DEVELOPING AN NLP-ENHANCED SUSTAINABILITY BALANCED SCORECARD FOR CONSTRUCTION COMPANIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ATEFEH AALI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

SEPTEMBER 2024

Approval of the thesis:

DEVELOPING AN NLP-ENHANCED SUSTAINABILITY BALANCED SCORECARD FOR CONSTRUCTION COMPANIES

submitted by **ATEFEH AALI** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Erdem Canbay
Head of Department, **Civil Engineering** _____

Assoc. Prof. Dr. Güzide Atasoy Özcan
Supervisor, **Civil Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Rifat Sönmez
Civil Engineering, METU _____

Assoc. Prof. Dr. Güzide Atasoy Özcan
Civil Engineering, METU _____

Assist. Prof. Dr. Aslı Akçamete Güngör
Civil Engineering, METU _____

Assoc. Prof. Dr. Hande Işık Öztürk
Civil Engineering, METU _____

Assist. Prof. Dr. Saman Aminbakhsh
Civil Engineering, Atılım University _____

Date:05.09.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Atefeh Aali

Signature :

ABSTRACT

DEVELOPING AN NLP-ENHANCED SUSTAINABILITY BALANCED SCORECARD FOR CONSTRUCTION COMPANIES

Aali, Atefeh
M.S., Department of Civil Engineering
Supervisor: Assoc. Prof. Dr. Güzide Atasoy Özcan

September 2024, 102 pages

Due to global challenges, construction companies have implemented sustainable practices and reported their contributions towards sustainable development through annual sustainability reports. Aligning sustainability initiatives with strategic objectives and continuously monitoring progress is crucial for informed decision making. However, the lack of a comprehensive framework to evaluate construction companies' sustainability performance has been a significant obstacle. This research introduces the sustainability balanced scorecard (SBSC), an innovative framework for evaluating sustainability performance in construction industry.

The SBSC enhances the traditional balanced scorecard (BSC) by expanding it according to the triple bottom line principles. The absence of widely accepted or standardized indicators makes it difficult to identify and integrate the SBSC category indicators. Hence, to populate these aspects, sustainability reports were used to analyze key sustainability topics in the construction sector. The lack of a standardized format in these reports necessitated integrating natural language processing (NLP) techniques to streamline and improve their analysis. This study utilizes BERTopic, an unsupervised machine learning algorithm, to perform topic modeling and uncover the most

common corporate social responsibility (CSR) topics and practices discussed in the sustainability reports of top international contractors. The topics and their representative keywords were then used to create a strategy map as a foundation for an SBSC and define metrics to achieve sustainability in the construction sector. The suggested tool was then verified in a case study on a sample Turkish company.
The research findings will assist companies in publishing more comprehensive sustainability reports and provide managers with a novel performance evaluation tool designed for the construction industry.

Keywords: Sustainability Balanced Scorecard, Topic Modeling, BERTopic, Sustainability Report, Natural Language Processing

ÖZ

İNŞAAT ŞİRKETLERİ İÇİN NLP DESTEKLİ SÜRDÜRÜLEBİLİRLİK DENGELİ SKORKARTI GELİŞTİRİLMESİ

Aali, Atefeh
Yüksek Lisans, İnşaat Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Güzide Atasoy Özcan

Eylül 2024 , 102 sayfa

Küresel zorluklar nedeniyle inşaat şirketleri, sürdürülebilir uygulamaları hayata geçirmiş ve sürdürülebilir kalkınmaya olan katkılarını yıllık sürdürülebilirlik raporları aracılığıyla raporlamışlardır. Sürdürülebilirlik girişimlerini stratejik hedeflerle uyumlu hale getirmek ve ilerlemeyi sürekli olarak izlemek, bilinçli kararlar almak için önemlidir. Ancak, inşaat şirketlerinin sürdürülebilirlik performansını değerlendirecek kapsamlı bir çerçevenin eksikliği önemli bir engel olmuştur. Bu araştırma, inşaat sektöründe sürdürülebilirlik performansını değerlendirmek için yenilikçi bir çerçeve olan sürdürülebilirlik dengeli skorkartını (SBSC) tanıtmaktadır.

SBSC, üçlü sorumluluk kavramı ilkelerine göre genişleterek geleneksel dengeli skorkartı geliştirir. Geniş çapta kabul görmüş veya standartlaştırılmış göstergelerin olmaması, SBSC kategorileri göstergelerini belirlemeyi ve entegre etmeyi zorlaştırmaktadır. Bu alanları doldurmak için inşaat sektöründeki önemli sürdürülebilirlik konularını analiz etmek amacıyla sürdürülebilirlik raporları kullanılmıştır. Bu raporlarda standart bir formatın bulunmaması, analizleri kolaylaştırmak ve iyileştirmek için doğal dil işleme (NLP) tekniklerinin entegrasyonunu gerektirmiştir. Bu çalışma, üst düzey

uluslararası müteahhitlerin sürdürülebilirlik raporlarında tartışılan en yaygın kurumsal sosyal sorumluluk konularını ve uygulamalarını ortaya çıkarmak için gözetimsiz bir makine öğrenmesi algoritması olan BERTopic'i kullanarak konu modellemesi gerçekleştirmektedir. Konular ve bunların temsilci anahtar kelimeleri, SBSC için bir strateji haritası oluşturmak ve inşaat sektöründe sürdürülebilirliği sağlamak için metrikler tanımlamak üzere kullanılmıştır. Önerilen araç, örnek bir Türk şirketi üzerinde bir vaka çalışması ile doğrulanmıştır.
Araştırma bulguları, şirketlerin daha kapsamlı sürdürülebilirlik raporları yayınlamalarına yardımcı olacak ve yöneticilere inşaat sektörü için tasarlanmış yeni bir performans değerlendirme aracı sunacaktır.

Anahtar Kelimeler: Sürdürülebilirlik Dengeli Skorkartı, Konu Modelleme, BERTopic, Sürdürülebilirlik Raporu, Doğal Dil İşleme

To My Beloved Family

ACKNOWLEDGMENTS

The work presented in this thesis would not have come to an end without the support of many people. I would like to express my gratitude and appreciation to all the people who assisted me directly or indirectly during this process and made this thesis possible.

First and foremost, I would like to express my deep and sincere gratitude to my precious advisor, Assoc. Prof. Dr. Güzide Atasoy Özcan for her valuable feedback, patience, and continuous support. I learned a lot from her and feel privileged to have had the opportunity to learn under her mentorship during my thesis journey.

I would also like to thank Prof. Dr. İrem Dikmen Toker and Dr. Açelya Ecem Yıldız for their valuable insights and comments throughout this thesis.

I owe my deepest gratitude to my loving parents and my dear sister for always being there for me and providing me with their unlimited kindness and support. Their belief in me has motivated me to overcome the challenges, and I am forever grateful to have such a wonderful family by my side along the way.

Last but not least, I would like to thank my committee members for their constructive feedback and suggestions that have greatly improved the quality of my work.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiv

LIST OF FIGURES xv

LIST OF ABBREVIATIONS xvii

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation and Problem Definition 1

1.2 Scope of the Research 3

1.3 Research Questions 5

1.4 Contributions and Novelties 6

1.5 The Outline of the Thesis 6

2 LITERATURE REVIEW 9

2.1 Balanced Scorecard 10

2.1.1 Balanced Scorecard Architecture 10

2.1.2 Balanced Scorecard in Construction Companies 12

2.2 Sustainability Balanced Scorecard 14
2.2.1 Sustainability Balanced Scorecard Architectures 15
2.2.2 Sustainability Balanced Scorecard in Construction 16
2.3 Sustainability in Construction 17
2.3.1 Most Common Standards in Sustainability Reporting 19
2.4 Natural Language Processing and its Applications in Construction . . 20
2.4.1 Analyzing Sustainability Reports With NLP In Sectors Other than Construction 21
2.4.2 Topic Modeling 23
3 METHODOLOGY 27
3.1 Dataset 29
3.1.1 Data Collection 29
3.1.2 Data Preprocessing 30
3.2 Topic Modeling with BERTopic 31
3.2.1 Embeddings 32
3.2.2 Dimensionality Reduction 33
3.2.3 Clustering 33
3.2.4 Tokenizer 33
3.2.5 Weighting Scheme and Topic Representations 34
3.3 Validation and Verification of the Results 35
3.3.1 Categorization of the Topics and Creating a Strategy Map . . . 35
3.3.2 Case study on a Turkish Construction Company 37
4 RESULTS 39

4.1 Topic Modeling Results 40
4.1.1 Discussion of the Initial Topics List 40
4.1.2 Merging the Similar Topics 49
4.2 The Creation of the Sustainability Balanced Scorecard (SBSC) 54
4.2.1 The Initial SBSC for Sustainable Construction 54
4.2.2 The proposed strategy map for achieving sustainable construction 57
4.2.3 Suggesting Key Performance Indicators (KPIs) for the SBSC . 60
4.3 Case Study 66
4.3.1 Customizing the SBSC for a Turkish Construction Company . 67
5 CONCLUSIONS 73
5.1 Overview 73
5.2 Contributions 75
5.3 Limitations and Future Research Directions 76
REFERENCES 79
APPENDICES
A. ENR's top 250 international contractors 91
B. The extended stop words list 96
C. Topic Word Scores 99

LIST OF TABLES

TABLES

Table 4.1 Topic Modeling Results 41

Table 4.2 Merging the topics to define general themes 51

Table 4.3 The Sustainability category of the SBSC and suggested KPIs 62

Table 4.4 The External Stakeholders category of the SBSC and suggested KPIs 63

Table 4.5 The Internal Process category of the SBSC and suggested KPIs . . . 64

Table 4.6 The Learning and Growth category of the SBSC and suggested KPIs 65

LIST OF FIGURES

FIGURES

Figure 1.1 Generic Balanced Scorecard [1] 2

Figure 2.1 Generic BSC adapted from [2] 10

Figure 2.2 A simplified form of generic BSC adapted from [3] 11

Figure 2.3 Example of a strategy map for a sample company derived from [4] 13

Figure 2.4 Different architectures of SBSC according to [5] 16

Figure 2.5 Publications of NLP in construction derived from [6] 21

Figure 3.1 Research methodology 28

Figure 3.2 Origin of input reports 30

Figure 3.3 BERTopic algorithm 32

Figure 4.1 Intertopic distance map 46

Figure 4.2 Hierarchical clustering 48

Figure 4.3 The most dominant themes in sustainable construction 53

Figure 4.4 The SBSC for sustainable construction 56

Figure 4.5 The proposed strategy map for achieving sustainable performance 59

Figure 4.6 The proposed strategy map for the sample company 68

Figure 4.7 SBSC for the sample company 71

Figure A.1 ENR's top 250 international contractors - Part 1 91

Figure A.2 ENR's top 250 international contractors - Part 2 92

Figure A.3 ENR's top 250 international contractors - Part 3 93

Figure A.4 ENR's top 250 international contractors - Part 4 94

Figure A.5 ENR's top 250 international contractors - Part 5 95

Figure C.6 Topic Word Scores Part 1 99

Figure C.7 Topic Word Scores Part 2 100

Figure C.8 Topic Word Scores Part 3 101

Figure C.9 Topic Word Scores Part 4 102

LIST OF ABBREVIATIONS

ABBREVIATIONS

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
BS	Balanced Scorecard
CE	Civil Engineering
CSR	Corporate Social Responsibility
ELMo	Embeddings from Language Models
ENR	Engineering News Record
EPR	Extended Producer Responsibility
ESG	Environmental, Social, and Governance
GloVE	Global Vectors for Word Representation
GRI	Global Reporting Initiative
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
ML	Machine Learning
NLP	Natural Language Processing
NMF	Non-negative Matrix Factorization
SBIC	Standard-Based Impact Classification
SBSC	Sustainability Balanced Scorecard
SDGs	Sustainable Development Goals
SMEs	Small and Medium-sized Enterprises

TBL	Triple Bottom Line
UMAP	Uniform Manifold Approximation and Projection for Dimension Reduction
UN	the United Nations
UNGC	United National Global Compact
WOS	Web Of Science

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

The idea of sustainable development has become increasingly popular over the past few decades in a variety of global industries, and many businesses have started to actively incorporate sustainability principles into their strategic decision-making processes [7]. The construction industry is no exception to this trend. The building industry is notorious for its adverse global effects [8].

According to the UN Environment Program report in 2023, the construction industry is responsible for 37 percent of global emissions, making it by far the largest emitter of greenhouse gases [9]. Therefore, the integration of sustainability factors into construction companies' strategic management, along with their financial goals, can significantly influence their ability to take the lead in the industry. In order to measure their sustainability performance and contribute to sustainable development, companies need an effective tool [10][11].

Balanced Scorecard (BSC), introduced by Kaplan and Norton in 1996 [12] [13], is a popular performance measurement method consisting of four primary categories: financial, customer, internal processes, and learning and growth (See Figure 1.1). These categories are directly in line with the strategic objectives of the companies [14]. The foundation for a BSC is a strategy map that demonstrates the strategic objectives of the organization [15]. There is a cause-and-effect relationship between these strategic objectives, and achieving them helps the business implementing the BSC succeed in its overall strategy and vision [1]. The success of the BSC highly depends on the way it connects the four categories in a casual chain to achieve the overall strategy [16].

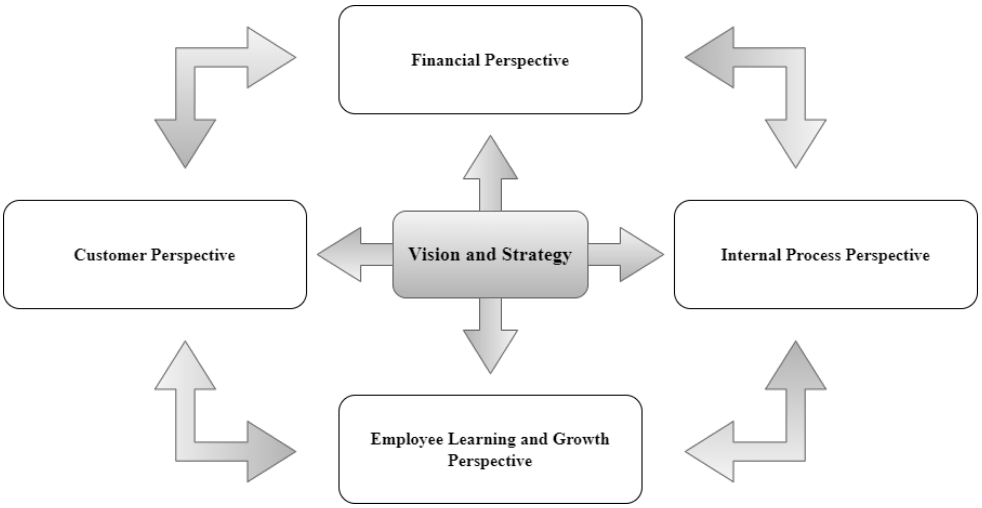

Figure 1.1: Generic Balanced Scorecard [1]

This framework has been introduced as a highly effective tool to evaluate the performance of companies in the construction industry [15]. However, in traditional BSC, the effect of sustainability in performance evaluation has not been taken into account [10]. As a result, many researchers have proposed different integration methods of sustainability principles into the generic BSC framework to build Sustainability Balanced Scorecards (SBSCs) [17], but the introduced SBSCs in the literature are not specific to the construction industry.

Some researchers have modified the BSC for specific purposes, such as promoting the circular economy [15] or mitigating delays in the construction industry [18]. However, the design of an SBSC for the sustainability performance evaluation of construction companies remains a gap that needs to be studied. It is a challenge to define strategic objectives and Key Performance Indicators (KPI) to evaluate the overall sustainability performance of a construction company. The top construction organizations have taken the lead in declaring their global impact in their sustainability reports, and some specific and common topics are discussed within these documents. However, while these voluminous reports are often based on the Global Reporting Initiative (GRI) standards [19], there is not a single formatting and reporting trend that each company follows in publishing them, and the content of these reports typically differs from company to company.

