited	12	5	4	21
2018	601390	China Railway Group Limited	12	5	2	19
2019	601390	China Railway Group Limited	12	5	4	21
2020	601390	China Railway Group Limited	12	5	4	21
2016	601398	Industrial and Commercial Bank of China Limited	11	4	5	23
2017	601398	Industrial and Commercial Bank of China Limited	10	3	4	20
2018	601398	Industrial and Commercial Bank of China Limited	10	5	5	23
2019	601398	Industrial and Commercial Bank of China Limited	11	5	5	24
2020	601398	Industrial and Commercial Bank of China Limited	10	5	3	21
2016	601618	Metallurgical Corporation of China Ltd.	10	5	3	18
2017	601618	Metallurgical Corporation of China Ltd.	10	5	5	21
2018	601618	Metallurgical Corporation of China Ltd.	8	4	5	17
2019	601618	Metallurgical Corporation of China Ltd.	11	5	4	21
2020	601618	Metallurgical Corporation of China Ltd.	10	5	4	20
2016	601628	China Life Insurance Company Ltd.	11	4	4	22
2017	601628	China Life Insurance Company Ltd.	11	5	4	22
2018	601628	China Life Insurance Company Ltd.	10	5	4	22
2019	601628	China Life Insurance Company Ltd.	9	4	3	19
2020	601628	China Life Insurance Company Ltd.	11	5	4	23
2016	601688	Huatai Securities Co., Ltd.	10	5	3	18
2017	601688	Huatai Securities Co., Ltd.	9	2	3	14
2018	601688	Huatai Securities Co., Ltd.	10	4	3	19
2019	601688	Huatai Securities Co., Ltd.	9	4	6	20
2020	601688	Huatai Securities Co., Ltd.	10	4	4	19
2016	601727	Shanghai Electric Group Co., Ltd.	10	3	4	17
2017	601727	Shanghai Electric Group Co., Ltd.	9	5	4	18
2018	601727	Shanghai Electric Group Co., Ltd.	8	5	4	17
2019	601727	Shanghai Electric Group Co., Ltd.	10	5	7	22
2020	601727	Shanghai Electric Group Co., Ltd.	9	5	6	20
2016	601788	Everbright Securities Co., Ltd.	7	4	2	13
2017	601788	Everbright Securities Co., Ltd.	9	4	4	18
2018	601788	Everbright Securities Co., Ltd.	9	4	2	15
2019	601788	Everbright Securities Co., Ltd.	7	2	1	10
2020	601788	Everbright Securities Co., Ltd.	7	2	2	11
2016	601800	China Communications Construction Company Limited	0	0	0	0
2017	601800	China Communications Construction Company Limited	0	0	0	0
2018	601800	China Communications Construction Company Limited	0	0	0	0
2019	601800	China Communications Construction Company Limited	0	0	0	0
2020	601800	China Communications Construction Company Limited	0	1	0	1

2016	601857	PetroChina Company Limited	11	4	5	22
2017	601857	PetroChina Company Limited	11	5	5	24
2018	601857	PetroChina Company Limited	12	5	5	25
2019	601857	PetroChina Company Limited	12	5	5	25
2020	601857	PetroChina Company Limited	10	5	5	23
2016	601898	China Coal Energy Company Limited	8	4	2	14
2017	601898	China Coal Energy Company Limited	9	4	3	16
2018	601898	China Coal Energy Company Limited	0	0	0	0
2019	601898	China Coal Energy Company Limited	0	0	0	0
2020	601898	China Coal Energy Company Limited	9	4	2	15
2016	601899	Zijin Mining Group Co., Ltd.	10	5	4	19
2017	601899	Zijin Mining Group Co., Ltd.	10	4	4	18
2018	601899	Zijin Mining Group Co., Ltd.	10	4	4	18
2019	601899	Zijin Mining Group Co., Ltd.	8	4	3	16
2020	601899	Zijin Mining Group Co., Ltd.	10	4	4	19
2016	601933	Yonghui Superstores Co., Ltd.	0	0	0	0
2017	601933	Yonghui Superstores Co., Ltd.	0	0	0	0
2018	601933	Yonghui Superstores Co., Ltd.	0	0	0	0
2019	601933	Yonghui Superstores Co., Ltd.	7	4	4	15
2020	601933	Yonghui Superstores Co., Ltd.	0	0	0	0
2016	601939	China Construction Bank Corporation	9	4	3	16
2017	601939	China Construction Bank Corporation	10	5	3	18
2018	601939	China Construction Bank Corporation	8	4	5	18
2019	601939	China Construction Bank Corporation	10	2	4	16
2020	601939	China Construction Bank Corporation	10	2	2	14
2016	601985	China National Nuclear Energy Power Co., Ltd.	0	0	0	0
2017	601985	China National Nuclear Energy Power Co., Ltd.	0	0	0	0
2018	601985	China National Nuclear Energy Power Co., Ltd.	9	4	3	17
2019	601985	China National Nuclear Energy Power Co., Ltd.	7	4	3	15
2020	601985	China National Nuclear Energy Power Co., Ltd.	10	5	4	20
2016	601988	Bank of China Limited	10	5	4	22
2017	601988	Bank of China Limited	10	4	3	20
2018	601988	Bank of China Limited	10	5	3	21
2019	601988	Bank of China Limited	11	4	3	21
2020	601988	Bank of China Limited	11	5	5	24
2016	603993	China Molybdenum Co., Ltd.	5	4	3	12
2017	603993	China Molybdenum Co., Ltd.	10	4	3	18
2018	603993	China Molybdenum Co., Ltd.	7	4	3	15
2019	603993	China Molybdenum Co., Ltd.	10	3	4	17
2020	603993	China Molybdenum Co., Ltd.	10	4	5	20

Appendix 8 Descriptive statistics

Descriptive Statistics	N	Range	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
SDG1 score	300	12.00	0.00	12.00	6.71	4.14	-0.76	0.14	-1.04	0.28
SDG7 score	300	5.00	0.00	5.00	3.13	1.90	-0.70	0.14	-1.07	0.28
SDG13 score	300	7.00	0.00	7.00	2.71	1.85	-0.13	0.14	-0.82	0.28
total SDG score	300	25.00	0.00	25.00	13.17	7.92	-0.72	0.14	-0.93	0.28
Belt and Road initiative	300	1.00	0.00	1.00	0.09	0.28	2.95	0.14	6.77	0.28
CDP reporting binary	300	1.00	0.00	1.00	0.13	0.33	2.26	0.14	3.11	0.28
SDG Innovation Demonstration Zone	300	0.00	0.00	1.00	1.00	0.00
Commercial SOEs	300	1.00	0.00	1.00	0.57	0.50	-0.27	0.14	-1.94	0.28
public-interest SOEs	300	1.00	0.00	1.00	0.15	0.36	1.97	0.14	1.89	0.28
other types of companies	300	1.00	0.00	1.00	0.27	0.44	1.06	0.14	-0.88	0.28
Political connection party committee	300	25.00	0.00	25.00	5.62	5.89	1.10	0.14	0.45	0.28
Political connection overall counts	300	10.00	0.00	10.00	1.89	1.88	1.33	0.14	2.41	0.28
Political connection overall binary	300	1.00	0.00	1.00	0.71	0.46	-0.93	0.14	-1.14	0.28
Political connection CEO executives count	300	4.00	0.00	4.00	0.39	0.76	2.50	0.14	7.53	0.28
Political connection CEO executives binary	300	1.00	0.00	1.00	0.28	0.45	0.99	0.14	-1.04	0.28
Political connection director counts	300	9.00	0.00	9.00	1.36	1.53	1.55	0.14	3.51	0.28
Political connection director binary	300	1.00	0.00	1.00	0.63	0.48	-0.53	0.14	-1.74	0.28
Green bond issuance binary	300	1.00	0.00	1.00	0.06	0.23	3.85	0.14	12.94	0.28
evaluation of accounting information quality	214	3.00	1.00	4.00	1.28	0.57	2.23	0.17	5.31	0.33
GICS Consumer Discretionary	300	1.00	0.00	1.00	0.08	0.28	3.03	0.14	7.23	0.28
GICS Consumer Staples	300	1.00	0.00	1.00	0.07	0.25	3.49	0.14	10.26	0.28
GICS Energy	300	1.00	0.00	1.00	0.12	0.32	2.40	0.14	3.79	0.28
GICS Financials	300	1.00	0.00	1.00	0.33	0.47	0.71	0.14	-1.51	0.28
GICS Industrials	300	1.00	0.00	1.00	0.18	0.39	1.65	0.14	0.71	0.28
GICS Information Technology	300	1.00	0.00	1.00	0.08	0.28	3.03	0.14	7.23	0.28
GICS Utilities	300	1.00	0.00	1.00	0.03	0.18	5.23	0.14	25.48	0.28
GICS Communication Service	300	1.00	0.00	1.00	0.02	0.13	7.59	0.14	55.97	0.28
GICS Materials	300	1.00	0.00	1.00	0.05	0.22	4.15	0.14	15.33	0.28
GICS Health care	300	1.00	0.00	1.00	0.02	0.13	7.59	0.14	55.97	0.28
GICS Real Estate	300	1.00	0.00	1.00	0.02	0.13	7.59	0.14	55.97	0.28
COVID-19	300	1.00	0.00	1.00	0.20	0.40	1.51	0.14	0.27	0.28
The number of independent directors	300	4.00	3.00	7.00	4.06	0.99	0.53	0.14	-0.70	0.28
The number of non-independent directors	300	32.00	10.00	42.00	23.06	6.84	0.48	0.14	-0.14	0.28
The number of directors	300	35.00	13.00	48.00	27.12	7.39	0.41	0.14	-0.32	0.28
The number of executive directors	156	6.00	1.00	7.00	2.71	1.48	0.56	0.19	-0.48	0.39
The number of non-executive directors	156	16.00	1.00	17.00	5.47	4.02	1.26	0.19	0.94	0.39
Institutional shareholding	300	346.64 billion	427.10 million	347.08billion	345.99 billion	765.68 billion	3.08	0.14	8.32	0.28
insider shareholding	295	14.55 billion	14.67	14.55billion	14.55 billion	14.67	6.76	0.14	46.10	0.28
top5 Holding Percentage CSMAR	300	74.87	23.60	98.48	66.57	19.31	-0.16	0.14	-0.85	0.28

top10 Holding Percentage CSMAR	300	69.02	29.56	98.59	72.59	16.85	-0.41	0.14	-0.53	0.28
the number of board meeting	299	42.00	2.00	44.00	12.54	6.99	1.95	0.14	4.82	0.28
Audit by Big 4	300	1.00	0.00	1.00	0.75	0.43	-1.18	0.14	-0.61	0.28
Going concern Audit Opinion	300	1.00	0.00	1.00	0.02	0.15	6.35	0.14	38.54	0.28
Altman Z score	190	21.13	-1.32	19.81	2.40	2.54	2.98	0.18	13.04	0.35
firm size market capitalisation	294	2077564.25	7839.65	2085403.90	301765.64	412216.47	2.31	0.14	4.77	0.28
analyst recommendation capitalIQ	278	1.00	1.00	3.69	2.24	0.60	0.46	0.15	-0.65	0.29
SynTAo GreenFinance ESG score	284	35.00	36.88	71.88	53.14	6.21	-0.05	0.15	-0.03	0.29
SynTAo GreenFinance Environment score	284	69.39	13.00	82.39	27.27	16.65	1.54	0.15	0.80	0.29
SynTAo GreenFinance Social score	284	65.44	8.75	74.19	26.12	16.56	1.64	0.15	1.05	0.29
SynTAo GreenFinance governance score	284	63.19	7.50	70.69	21.51	14.62	1.65	0.15	1.26	0.29
SynTAo GreenFinance ESG Management score	284	58.76	4.38	63.13	23.06	10.70	1.04	0.15	1.11	0.29
SynTAo GreenFinance ESG Risk Score	284	78.94	18.88	97.81	42.63	19.72	1.55	0.15	0.83	0.29
Total revenue	295	2.96 million	2210.63	2.96 million	0.25 million	0.45 million	3.72	0.14	15.64	0.28
Gross profit	190	0.83 million	-12317.00	0.82 million	0.06 million	0.14 million	4.00	0.18	15.37	0.35
total receivables	195	0.39 million	613.77	0.39 million	0.05 million	0.08 million	2.35	0.17	4.98	0.35
total equity	295	2.93 million	-22253.90	2.91 million	0.31 million	0.53 million	2.70	0.14	7.08	0.28
cashflow from Operations	295	2.53 million	-1496381.00	1.03 million	-0.02 million	0.27 million	-2.55	0.14	12.87	0.28
cashflow from investing	295	1.79 million	-1190453.00	0.60 million	-0.06 million	0.17 million	-2.91	0.14	15.33	0.28
cashflow from financing	295	2.41 million	-211787.00	2.20 million	0.09 million	0.33 million	4.30	0.14	19.00	0.28
enterprise value	214	3.24 million	10858.82	3.25 million	0.31 million	0.47 million	3.92	0.17	17.63	0.33
Cash and Cash Equivalents	260	569.68 billion	0.00 billion	569.68 billion	52.89 billion	81.12 billion	3.94	0.15	20.02	0.30
Net Accounts Receivable	260	2,102.76 billion	0.04 billion	2,102.80 billion	83.01 billion	256.00 billion	4.98	0.15	29.05	0.30
Net Inventories	206	1,002.06 billion	0.00 billion	1,002.06 billion	49.91 billion	127.03 billion	5.07	0.17	29.95	0.34
Total Current Assets	206	1,547.39 billion	0.00 billion	1,547.39 billion	156.44 billion	246.70 billion	2.95	0.17	10.38	0.34
Net Intangible Assets	296	235.78 billion	0.00 billion	235.78 billion	15.33 billion	30.74 billion	4.23	0.14	21.88	0.28
Total Assets	300	33,333.89 billion	11.21 billion	33,345.10 billion	2,747.94 billion	6,206.06 billion	3.02	0.14	8.58	0.28
Short-term Loans	250	134.75 billion	0.00 billion	134.75 billion	16.37 billion	21.22 billion	2.30	0.15	6.52	0.31
Total Current Liabilities	206	1,317.49 billion	0.00 billion	1,317.49 billion	149.09 billion	228.94 billion	2.58	0.17	7.28	0.34
Total Long-term Liabilities	288	1,371.85 billion	0.00 billion	1,371.85 billion	171.06 billion	275.19 billion	2.18	0.14	4.19	0.29
Total Non-current Liabilities	206	627.19 billion	0.00 billion	627.19 billion	64.21 billion	97.86 billion	2.97	0.17	10.78	0.34
Total Liabilities	300	30,430.09 billion	5.41 billion	30,435.50 billion	2,443.04 billion	5,718.09 billion	3.01	0.14	8.48	0.28
Total Equity to Parent Company	300	2,921.87 billion	-28.37 billion	2,893.50 billion	284.03 billion	509.27 billion	2.86	0.14	8.13	0.28
Total Shareholders' Equity	300	2,931.77 billion	-22.25 billion	2,909.52 billion	304.90 billion	524.48 billion	2.73	0.14	7.25	0.28
Total Liabilities and Shareholder Equity	300	33,333.89 billion	11.21 billion	33,345.10 billion	2,747.94 billion	6,206.06 billion	3.02	0.14	8.58	0.28
Total Operating Revenue	300	2,963.98 billion	2.21 billion	2,966.19 billion	256.97 billion	441.75 billion	3.70	0.14	15.83	0.28
Total Operating Costs	300	2,885.13 billion	1.76 billion	2,886.89 billion	219.24 billion	420.55 billion	4.09	0.14	18.46	0.28
R&D Expenses	116	21.83 billion	0.01 billion	21.84 billion	4.15 billion	5.27 billion	1.38	0.23	1.16	0.45
Total Profit	300	463.43 billion	-71.30 billion	392.13 billion	40.41 billion	76.33 billion	2.88	0.14	8.23	0.28
Net Profit	300	386.43 billion	-68.74 billion	317.69 billion	32.47 billion	61.69 billion	2.85	0.14	8.05	0.28
Total debt	297	4,788.04 billion	0.00 billion	4,788.04 billion	351.77 billion	659.38 billion	3.35	0.14	13.90	0.28
EBIT	297	434.22 billion	-44.14 billion	390.08 billion	41.29 billion	77.52 billion	2.80	0.14	7.57	0.28
EBITDA	297	457.49 billion	-39.36 billion	418.13 billion	53.08 billion	90.11 billion	2.45	0.14	5.27	0.28
Current Ratio	249	4.56	0.11	4.66	1.26	0.72	1.72	0.15	4.28	0.31