The goal of this research is to develop an SBSC tailored especially for the construction

sector, building up on the sustainability topics focused on by the major construction companies. Such an approach will allow construction companies to align their strategic objectives in relation to sustainability and assess their sustainability performance. The research findings will aid businesses in contributing to sustainable development by lowering their annual negative effect and tracking their performance through effective sustainability reporting. The practical implications for managers include effective means to set sustainability objectives and a cutting-edge performance assessment tool created especially for the construction sector.

1.2 Scope of the Research

In response to the lack of a comprehensive and standardized tool for performance evaluation of the construction companies that takes the sustainability principles into account, the present study attempts to develop a Sustainability Balanced Scorecard specifically for the construction sector. To achieve this objective, the best practices followed by top global construction companies were explored. Since these companies prefer to publish their contributions in their sustainability reports, these documents were considered valuable resources for finding the most significant sustainability themes and topics discussed in the construction sector. These reports range from tens to hundreds of pages, and they focus on various and mostly common topics such as waste and water management, diversity and inclusion, social responsibility activities, and employee training. The sustainability reports are often based on the Global Reporting Initiative (GRI) standards [19]. However, although adhering to this standard results in semi-structured reports, there is no uniform formatting or reporting trend that all companies follow when publishing them. Instead, the content of these reports varies from company to company depending on their focus areas, the data they currently measure, or the subjects they wish to have more emphasis on when declaring their sustainability impact.

The analysis of the contents of these documents can reveal insights from their textual data. However, since the manual analysis of these voluminous reports is time-consuming and prone to error and bias, the utilization of text mining was suggested. Text mining refers to analyzing large collections of structured or unstructured textual

data to find relevant information that might be hidden within the text [6] [20]. In this study, Natural Language Processing (NLP), which is a component of text mining, was used to analyze the textual information in the sustainability reports.

In this process, the original BSC framework was adapted by incorporating the three principles of the "Triple Bottom Line" - environmental, financial, and social aspects. In order to identify the strategic objectives for each of the SBSC categories, an unsupervised topic modeling approach was employed to analyze the text content of sustainability reports published by the top international contractors according to Engineering News-Record (ENR).

For doing the topic modeling task, BERTopic [21] that is based upon the Bidirectional Encoder Representations from Transformers (BERT) architecture [22], was chosen due to its promising performance in comparison to older topic modeling techniques like Latent Dirichlet Allocation (LDA) [23], Non-negative Matrix Factorization (NMF) [24], and Top2Vec [25] [26]. Furthermore, in contrast to other approaches, BERTopic does not require the determination of a limit for the hyperparameter that defines the number of topics during the topic modeling, which leads to a more thorough and in-depth text analysis for identifying the hidden themes and topics within the dataset [27].

This topic modeling approach resulted in the identification of 90 distinct topics within the corpus of sustainability reports. These findings were then presented to a small focus group of specialists and academics in construction management and sustainability in order to aggregate related subjects and find the most dominant topics discussed for achieving sustainable construction. The focus group then used these subjects to develop a strategy map that will help construction companies achieve sustainability in their plans and operations. This strategic map was then used as the foundation for designing and implementing a SBSC for the construction companies. The key performance indicators (KPIs) for the suggested SBSC template were also determined using the representative keywords of the topics found during the topic modeling process and the recommendations from the expert group. According to the literature, every organization determines its unique objectives in its SBSC and has its own metrics to reach them [28]. Therefore, the strategic objectives and their suggested KPIs in the proposed SBSC template served merely as a base and reference and, therefore, could be customized for different companies with different strategies and capabilities.

To verify the applicability of this framework, a case study was conducted on a construction firm in Türkiye. In this regard, the SBSC was customized for this company according to the topics discussed in their sustainability report. The management of the organization provided input on the framework's applicability and usefulness and specified their metrics in line with the overall strategy map of sustainable constructions. The managers stated that this SBSC tool provided a highly effective, valuable and practical framework for classifying and arranging their strategies while taking the most important sustainability issues into consideration. They added that this framework has the potential to be used globally and in a variety of settings since it offers a solid foundation for attaining sustainability in the construction sector and can be customized to each company's strategic objectives and vision.

1.3 Research Questions

This thesis aims to address the following research questions:

1. *What are the prevailing topics in the sustainability reports of construction companies that can serve as strategic objectives for the SBSC perspectives in the construction industry?*

   This question was addressed by employing BERTopic, a topic modeling technique, to analyze the content of the sustainability reports of the leading construction companies in order to identify the most dominant themes and topics discussed within these documents. The topics were then evaluated, agreed and categorized to construct the various SBSC sections with the guidance of an expert group.

2. *How can an SBSC framework be developed using the prevailing topics identified from sustainability reports?*

   In order to address this question, a strategy map was developed with the expert group using strategic objectives that corresponded with the recognized topics to serve as a foundation for a SBSC. The most representative keywords of the topics and the expert recommendations were also used to define metrics to achieve sustainability in the construction sector. These objectives and metrics could be

customized by each company according to their specific goals and indicators. The suggested tool was then verified through a case study on a sample Turkish company.

1.4 Contributions and Novelties

The contributions of this research are as follows:

- Application of NLP for topic modeling on sustainability reports of construction companies. In this way, the most common themes and topics related to sustainable construction are uncovered, delineated, and categorized.
- Integration of sustainability to original BSC framework and introduction of a novel SBSC, specifically designed for the construction industry. In this way, construction companies can align their strategic objectives with the way leading construction companies communicate their sustainable practices while performing customized performance evaluations.

1.5 The Outline of the Thesis

This thesis consists of five chapters. The motivation and problem definition, the scope of the research, the research questions, and the contribution to the body of knowledge are explained in brief in the first chapter. The second chapter covers the previous literature on the Balanced Scorecard (BSC) and Sustainability Balanced Scorecard (SBSC), their architectures and applications in the construction industry and other fields, sustainability in the industry, and the most widely used sustainability reporting standards. This chapter then dives into the use of Natural Language Processing (NLP), a branch of Machine Learning (ML), in construction-related documents.
The third chapter explains the methodology used to implement this study. Every aspect of data collection, data preprocessing, the different steps of topic modeling with BERTopic, and categorizing and merging the topics within a focus group to create a strategy map for achieving sustainability in the construction industry is covered in detail. Ultimately, this strategy map is used as a foundation to create an SBSC for a

Turkish construction company to validate and verify the applicability and usefulness of the final framework on feedback from the management division of the business.

The research findings, including topic modeling results using BERTopic and related visualizations such as topic word scores, hierarchical clustering, and intertopic distance map, are presented and discussed in the fourth chapter. A focus group comprised of experts finalized the design of a strategy map for achieving sustainable performance within the construction industry. The most representative keywords of the revealed topics were used in combination with expert judgment to prepare a list of suggested key performance indicators (KPIs) for the SBSC objectives. This chapter also includes the outcomes of customizing the SBSC for a Turkish construction business with reference to the strategy map, their own metrics, and their own strategies according to the subjects discussed within their sustainability reports. Finally, feedback from the management of the construction company was taken to get their thoughts and insights on the proposed framework and how the outcomes of this tool may help them in shaping their strategies to improve their sustainability performance.

The final chapter includes the conclusion of this research. The overview of the study and its key findings are summarized in this part. Additionally, this chapter states the limitations of the research and provides an outline for future studies.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a comprehensive overview of the present research setting on the Balanced Scorecard (BSC), Sustainability Balanced Scorecard (SBSC), its architectures and applications in the construction industry, sustainability in the industry, and the most widely used sustainability reporting standards. This chapter then explains the use of Natural Language Processing (NLP) in construction-related documents and covers the different topic modeling approaches and how they are used for text analysis purposes.

The literature review process was conducted in two distinct stages. In the initial phase, relevant research articles were identified and collected through systematic searches of databases such as "Scopus," "Web of Science (WOS)," and "Google Scholar." Different combinations of relevant terms were used in the search process. Subsequently, the retrieved articles were categorized into multiple groups according to their titles and research objectives. The first section of the chapter gives a short overview of the BS's development and its different perspectives and applications within the construction industry. A description of the SBSC literature and its different architectures and applications comes next. Furthermore, the development of the sustainable construction concept is investigated, including an examination of the idea of sustainable reporting in the construction industry and the most common sustainability reporting standards. Finally, the chapter explores the broad concept of NLP and its various applications in Civil Engineering (CE), especially with regard to sustainability reports in the construction industry as well as other sectors. The chapter concludes with a look at using topic modeling as an NLP technique for text analysis.

2.1 Balanced Scorecard

Firstly introduced by Robert Kaplan, a university professor at Harvard and David Norton, a specialist and consultant in performance measurement and strategy management [1], the Balanced Scorecard (BSC) has been extensively used as a strategic management tool in different fields, and it has been one of the most influential ideas of the twentieth century according to Harvard Business Review [29]. According to [29], BSC can be viewed as a variety of tools, including a measurement system, a strategic management system, and a communication tool.

2.1.1 Balanced Scorecard Architecture

The original BSC consists of four main aspects: Financial, Customer, Internal Process, and Learning and Growth. Each of these aspects contains strategic objectives or goals, measures or key performance indicators (KPIs), targets, and related initiatives in order to achieve those goals. The generic BSC architecture is depicted in Figure 2.1.

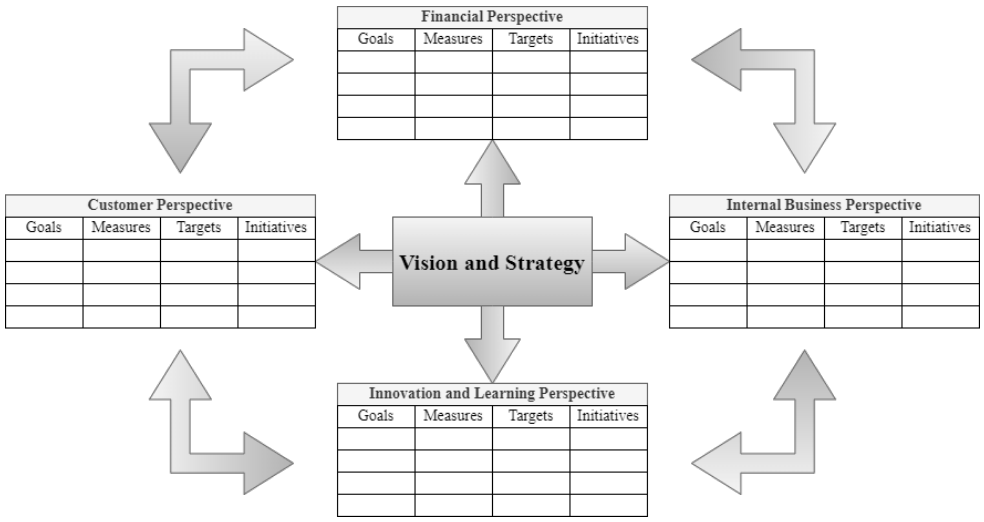

Figure 2.1: Generic BSC adapted from [2]

In order to fill each perspective, the company should answer some questions [3]. These questions are depicted in Figure 2.2. It should also be noted that each per-

spective on the Balanced Scorecard would be represented by a combination of performance drivers (lead measures) and core outcome (lag) measures. The results of previous actions are the lag indicators. The performance indicators that drive the achievement of the lag indicators are known as lead indicators. For instance, for the lagging measure of customer satisfaction, on-time delivery may serve as a leading indicator [29].

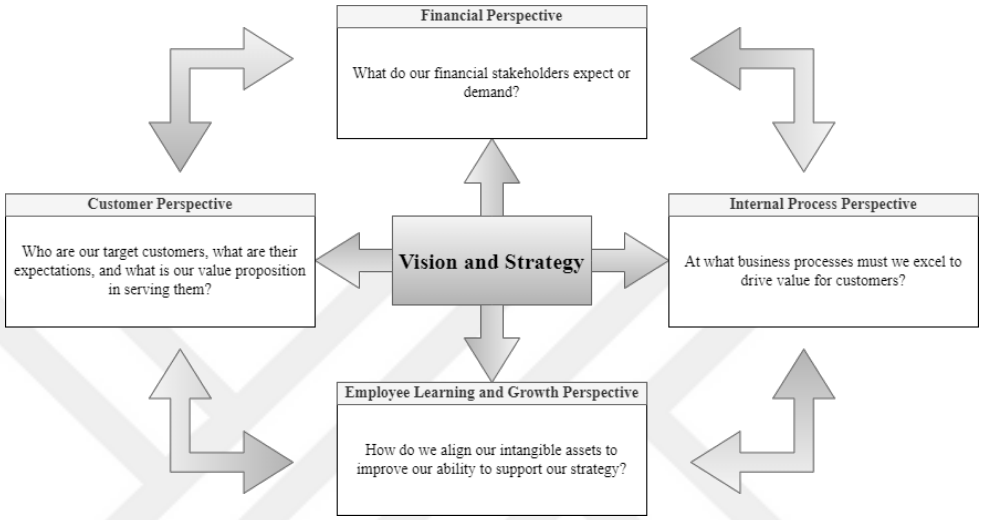

Figure 2.2: A simplified form of generic BSC adapted from [3]

Organizations must provide answers to two important questions when selecting measures for the Scorecard's Customer perspective. The first question specifies the target customers and the second one addresses how to best serve these customers using the organization's value proposition. Although these questions seem straightforward, there are a lot of challenges for organizations to overcome since it seems like they usually try to please everyone in their strategies [29][1]. Several common metrics are included in this perspective, including market share, customer satisfaction, customer value, customer profitability, and customer retention. By taking these steps, a company can clearly define the customers it should target and determine their needs and expectations from the business [1].

The Internal Process perspective focuses on the internal processes that the business needs to succeed in order to keep adding the value that customers and, eventually, shareholders expect in a productive and efficient manner. This perspective may be used to represent product development, production, manufacturing, delivery, and

post-sale services [29][1].
The measures in the Balanced Scorecard's Learning and Growth perspective serve as the basis and foundation for the other three perspectives. In fact, after the measures for the other perspectives of the BSC are determined, there will definitely be some gaps between the current organizational infrastructure and the skill level of the staff and information systems. Measures like employee skills, employee satisfaction, information availability, and alignment can be included in this perspective to help the business reach the level of expertise and foundation needed to meet its objectives and expectations [29][1].
The financial perspective is of particular importance in businesses that are profit-oriented. The typical metrics from this perspective include profitability, revenue growth, and economic value added, which can enable businesses to determine any bottom-line improvements as a result of executing their metrics from other perspectives of the BSC [29][1].

Another notion that should be stated is the link between the metrics inside these perspectives. Strategic maps are typically used to illustrate the cause-and-effect links between the strategic objectives found in each perspective of the BSC. A corporate strategy map is shown in Figure 2.3. A cause-and-effect reasoning is used to specify these arrows connecting various strategic objectives, and this work is typically completed with the assistance of experts in the relevant field. Furthermore, a mix of lead and lag indicators ought to be included in the strategy map, as per [1] and [29]. The lead indications are what determine future performance, whereas the lag indicators are the results of past activities.