Quick Ratio	249	4.30	0.07	4.38	1.00	0.66	1.85	0.15	4.62	0.31
Cash Ratio	205	2.58	0.01	2.59	0.39	0.45	2.98	0.17	10.25	0.34
Working Capital to Loans	199	138.32	-10.29	128.02	1.86	10.46	9.58	0.17	109.48	0.34
Working Capital	206	424.40 billion	-194.51 billion	229.89 billion	7.35 billion	51.98 billion	0.14	0.17	3.95	0.34
Net Cash Flow from Operating Activities to Current Liabilities	205	1.64	-0.16	1.49	0.26	0.28	1.67	0.17	3.88	0.34
Liabilities to Assets Ratio	300	0.97	0.16	1.14	0.67	0.20	-0.44	0.14	-0.63	0.28
Long-term Loans to Total Assets	300	0.57	0.00	0.57	0.06	0.10	2.94	0.14	10.82	0.28
Liabilities to Tangible Assets	300	0.97	0.18	1.15	0.71	0.20	-0.69	0.14	-0.32	0.28
Equity Multiplier	300	46.61	-29.44	17.17	4.94	4.66	-0.46	0.14	9.32	0.28
Debt to Equity	300	46.61	-30.44	16.17	3.94	4.66	-0.46	0.14	9.32	0.28
Equity to Debt	300	5.25	-0.12	5.14	0.68	0.76	2.36	0.14	7.01	0.28
Debt to Long-term Capital	300	1.26	-0.56	0.71	0.19	0.21	0.40	0.14	-0.07	0.28
Long-term Debt to Working Capital	205	435.48	-192.94	242.55	2.10	31.87	2.76	0.17	31.95	0.34
EBITDA to Total Liabilities	208	1.83	-0.36	1.47	0.22	0.26	2.35	0.17	7.45	0.34
Net Cash Flow from Operating Activities Total Liabilities	300	1.47	-0.30	1.17	0.12	0.20	2.31	0.14	7.75	0.28
Net cashflow from Operating Activities to Interest-bearing Liabilities	245	427.78	-156.40	271.38	5.28	36.48	5.11	0.16	36.13	0.31
Liabilities to Equity Market Cap	298	0.94	0.02	0.96	0.61	0.26	-0.48	0.14	-0.86	0.28
Total Liabilities to Net Tangible Assets	294	15.99	0.21	16.20	4.44	4.31	1.16	0.14	-0.04	0.28
Basic Earnings per Share after Deducting Extraordinary Items	295	13.57	-5.45	8.12	0.97	1.38	1.68	0.14	7.87	0.28
Basic Earnings per Share	300	13.65	-5.24	8.41	1.02	1.40	1.64	0.14	7.94	0.28
Diluted Earnings per Share	281	13.62	-5.24	8.38	1.06	1.42	1.57	0.15	7.57	0.29
Ratio of Current Assets	206	0.91	0.00	0.91	0.46	0.24	0.11	0.17	-1.02	0.34
Ratio of Cash Assets	300	0.59	0.00	0.59	0.12	0.11	1.80	0.14	4.08	0.28
Ratio of Receivable Assets	300	0.36	0.00	0.36	0.07	0.08	1.49	0.14	1.46	0.28
Ratio of Working Capital to Current Assets	205	9.18	-8.39	0.79	-0.20	1.07	-3.74	0.17	19.21	0.34
Ratio of Working Capital	206	1.36	-0.68	0.67	0.05	0.21	0.19	0.17	1.97	0.34
Ratio of Working Capital To Net Assets	206	18.62	-3.63	14.99	0.14	1.30	7.16	0.17	85.34	0.34
Ratio of Non-current Assets	206	0.95	0.00	0.95	0.54	0.24	-0.14	0.17	-1.00	0.34
Ratio of Fixed Assets	300	0.74	0.00	0.74	0.16	0.18	1.22	0.14	0.82	0.28
Ratio of Intangible assets	300	0.28	0.00	0.28	0.04	0.05	2.33	0.14	5.66	0.28
Tangible Assets Ratio	300	0.29	0.71	1.00	0.95	0.06	-1.91	0.14	3.16	0.28
Ratio of Owners Equity	300	0.97	-0.14	0.84	0.33	0.20	0.44	0.14	-0.63	0.28
Ratio of Retained Earnings to Total Assets	300	1.14	-0.41	0.73	0.14	0.16	0.85	0.14	3.37	0.28
Ratio of Shareholders Equity to Fixed Assets	300	150.09	-0.43	149.65	13.01	23.51	3.62	0.14	14.83	0.28
Ratio of Current Liabilities	206	0.99	0.00	0.99	0.70	0.22	-0.67	0.17	-0.31	0.34
Ratio of Operating liabilities	206	1.07	-0.08	0.99	0.51	0.25	-0.09	0.17	-0.94	0.34
Ratio of Accounts Receivable to Income	300	14.91	0.00	14.91	0.42	1.55	6.82	0.14	52.14	0.28

Return on Assets	300	0.62	-0.42	0.20	0.03	0.06	-1.51	0.14	15.93	0.28
Return on Equity	300	3.88	-3.51	0.36	0.08	0.31	-10.04	0.14	111.49	0.28
EBIT	208	211.17 billion	-65.17 billion	146.00 billion	17.59 billion	26.74 billion	2.24	0.17	6.66	0.34
EBITDA	208	413.40 billion	-60.40 billion	353.00 billion	30.50 billion	55.68 billion	3.68	0.17	14.89	0.34
Ratio of Net Profit to Total Profit	285	2.37	-1.25	1.11	0.79	0.18	-8.29	0.14	85.32	0.29
Ratio of Total Profit to EBIT	195	6.25	-3.83	2.43	0.79	0.51	-7.21	0.17	64.42	0.35
Ratio of EBIT to Total Assets	208	0.66	-0.40	0.26	0.06	0.07	-1.13	0.17	9.44	0.34
Return on Long-term Capital	208	2.80	-2.45	0.35	0.09	0.26	-7.45	0.17	66.08	0.34
Gross Operating Margin	205	1.16	-0.41	0.74	0.24	0.17	0.58	0.17	2.16	0.34
Total Leverage	193	41.69	0.96	42.65	2.51	3.66	8.03	0.18	79.46	0.35
Market Value	298	32,482.90 billion	17.10 billion	32,500.00 billion	2,770.66 billion	6,128.71 billion	2.99	0.14	8.41	0.28
Tobin's Q B	298	9.20	0.81	10.01	1.44	0.99	4.17	0.14	23.63	0.28
Book-to-Market Ratio	298	1.18	0.10	1.28	0.87	0.24	-1.33	0.14	1.21	0.28
EV to EBITDA	194	145.30	4.05	149.35	19.32	15.22	4.45	0.18	30.51	0.35
EPS estimate GAAP	277	10.32	-1.53	8.79	1.14	1.30	2.34	0.15	7.82	0.29
actual EPS (GAAP)	282	13.62	-5.24	8.38	1.05	1.42	1.60	0.15	7.64	0.29
EPS normalised estimate	261	10.24	-1.49	8.75	1.12	1.33	2.32	0.15	7.47	0.30
EPS normalised actual	262	13.34	-5.45	7.89	1.01	1.41	1.82	0.15	7.64	0.30
beta for 5 years	300	1.89	0.21	2.10	1.02	0.37	0.64	0.14	0.02	0.28
Yearly Beta Value of Composite Market	300	3.43	-0.38	3.06	0.86	0.46	0.46	0.14	1.66	0.28
Beta1	300	5.57	-1.01	4.56	-0.49	0.42	6.28	0.14	71.28	0.28
Beta2	300	5.77	-5.22	0.55	-0.62	0.60	-2.27	0.14	11.70	0.28
Sum Beta	300	3.07	-2.93	0.14	-1.11	0.45	-0.84	0.14	4.32	0.28
Combined Alpha Model Region Rank	295	99.00	1.00	100.00	58.03	26.03	-0.42	0.14	-0.79	0.28
Combined Alpha Mode Country Rank	295	99.00	1.00	100.00	72.03	26.01	-1.02	0.14	0.06	0.28
Combined Alpha Model Sector Rank	295	99.00	1.00	100.00	56.20	26.37	-0.33	0.14	-0.87	0.28
Combined Alpha Model Industry Rank	295	98.00	2.00	100.00	55.90	25.12	-0.30	0.14	-0.81	0.28

Appendix 9 Interaction effect

TreeNet (Gradient boosting machine) Interactions	
Variables tested for 2-way interactions: Top 5 Whole Variable Interactions	
Score	Predictor
5.02	Ratio of fixed assets
4.65	Beta for 5 years
4.44	Ratio of net profit to total profit
4.22	Total revenue
4.21	Working capital
4.07	Debt to long-term capital
4.00	Ratio of total profit to EBIT
3.88	SYNTAOGREENFINANCE ESG management score
3.38	Tangible assets ratio
3.26	Institutional shareholding
3.23	TOP5 holding percentage
3.23	Total leverage
3.13	Audit by BIG4
3.00	Ratio of cash assets
2.91	EV TO EBITDA
2.89	Insider shareholding
2.83	Ratio of retained earnings to total assets
2.80	Ratio of shareholders equity to fixed assets
2.77	Ratio of operating liabilities
2.74	Cashflow from Investing
2.60	Gross operating margin
2.57	Other types of companies
2.54	EPS estimate GAAP
2.53	Liabilities to tangible assets
2.41	SYNTAOGREENFINANCE governance score
2.38	Cash ratio
2.36	SYNTAOGREENFINANCE ESG score
2.30	Yearly beta value of composite market
2.22	Ratio of receivable assets
2.20	Number of non-independent directors
2.15	Quick ratio
2.13	Number of independent directors
2.08	Current ratio
2.02	Short-term loans
2.01	Cash and cash equivalents
2.01	Cashflow from operations
1.98	Long-term loans to total assets
1.97	EPS normalised estimate
1.94	Actual EPS GAAP
1.92	Ratio of intangible assets

1.89	Return on equity
1.87	Total debt
1.86	Firm size market capitalisation
1.81	Top10 holding percentage
1.80	Net cashflow from operating activities total liabilities
1.77	No directors
1.77	Net intangible assets
1.76	Tobins Q
1.75	Total operating costs
1.74	Total receivables
1.72	Ratio of working capital to net assets
1.72	EPS normalised actual
1.68	Total current assets
1.67	Analyst recommendation
1.66	Diluted earnings per share
1.61	The number of board meeting
1.58	No non-executive directors
1.57	Net accounts receivable
1.56	Ratio of working capital
1.55	Ratio of current liabilities
1.55	Return on assets
1.55	Enterprise value
1.55	SYNTAOGREENFINANCE Environmental score
1.54	Ratio of accounts receivable to income
1.53	EBITDA
1.51	Total profit
1.45	SYNTAOGREENFINANCE Social score
1.45	Ratio of current assets
1.44	EBIT
1.44	Cashflow from financing
1.43	Political connection CEO/Executives counts
1.41	Total operating revenue
1.41	Long term debt to working capital
1.40	Political connection director counts
1.39	Total non-current liabilities
1.37	Liabilities to equity marketcap
1.36	Working capital to loans
1.34	Total current liabilities
1.32	Sumbeta
1.31	Net cashflow from operating activities to interest-bearing liabilities
1.24	Book to market ratio
1.24	R d expenses
1.23	Altman Z score
1.20	Basic earnings per share after deducting extraordinary items

Appendix 10 Definition of predictors

Dimension	Indicator	Measurement	Data source
Regulatory influence	SDG regulatory influence	whether companies are in cities that participated in the Innovation Demonstration Zone project. A company will receive a score of one if it is located in one of these cities; otherwise, it will receive a score of zero.	CSMAR & manual classification
	Belt and Road Initiative	whether companies have BRI deals	Thomson Reuters database
Political connection (counts)	Political connection overall counts	The total number of senior management team members that are politically connected	CSMAR & automated data classification
	Political connection CEO/executives counts	The total number of CEO and executives that are politically connected	CSMAR & automated data classification
	Political connection director counts	The total number of directors that are politically connected	CSMAR & automated data classification
	The number of senior management team in the party committee	the number of senior management team members in the party committee	CSMAR & automated data classification
The presence of political connection	Political connection overall binary	A score of one will be assigned if the company has at least one politically connected member in the senior management team; otherwise, a score of zero will be assigned	CSMAR & automated data classification
	Political connection CEO/executives binary	A score of one will be assigned if the company's CEO or executive is politically connected; otherwise, a score of zero will be assigned	CSMAR & automated data classification
	Political connection director binary	A score of one will be assigned if the company has at least one politically connected director; otherwise, a score of zero will be assigned	CSMAR & automated data classification
	The presence of party committee	whether the company has party committee embedding	CSMAR & automated data classification
Shareholding influence	Commercial SOE	Commercial SOEs include 'perfect competitive sectors and strategic sectors (i.e. key industries related to national security and national economic lifelines)' (Lin et al., 2020, p. 39).	Definition from 'Opinions of the State Council on Development of An Economy of Mixed-Ownership of State-Run Enterprise'
	Public-interest SOE	Public interest state-owned enterprises target the quality and efficiency of public goods and services, to ensure people's livelihood and serve society.	SOE data from the CSMAR database
	Other types of companies	All other companies, except Commercial and public-interest SOEs	Manual data classification
Incentivising influence	Green bond	The issuance of green bond, including green bonds issued in China (Day) and green bonds issued overseas by domestic entities (Day).	Thomson Reuters database
Financial performance	Return on asset	Calculation formula: $\text{Net Profit} / \text{Balance of Total Assets}$. It will be denoted by Null if the denominator is not disclosed or zero.	CSMAR
	EBIT	Calculation formula: $\text{Net Profit} + \text{Income Tax Expenses} + \text{Financial Expenses}$	CSMAR
	Cash flow from financing	Cash from Financing [2004] is a subtotal line item across all templates with the following components: Standard template:	Capital IQ

		<p>Total Debt Issued [2161] Total Debt Repaid [2166] Issuance of Common Stock [2169] Repurchase of Common Stock [2164] Issuance of Preferred Stock [2181] Repurchase of Preferred Stock [2172] Total Dividends Paid [2022] Other Financing Activities, (Summary Subtotal) [2175]</p> <p>Non-standard templates: Total Debt Issued - (Template Specific) [2162] Total Debt Repaid - (Template Specific) [2163] Issuance of Common Stock - (Template Specific) [2159] Repurchase of Common Stock - (Template Specific) [2165] Issuance of Preferred Stock - (Template Specific) [2174] Repurchase of Preferred Stock - (Template Specific) [2170] Total Dividends Paid [2022] Other Financing Activities, (Summary Subtotal) - (Template Specific) [2176]</p>	
	Cash flow from investing	<p>Cash From Investing [2005] is a subtotal line item across all templates with the following components:</p> <p>Standard template: Capital Expenditure [2021] Sale of Property, Plant, and Equipment [2042] Cash Acquisitions [2057] Divestitures [2077] Other Investing Activities, (Summary Subtotal) [2177]</p> <p>Non-standard templates: Capital Expenditure [2114] Sale of Property, Plant and Equipment [2042] Cash Acquisitions [2057] Divestitures [2077] Other Investing Activities, (Summary Subtotal) - (Template Specific) [2178]</p> <p>Real Estate template - additional components: Acquisition of Real Estate Assets, Total [2124] Sale of Real Estate Assets, Total [2125] Net Sale / Acquisition of Real Estate Assets [2009] Purchase / Sale of Intangible Assets - (Template Specific) [2030]</p> <p>Real Estate template - components excluded: Capital Expenditure [2114] Sale of Property, Plant and Equipment [2042]</p>	Capital IQ
	Cash flow from operations	<p>Cash From Operations [2006] is a subtotal line item across all templates with the following components:</p> <p>Standard template: Net Income [2150] Depreciation and Amortization, Total [2160] Amortization of Deferred Charges, Total - (CF) [2014] Other Non-Cash Items, Total [2179] Change in Net Operating Assets [2010]</p> <p>Non-standard templates:</p>	Capital IQ

		Net Income [2150] Depreciation and Amortization, Total [2185] Amortization of Deferred Charges, Total - (Template Specific) [2015] Other Operating Activities, Total (Template Specific) [2180]	
	Tobin's Q	Market value/ total assets	CSMAR
	Enterprise value	For companies in Standard, REIT and Utility templates S&P Capital IQ calculates Total Enterprise Value as: Filing Date MarketCap + Book Value of Total Debt + Book Value of Preferred Stock + Book Value of Minority Interest - Cash & Short Term Investments For Insurance companies, S&P Capital IQ calculates Total Enterprise Value as: Filing Date MarketCap + Book Value of Total Debt + Book Value of Preferred Stock + Book Value of Minority Interest - Cash & Equivalents	Capital IQ
	Total revenue	Total revenue of the company	Thomson Reuters
	Short-term loan	Defined under the indicated accounting principles	CSMAR
	Ratio of receivable assets	"Calculation formula: (Net Notes Receivable + Net Accounts Receivable)/ (Total Assets). In case each item in denominator or numerator is Null, the item takes the value zero. It will be denoted by Null if the total denominator is equal to zero."	CSMAR
	Cash and cash equivalents	Defined under the indicated accounting principles	
	Net accounts receivable	Defined under the indicated accounting principles	
	Total accounts receivable	Defined under the indicated accounting principles	
	Total operating costs	Defined under the indicated accounting principles	