2.1.2 Balanced Scorecard in Construction Companies

In the construction industry, as in other industries, BSC has been thoroughly studied. [30] looked into how the BSC method was being used by Slovakian construction companies as a strategic management tool. Less than 15 percent of Slovakia's construction companies employed the BSC concept, despite the fact that over half of all businesses worldwide now successfully use BSC or its tools. The authors also underlined the positive effects of utilizing performance management systems and supported

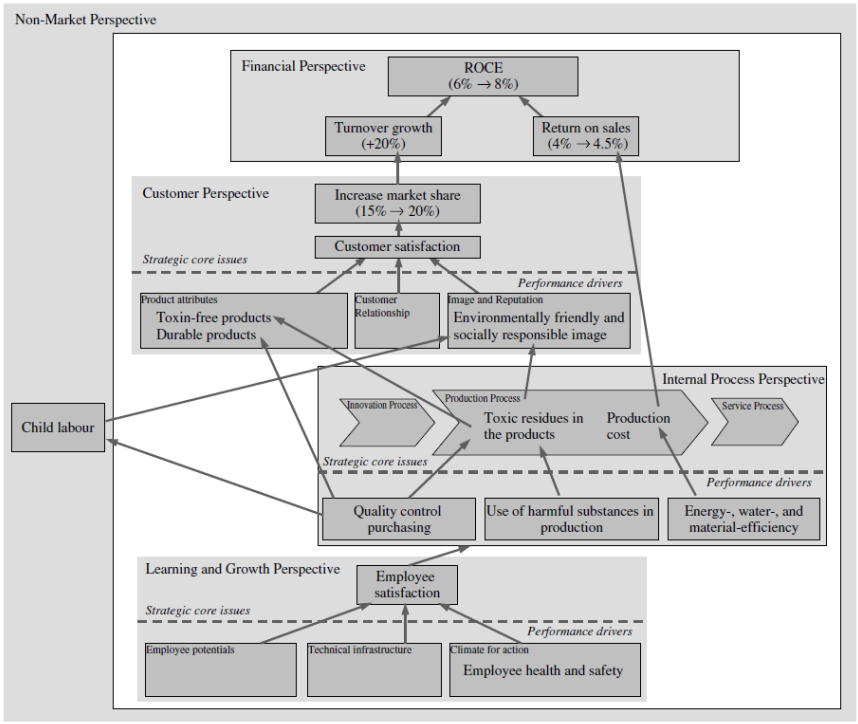

Figure 2.3: Example of a strategy map for a sample company derived from [4]

the BSC concept to raise managers' awareness of it. Using the examples of businesses like STRABAG and ARUP, [31] defined a BSC for construction companies and came to the conclusion that investing in the development of long-term intangible assets can decrease short-term profitability while creating long-term value and improving the company's market positioning. [32] added two more categories called project and sub-contractor perspectives to the original BSC and introduced this framework for the construction industry. In a different study, [15] created a BSC tool to reach a high degree of circular economy in the construction sector using the example of Kazakhstan. The research's findings, according to the authors, might also apply to comparable construction firms in developing nations like Central Asia.

In order to evaluate the IT implementation in the construction industry, [33] employed BSC as a tool. [34] reviewed the Construction Industry Master Plan, which was introduced by stakeholders in the Malaysian construction industry, to link its recommendations to the perspectives of the BSC. The authors noted that while the learning and growth perspective has received significant attention, the critical success factors and strategic thrusts have addressed all four BSC perspectives, with a weak emphasis on

customer management.

BSC has also been combined with numerous other tools, including the SWOT matrix [35] and the Business Excellence Model (BEM) [36], to measure the overall performance of the constructions, per the recommendations of academics like [37] and [38]. In order to reduce construction delays during the project's construction phase, [18] created a framework that combines Quality Function Deployment (QFD) and the BSC. A unique Performance Management Model (PMM) based on the advantages of BSC and the European Foundation for Quality Management (EFQM) was presented by [39] in 2013. The Analytic Hierarchy Process (AHP) was used to link these two models. The authors then went through ten steps in a Croatian construction company to verify their framework.

2.2 Sustainability Balanced Scorecard

Concerns about environmental preservation, corporate social responsibility, and legal requirements have made sustainability a significant issue today [40]. Many organizations worldwide feel responsible for protecting the environment and advancing the welfare of current and future generations. To address these issues, businesses have started integrating sustainability into their operations to maintain sustainable integrity while aiming for financial gains [41]. Being successful from a single sustainability perspective is insufficient for a company to be considered a contributor to sustainable development, and corporate sustainability seeks to advance businesses from economic, environmental, and social aspects [42]. In response to the need for a useful tool to meet sustainability objectives and evaluate sustainability performance [10][11], many academics have suggested Sustainability Balanced Scorecard (SBSC) as a strategic management tool for measuring sustainability performance. Although the four generic perspectives of the BSC are considered sufficient for many sectors, they should only be viewed as a template, and companies may use three of the four perspectives or add one or more perspectives, depending on the industry they operate in [1]. BSC's ability to integrate financial and non-financial metrics into each other[43] makes it a promising tool for integrating strategy and sustainability in businesses [5].

The concept of SBSC originates from traditional BSC, which integrates sustainability perspectives with the four BSC perspectives to clearly identify economic, social, and environmental issues as well as to incorporate performance metrics and sustainability-related objectives [17]. Understanding and evaluating the incorporation of sustainability principles into the BSC framework has been the subject of extensive research [17]. In a study in 2020, [44] proposed an SBSC for a Brazilian agriculture and livestock company by adding an additional perspective to BSC. They came to the conclusion that SBSC is a beneficial tool for incorporating sustainability into a business's strategic management. [45] used Global Reporting Initiative (GRI) indicators and scoring-benchmarking techniques to develop an SBSC and applied their proposed framework to a group of Greek companies. [46] introduced a new SBSC with four new perspectives, including "triple bottom line," "stakeholders," "process/products," and "learning and innovation" to measure environmental performance. This framework was then applied to three Portuguese firms. In their future research, inspired by the structure of SBSC, they performed a comparative CSR benchmarking on 23 real estate firms with shopping centers on various continents. [47] and [48] used SBSC to perform a case study on an Austrian brewery for sustainability management in Small and Medium-sized Enterprises (SMEs).

2.2.1 Sustainability Balanced Scorecard Architectures

This section focuses on multiple research studies that classified various SBSC architectures found in the literature. According to these studies, there are numerous approaches for incorporating sustainability principles into BSC, and multiple viewpoints exist in the literature regarding the superiority of various SBSC architectures [11]. However, there is no agreement regarding which architecture is better than the other [42]. Businesses select the SBSC architecture that most effectively addresses their requirements, stakeholder demands, and challenges [28].

According to [4], there are three primary methods for integrating sustainability into the BSC framework. The first method integrates social and environmental factors into the four categories of traditional BSC. In this method, the arrangement of the four categories remains unchanged, and the environmental and social aspects com-

plement the other strategic elements. The second method introduces a fifth category to the four categories of conventional BSC [14]. This additional category considers environmental and social factors. The final method is adding a non-market perspective to the BSC. This method becomes necessary when social or environmental factors are not strategically relevant enough to be included in one of the four standard BSC perspectives [4]. In another comprehensive study by [5], the authors categorized the SBSC architectures into twelve distinct models. These architectures are presented in Figure 2.4. According to their research, there are three different kinds of SBSC architectures: non-hierarchical, semi-hierarchical, and strictly hierarchical, and they differ in terms of full or partial integration of the sustainability principles into the generic BSC, adding the sustainability aspect as an add-on, or extending the original BSC in order to include sustainability matters.

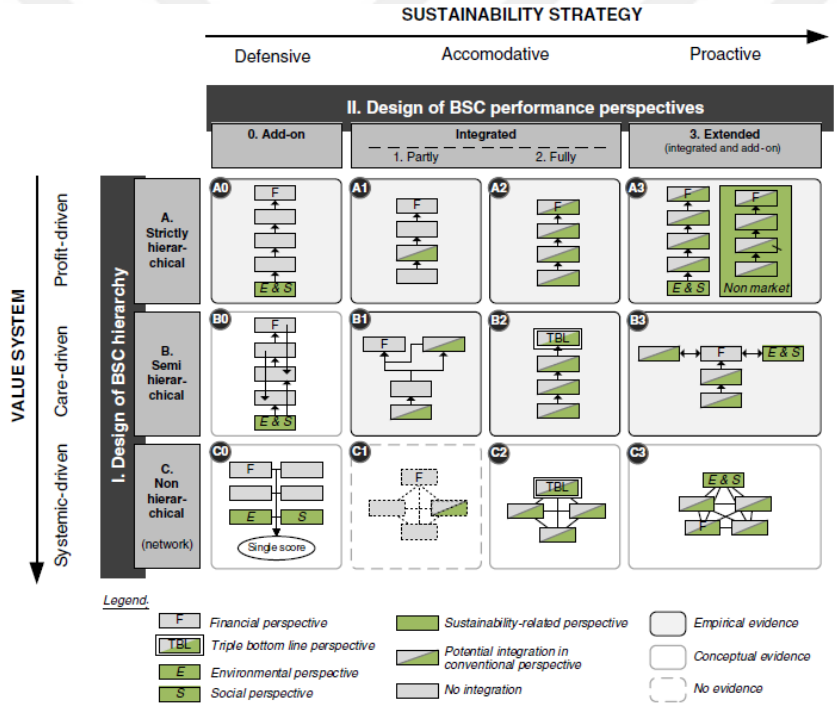

Figure 2.4: Different architectures of SBSC according to [5]

2.2.2 Sustainability Balanced Scorecard in Construction

Many studies have been performed on the applications of SBSC in different industries, such as the manufacturing sector [11][49], oil industry [50], agriculture [44],

thermal power enterprises [10], hotel industry [51][52], and glass industry [28]. In these studies, the proposed SBSC was tailored according to the specific needs and characteristics of the industry for which it was created. For instance, in designing the SBSC for the glass industry, [28] defined some metrics like glass cutting waste rate and glass processing waste rate as KPIs for the strategic objective of increasing product quality. These metrics are mainly specific to the glass industry and they make the SBSC customized for performance evaluation in this specific field. Similarly, in an SBSC designed for the Peruvian small and medium manufacturing enterprises in the plastic sector by [49], defining strategic objectives such as the use of non-toxic materials and clean technologies, and increasing new product offerings makes the framework specific for the manufacturing sector. Thus, SBSCs in the literature are mainly tailored according to the specific needs of different sectors.

Additionally, different scholars have implemented various methodologies in developing the industry-specific SBSCs. For example, [28] used a combination of strategic objectives from the literature and group meetings, [11] used the literature and secondary sources (e.g. industry investigations), and since the interaction of modules is significant, scholars like [10], [53], and [54] used a case study in their research process. Overall, scholars presented that they developed SBSCs representative of their domains. However, a SBSC that is designed according to the needs and unique characteristics of the construction industry is missing and needs more investigation.

There are various ways to incorporate sustainability into the general BSC architecture, and some researchers have conducted in-depth literature reviews focused only on the various approaches to creating an SBSC [5][4]. The subsection that follows provides a brief discussion of these studies.

2.3 Sustainability in Construction

The United Nations World Commission on Environment and Development originally defined "sustainable development" as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” [55]. On top of this idea, John Elkington defined the Triple Bottom Line (TBL) years later. According to Elkington [56], a company must ensure a triple result

in order to be considered sustainable: it must be financially viable, socially beneficial, and environmentally responsible. The construction industry has been notorious for its adverse impacts on the global aspect. Therefore, governments, practitioners, and the academic community have placed a high priority on taking the sustainability of the construction industry into account [57]. As a result, building firms have begun to publish yearly sustainability reports outlining their global influence and role in sustainable development.

Sustainability reporting is now a worldwide phenomenon, and developed countries with advanced economic states take it seriously. These sustainability reports can help companies stand out among their competitors and gain a competitive advantage. Several studies have proved that by publishing sustainability reports, companies even try to create a positive image for themselves in times of crisis [58]. Sustainability reports are considered one of the best references to showcase a company's sustainability strategies and its contribution toward a more sustainable future. However, these reports usually consist of hundreds of pages consisting of a variety of topics from diversity and inclusion to combating climate change and human rights and ethics, providing valuable insights into the latest trends, strategies, objectives, and performance of construction companies in terms of sustainability. But still, it is not easy to scroll each and every one of these voluminous documents to delineate the latest trends and depict the general picture of the industry. Therefore, the need for the utilization of AI to ease this task arises [58].

Sustainability reports do not have a single format or style, and they consist of many graphics, tables, and text embedded inside different shapes and pictures. Therefore, extracting textual information from voluminous ESG reports is a very time-consuming practice. In addition, when the textual information is embedded in pictures, the results obtained from analyzing the extracted text might not be completely reliable, which might cause an additional limitation [59]. The most widely used standards, like the United Nations Sustainable Development Goals (UN) and the Global Reporting Initiative (GRI), are typically the foundation for these reports. The subsection that follows provides a more thorough explanation of these standards.

2.3.1 Most Common Standards in Sustainability Reporting

In light of global concerns many organizations, though not all of them, have started publishing annual sustainability reports to show that they are supporting sustainable development and to make their environmental, social, and financial activities more transparent. These reports require construction companies to disclose their impact on the global stage. Taking this into consideration, investors prefer to invest in construction firms that consider themselves responsible in terms of sustainability [60].

While there are differences in the structure and template used by construction businesses to publish their reports, the majority of the world's leading construction companies follow certain guidelines when organizing these publications. According to the European Commission, there are numerous global reporting frameworks available for businesses to use when presenting non-financial information [61]. Global Reporting Initiative (GRI), which was developed in 1997, is the most well-known and well-established reporting standard [19] [62]. The GRI Standards provide a modular and interrelated framework for organizations to report on a wide range of Environmental, Social, and Governance (ESG) impacts. It includes four main standards: GRI100, containing universal standards; GRI200, containing economical standards; GRI300, containing environmental standards; and GRI400, containing social standards. The companies are provided with a wide selection of topics to prepare their reports accordingly and each company determines their material topics to reflect their priorities within their publications [60]. The GRI Standards and the UN Sustainable Development Goals (SDGs) are highly complementary and connected. The 17 UN SDGs and their representative targets are directly mapped to some of the topic-specific GRI Standards. Organizations can utilize the GRI Standards to measure and report on their contributions towards the accomplishment of the SDGs to show how their actions and impacts connect to the global sustainable development agenda.

Another framework is the United Nations Global Compact (UNGC), founded in 2000, which offers a sustainability framework to all companies, regardless of size or location. Ten optional principles that guide corporate responsibility form the basis of this framework. The principles of UNGC are about human and labor rights, environmental protection, and transparency [63].

Last but not least, the 2010 ISO 26000-Guidance on Social Responsibility aims to

direct businesses on how to communicate their successes and advancements to their customers. It provides a thorough disclosure of the most important aspects of CSR. Therefore, it is a tool that may be used globally to encourage sustainability reporting [62]. It should be mentioned that while GRI offers a thorough template, not all of the sections are mandatory. Companies, therefore, choose what to report and give greater attention to based on their activities and priorities.

2.4 Natural Language Processing and its Applications in Construction

The need for big data analytical tools has increased due to the construction industry's rapidly expanding textual information [6]. NLP is a branch of machine learning and linguistics that uses computational methods to process, comprehend, and mimic human language abilities [20].

This tool has been widely used in the construction industry for a variety of purposes. In a comprehensive review paper, [6] has investigated the various construction management domains where text mining and natural language processing have been employed for various objectives. According to this research, the number of publications that combined text mining and NLP with the construction industry has significantly increased since 2018 with the introduction of Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from Language Models (ELMo). Additionally, the authors disclosed that construction safety and management is the focus of one-fifth of the publications in the literature regarding the integration of text mining and natural language processing into construction management. The majority of research attention is directed toward this field. Text mining and Natural Language Processing (NLP) applications in construction management also focus on automated compliance checking, contract management, and knowledge, document, and information management. The number of publications in each focus area is displayed in Figure 2.5, as per [6].

NLP is used for a wide range of tasks, including text analysis, topic modeling, semantic analysis, and text classification. The body of research demonstrates that NLP is widely applied across the building industry. There is a gap in the literature, though, regarding the use of topic modeling—an unsupervised machine learning algorithm

that makes use of natural language processing—on the sustainability reports of building companies. An examination of earlier studies shows that many NLP techniques have been used to analyze reports, particularly sustainability reports, of businesses in other industries. The previous research on the application of various NLP techniques to the analysis of reports in industries other than construction is summed up in the following subsection.