	Total operating revenue	Defined under the indicated accounting principles
	Ratio of current liabilities	Calculation formula: Current liabilities / Total liabilities. It will be denoted by Null if the denominator is not disclosed or equal to zero.
	Total-long-term liabilities	Defined under the indicated accounting principles
	Total profit	Defined under the indicated accounting principles
	Total liabilities	Defined under the indicated accounting principles
	Total equity	Equity as defined under the indicated accounting principles. Includes par value, paid in capital, retained earnings, and other adjustments to equity.
	net cashflow from operating activities to interest-bearing liabilities	Calculation formula: net cashflow from operating activities / interest-bearing liabilities. It will be denoted by Null if the denominator is not disclosed or equal to zero.
	Net profit	Defined under the indicated accounting principles
	Net inventories	Defined under the indicated accounting principles
	Total equity to parent company	Defined under the indicated accounting principles
	Total current liabilities	Defined under the indicated accounting principles
	R&D expense	Defined under the indicated accounting principles
	Market value	Defined under the indicated accounting principles
	Gross profit	Defined under the indicated accounting principles
	Total shareholder equity	Defined under the indicated accounting principles
	Total non-current liabilities	Defined under the indicated accounting principles
	Ratio of EBIT to total assets	EBIT/total asset
	Ratio of working capital to current asset	"Calculation formula: (Total Current Assets Total Current Liabilities) / (Total Current Assets). In case each item in denominator or numerator is Null, the item takes the value zero. It will be denoted by Null if the total denominator is equal to zero."
	Total liabilities and shareholder equity	

Ratio of non-current assets	"Calculation formula: Non current Assets / Total Assets or (1 Ratio of Cash to Assets). It will be denoted by Null if the denominator is not disclosed or equal to zero."
Evaluation of accounting information quality	1=Excellent; 2=Good; 3=Qualified; 4=Unqualified.
Total liabilities to net tangible assets	Total liabilities/net tangible asset
Ratio of working capital	"Calculation formula: (Current Assets Current Liabilities) / Total Assets. It will be denoted by Null if the denominator is not disclosed or zero."
net cashflow from operating activities to current liabilities	Calculation formula: net cashflow from operating activities current liabilities. It will be denoted by Null if the denominator is not disclosed or equal to zero.
Liabilities to asset ratio	Calculation formula: liabilities / assets. It will be denoted by Null if the denominator is not disclosed or equal to zero.
Liabilities to equity marketcap	Calculation formula: liabilities / equity marketcap. It will be denoted by Null if the denominator is not disclosed or equal to zero.
Ratio of operating liabilities	"Calculation formula: (Total Current Liabilities Short term Loans Non current Liabilities Due within One Year Trading Financial Liabilities Derivative Financial Liabilities) / Total Liabilities. It will be denoted by Null if the denominator is not disclosed or zero."
Ratio of intangible asset	"Calculation formula: Net Intangible Assets / Total Assets. It will be denoted by Null if the denominator is not disclosed or zero."
Net Intangible Assets	
Ratio of current asset	Calculation formula: current assets / total assets It will be denoted by Null if the denominator is not disclosed or zero."
Book to market ratio	Calculation formula: Book value/ market value
Long-term loans to total assets	"Calculation formula: Long term Loans / Total Assets. In case the numerator is null, it takes the value zero. It will be denoted by Null if the denominator is null or zero."
Ratio of accounts receivable to income	"Calculation formula: Accounts Receivable / Operating Revenue. It will be denoted by Null if the denominator is not disclosed or equal to or below zero."
Ratio of retained earnings to total assets	"Calculation formula: (Surplus Reserves + Undistributed Profit) / Total Assets. It will be denoted by Null if the denominator is not disclosed or zero."

	Ratio of cash asset	Calculation formula: Ending Balance of Cash and Cash Equivalents / Total Assets. It will be denoted by Null if the denominator is not disclosed or equal to zero."
	Liabilities to tangible assets	"Calculation formula: (Total liabilities) / (Total Assets Net Intangible Assets Net Goodwill). In case the numerator is null, it takes the value zero. When the Net Intangible Assets or Net Goodwill is null, the item takes the value zero; when the Total Assets is null or equal to zero, it will be denoted by Null."
	EBITDA	Calculation formula: Net Profit + Income Tax Expenses + Financial Expenses + Depreciation of Fixed Assets, Oil and Gas Assets, and Bearer Biological Assets + Amortization of Intangible Assets + Amortization of Long term Deferred Expenses
	Return on equity	Calculation formula: Net Profit / Average Balance of Shareholders' Equity. It will be denoted by Null if the denominator is not disclosed or equal to or below zero. Average Balance of Shareholders' Equity = (Beginning Balance of Shareholders' Equity + Ending Balance of Shareholders' Equity) / 2.
	Ratio of fixed asset	Calculation formula: Net Fixed Assets / Total Assets. It will be denoted by Null if the denominator is not disclosed or zero.
	Ratio of net profit to total profit	Calculation formula: Net Profit / Total Profit. The ratio is a sub ratio of the Net Operating Margin. Net Operating Margin = Ratio of Net Profit to Total Profit * Ratio of Total Profit to EBIT * Ratio of EBIT to Operating Revenue. It will be denoted by Null if the denominator is not disclosed or equal to or below zero.
	Ratio of Total Profit to EBIT	Calculation formula: Total Profit / EBIT. The ratio is a sub ratio of the Net Operating Margin. Net Operating Margin = Ratio of Net Profit to Total Profit * Ratio of Total Profit to EBIT * Ratio of EBIT to Operating Revenue. It will be denoted by Null if the denominator is not disclosed or equal to or below zero.
	Ratio of EBIT to Total Assets	Calculation formula: EBIT / Total Assets. It will be denoted by Null if the denominator is not disclosed or equal to or below zero.
	Return on Long-term Capital	Calculation formula: (Net Profit + Income Tax Expenses + Financial Expenses) / Long term Capital. Long term Capital = Average Balance of Non current Liabilities + Average Balance of Owners' Equity = (Beginning Non current Liabilities+ Ending Non current Liabilities) / 2 + (Beginning Owners' Equity + Ending Owners' Equity) / 2. It will be denoted by Null if the denominator is not disclosed or equal to or below zero.
	Gross Operating Margin	Calculation formula: After 2007: (Operating Revenue Operating Costs) / Operating Revenue. It will be denoted by Null if the denominator is not disclosed or equal to or below zero.

	Total Leverage	Calculation formula: $(\text{Net Profit} + \text{Income Tax Expenses} + \text{Financial Expenses} + \text{Depreciation of Fixed Assets, Oil and Gas Assets, and Bearer Biological Assets} + \text{Amortization of Intangible Assets} + \text{Amortization of Long term Deferred Expenses}) / (\text{Net Profit} + \text{Income Tax Expenses})$. It will be denoted by Null if the denominator is not disclosed or equal to zero, or financial expenses are below zero. Or Financial Leverage * Operating Leverage.
	EV to EBITDA	Calculation formula: Market Capitalization / EBITDA. It will be denoted by Null if the denominator is not disclosed or equal to or below zero.
	Net Cash Flow from Operating Activities to Current Liabilities	Calculation formula: Net Cash Flow from Operating Activities / Total Current Liabilities. It will be denoted by Null if the denominator is not disclosed or zero.
	Working Capital	Calculation formula: Total Current Assets Total Current Liabilities. In case the Total Current Liabilities is null, it takes the value zero. It will be denoted by Null if the Total Current Assets is null.
	Cash Flow-based Interest Coverage Ratio	Net Cash Flow from Operating Activities / Financial Expenses. It will be denoted by Null if the numerator is null, or the denominator is not disclosed or equal to or below zero
	equity multiplier	Total Assets / Total Owners' Equity. It will be denoted by Null if either the numerator or the denominator is null or zero
	Liabilities to Assets Ratio	Calculation formula: Total Liabilities / Total Assets. In case the numerator is null, it takes the value zero. It will be denoted by Null if the denominator is null or zero.
	Working Capital to Loans	Calculation formula: $(\text{Total Current Assets Total Current Liabilities}) / (\text{Short term Loans} + \text{Long term Loans})$. In case each item in denominator or numerator is null, it takes the value zero. It will be denoted by Null if the total denominator is equal to zero.
	Long-term Loans to Total Assets	Calculation formula: Long term Loans / Total Assets. In case the numerator is null, it takes the value zero. It will be denoted by Null if the denominator is null or zero.
	Liabilities to Tangible Assets	Calculation formula: $(\text{Total liabilities}) / (\text{Total Assets Net Intangible Assets Net Goodwill})$. In case the numerator is null, it takes the value zero. When the Net Intangible Assets or Net Goodwill is null, the item takes the value zero; when the Total Assets is null or equal to zero, it will be denoted by Null.
	Equity to Debt	Calculation formula: Total Owners' Equity / Total Liabilities. In case the numerator is null, it takes the value zero. The result will be Null if the denominator is null or zero.
	Debt to Long-term Capital	Calculation formula: Total Non-current liabilities / (Total Owners' Equity + Total Non-current liabilities). In case each item in denominator or numerator is null, the item takes the value zero. It will be denoted by Null if the total denominator is equal to zero.
	Long-term Debt to Working Capital	Calculation formula: Total Non-current liabilities / (Total Current Assets Total Current Liabilities). In case each item in denominator or numerator is null, the item takes the value zero. It will

		be denoted by Null if the total denominator is equal to zero.
	EBITDA to Total Liabilities	Calculation formula: (Net Profit + Income Tax Expenses + Financial Expenses + Amortization of Long term Deferred Expenses + Amortization of Intangible Assets + Depreciation of Fixed Assets, Oil and Gas assets, and Bearer Biological Assets) / Average Balance of Total Liabilities. The denominator is the average of Total Liabilities, that is, (Ending Value of Current Year + Ending Value of Previous Year) / 2. In case each item in numerator is null, it takes the value zero. It will be denoted by Null if the total numerator is equal to zero or the denominator is null or equal to zero.
	Long-term Debt to Working Capital	Calculation formula: Total Non-current liabilities / (Total Current Assets Total Current Liabilities). In case each item in denominator or numerator is null, the item takes the value zero. It will be denoted by Null if the total denominator is equal to zero.
	Debt to Long-term Capital	Calculation formula: Total Non-current liabilities / (Total Owners' Equity + Total Non-current liabilities). In case each item in denominator or numerator is null, the item takes the value zero. It will be denoted by Null if the total denominator is equal to zero.
	Current Ratio	Calculation formula: Current Assets / Current Liabilities. It will be denoted by Null if the denominator is not disclosed or zero.
	Quick Ratio	Calculation formula: (Current Assets Inventories) / Current Liabilities. It will be denoted by Null if the denominator is not disclosed or zero.
	Cash Ratio	Calculation formula: Ending Balance of Cash and Cash Equivalents / Current Liabilities. It will be denoted by Null if the denominator is not disclosed or zero.
	Working Capital to Loans	Calculation formula: (Total Current Assets Total Current Liabilities) / (Short term Loans + Long term Loans). In case each item in denominator or numerator is null; it takes the value zero. It will be denoted by Null if the total denominator is equal to zero.
	Working Capital	Calculation formula: Total Current Assets Total Current Liabilities. In case the Total Current Liabilities is null, it takes the value zero. It will be denoted by Null if the Total Current Assets is null.
	Earnings per share 1	Net Profit / Ending Paid-in Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero
	Earnings per Share TTM1	(Net Profit) TTM / Ending Paid-in Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero.
	Earnings per Share 2	Net Profit in Current Period/ Latest Share Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero
	Earnings per Share TTM2	(Net Profit) TTM/ Latest Share Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero

	Basic Earnings per Share after Deducting Extraordinary Items	The data are directly from the financial summaries of annual reports of listed companies.	
	Earnings per Share 3	(Net Profit-Non-operating Income + Non-operating Expenses)/ Ending Paid-in Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero.	
	Earnings per Share TTM3	(Net Profit-Non-operating Income + Non-operating Expenses) TTM / Ending Paid-in Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero.	CSMAR
	Earnings per Share 4	(Net Profit - Non-operating Income + Non-operating Expenses) / Latest Share Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero	CSMAR
	Earnings per Share TTM4	(Net Profit - Non-operating Income + Non-operating Expenses) TTM / Latest Share Capital. It will be denoted by Null if the denominator is not disclosed or equal to zero	CSMAR
	EPS estimate GAAP	Earnings Per Share GAAP. Analysts calculate EPS according to Generally Accepted Accounting Principles, as reported by companies. Earnings Per Share Excluding Exceptional Items. Most analysts define EPS Normalized as the adjusted EPS for pre-and after-tax unusual gains or losses (Exceptional Items) divided by the number of Diluted Weighted Average Shares Outstanding (WASO).	Capital IQ
	actual EPS (GAAP)		
	EPS normalised estimate		
	EPS normalised actual		
	Beta 1	Beta1: sensitivity of accounting earnings to “good news”	CSMAR
	Beta2	Beta2: sensitivity of accounting earnings to “bad news” over “good news”	CSMAR
	Total debts		Thomson Reuters
	Beta for 5 years	“CAPM Beta. A measure of how much the stock moves for a given move in the market. It is the covariance of the security's price movement in relation to the market's price movement. Based on data availability, various look back periods can be used to calculate it.”	Thomson Reuters
	Yearly Beta Value of Composite Market		
	alpha based on regional ranking	The regional 1-100 percentile rank based on StarMine Combined Alpha Model, which combines the available StarMine models in the optimal, static, linear weighting scheme for each region. Scores of 100 indicate the highest rated stocks in the region; scores of 1 indicate the lowest rated stocks in the region. StarMine models included in the Combined Alpha Model are: Analyst Revisions, Earnings Quality, Relative Valuation, Intrinsic Valuation, Price Momentum, Smart Holdings, Insider Filings (US only), and Short Interest (US only).	
	alpha based on country ranking	The country level 1-100 percentile rank based on StarMine Combined Alpha Model, which combines the available StarMine models in the optimal, static, linear weighting scheme for each region. Scores of 100 indicate the highest rated stocks in the country; scores of 1 indicate the lowest rated stocks in the country. StarMine models included in the Combined	

		Alpha Model are: Analyst Revisions, Earnings Quality, Relative Valuation, Intrinsic Valuation, Price Momentum, Smart Holdings, Insider Filings (US only), and Short Interest (US only).	
	alpha based on sector ranking	The sector level 1-100 percentile rank based on StarMine Combined Alpha Model, which combines the available StarMine models in the optimal, static, linear weighting scheme for each region. Scores of 100 indicate the highest rated stocks among firms in the same region and TRBC Economic Sector; scores of 1 indicate the lowest rated stocks in the same region and TRBC Economic Sector. The TRBC Economic Sector is analogous to the GICS sector. StarMine models included in the Combined Alpha Model are: Analyst Revisions, Earnings Quality, Relative Valuation, Intrinsic Valuation, Price Momentum, Smart Holdings, Insider Filings (US only), and Short Interest (US only).	Thomson Reuters
	alpha based on industry ranking	The industry level 1-100 percentile rank based on StarMine Combined Alpha Model, which combines the available StarMine models in the optimal, static, linear weighting scheme for each region. Scores of 100 indicate the highest rated stocks among firms in the same region and TRBC Industry Group; scores of 1 indicate the lowest rated stocks in the same region and TRBC Industry Group. The TRBC Industry Group is analogous to the GICS industry. StarMine models included in the Combined Alpha Model are: Analyst Revisions, Earnings Quality, Relative Valuation, Intrinsic Valuation, Price Momentum, Smart Holdings, Insider Filings (US only), and Short Interest (US only).	
Corporate governance	The number of executives	the total number executives in a company	CSMAR database
	The number of non-executives	the total number non-executives in a company	
	The number of directors	the total number of directors in a company	
	The number of independent directors	the total number of independent directors in a company	
	The number of non-independent directors	the total number of non-independent directors in a company	
	Board meeting	the frequency of board meetings	
	Institutional ownership	the number of shares of the company owned by institutional shareholders	
	Insider ownership	the number of shares owned by insider shareholders	
	Shares owned by top 5 shareholders	Proportion of shares held by top 5 and top 10 shareholders from the shareholder database	
	Shares owned by top 10 shareholders		
ESG performance	ESG score	It covers environmental, social, and governance scores.	SynTao Green Finance

	Environmental score	It covers environmental policy, energy& resource consumption, pollutant discharge, climate actions, and biodiversity.		
	Social score	It covers human capital development, supply chain management, customer rights, product management, data security, and community.		
	Governance score	It covers business ethics, governance structure and compliance.		
	ESG management score	It measures corporate active ESG management, by weighting environmental, social, and governance management indicators. The higher the score, the better the ESG management.		
	ESG risk score	It weights environmental, social and governance risk indicators. The higher the score, the higher the risk level.		
Other variables	Firm size	market capitalisation	CSMAR & manual classification	
	Audit assurance	CSMAR: In corporate social responsibility section -->General CSR information --> auditor from Big four accounting firms or not	CSMAR	
	Industry background	from CSMAR database: The industry classification of listed companies promulgated by the China Securities Regulatory Commission Manually further classify the companies based on the GICS 11 section guideline		
	Analyst recommendation	"Capital IQ: It is the opinion expressed by the analysts on the stock performance over a short-term period. Capital IQ provides both the Broker Recommendations and Standardized Recommendations. Capital IQ will standardize all the individual recommendations across the brokers on a 1- 5-point scale 1: Buy recommendation 2: outperform recommendation 3: Hold recommendation 4: Underperform recommendation 5: Sell recommendation".	Capital IQ database	
	Financial risk	Altman Z score		
		Credit rating: high risk vs low risk (binary variable)		S&P Global database
		Audit opinion: Going concern opinion in financial reports information related to going concern problem: Significant uncertainties relating to going concern Material uncertainties relating to going concern Basis for forming qualified opinion None		CSMAR
	CDP reporting	CDP scores-- the CDP score that the company acquires		CDP database
COVID-19 Pandemic	whether COVID-19 happened in a particular year			

Appendix 11 Post hoc tests

This thesis conducted post hoc tests. It is widely employed as additional tests for Kruskal Wallis test to see whether significant differences can be identified in homogeneous groups. This thesis intends to observe the significant variations of SDG disclosures based on homogeneous year groups.