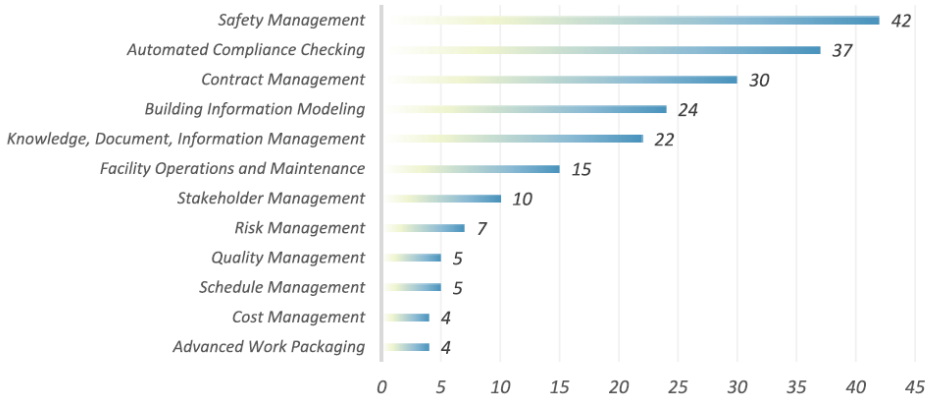

Figure 2.5: Publications of NLP in construction derived from [6]

2.4.1 Analyzing Sustainability Reports With NLP In Sectors Other than Construction

Integrating Natural Language Processing (NLP) applications into sustainability reports has been a primary focus of research. With a text classification purpose, [64] finetuned [65]'s RoBERTa model to automatically process and analyze large documents and determine whether or not their content is relevant to the SDGs. In a separate study, [59] evaluated the reports of Nordic businesses using various NLP techniques to determine which one was most effective in rating sustainability reports. Based on the study's findings, Latent Semantic Analysis (LSA) and Global Vectors for Word Representation (GloVE) produced the best outcomes. In order to analyze sustainability reports from large German companies and match pertinent text passages from sustainability reports to the corresponding legal regulations from the Global Reporting Initiative (GRI) standards, [66] developed a recommender system called Sustain.AI.

In the banking industry, [67] employed a rule-based NER approach to apply text min-

ing, examine the financial reports of significant Spanish companies, and estimate an index that determines how closely the TCFD recommendations are followed in relation to climate-related financial disclosures. The authors came to the conclusion that the number of disclosures pertaining to climate change is increasing annually and that certain disclosures are exclusive to reports other than annual and ESG reports. In a different study, [68] conducted a textual analysis of the sustainability reports of 98 listed European banks in order to discover the sustainability reports' thematic evolution. They discovered that the most important focus theme was climate action and that different European regions had distinct focal points. Through a semi-supervised learning methodology, [69] examined how banks communicate about their actions in connection to climate change. The statements from sustainability and the annual reports of commercial EU banks from 2015 to 2020 formed their dataset. In this study, a data-driven framework called climateBUG, with an accuracy of 91.42 percent, was trained.

In the smartphone manufacturing industry, [70] employed NLP to perform sentiment analysis and topic modeling on 300,000 tweets gathered via the Twitter API, in addition to 120 sustainability reports from major manufacturers such as Apple, Samsung, and others. A roadmap for Extended Producer Responsibility (EPR) plans was formed by dividing the extracted topics into six general categories after analyzing the interactions between manufacturers and consumers. In a different study, [58] used sentiment analysis and sentence similarity to evaluate and visualize the textual data in sustainability reports from globally recognized sustainability leader companies. The writers determined the positive and negative comment rates of these reports and developed a thematic structure. By performing a supervised text classification on the small corporate sustainability reports, [71] aimed to identify historical trends in ESG discussions. The study discovered that approximately 15 percent of the discussions were about ESG factors, which are essential to business strategy. In different supervised text classification, [72] created a Standard-Based Impact Classification method (SBIC) by integrating the flexibility of LDA with the rigidity of a predefined topic dictionary approach to extract social impact information from non-financial statements in 151 reports from 127 companies.

In order to find out the current ESG focus of the IT industry, [73] employed MLM and NSP methods of NLP on the sustainability reports of Forbes India's top compa-

nies in 2021. They extracted keywords and tokenized them using BERT and YAKE techniques to conduct a thematic analysis of the sustainability reports and detect their current ESG focus. The study successfully demonstrated a high prediction accuracy of 98 percent in classifying ESG themes. Another research was done by [74] on the CSR-related data of annual reports from different Indian companies in order to determine their primary CSR focus. The five main themes in these textual documents, according to the authors' text classification and topic modeling using LDA, were agriculture, healthcare, societal development, education, and environmental sustainability. Similarly, [75] discovered 42 topics that are connected to sustainability after performing trend analysis and topic modeling with LDA on 9514 sustainability reports that were published between 1999 and 2015. As a potential future project, the authors indicated the need to focus on the sustainability reports of a particular sector rather than analyzing general sustainability reports.

According to the current body of knowledge, NLP applications have been extensively utilized in a variety of industries, including banking, manufacturing, IT, smartphones, and others, to analyze textual sustainability-related information found in reports. Nevertheless, a research gap still exists in the analysis of sustainability reports using NLP within construction-related fields. In addition, given the distinct and unique characteristics of the construction industry, generalizing results from text analysis of sustainability reports of other sectors may not provide accurate and industry-specific information. Thus, there is a need for further research in the text analysis of sustainability reports within construction companies using NLP techniques.

2.4.2 Topic Modeling

Topic modeling is a natural language processing technique for identifying themes and hidden contextual patterns in a corpus of documents [76]. This process enables the information to be efficiently arranged and analyzed [26]. In order to determine the most common themes, topic modeling has been extensively used across a variety of text corpora, including books, newspapers, journal articles, and social media postings [77]. According to the literature, some of the most common topic modeling techniques include Non-Negative Matrix Factorization (NMF) [24], Top2Vec [25], Latent Dirichlet Allocation (LDA) [23], BERTopic [21], which is a recently developed

promising topic modeling algorithm based on the Bidirectional Encoder Representations from Transformers (BERT) architecture [22].

Among these methods, LDA is the most established and common technique and it has been widely used in different fields such as smartphone manufacturing [70], medicine [78], fashion [79], and construction safety [80] and waste [81]. However, several researchers have critiqued this method because of its limitations, including handling sparse and noisy datasets, topic overlapping, and poor interpretability [26] [82] [83]. Consequently, advanced methods such as BERTopic have emerged as a potentially useful tool to tackle such challenges [21].

Numerous scholars have conducted comparative analyses of various topic modeling techniques in order to determine which performs best. In one of the highly-cited papers about topic modeling, [84] classified the tweeter posts using four topic modeling techniques: LDA, NMF, Top2Vec, and BERTopic. In this study, LDA and NMF were examined together because the preprocessing processes were identical, and BERTopic's performance was compared to Top2Vec in a similar manner. The authors found that, when comparing the effectiveness of each technique, NMF produces more effective results than LDA, and that BERTopic outperforms Top2Vec in tweeter post analysis. In a different comparison study on topic modeling implementation in marketing research, [85] reported that the results show the effectiveness of using BERTopic in customer review analysis based on the complexity and various stages of implementation of four different topic modeling techniques: LDA, SBERT, BERTopic, and Top2Vec. Similarly, a study by [86] found that the BERTopic model in the experiment performs at least 34.2 percent better in Chinese and English clustering than the other algorithm models, namely LDA and Top2Vec, and yields a better topic clustering effect. The data for this research was obtained from Weibo and Twitter, which are among the largest social media platforms in China.

In a different study, [87] compared the outcomes of three methods—LDA, NMF, and BERTopic—and conducted topic modeling on datasets from Amazon product reviews. The results show that all three algorithms are effective and practical. However, the results given by BERTopic were more meaningful than those of other methods according to the consistency calculation metric.

Because of the fact that BERTopic is an emerging method, it hasn't been used in as many construction-related studies as other algorithms. In order to investigate the

three pillars of construction robotics, researchers such as [88] conducted topic modeling using BERTopic on the pertinent papers in a thorough review paper. However, there is still a gap in the analysis of the sustainability domain within construction-related reports and documents to address the most common themes covered in these documents. This research aims to fill that gap and lay the groundwork for the creation of an SBSC that is specifically designed for the building industry.

CHAPTER 3

METHODOLOGY

Sustainability reports of the leading construction companies showcase the latest trends in the construction industry and reflect their priorities and the criteria they use to assess their performance in terms of sustainability. Hence, this methodology examines these reports as a valuable source and puts forward the trendy topics that can be used in the SBSC. This chapter provides an overview of the four primary phases undertaken in this research, as shown in Figure 3.1. During the first phase, data collection was performed by investigating the websites of the companies on ENR's top international contractors list to build a corpus of PDF reports. In the next phase, the preprocessing of these PDF files to extract the relevant textual data and exclude unnecessary sections, such as tables of contents and figures, was completed. The output of this phase was a structured data frame saved in a CSV file to serve as input for the subsequent text analysis phase. In the third phase, topic modeling using BERTopic, an unsupervised machine learning algorithm, was used because of the justifications stated in the previous chapter. In this stage, the algorithm of BERTopic, consisting of several steps, namely embeddings, dimensionality reduction, clustering, tokenization, and weighting schemes, was used to categorize the input data into distinct groups alongside the topic names, counts, and representative words. The last phase includes the validation and categorization of the topics by experts to define the most dominant aspects of achieving sustainable performance. The categorized topics were used by the expert group to develop a strategy map as a base for the SBSC and suggest key performance indicators (KPIs) derived from the most representative keywords of the identified topics and their insights. Finally, to verify the applicability of this framework, the SBSC was customized for this company according to the topics discussed in their sustainability reports and the feedback from them management was taken.

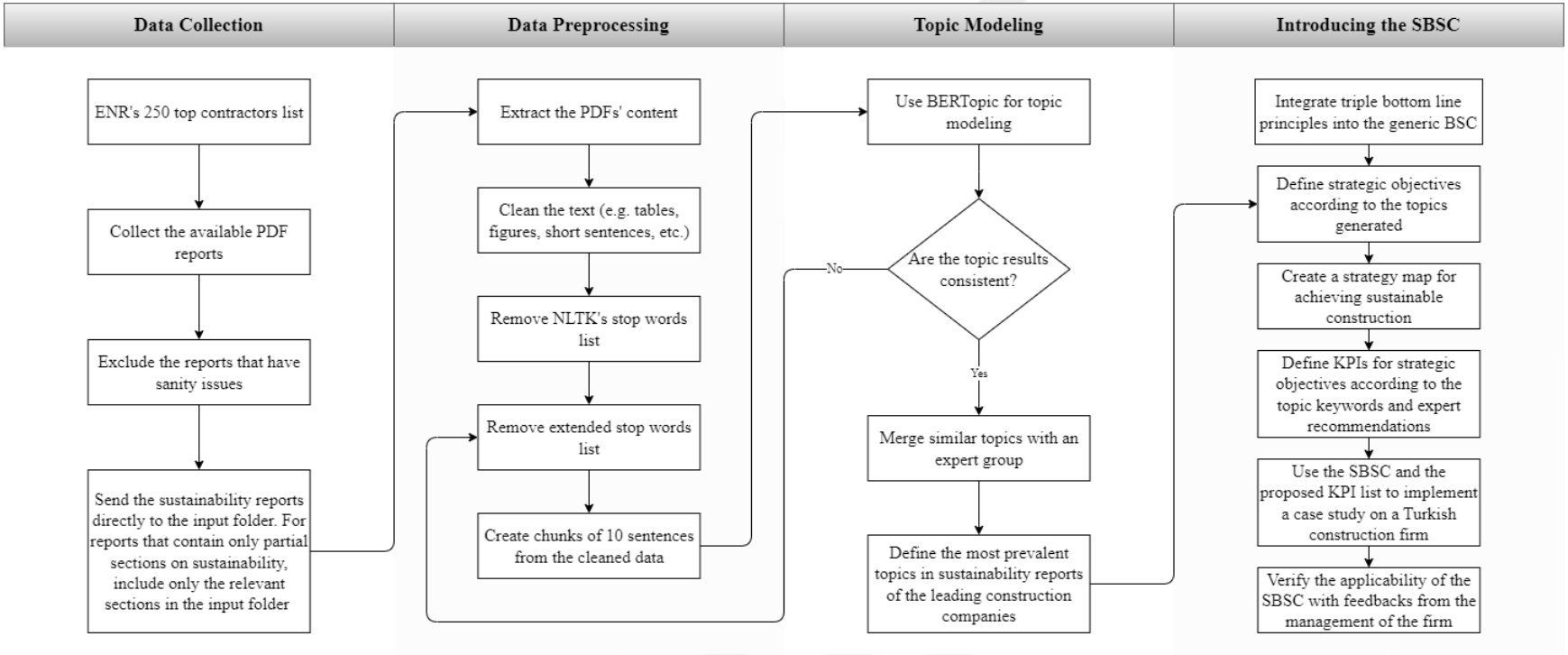

Figure 3.1: Research methodology

3.1 Dataset

The dataset for this study was the sustainability reports of the world's leading construction companies (See Appendix A). These companies were chosen according to the Engineering News Record's (ENR) list of the top 250 international contractors in 2021 [89]. The data collection and data preprocessing steps are explained in detail in the following subsections.

3.1.1 Data Collection

The data collection process consisted of several steps. First, the names of 250 top international contractors were derived from the ENR report for the year 2021, which was the most recent list when this research study started in 2022. Then, the websites of each company were investigated to find and download the latest sustainability reports. During this step, it was observed that while some companies prefer to publish sustainability reports, others publish annual reports, and sustainability-related issues are mostly integrated into the report. After excluding the companies that published either no reports or reports in a language other than English, a total of 128 reports were downloaded. However, after careful inspection, it was observed that some of these reports consisted primarily of graphics and photos, which made them unsuitable for text retrieval. Additionally, a sanity check process was performed on the reports to detect the problematic PDF files that were in a format that the textual information could not be extracted.

Among these reports, those that were sustainability reports were directly moved to the final dataset folder, and for annual reports with partial sustainability-related sections, only the relevant parts were derived and included in the dataset folder. Ultimately, the final dataset folder included 113 PDF files, which were ready for the preprocessing phase of the research. During the data collection phase, care was taken to ensure that the most recent report published by the company was included. In this regard, while many companies publish a report on an annual basis, some companies do not necessarily release their sustainability or annual reports on a regular time frame. The research data was gathered towards the end of 2022, and the final dataset included 74

reports that were published in the same year, 30 reports that were published in 2021, 5 reports that were published in 2020, and 4 documents that were published prior to 2020. The number of PDF reports used by nation and continent in the final dataset is shown in Figure 3.2.

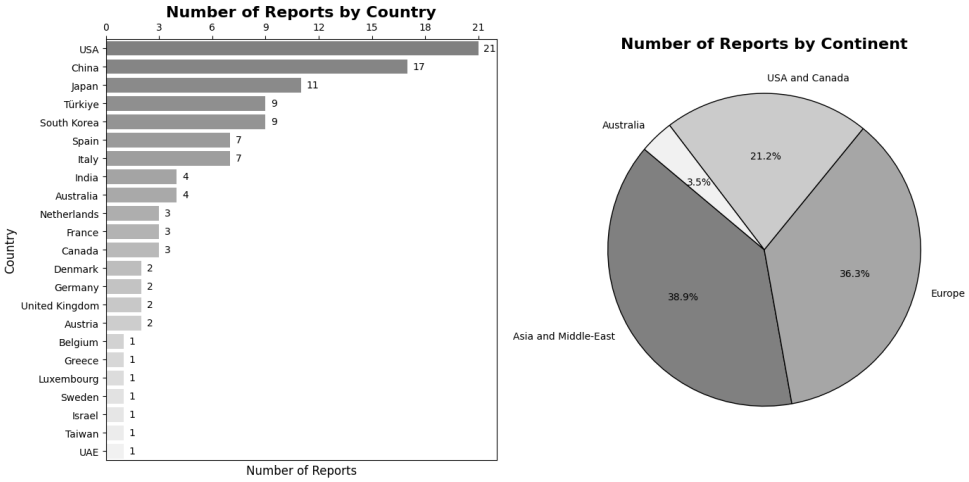

Figure 3.2: Origin of input reports

3.1.2 Data Preprocessing

While traditional topic modeling methods like LDA and NMF require special preprocessing steps like stemming and lemmatization, BERTopic does not need such steps since it is designed to use BERT embeddings [84]. However, since the PDF files used as the dataset in this research do not have a single formatting and include many visualizations, tables, and sections that are not required in the text analysis, there is a need to clean these PDF files and extract only the textual information that is required for the topic modeling. In addition, the preprocessing step highly impacts the quality of the results in topic modeling [90]. A portion of the code from GitHub [91] was used for the preprocessing of the PDF documents, and it was modified in accordance with the research's objectives. In this phase, first, the PDF documents' contents were gathered using the tika library [92]. Then, the sentences were extracted from these contents and were cleaned using regular expressions. During this process, sections like the URLs, headers, punctuations, figures, and tables of contents were removed, and the sentences were split using NLTK's sentence tokenizer [93]. Subsequently, the

most frequently occurring words in the English language, like we, and, is, are, etc., were eliminated from the dataset using the NLTK's English stop words list. Furthermore, to eliminate words that might act as noise during the topic modeling process in the particular field in which the research was conducted, an extended stopword list was employed. This list contained words like the names of cities and countries, non-English words, organization names, etc. Appendix B contains the list of extended stop words, which consists of 899 words that were added based on trial and error, taking into account the outcomes of subsequent steps after several iterations of the code.