Table 8.2 shows the post hoc test results for overall SDG disclosures —whether significant differences can be detected when disclosures are grouped by firm-year. Among all paired comparisons, significant differences have been observed in SDG disclosure variations between 2016 and other years respectively, based on ‘sig’ column²⁷. However, the Bonferroni correction substantially affects the interpretation of significance, where some significant results become insignificant after adjustments. Table 8.2 displays significant differences between 2016 and 2020, with the significance level at 0.04. That is, a significant difference has only been identified in the distributions of total SDG disclosures between 2016 and 2020.

Table 8.2 Post hoc test for overall SDG disclosures

	Test Statistic	Std.Error	Std.Test Statistic	Sig.	Adj.Sig. ^a
2016-2017	-17.41	15.74	-1.11	0.27	1.00
2016-2018	-27.03	15.74	-1.72	0.09	0.86
2016-2019	-35.50	15.74	-2.26	0.02	0.24
2016-2020	-45.89	15.74	-2.92	0.00	0.04
2017-2018	-9.63	15.74	-0.61	0.54	1.00
2017-2019	-18.09	15.74	-1.15	0.25	1.00
2017-2020	-28.48	15.74	-1.81	0.07	0.70
2018-2019	-8.47	15.74	-0.54	0.59	1.00
2018-2020	-18.86	15.74	-1.20	0.23	1.00
2019-2020	-10.39	15.74	-0.66	0.51	1.00

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .050.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Table 8.3 displays the post hoc test for poverty alleviation disclosures. Consistent with overall SDG disclosures, significant differences only occur between 2016 and 2020, with the significance level at 0.04. This indicates that a significant difference has been identified in the distribution of poverty alleviation disclosures between 2016 and 2020.

²⁷ Specifically, there is significant difference between 2016 and 2018, with the significance level at 0.1 (sig = 0.086). This indicates that the distribution of total SDG disclosures in 2016 is significantly different from the distribution in 2018. Similarly, with a value of 0.024 (at 0.05 significance level), there is a significant difference between the distributions of SDG disclosures in 2016 and in 2019. The significance level is at 0.004 between 2016 and 2020. Overall, among all paired comparisons, significant differences have only been identified between the distribution of SDG disclosures in 2016 and that of the rest of years, with the significance level becoming stronger over time.

Table 8.3 Post hoc test for poverty alleviation disclosure

	Test Statistic	Std.Error	Std.Test Statistic	Sig.	Adj.Sig. ^a
2016-2017	-20.76	15.75	-1.32	0.19	1.00
2016-2018	-29.46	15.75	-1.87	0.06	0.62
2016-2019	-36.57	15.75	-2.32	0.02	0.20
2016-2020	-45.96	15.75	-2.92	0.00	0.04
2017-2018	-8.69	15.75	-0.55	0.58	1.00
2017-2019	-15.80	15.75	-1.00	0.32	1.00
2017-2020	-25.19	15.75	-1.60	0.11	1.00
2018-2019	-7.11	15.75	-0.45	0.65	1.00
2018-2020	-16.50	15.75	-1.05	0.30	1.00
2019-2020	-9.39	15.75	-0.60	0.55	1.00

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Table 8.4 displays the post hoc test for energy disclosures during the period between 2016 and 2020. Consistent with overall SDG disclosures and poverty alleviation disclosures, significant differences in energy disclosures can be observed between 2016 and 2020, after the Bonferroni correction.

Table 8.4 Post hoc test for energy disclosures

	Test Statistic	Std.Error	Std.Test Statistic	Sig.	Adj.Sig. ^a
2016-2017	-17.90	15.33	-1.17	0.24	1.00
2016-2018	-30.22	15.33	-1.97	0.05	0.49
2016-2019	-34.77	15.33	-2.27	0.02	0.23
2016-2020	-45.28	15.33	-2.95	0.00	0.03
2017-2018	-12317.00	15.33	-0.80	0.42	1.00
2017-2019	-16.87	15.33	-1.10	0.27	1.00
2017-2020	-27.38	15.33	-1.79	0.07	0.74
2018-2019	-4.55	15.33	-0.30	0.77	1.00
2018-2020	-15.07	15.33	-0.98	0.33	1.00
2019-2020	-10.52	15.33	-0.69	0.49	1.00

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Table 8.5 displays the post hoc test for climate change disclosures during the period between 2016 and 2020. Significant differences in the distribution of climate change disclosures between 2016 and the rest of years (2019, and 2020 respectively). Specifically, with a p-value at 0.02, there is a significant

difference in the distribution of climate change disclosures between 2016 and 2019. Similarly, the p-value for the 2016-2020 group is at 0.01, indicating that a significant difference has been identified in climate change disclosures between 2016 and 2020. Therefore, the test suggests significant variation on climate change disclosures between 2016 and 2019, as well as between 2016 and 2020. This may indicate that after the introduction of SDGs, corporate climate change disclosures have significant changed in 2019 and 2020, compared to 2016.

Table 8.5 Post hoc test for climate change disclosures

	Test Statistic	Std.Error	Std.Test Statistic	Sig.	Adj.Sig. ^a
2016-2017	-17.19	15.49	-1.11	0.27	1.00
2016-2018	-35.70	15.49	-2.31	0.02	0.21
2016-2019	-48.84	15.49	-3.15	0.00	0.02
2016-2020	-52.39	15.49	-3.38	<.001	0.01
2017-2018	-18.51	15.49	-1.20	0.23	1.00
2017-2019	-31.65	15.49	-2.04	0.04	0.41
2017-2020	-35.20	15.49	-2.27	0.02	0.23
2018-2019	-13.14	15.49	-0.85	0.40	1.00
2018-2020	-16.69	15.49	-1.08	0.28	1.00
2019-2020	-3.55	15.49	-0.23	0.82	1.00

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .050.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix 12 Relative Variable Importance Scores

Variable	RVI score	
TOTAL_DEBT	100.00	
BETA_FOR_5_YEARS	99.83	
RATIO_OF_NET_PROFIT_TO_TOTAL_PROFIT	98.98	
RATIO_OF_FIXED_ASSETS	96.91	
YEARLY_BETA_VALUE_OF_COMPOSITE_MARKET	96.36	
ANALYST_RECOMMENDATION_CAPITALIQ	90.91	
CASHFLOW_FROM_FINANCING	89.47	
TOBINS_Q	87.17	
CASHFLOW_FROM_INVESTING	87.04	
RETURN_ON_EQUITY	86.94	
SYNTAOGREENFINANCE_ESGSCORE	86.40	
CASHFLOW_FROM_OPERATIONS	85.24	
TOP5_HOLDING_PERCENTAGE_CSMAR	85.07	
INSITUTIONAL_SHAREHOLDING_CSMAR	84.27	
ENTERPRISE_VALUE	84.18	
TOTAL_REVENUE	84.10	
SYNTAOGREENFINANCE_GSCORE	83.84	
RETURN_ON_ASSETS	83.40	
FIRM_SIZE_MARKET_CAPITALISATION	82.47	
SYNTAOGREENFINANCE_ESGMANAGEMENTSCORE	82.26	
DEBT_TO_LONG_TERM_CAPITAL	82.20	
RATIO_OF_CASH_ASSETS	81.65	
LIABILITIES_TO_TANGIBLE_ASSETS	81.61	
RATIO_OF_SHAREHOLDERS_EQUITY_TO_FIXED_ASSETS	81.38	
RATIO_OF_RETAINED_EARNINGS_TO_TOTAL_ASSETS	80.63	
INSIDER_SHAREHOLDING	80.58	
SHORT_TERM_LOANS	78.93	
CASH_AND_CASH_EQUIVALENTS	78.70	
RATIO_OF_INTANGIBLE_ASSETS	76.36	
SYNTAOGREENFINANCE_SSCORE	76.24	
TANGIBLE_ASSETS_RATIO	76.18	
NET_INTANGIBLE_ASSETS	76.16	