The preprocessing step was done on each PDF report one by one, and after the text cleaning, sentences were separated from each other to create each row of an Excel file. Finally, a data frame consisting of the sentences was created and saved into a CSV file. Subsequently, these CSV files were combined to create a single input data file to use in the next steps. The initial input CSV file consisted of 94312 rows, meaning that each cleaned sentence was placed in each row of the file. These sentences were then used to create batches of ten sentences to create each input document for the BERTopic. This number was decided according to the approximate number of sentences in specific sections of the reports before the beginning of the following title, and it was fixed on a try-and-error basis. The reason behind the creation of these batches was the fact that each sentence in a report did not necessarily represent a specific topic, and this created a lot of noise in the final results after the text analysis. Therefore, creating batches of sentences that all together gave meaning to a specific subject improved the results of the text analysis. After this step, the final CSV file was prepared for the analysis in the next phase, which was doing the topic modeling on the dataset and determining the most prevalent topics and themes discussed within this textual information that was derived from the sustainability reports of the leading construction firms.

3.2 Topic Modeling with BERTopic

As shown in Figure 3.3, the BERTopic algorithm consists of five steps: Word embeddings, Dimensionality Reduction, Clustering, Tokenization, and Weighting Scheme.

Each part of this process is explained in detail in the following sections. The values of the hyperparameters in the topic modeling process are based on the BERTopic's default settings. "top_n_words" is set to ten by default to find the ten most significant and representative words within each topic. Additionally, in order to reveal all the hidden topics within the input dataset and perform comprehensive and detailed text analysis, no specific number was given to the "nr_topics" parameter to limit the number of topics generated.

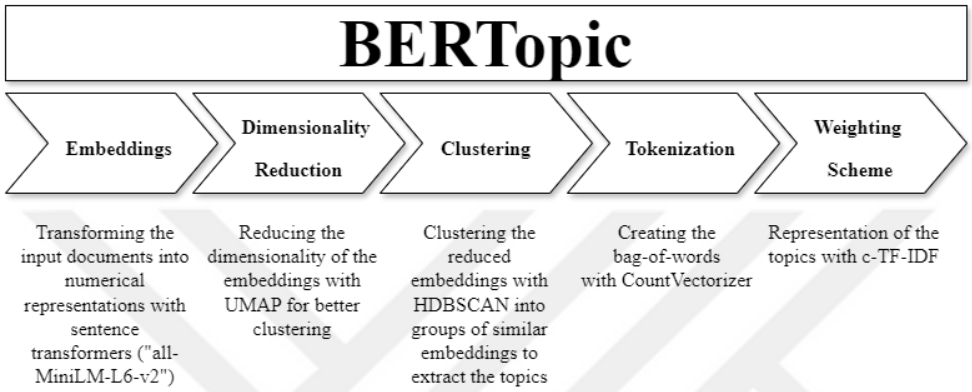

Figure 3.3: BERTopic algorithm

3.2.1 Embeddings

In the first step of the BERTopic algorithm, the input data is transformed into numerical representations using a pre-trained language model. For the objective of this research, the "all-MiniLM-L6-v2" pre-trained sentence transformer model [94] was employed due to its capability of capturing the semantic similarity between the input documents [21]. This model is suited for a variety of use scenarios and was trained to detect document similarities in the English language [21]. It should be highlighted, of course, that no embedding model is ideal for converting input documents into numerical representations; rather, because BERTopic allows a variety of embedding strategies, the choice of the best pre-trained model is specific to the study's goal. Furthermore, BERTopic will be able to use newly developed state-of-the-art pre-trained embedding models. Thus, the results gained using this algorithm might improve with any newly-released model [21]

3.2.2 Dimensionality Reduction

The curse of dimensionality causes difficulty in document analysis and clustering once the input data has been embedded since embeddings are often highly dimensional [21]. Consequently, following the embeddings, BERTopic uses a dimensionality reduction phase. As a result of its capacity to represent both the local and global high-dimensional space in lower dimensions, BERTopic employs by default the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) [95]. Naturally, similar to the embeddings step, BERTopic provides alternative models for this step as well as additional methods of dimensionality reduction based on research requirements [21].

3.2.3 Clustering

Following the embeddings and dimensionality reduction stages, BERTopic clusters similar embeddings using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) method [96] to extract topics in the subsequent phase. While HDBSCAN's capability to capture structures with varying densities is impressive, similar to the first two stages, BERTopic provides using different clustering alternatives and, with the introduction of additional models, there's always a chance for even better clustering outcomes [21].

3.2.4 Tokenizer

In this part bag-of-words representations are generated to tokenize the created clusters [97]. This means that each document is treated as a collection of words or tokens, taking into consideration the frequency of each word's occurrence rather than having grammatically correct and complete sentences. For instance, after cleaning the textual data, a sentence like "We aim to control our carbon footprint to achieve carbon neutrality" in a company's sustainability report becomes tokens at the end of the tokenization process such as "aim," "control," "carbon," "footprint," "achieve," "carbon," and "neutrality." CountVectorizer was used to achieve this goal and turn the sentences

in each data document, which was used as input data for the BERTopic, into tokens or a collection of words. To build topic representations, this procedure is combined with the following phase that involves the c-TF-IDF computation [21].

3.2.5 Weighting Scheme and Topic Representations

By combining two statistics of term frequency and inverse document frequency, TF-IDF can determine the relevance of a word to a document [98]. Equation 3.1 shows this process. In this equation, the term frequency models the term frequency t in document d. The logarithm of the number of documents in a corpus N divided by the total number of documents that include t gives the inverse document frequency, which expresses how much information a word adds to a document.

$$W_{t,d} = \text{tf}_{t,d} \cdot \log\left(\frac{N}{\text{df}_t}\right) \tag{3.1}$$

In BERTopic, to compute TF-IDF at the cluster level instead of the document level, this formula was modified to provide topic-word distributions for every cluster of documents. The modified version of TF-IDF is shown in equation 3.2. The average amount of words per class is represented by A in this equation [21].

$$W_{t,c} = \text{tf}_{t,c} \cdot \log\left(1 + \frac{A}{\text{tf}_t}\right) \tag{3.2}$$

In this equation, "class c" corresponds to the cluster that was previously formed, and the frequency of word x within that class is extracted. This yields the TF representation that is class-based. The average number of words in class A is then divided by the frequency of word x in all classes, and the logarithm of one is added. To make values positive, it is required in the logarithm to include a plus one. This leads to the IDF representation, which is class-based. The relevance score for each word in each class is then obtained by multiplying tf by IDF, as in the traditional TF-IDF. In other words, instead of utilizing the traditional TF-IDF procedure, a modified version of the technique that yields better representations was used in BERTopic. It should also be noted that the topic representation is independent of the clustering phase; thus,

while c-TF-IDF is enabled by default in BERTopic for the topic representations, the topic extraction techniques can be customized by parameter tuning and using different weighting schemes.
This step helps in identifying the key terms that best describe the topics represented by each cluster. Finally, all the identified topics are represented as titles in the form of word combinations with their representative keywords and their corresponding word scores. In addition, BERTopic offers a number of visualizations to aid in the analysis of the revealed topics and the relationships between these topics, including topic word scores, hierarchical clustering, and intertopic distance maps, which will be discussed in detail in the next chapter.

3.3 Validation and Verification of the Results

Every research result should be validated in order to prove its applicability. In this section, the topics found by BERTopic are presented to a focus group consisting of university professors, sustainability experts, contractors, and construction managers from several construction companies and educational institutions. Prior to the data collection from participants, human studies ethics committee approval (0384-ODTUİAEK-2024) was acquired. This section was completed in two stages. The first section involved validating the results of topic modeling, as well as establishing a strategic map for attaining sustainable structures with the help of specialists in construction management and sustainability. In this step, a proposed list of key performance indicators (KPIs) was also created according to the most representative keywords of the revealed topics and the recommendations of the expert group. The second stage involved implementing the suggested SBSC framework in a Turkish construction company in order to verify its effectiveness and applicability. These processes are explained in the following sections, and the results are demonstrated in the next chapter.

3.3.1 Categorization of the Topics and Creating a Strategy Map

In the first part, the most dominant topics discussed within the sustainability reports of top global contractors were derived from BERTopic. The next chapter contains a

full list of these topics, along with all of the associated visuals. Many of these topics shared a lot of relevance and could be combined to create a topic with a more inclusive title. It should be mentioned that because the sustainability principles might change based on the industry being covered, consulting with specialists in the field of construction sustainability and management was important for the formation of these topics. As a result, a meeting was scheduled to examine the topics that were generated and group them into broad categories that would represent sustainability in the construction industry. The group consisted of university professors and professionals in the field of sustainability and construction management. After classifying the topics and thereby narrowing the scope of the study, the topics were presented as strategic objectives in an SBSC, as the goal of this research is to create a specific SBSC for the construction industry.

To determine the best architecture for the SBSC, a widely referenced review study analyzing various SBSC architectures in the literature was taken into consideration [5]. This study states that the general BSC can incorporate sustainability or part of its concepts in twelve distinct ways. A SBSC may be strictly hierarchical, semi-hierarchical, or non-hierarchical. The sustainability principles may be added to the BSC as separate categories (add-on), fully or partially integrated into each SBSC category, or the SBSC may be expanded to include an add-on devoted to sustainability matters in addition to the integration of the sustainability principles within its categories. There is no one-size-fits-all framework for creating an SBSC; instead, each framework should be created with consideration for specifics of the industry in question [42]. In a recent study, [49] used a fully integrated and semi-hierarchical architecture to create an SBSC for the manufacturing sector. The triple bottom line principles were added to the generic BSC's financial category in this study, and other sustainability-related concerns were incorporated into the remaining categories as well. This architecture was also found to be suitable for the construction industry since, in this sector, sustainability could be included in every facet of the project. Therefore, this architecture was selected to create the foundation on which the proposed template was going to be created. As a starting point, the template SBSC was created according to the semi-hierarchial and partly integrated architecture of SBSC discussed in the second chapter. To do this, the financial category of the generic BSC was expanded to include triple bottom line principles, which are financial, environmental, and social aspects.

In addition, the customer perspective of the BSC was modified as external stakeholders to contain both customers and stakeholders in the specific construction industry. Every SBSC should be presented in terms of cause-and-effect relationships between its strategic objectives. In other words, every BSC/SBSC is linked to a strategy map with strategic objectives and the causal relationships between these objectives. Therefore, a strategy map was developed by the expert group with the purpose of achieving sustainable performance. Each strategic target was assigned a proper category, and the cause-and-effect links between these objectives were defined. According to the professional judgment of the group, a few strategic objectives were also added in order to complete the strategy map for achieving sustainable performance within construction companies. Finally, a list of suggested key performance indicators (KPIs) was prepared with regard to the most representative keywords of the topics and any recommendations from the specialists that could represent the unique construction industry.

3.3.2 Case study on a Turkish Construction Company

After finalizing the strategy map with the assistance of the experts and preparing a suggested list of KPIs, the next step was customizing the SBSC for a construction firm to verify its applicability. The keywords that were most representative of the topics were combined with expert insights to suggest KPIs for the SBSC objectives. However, according to [28], every organization has its own unique strategies and metrics for self-evaluation. Therefore, the strategy map created in this study and the suggested KPI list were used to customize the SBSC for a Turkish construction company in accordance with their own objectives and metrics. In this instance, the strategic map acted merely as a foundation for helping construction companies move toward greater sustainability; it was developed by examining the best sustainability practices of leading construction organizations, which could be a valuable reference for construction companies to align their strategic initiatives with the way the industry leaders communicate their sustainability actions. This part of the research aimed to demonstrate that the proposed framework is a dynamic tool and offers the base and the reference to the main goal of achieving sustainable performance, and each company can modify this framework in order to set its own targets and measure its

performance using the most suitable metrics that represent their main priorities.
The top management of the Turkish construction company was consulted in the last step to get their input on the suggested tool, its applicability, and how much it represented the sustainability issues within the construction industry. They were then requested to offer any recommendations for how to change and enhance the framework for a more thorough and effective use.

CHAPTER 4

RESULTS

This chapter presents the findings of this study. After collecting the available sustainability reports of ENR's top global contractors, the textual information was extracted, cleaned, and subjected to a topic modeling process with BERTopic. Such information portrays the latest sustainability perspectives of the leading construction companies that can show the strong suits of the industry and the weaknesses that should be addressed. As a result, the ninety most dominant themes and topics discussed within the best sustainable practices of the top global construction companies were identified. This chapter presents the list of these topics together with the representative words of each topic and the visualizations that best capture their meaning. A mini focus group was then formed to analyze and merge the similar topics. The most dominant topics that help the building sector to become more sustainable were thus identified as sixteen major themes. This chapter displays the last set of topics in a figure. Using a SBSC template that was enhanced in accordance with the triple bottom line principles, these topics were then used by the expert group to identify strategic objectives and develop a strategy map as a foundation toward sustainable construction. In addition, relevant key performance indicators (KPIs) according to the most representative keywords of the defined topics were suggested to finalize the SBSC template that can be customized according to the different goals and metrics of each company. Finally, the suggested tool was verified through a case study on a Turkish construction company, and the top management of the company was asked to give feedback about the relevance and applicability of this tool in making them become more sustainable. They believed that this framework is a useful tool that would help construction companies to plan and organize their sustainability initiatives and the best aspect of it is that it can be modified according to different scenarios and different companies.

4.1 Topic Modeling Results

Following the collection of the sustainability reports from the leading international contractors according to ENR, as was detailed in the previous chapter, the textual data was extracted, cleaned, and then put through a topic modeling process using BERTopic. This process identified ninety most dominant themes and topics discussed within the best sustainable practices of the top global construction companies. The output from BERTopic shows the identified topics with a keyword combination. To have a deeper understanding of the information from each topic, it is also possible to generate the most representative words for each topic. The topic word scores related to these topics are shown in Appendix C and will be discussed in detail in the following subsection.

4.1.1 Discussion of the Initial Topics List

The complete list of the topics and their top ten representative words are shown in Table 4.1. The appearance of some topics associated with the latest trends such as combating climate change and integrating digitalization in the industry was significant. Additionally, some topics that were determined by the topic modeling were highly expected since they are the usual topics that are strongly connected with sustainability principles. For instance, due to the high accident potential in the building industry, topic 0 (safety_health_occupational_accidents) is among the most important topics that almost all construction companies are eager to improve in order to become more sustainable. Similarly, topics 8, 9, 82, 76, and 64, which represent the general topic of combating climate change and reducing carbon emission, were highly expected to appear since the construction industry is among the top emitters of greenhouse gases and construction companies are mostly aligned to improve themselves by targeting goals like reaching Net Zero by a certain year. As it is understood from the top ten representative words of topic 8 ([’emissions’, ’scope’, ’ghg’, ’carbon’, ’data’, ’energy’, ’reduction’, ’gas’, ’electricity’, ’greenhouse’]) one of the current focus areas in reducing these emissions are done by calculation of the different scopes (namely scope 1, 2, and 3) of the construction companies which are the different sources of carbon emissions for an organization. The scope 1 carbon emission represents the

direct emissions of the company from the sources that are owned or controlled by the firm. The scope 2 emissions represent the indirect emissions that the company produces from its energy use and the scope 3 emissions are the other indirect carbon emissions that are the results of the company's activities but occur from sources that are not directly controlled by the organization (e.g. carbon emission by employee commuting).