TOP10_HOLDING_PERCENTAGE_CSMAR	76.14	
RATIO_OF_RECEIVABLE_ASSETS	73.63	
RATIO_OF_ACCOUNTS_RECEIVABLE_TO_INCOME	72.95	
LONG_TERM_LOANS_TO_TOTAL_ASSETS	72.77	
NO_NON_INDEPENDENT_DIRECTORS	72.37	
RATIO_OF_TOTAL_PROFIT_TO_EBIT	70.70	
NET_CASHFLOW_FROM_OPERATING_ACTIVITIES_TOTAL_LIABILITIES	69.90	
SYNTAOGREENFINANCE_ESCORE	68.85	
NET_ACCOUNTS_RECEIVABLE	68.53	
THE_NUMBER_OF_BOARD_MEETING	68.40	
BETA2_CSMAR	67.60	
QUICK_RATIO	67.34	
WORKING_CAPITAL	67.22	
SUMBETA_CSMAR	66.98	
CURRENT_RATIO	65.91	
COMBINED_ALPHA_MODEL_INDUSTRY_RANK	65.67	
BETA1_CSMAR	64.19	
SYNTAOGREENFINANCE_ESGRISKSCORING	64.13	
GROSS_OPERATING_MARGIN	63.90	
BASIC_EARNINGS_PER_SHARE_AFTER_DEDUCTING_EXTRAORDINARY_ITEMS	63.28	
NO_DIRECTORS	63.12	
COMBINED_ALPHA_MODEL_REGION_RANK	63.09	
RATIO_OF_OPERATING_LIABILITIES	62.71	
COMBINED_ALPHA_MODE_COUNTRY_RANK	62.58	
TOTAL_OPERATING_COSTS	60.91	
BOOK_TO_MARKET_RATIO	60.07	
NO_NON_EXECUTIVE_DIRECTORS	59.72	
RATIO_OF_CURRENT_LIABILITIES	59.71	
TOTAL_LONG_TERM_LIABILITIES	59.47	
POLITICALCONNECTION_PARTY_COMMITTEE	59.15	
BASIC_EARNINGS_PER_SHARE	58.65	
EARNINGS_PER_SHARE_1	58.14	
LIABILITIES_TO_EQUITY_MARKETCAP	57.69	

TOTAL_PROFIT	57.11	
COMBINED_ALPHA_MODEL_SECTOR_RANK	56.70	
CASH_RATIO	56.65	
POLITICALCONNECTION_DIRECTOR_COUNTS	54.58	
POLITICALCONNECTION_CEO_EXECUTIVES_COUNTS	52.12	
LONG_TERM_DEBT_TO_WORKING_CAPITAL	52.03	
EARNINGS_PER_SHARE_3	51.86	
EV_TO_EBITDA	51.77	
TOTAL_OPERATING_REVENUE	51.36	
LIABILITIES_TO_ASSETS_RATIO	51.12	
EPS_ESTIMATE_GAAP	51.05	
TOTAL_LIABILITIES	50.74	
TOTAL_LEVERAGE	50.13	
NET_CASHFLOW_FROM_OPERATING_ACTIVITIES_TO_INTERESTBEARING_LIABILITIES	50.00	
NET_CASHFLOW_FROM_OPERATING_ACTIVITIES_TO_CURRENT_LIABILITIES	49.89	
RATIO_OF_CURRENT_ASSETS	49.81	
TOTAL_EQUITY	49.68	
TOTAL_CURRENT_ASSETS	49.65	
TOTAL_RECEIVABLES	49.55	
NET_PROFIT	49.36	
EPS_NORMALISED_ACTUAL	48.64	
NO_INDEPENDENT_DIRECTORS	48.45	
NET_INVENTORIES	47.12	
TOTAL_EQUITY_TO_PARENT_COMPANY	46.63	
POLITICALCONNECTION_OVERALL_COUNTS	46.45	
RATIO_OF_WORKING_CAPITAL_TO_NET_ASSETS	46.14	
TOTAL_CURRENT_LIABILITIES	45.90	
R_D_EXPENSES	45.34	
RATIO_OF_WORKING_CAPITAL	43.28	
EPS_NORMALISED_ESTIMATE	43.08	
RETURN_ON_LONG_TERM_CAPITAL	41.98	
OTHER_TYPES_OF_COMPANIES	41.88	
ALTMAN_ZSCORE	41.81	

EBITDA_TO_TOTAL_LIABILITIES	40.70	
RATIO_OF_NON_CURRENT_ASSETS	40.14	
TOTAL_ASSETS	39.78	
EARNINGS_PER_SHARE_2	39.63	
EARNINGS_PER_SHARE_4	39.58	
MARKET_VALUE_A	39.46	
TOTAL_LIABILITIES_TO_NET_TANGIBLE_ASSETS	39.43	
AUDIT_BY_BIG4	39.36	
GROSS_PROFIT	39.20	
EBITDA_CSMAR	38.80	
WORKING_CAPITAL_TO_LOANS	38.73	
TOTAL_NON_CURRENT_LIABILITIES	38.05	
TOTAL_SHAREHOLDERS_EQUITY	36.90	
BELT_AND_ROAD_INITIATIVE	34.94	
RATIO_OF_EBIT_TO_TOTAL_ASSETS	34.31	
NO_EXECUTIVE_DIRECTORS	34.03	
EQUITY_MULTIPLIER_Y	33.80	
GICS_CONSUMER_DISCRETIONARY	33.56	
EVALUATION_OF_ACCOUNTING_INFORMATION_QUALITY	33.35	
ACTUAL_EPS_GAAP	32.53	
DILUTED_EARNINGS_PER_SHARE	28.73	
EBIT_CSMAR	25.92	
CREDITRATING_HIGHER_LT	24.25	
PUBLIC_INTEREST_SOES	21.60	
CREDITRATING_HIGHER_ST	18.53	
CDP_REPORTING_LOWRANKING	17.52	
GICS_ENERGY	16.43	
COMMERCIAL_SOES	14.43	
COVID_19	11.62	
CDP_REPORTING_BINARY	11.44	
SDG_INNOVATION_DEMONSTRATION_ZONE	10.93	
GICS_INDUSTRIALS	9.79	
CREDITRATING_LOWER_ST	7.73	
GICS_INFORMATION_TECHNOLOGY	7.60	

GICS_CONSUMER_STAPLES	7.45	
GICS_MATERIALS	6.81	
GREEN_BOND_ISSUANCE_BINARY	6.30	
EARNINGS_PER_SHARE_TTM1	0.00	
GICS_UTILITIES	0.00	
EARNINGS_PER_SHARE_TTM2	0.00	
TYPE_OF_AUDIT_OPINION_UNQUALIFIED_OPINION_WITH_EMPHASIS_OF_MATTER_PARAGRAPH	0.00	
EARNINGS_PER_SHARE_TTM3	0.00	
TYPE_OF_AUDIT_OPINION_STANDARD_UNQUALIFIED_OPINION	0.00	
EARNINGS_PER_SHARE_TTM4	0.00	
GICS_FINANCIALS	0.00	
GOING_CONCERN_AUDIT_OPINION_BINARY	0.00	
GICS_REAL_ESTATE	0.00	
CREDITRATING_LOWER_LT	0.00	
RATIO_OF_OWNERS_EQUITY	0.00	
TYPE_OF_AUDIT_OPINION_QUALIFIED_OPINION_WITH_EMPHASIS_OF_MATTER_PARAGRAPH	0.00	
GICS_HEALTH_CARE	0.00	
POLITICALCONNECTION_DIRECTOR_BINARY	0.00	
POLITICALCONNECTION_CEO_EXECUTIVES_BINARY	0.00	
POLITICALCONNECTION_OVERALL_BINARY	0.00	
DEBT_TO_EQUITY	0.00	
EQUITY_TO_DEBT	0.00	
TYPE_OF_AUDIT_OPINION_DISCLAIMER_OF_OPINION	0.00	
GICS_COMMUNICATION_SERVICE	0.00	
CDP_REPORTING_HIGHRANKING	0.00	
RATIO_OF_WORKING_CAPITAL_TO_CURRENT_ASSETS	0.00	
TOTAL_LIABILITIES_AND_SHAREHOLDER_EQUITY	0.00	