The appearance of topics related to digitalization and innovation such as 18, 59, 75, and 46 in the technology was expected due to the wide integration of digital technologies in almost all the industries in the recent years. However, topics 17 (security_information_cyber_protection) and 45 (noise_dust_air_pollution) were among the topics that were less expected. However, the appearance of such subjects demonstrate that the leading construction companies not only consider their cyber protection as part of their digitalization growth, but they also try to improve their adverse environmental impacts other than merely carbon emission. The results show that reducing any kind of disturbance such as noise, dust, or any other issue that may affect the environment and the people living there is also significant in evaluating the sustainability performance of a construction company.

Another interesting result was the appearance of topics such as topic 12 (sustainability_sustainable_report_social), 20 (report_gri_reporting_assurance), and 60 (disclosure_gri_description_aspect) with representative words like "gri," "reporting," "standards," "responsibility," and "report". While some of the leader construction firms such as SKANSKA, a Swedish multinational construction and development company, started publishing environmental reports in the 1990s to disclose their commitment to sustainability and environmental management [99], when considered in general, it has not been a very long time since most of the companies started reporting their sustainability impact. Therefore, speaking about the reporting itself and how the companies are committed to disclose their impact are among utmost importance nowadays.

Table 4.1: Topic Modeling Results

Count	Name	Representations
3701	-1_energy_construction_business_development	['energy', 'construction', 'business', 'development', 'project', 'sustainability', 'projects', 'new', 'management', 'employees']
743	0_safety_health_occupational_accidents	['safety', 'health', 'occupational', 'accidents', 'employees', 'work', 'training', 'hse', 'management', 'medical']
400	1_cash_assets_liabilities_financial	['cash', 'assets', 'liabilities', 'financial', 'note', 'income', 'interest', 'value', 'lease', 'net']

Table 4.1: Topic Modeling Results (continued)

Count	Name	Representations
347	2_compliance_ethics_corruption_anti	['compliance', 'ethics', 'corruption', 'anti', 'ethical', 'code', 'conduct', 'employees', 'bribery', 'integrity']
252	3_suppliers_supplier_procurement_supply	['suppliers', 'supplier', 'procurement', 'supply', 'chain', 'evaluation', 'companies', 'management', 'partner', 'partners']
247	4_board_directors_committee_supervisory	['board', 'directors', 'committee', 'supervisory', 'executive', 'members', 'governance', 'independent', 'director', 'audit']
203	5_remuneration_performance_shares_executive	['remuneration', 'performance', 'shares', 'executive', 'vesting', 'board', 'share', 'salary', 'variable', 'fixed']
181	6_risk_risks_management_business	['risk', 'risks', 'management', 'business', 'project', 'committee', 'financial', 'control', 'board', 'internal']
170	7_waste_recycling_materials_hazardous	['waste', 'recycling', 'materials', 'hazardous', 'recycled', 'disposal', 'circular', 'use', 'reuse', 'economy']
142	8_emissions_scope_ghg_carbon	['emissions', 'scope', 'ghg', 'carbon', 'data', 'energy', 'reduction', 'gas', 'electricity', 'greenhouse']
133	9_climate_risks_change_opportunities	['climate', 'risks', 'change', 'opportunities', 'related', 'risk', 'scenario', 'describe', 'physical', 'carbon']
129	10_biodiversity_species_nature_areas	['biodiversity', 'species', 'nature', 'areas', 'conservation', 'ecological', 'protected', 'environmental', 'area', 'protection']
125	11_quality_customer_satisfaction_customers	['quality', 'customer', 'satisfaction', 'customers', 'service', 'management', 'system', 'inspection', 'iso', 'products']
124	12_sustainability_sustainable_report_social	['sustainability', 'sustainable', 'report', 'social', 'governance', 'development', 'society', 'management', 'environmental', 'responsibility']
112	13_environmental_management_protection_environment	['environmental', 'management', 'protection', 'environment', 'system', 'waste', 'pollution', 'ecological', 'eco', 'regulations']
109	14_rights_human_labour_respect	['rights', 'human', 'labour', 'respect', 'labor', 'principles', 'harassment', 'discrimination', 'forced', 'compliance']
106	15_taxonomy_eligible_eu_activities	['taxonomy', 'eligible', 'eu', 'activities', 'aligned', 'criteria', 'activity', 'economic', 'regulation', 'capex']
100	16_water_wastewater_groundwater_treatment	['water', 'wastewater', 'groundwater', 'treatment', 'discharge', 'consumption', 'use', 'used', 'sewage', 'waste']
99	17_security_information_cyber_protection	['security', 'information', 'cyber', 'protection', 'cybersecurity', 'personal', 'data', 'threats', 'attacks', 'privacy']
93	18_data_construction_system_digital	['data', 'construction', 'system', 'digital', 'bim', 'smart', 'automation', 'technology', 'platform', 'engineering']
89	19_director_executive_member_officer	['director', 'executive', 'member', 'officer', 'outside', 'board', 'division', 'manager', 'general', 'university']
88	20_report_gri_reporting_assurance	['report', 'gri', 'reporting', 'assurance', 'standards', 'data', 'materiality', 'sustainability', 'information', 'button']
76	21_children_school_students_education	['children', 'school', 'students', 'education', 'schools', 'support', 'foundation', 'families', 'orchestra', 'program']
74	22_audit_financial_assurance_statements	['audit', 'financial', 'assurance', 'statements', 'auditors', 'board', 'auditor', 'matters', 'information', 'procedures']
69	23_stakeholders_stakeholder_sustainability_materiality	['stakeholders', 'stakeholder', 'sustainability', 'materiality', 'engagement', 'dialogue', 'communication', 'topics', 'issues', 'impact']
67	24_offshore_wind_dredging_vessel	['offshore', 'wind', 'dredging', 'vessel', 'seaspan', 'vessels', 'marine', 'installation', 'market', 'farm']
66	25_training_talents_talent_employees	['training', 'talents', 'talent', 'employees', 'courses', 'skills', 'job', 'learning', 'professional', 'employee']
61	26_power_energy_photovoltaic_electric	['power', 'energy', 'photovoltaic', 'electric', 'grid', 'generation', 'storage', 'wind', 'solar', 'carbon']
59	27_concrete_materials_cement_wood	['concrete', 'materials', 'cement', 'wood', 'building', 'carbon', 'piles', 'buildings', 'construction', 'material']
50	28_leave_childcare_employees_family	['leave', 'childcare', 'employees', 'family', 'care', 'work', 'support', 'working', 'hours', 'children']
47	29_county_rural_poverty_revitalization	['county', 'rural', 'poverty', 'revitalization', 'alleviation', 'assistance', 'agricultural', 'counties', 'village', 'targeted']
42	30_esg_governance_management_strategy	['esg', 'governance', 'management', 'strategy', 'committee', 'social', 'issues', 'company', 'responsible', 'sustainability']

Table 4.1: Topic Modeling Results (continued)

Count	Name	Representations
42	31_volunteer_rescue_local_relief	['volunteer', 'rescue', 'local', 'relief', 'free', 'service', 'community', 'welfare', 'people', 'medical']
42	32_foundation_community_communities_social	['foundation', 'community', 'communities', 'social', 'support', 'giving', 'local', 'people', 'volunteering', 'donations']
36	33_rail_city_river_metro	['rail', 'city', 'river', 'metro', 'bridge', 'project', 'highway', 'airport', 'transportation', 'line']
34	34_learning_training_courses_skills	['learning', 'training', 'courses', 'skills', 'employees', 'online', 'knowledge', 'talent', 'development', 'offers']
31	35_efficiency_principle_must_un	['efficiency', 'principle', 'must', 'un', 'formations', 'water', 'consumption', 'increase', 'waste', 'kinds']
29	36_human_resources_employees_hr	['human', 'resources', 'employees', 'hr', 'employee', 'hro', 'job', 'culture', 'personnel', 'satisfaction']
28	37_women_female_gender_positions	['women', 'female', 'gender', 'positions', 'diversity', 'senior', 'equality', 'career', 'womens', 'recruitment']
28	38_hydrogen_stations_mch_energy	['hydrogen', 'stations', 'mch', 'energy', 'lohc', 'green', 'production', 'supply', 'carbon', 'station']
27	39_pension_defined_obligations_benefit	['pension', 'defined', 'obligations', 'benefit', 'plans', 'plan', 'assets', 'funds', 'return', 'assumptions']
26	40_reef_boskalis_reefs_mussel	['reef', 'boskalis', 'reefs', 'mussel', 'coastbusters', 'nature', 'ocean', 'coastal', 'marine', 'arp']
26	41_sustainability_un_sustainable_companies	['sustainability', 'un', 'sustainable', 'companies', 'social', 'goals', 'must', 'development', 'satisfaction', 'goal']
25	42_accountability_memberships_contribution_organizational	['accountability', 'memberships', 'contribution', 'organizational', 'product', 'esg', 'sasb', 'social', 'verification', 'sdgs']
25	43_tax_deferred_income_taxable	['tax', 'deferred', 'income', 'taxable', 'assets', 'losses', 'taxes', 'differences', 'liabilities', 'temporary']
24	44_inflation_growth_european_ihs	['inflation', 'growth', 'european', 'ihs', 'markit', 'industry', 'prices', 'construction', 'sector', 'economy']
24	45_noise_dust_air_pollution	['noise', 'dust', 'air', 'pollution', 'environmental', 'construction', 'sites', 'water', 'vibration', 'impact']
24	46_machine_manufacturing_equipment_industry	['machine', 'manufacturing', 'equipment', 'industry', 'tool', 'plants', 'steel', 'intelligent', 'standards', 'technology']
23	47_lighting_energy_heating_electricity	['lighting', 'energy', 'heating', 'electricity', 'kwh', 'saving', 'consumption', 'photovoltaic', 'heat', 'gas']
23	48_technological_innovation_research_scientific	['technological', 'innovation', 'research', 'scientific', 'science', 'technology', 'scientifi', 'national', 'tech', 'sci']
23	49_vessels_dredgers_hopper_vessel	['vessels', 'dredgers', 'hopper', 'vessel', 'fleet', 'offshore', 'suction', 'fuel', 'methanol', 'dredging']
22	50_indigenous_reconciliation_communities_aboriginal	['indigenous', 'reconciliation', 'communities', 'aboriginal', 'peoples', 'people', 'nations', 'traditional', 'community', 'relations']
22	51_tax_taxes_authorities_compliance	['tax', 'taxes', 'authorities', 'compliance', 'taxation', 'groups', 'evasion', 'countries', 'transactions', 'paid']
21	52_gmbh_energy_consumption_vehicles	['gmbh', 'energy', 'consumption', 'vehicles', 'electricity', 'innolation', 'renewable', 'emissions', 'fleet', 'efficiency']
21	53_collective_discrimination_employees_bargaining	['collective', 'discrimination', 'employees', 'bargaining', 'agreements', 'workers', 'rights', 'labour', 'equality', 'labor']
21	54_sustainable_energy_decarbonization_buildings	['sustainable', 'energy', 'decarbonization', 'buildings', 'sustainability', 'cities', 'solutions', 'clients', 'transition', 'infrastructure']
20	55_development_csr_member_construction	['development', 'csr', 'member', 'construction', 'business', 'engineering', 'innovation', 'party', 'responsibility', 'chairman']
19	56_brand_brands_fashion_fiber	['brand', 'brands', 'fashion', 'fiber', 'building', 'resort', 'image', 'golf', 'business', 'voice']
19	57_business_new_space_businesses	['business', 'new', 'space', 'businesses', 'holding', 'power', 'energy', 'construction', 'renewable', 'privatization']
19	58_construction_building_projects_center	['construction', 'building', 'projects', 'center', 'design', 'tata', 'infrastructure', 'completed', 'accomodation', 'technological']
18	59_digital_data_information_processes	['digital', 'data', 'information', 'processes', 'security', 'business', 'cloud', 'transformation', 'devices', 'bim']

Table 4.1: Topic Modeling Results (continued)

Count	Name	Representations
18	60_disclosure_gri_description_aspect	['disclosure', 'gri', 'description', 'aspect', 'kpi', 'general', 'policies', 'taken', 'impact', 'impacted']
18	61_offshore_oil_wind_energies	['offshore', 'oil', 'wind', 'energies', 'energy', 'billion', 'market', 'transition', 'gas', 'markets']
17	62_sustainability_goals_esg_sustainable	['sustainability', 'goals', 'esg', 'sustainable', 'aecom', 'governance', 'progress', 'committee', 'plan', 'political']
17	63_diversity_inclusion_gender_inclusive	['diversity', 'inclusion', 'gender', 'inclusive', 'diverse', 'equality', 'culture', 'women', 'discrimination', 'equal']
16	64_climate_related_opportunities_change	['climate', 'related', 'opportunities', 'change', 'risks', 'sustainability', 'committee', 'board', 'risk', 'target']
16	65_labor_employees_employment_remuneration	['labor', 'employees', 'employment', 'remuneration', 'salary', 'training', 'democratic', 'workers', 'gri', 'recruitment']
16	66_treatment_water_sewage_technology	['treatment', 'water', 'sewage', 'technology', 'wastewater', 'using', 'tmah', 'membrane', 'desulfurization', 'energy']
16	67_shares_share_dividend_ordinary	['shares', 'share', 'dividend', 'ordinary', 'preference', 'class', 'price', 'issued', 'treasury', 'capital']
15	68_carbon_oil_ccus_dioxide	['carbon', 'oil', 'ccus', 'dioxide', 'gas', 'energy', 'methane', 'efficiency', 'refining', 'capture']
15	69_traffic_heathrow_ebitda_revenues	['traffic', 'heathrow', 'ebitda', 'revenues', 'passengers', 'impacted', 'higher', 'reached', 'var', 'etr']
14	70_oil_gas_meters_chemical	['oil', 'gas', 'meters', 'chemical', 'refining', 'shale', 'billion', 'exploration', 'energy', 'output']
14	71_seed_immobilier_rating_housing	['seed', 'immobilier', 'rating', 'housing', 'office', 'building', 'green', 'buildings', 'higher', 'spaces']
14	72_voting_shareholders_postal_meeting	['voting', 'shareholders', 'postal', 'meeting', 'shareholder', 'vote', 'shares', 'proxy', 'euroclear', 'participate']
14	73_completed_works_corporation_sekkei	['completed', 'works', 'corporation', 'sekkei', 'city', 'building', 'project', 'design', 'area', 'construction']
14	74_ton_tonne_waste_consumption	['ton', 'tonne', 'waste', 'consumption', 'total', 'emissions', 'hazardous', 'herein', 'refers', 'mwh']
13	75_innovative_world_business_economic	['innovative', 'world', 'business', 'economic', 'contribute', 'approach', 'value', 'transformation', 'development', 'solutions']
13	76_carbon_green_energy_neutrality	['carbon', 'green', 'energy', 'neutrality', 'low', 'clean', 'emissions', 'citic', 'forest', 'greening']
13	77_donation_contribution_social_donations	['donation', 'contribution', 'social', 'donations', 'activities', 'points', 'campaign', 'briquettes', 'love', 'volunteering']
13	78_person_male_number_female	['person', 'male', 'number', 'female', 'turnover', 'rate', 'persons', 'employees', 'count', 'total']
13	79_estate_investment_real_transportation	['estate', 'investment', 'real', 'transportation', 'initiation', 'concessions', 'development', 'acquired', 'rental', 'afikim']
12	80_dei_diversity_inclusion_diverse	['dei', 'diversity', 'inclusion', 'diverse', 'workforce', 'racialized', 'hispanic', 'asian', 'ergs', 'inclusive']
12	81_leadership_programme_people_inclusive	['leadership', 'programme', 'people', 'inclusive', 'organisation', 'leaders', 'day', 'ohi', 'secure', 'diversity']
11	82_emissions_carbon_trucks_electric	['emissions', 'carbon', 'trucks', 'electric', 'fleet', 'reduction', 'emission', 'energy', 'machinery', 'sipc']
11	83_traffic_students_teachers_school	['traffic', 'students', 'teachers', 'school', 'drivers', 'bus', 'parents', 'primary', 'action', 'education']
11	84_students_career_engineering_opportunities	['students', 'career', 'engineering', 'opportunities', 'training', 'vocational', 'school', 'technical', 'internships', 'inclusion']
10	85_een_students_university_scholarship	['een', 'students', 'university', 'scholarship', 'egypt', 'scholarships', 'un', 'education', 'support', 'cooperat']
10	86_properties_property_residential_markets	['properties', 'property', 'residential', 'markets', 'divestments', 'sustainable', 'commercial', 'continues', 'investment', 'kingdom']
10	87_control_internal_audit_supervision	['control', 'internal', 'audit', 'supervision', 'legal', 'risk', 'system', 'law', 'special', 'prevention']
10	88_sustainability_stakeholders_business_growth	['sustainability', 'stakeholders', 'business', 'growth', 'sustainable', 'strategy', 'social', 'companies', 'sdgs', 'responsible']

BERTopic offers various visualization tools that can ease the analysis of topic arrays and their relationships with each other. As an example, BERTopic provides an interactive intertopic distance map for examining individual topics, which aids in the better analysis of the potentially vast array of subjects [21][84].

The intertopic distance map is depicted in Figure 4.1. This figure shows the proximity of the topics and how they create clusters of topics that are close to each other in terms of content. For instance, the topics 37 (women_female_gender_positions), 63 (diversity_inclusion_gender_inclusive), 80 (dei_diversity_inclusion_diverse) are incorporated into a single cluster in this map since they have contents that have an overlap with each other, and this similarity results in the representation of these topics as a single group. Similarly, topics 16 (water_wastewater_groundwater_treatment) and 66 (treatment_water_sewage_technology) were represented next to each other, demonstrating a cluster specific to water management. Topics 2, 14, 53, 28, 78, and 65 were also grouped together in a single cluster representing the matters related to human rights, ethics, and employee rights. The presence of representative words such as "labor," "respect," "ethical," "bribery," "democratic," "employees," and "right" shows the relevance of these topics to each other in terms of content.

As can be inferred from the presence of words like "biodiversity," "species," "conservation," "protection," "nature," and "reefs," topics 10 (biodiversity_species_nature_ areas) and 40 (reef_boskalis_reefs_mussel) also appeared in the same cluster because they carried the same theme of biodiversity. Topics 25 (training_talents_talent_ employees), 34 (learning_training_courses_skills), and 84 (students_career_engineering_ opportunities) also had the word “training” in common and carried the general topic of training and talent management. The presence of words like “talent,” “career,” and “skills” also strengthened this hypothesis.

Topics 9 (climate_risks_change_opportunities) and 64 (climate_related_opportunities _change) also appeared in the same cluster with common words of “change,” “climate,” “risk,” and “opportunities,” conveying the theme related to climate change. Overall, it is noted that roughly twenty-one separate themes were addressed within the textual dataset, based on the number of clusters produced by BERTopic, and these themes can be obtained by combining the comparable topics in the ninety output topic findings.

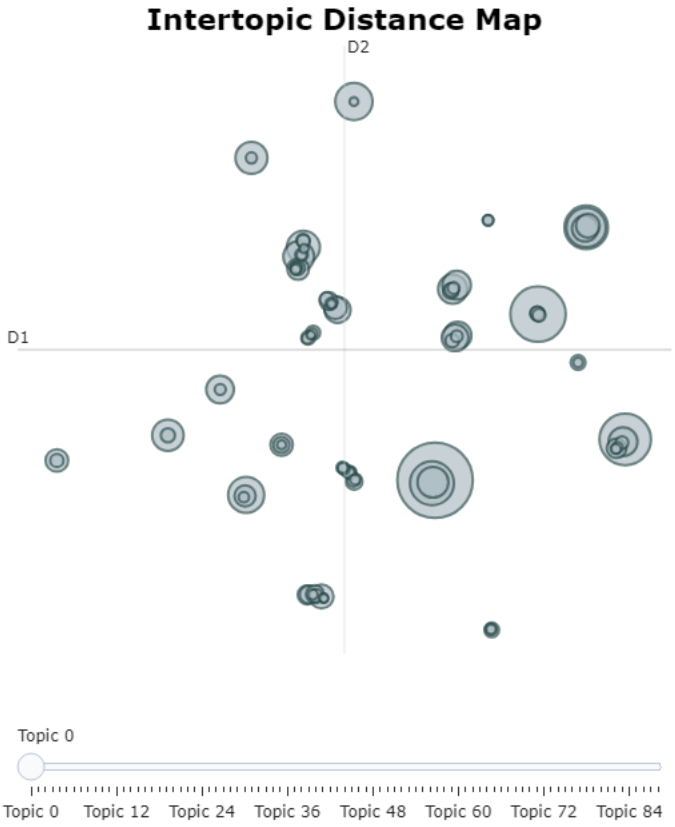

Figure 4.1: Intertopic distance map

Additionally, each determined topic can be analyzed by its most representative keywords. In the scope of this thesis, the ten most representative words of each topic were represented. This number can, of course, be modified by altering the argument for the top_n_words parameter in BERTopic. As an alternative, these representative words can be shown with their c-TF-IDF scores. Appendix C shows the top five words for each determined topic with their c-TF-IDF scores. This score is a measure of how relevant a word is to a specific topic relative to its frequency in other topics. Words with higher c-TF-IDF scores are more representative of the topic, meaning they are more distinctive or unique to that topic. For example, in topic 8, the words "emission," "scope," and "ghg" have the highest c-TF-IDF scores meaning that these words are the most relevant, representative, and unique to topic 8, which is shown by the emissions_scope_ghg_carbon word combination. Similarly, words such as "rights" and "human" in topic 14 (rights_human_labour_respect) have the highest scores showing

that these words are mainly unique to this determined topic. In other words, the word combinations that create the topic representations are built according to the words with the highest c-TF-IDF scores within a determined topic. In contrast, words with lower c-TF-IDF scores are less distinctive to that topic and might be more common within other topics.

Another feature of the BERTopic output is the count amount that demonstrates the number of input documents that are assigned to each topic. As it can be understood from the results, the number of input documents that are associated with occupational health and safety, financial performance, ethics, compliance, and anti-corruption topics is significant. These results are followed by the highest assigned documents to the topics related to supply chain, governance, risk management, waste management, carbon emission and climate change, biodiversity, customer satisfaction, sustainability reporting, environmental management, human rights, taxonomy, water management, digitalization, and so on. These statistics are beneficial when it's critical to identify the most frequently discussed topics inside a textual dataset from sustainability reports which can differ depending on the industry in question.

The other visualization tool that BERTopic offers is the hierarchical clustering dendrogram (See Figure 4.2). The level zero of the dendrogram shows how related topics that are shown with the same colors have been grouped together [84]. For instance, topic 46 (machine_manufacturing_equipment_industry), 48 (technological_ innovation _research_scientific), 18 (data_construction_system_digital), 59 (digital_data _information_processes), and 17 (security_information_cyber_protection) were grouped together because of their proximity in the content. Similarly, topics related to water management, waste management, biodiversity, and environmental impacts were categorized as one cluster comprising the environmental issues in the sustainability principles.

Essentially, a representation of this kind can aid academics in better understanding the criteria used by the algorithm to classify topics, and they can apply their specialized knowledge to these outcomes to provide more thorough categorizations. The results obtained from these visualizations were shared with the expert group to provide them with a reference in categorization and merging similar topics. The following section explains this process in detail.

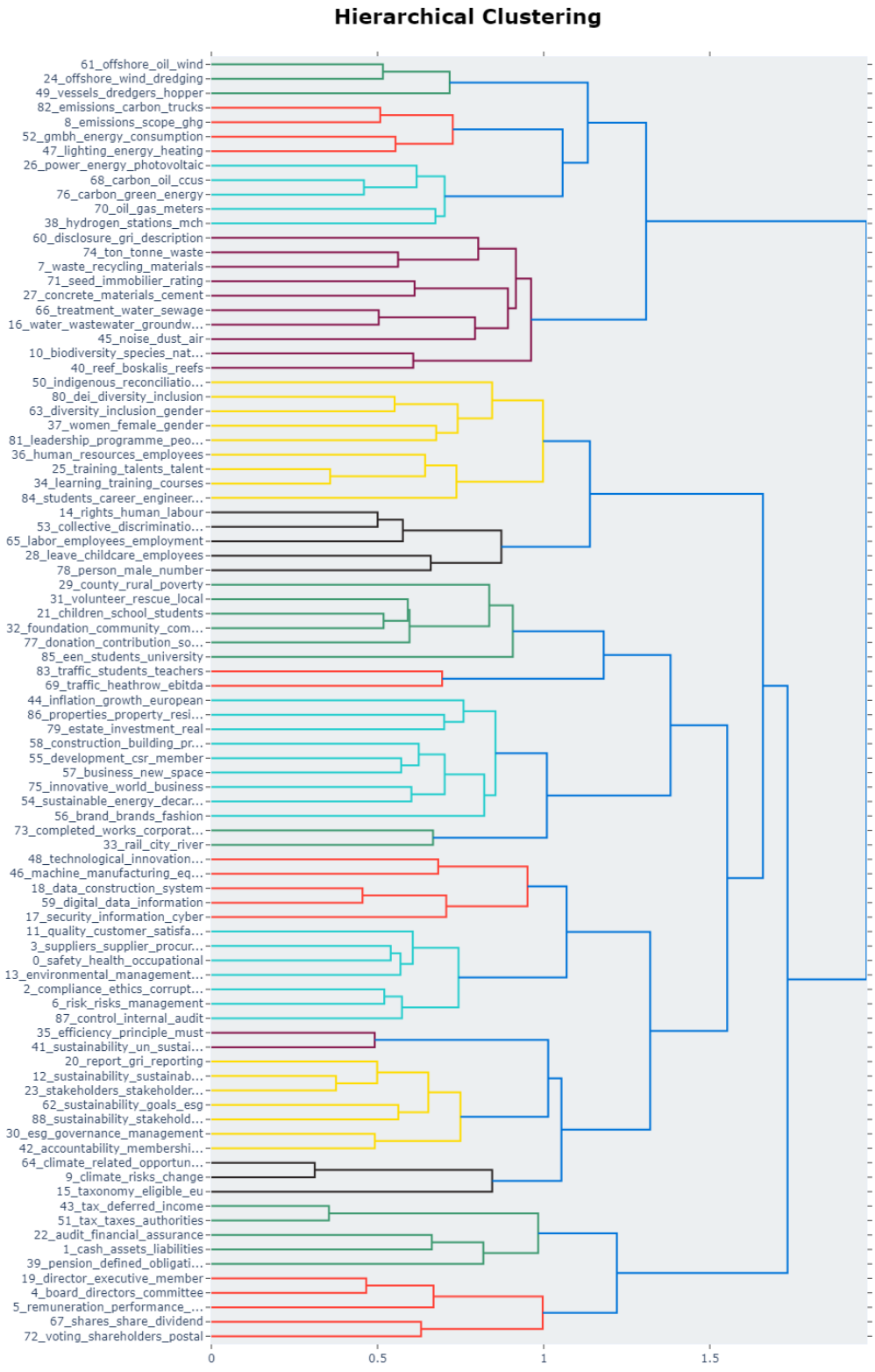

Figure 4.2: Hierarchical clustering

4.1.2 Merging the Similar Topics

As it is seen in the figure, although ninety topics were identified, some of these topics are in close proximity to each other, and when analyzed in a big picture, they create approximately twenty-one clusters of general topics. It is worth mentioning that this is the advantage of using BERTopic since there was no obligation to limit the number of topics to be identified before the text analysis. Actually, having no limitation on the topic number makes the text analysis more comprehensive and may result in finding all the hidden themes and subjects in a text corpus. Even though these topics may be numerous and highly detailed, they can subsequently be narrowed down by incorporating the expert opinion in the particular field under discussion and applying suggestions from the BERTopic visualization findings.

Additionally, unsupervised topic modeling results usually contain a few determined topics that are unrelated or irrelevant to the main focus area of the research, and this is inevitable when doing text analysis in a voluminous and diverse corpus of data. As a result, some topics that are unrelated should be identified by experts and be excluded. While analyzing and categorizing each topic, the experts considered topics 33 (rail_city_river_metro) and 73 (completed_works_corporation_sekkei) as unrelated topics and they suggested that they could be eliminated from the topic list. Topic -1 is also an outlier, and it contains documents that could not be placed in any other topic within the final results [21]. Therefore, this topic, which was represented as (energy_construction_business_development), was also eliminated from the final topic list.

This being said, some of the detected themes shared a lot of common content. Topics 25 and 34, for instance, which comprise training_talents_talent_employees and learning_training_courses_skills, have a lot in common and can be combined into one group called Employee Training and Talent Management. In the same way, topics 63 (diversity_inclusion_gender_inclusive), 78 (person_male_number_female), 37 (women_female_gender_positions), 80 (dei_diversity_inclusion_diverse), 81 (leadership_programme_people_inclusive) and 50 (indigenous_reconciliation_communities_aboriginal) all fall within the broad category of diversity, equal opportunities, and inclusion. This can lead to a reduction in the number of topics since many of the 90 topics can be combined into one.

Regarding this, the focus group examined every topic identified by BERTopic, as well as the significance of subjects observed in the hierarchical clustering and the intertopic distance map. Based on their professional judgment, they then categorized these topics in accordance with how similar they were to one another. As a result, the number of broad themes and topics for attaining sustainability within the construction industry was reduced to sixteen after related topics were combined. Table 4.2 shows the topics that are combined to create the general themes. Additionally, Figure 4.3 shows the final topic modeling results categorization. The figure was created using the Mind Manager software [100].

The validated final result shows that sixteen themes are the most important ones for the top global construction corporations to focus on in order to achieve sustainability within the construction sector. The following are the final validated topics by the expert group:

Environmental Impact of Construction Activities (including topics 71, 45, and 27), Sustainability Reporting (including topics 12, 15, 20, 60, 62, 56, and 41) Social Impact and Responsibility (including topics 21, 77, 85, 31, 32, 83, 29, and 42), Energy Efficiency and Renewable Energy (including topics 26, 68, 52, 38, 61, 54, 24, 47, 70, 57, and 49), Financial Performance (including topics 1, 22, 44, 39, 67, 69, 79, and 86), Waste Management (including topics 7, 74, and 13), Water Management (including topics 16, 66, and 35), Diversity, Equal Opportunities and Inclusion (including topics 37, 63, 78, 80, 81, and 50), Digitalization and Innovation (including topics 18, 59, 17, 48, 75, 46, 55, and 58), Employee Training and Talent Management (including topics 25, 34, and 84), Protecting Biodiversity and Habitat Management (including topics 10 and 40), Ethics, Compliance, Anti-corruption and Human Rights (including topics 2, 28, 36, 51, 43, 53, 65, and 14), Occupational Health and Safety (including topic 0), Combating Climate Change and Emission Reduction (including topics 64, 76, 82, 9, and 8), Customer Satisfaction and Service Quality (including topics 88, 72, 11, 23, and 3), and Governance (including topics 4, 5, 19, 30, 87, and 6).

It should be mentioned that these classifications were made using the intertopic distance map and hierarchical clustering results from BERTopic, along with some adjustments made from the expert perspective of the specialized group.

Table 4.2: Merging the topics to define general themes

General Themes	Topics
Environmental Impact of Construction Activities	71_seed_immobilier_rating_housing, 45_noise_dust_air_pollution, 27_concrete_materials_cement_wood
Sustainability Reporting	12_sustainability_sustainable_report_social, 15_taxonomy_eligible_eu_activities, 20_report_gri_reporting_assurance, 60_disclosure_gri_description_aspect, 62_sustainability_goals_esg_sustainable, 56_brand_brands_fashion_fiber, 41_sustainability_un_sustainable_companies
Social Impact and Responsibility	21_children_school_students_education, 77_donation_contribution_social_donations, 85_een_students_university_scholarship, 31_volunteer_rescue_local_relief 32_foundation_community_communities_social 83_traffic_students_teachers_school 29_county_rural_poverty_revitalization 42_accountability_memberships_contribution_organizational
Energy Efficiency and Renewable Energy	26_power_energy_photovoltaic_electric 68_carbon_oil_ccus_dioxide 52_gmbh_energy_consumption_vehicles 38_hydrogen_stations_mch_energy 61_offshore_oil_wind_energies 54_sustainable_energy_decarbonization_buildings 24_offshore_wind_dredging_vessel 47_lighting_energy_heating_electricity 70_oil_gas_meters_chemical 57_business_new_space_businesses 49_vessels_dredgers_hopper_vessel
Financial Performance	1_cash_assets_liabilities_financial 22_audit_financial_assurance_statements 44_inflation_growth_european_ihs 39_pension_defined_obligations_benefit 67_shares_share_dividend_ordinary 69_traffic_heathrow_ebitda_revenues 79_estate_investment_real_transportation 86_properties_property_residential_markets
Waste Management	7_waste_recycling_materials_hazardous 74_ton_tonne_waste_consumption 13_environmental_management_protection_environment
Water Management	16_water_wastewater_groundwater_treatment 66_treatment_water_sewage_technology 35_efficiency_principle_must_un
Diversity, Equal Opportunities and Inclusion	37_women_female_gender_positions 63_diversity_inclusion_gender_inclusive 78_person_male_number_female 80_dei_diversity_inclusion_diverse 81_leadership_programme_people_inclusive 50_indigenous_reconciliation_communities_aboriginal
Digitalization and Innovation	18_data_construction_system_digital 59_digital_data_information_processes 17_security_information_cyber_protection 48_technological_innovation_research_scientific 75_innovative_world_business_economic 46_machine_manufacturing_equipment_industry 55_development_csr_member_construction 58_construction_building_projects_center

Table 4.2: Merging the topics to define general themes (continued)

General Themes	Topics
Employee Training and Talent Management	25_training_talents_talent_employees
	34_learning_training_courses_skills
	84_students_career_engineering_opportunities
Protecting Biodiversity and Habitat Management	10_biodiversity_species_nature_areas
	40_reef_boskalis_reefs_mussel
Ethics, Compliance, Anti-corruption and Human Rights	2_compliance_ethics_corruption_anti
	28_leave_childcare_employees_family
	36_human_resources_employees_hr
	51_tax_taxes_authorities_compliance
	43_tax_deferred_income_taxable
	53_collective_discrimination_employees_bargaining
	65_labor_employees_employment_remuneration
	14_rights_human_labour_respect
Occupational Health and Safety	0_safety_health_occupational_accidents
Combating Climate Change and Emission Reduction	64_climate_related_opportunities_change
	76_carbon_green_energy_neutrality
	82_emissions_carbon_trucks_electric
	9_climate_risks_change_opportunities
	8_emissions_scope_ghg_carbon
Customer Satisfaction and Service Quality	88_sustainability_stakeholders_business_growth
	72_voting_shareholders_postal_meeting
	11_quality_customer_satisfaction_customers
	23_stakeholders_stakeholder_sustainability_materiality
	3_suppliers_supplier_procurement_supply
Governance	4_board_directors_committee_supervisory
	5_remuneration_performance_shares_executive
	19_director_executive_member_officer
	30_esg_governance_management_strategy
	87_control_internal_audit_supervision
	6_risk_risks_management_business

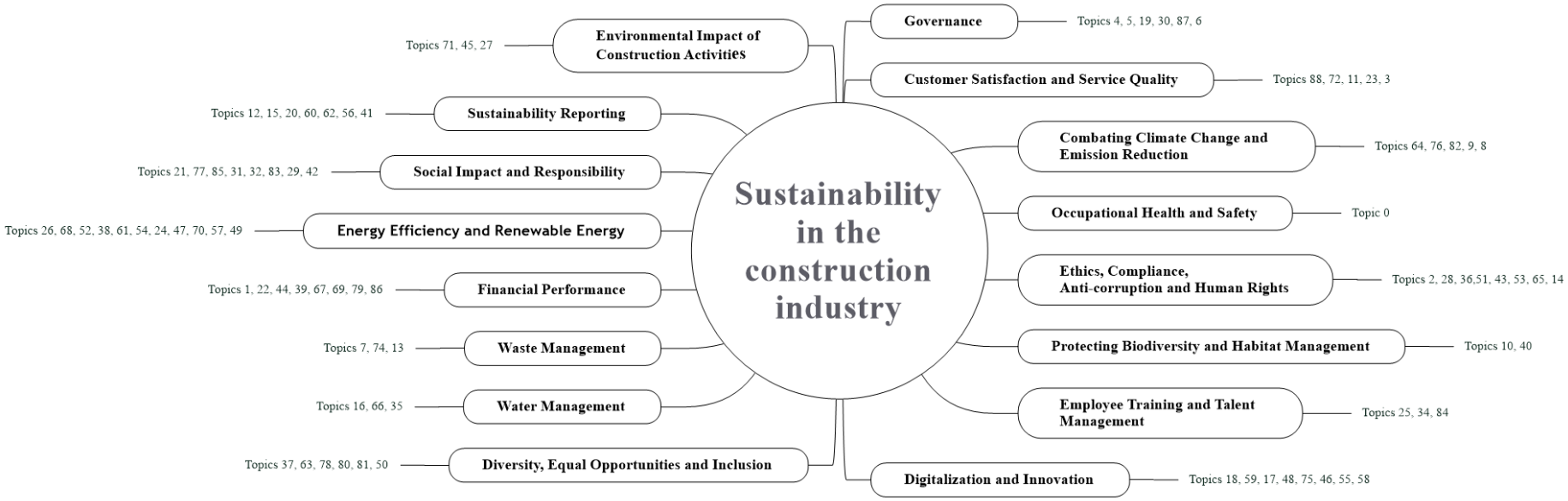

Figure 4.3: The most dominant themes in sustainable construction

4.2 The Creation of the Sustainability Balanced Scorecard (SBSC)

The creation of the framework included three steps. First, the prevailing topics discussed within the sustainability practices of the leading construction companies that were found as the results of the topic modeling with BERTopic were written in the form of strategic objectives to create an initial sustainability balanced scorecard (SBSC) representation. In the second step, since every BSC/SBSC is linked to a relative strategy map [15], the SBSC prepared in the previous step was presented to an expert group to create a strategy map as a basis for the implementation of the SBSC on the companies. Finally, the top representative keywords of the generated topics were combined with an expert view to suggest a list of key performance indicators (KPIs) for achieving sustainable performance within construction companies. The construction companies can utilize the proposed strategy map with the suggested KPIs as a reference to align their strategic objectives with the best practices of the leading construction companies in order to integrate sustainability into their strategy management.

4.2.1 The Initial SBSC for Sustainable Construction

With regard to a highly cited paper related to the analysis of the different architectures of SBSC [5], and with regard to the topics generated, the financial perspective of the generic balanced scorecard was expanded to include the three principles of the triple bottom line namely economic, environmental, and social matters (See Figure 4.4). According to the topics and the most prevalent themes in sustainable construction shown in figure 4.3, the objectives were added in their relevant place. For instance, water management was written in the environmental section as a strategic objective in the form of “optimize water management systems.” Similarly, diversity, equal opportunity, and inclusion were written as “promote diversity, equal opportunity, and inclusion” in the social section of the sustainability perspective in the SBSC. Financial performance was also stated in the form of “increase revenue” and “increase profit” according to the topics assigned to the financial performance.

Similar to the sustainability category, which included three different themes, the other categories were also divided to include different groups of topics according to the

strategic objectives that were supposed to be placed within the SBSC. The customer perspective in the generic BSC was changed to external stakeholders to include a wider array of customers in the construction sector. This perspective was divided into two main sections community (including objectives related to contributing to the economic development of the community and improving the company image) and customers and stakeholders (including the topics related to quality, customer satisfaction, stakeholder engagement, and enhancing market share).

The topics related to the internal process perspective could be stated in the form of three different capitals, namely human capital, technological capital, and organizational capital. The human included Occupational Health and Safety (OHS) and hiring the local people. It is worth mentioning that human capital can be defined with many subsections rather than these two; however, these are the themes that were mostly addressed in the reports of the leading contractors.

The organizational capital consisted of the objectives of supply chain, risk management, and governance. Similarly, the technological capital included cybersecurity, integration of the technology into the construction equipment, and utilizing digital platforms in the projects. These sections imply that the proposed SBSC is mostly suitable for construction firms with high-level and up-to-date capital that try to leverage their capabilities to integrate sustainability within their decision-making process and become more sustainable while excelling in merely financial aspects.

Finally, the learning and growth perspective was also presented in two different sections related to training and development, and technological capital. The training and development section consisted of employee training programs, talent acquisition, and offering student internships. Similar to the internal process, this perspective also included a technological capital section with objectives related to scientific activities and R&D projects.

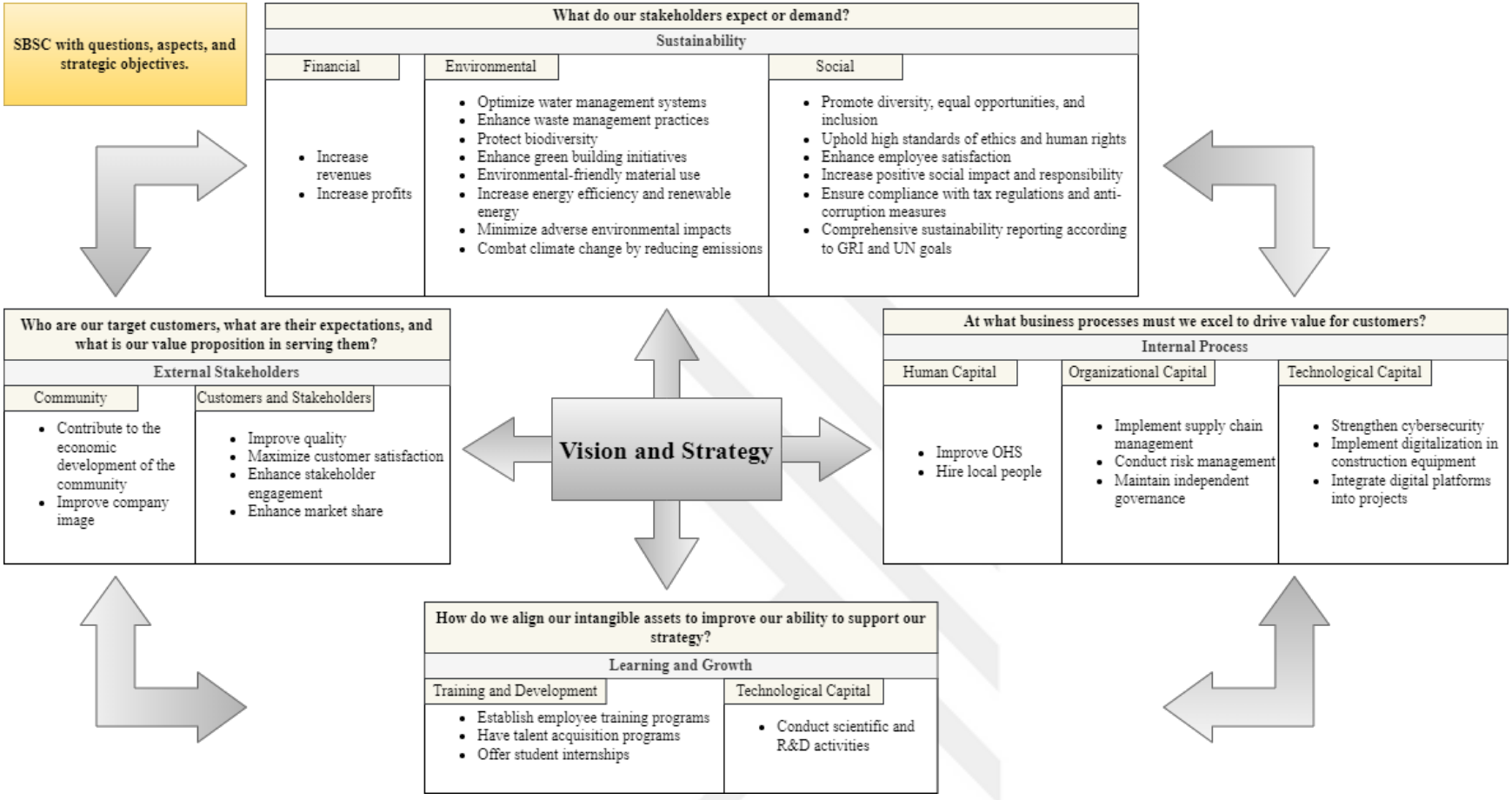

Figure 4.4: The SBSC for sustainable construction

4.2.2 The proposed strategy map for achieving sustainable construction

Every BSC/SBSC is linked to a strategy map [15]. Thus, it is suggested to create a strategy map using the defined objectives. For this purpose, the expert group used the strategic objectives to create a representative strategy map for achieving sustainable performance within construction companies. The strategic objectives for the creation of the strategy map were determined according to the most dominant themes and topics discussed in the best sustainability practices of the leading contractors according to the ENR list. Through the creation of this map, care was taken to include a combination of both lag and lead indicators. Moreover, the cause-and-effect relationships between the strategic objectives were determined using expert opinion in the sustainability and construction management field. Figure 4.5 depicts the proposed strategy map for achieving sustainable performance.

The lowest section of the strategy map included the learning and growth perspective with strategic objectives in two groups. Establishing training programs for the employees, talent acquisition programs, and offering student internships created the training and development group. The technological capital group included conducting scientific and R&D activities within this perspective. These groups were connected to all other categories in the higher parts with arrows since they either directly or indirectly contributed to the success of all other perspectives.

Three categories were included in the internal process perspective: extended technological capital, organizational capital, and human capital. The parts of the external stakeholders' perspective in the upper portion of the strategy map were impacted by the accomplishment of the strategic objectives within these groupings. Hiring local people in the human capital section contributed to the community's economic growth. Additionally, enhancing organizational capital, enhancing OHS, and integrating technology into internal processes raise quality achievement among customers and stakeholders, which raises customer satisfaction levels, increases market share and stakeholder engagement, and ultimately boosts sustainable increases in revenues from a future point of view. From the perspective of external stakeholders, contributions to the community's economic growth enhance the company's reputation and ultimately lead to higher revenues in the sustainability category's financial area. Similarly, positive social and environmental impact improves the company image and contributes to

the improvement of the financial section.

To complete the map, the experts added a few strategic objectives and made minor revisions to the overall objectives. Reducing costs, improving working capital, and cash management were defined as significant objectives in the construction sector since many construction companies have high capital, but they eventually end up being broke because they do not know how to manage their money. Therefore, this strategic objective was recommended to be added in the financial section because of its importance in the sector. Additionally, the experts stated that while integrating technology in the internal process results in enhancing quality in the customers section, it also improves project management and efficiency within a construction project. Therefore, improving project management and efficiency was also added to the organizational capital as a strategic objective and it was presented as one of the results of technological capital and as a causal objective for improving quality and reducing costs, improving working capital and cash management. Similarly, a stronger company image and a larger market share would result in attracting more customers and thus increasing revenues. Therefore, attracting customers was also added as an objective based on the expert recommendations. Talent retention as part of the human capital and becoming a sustainable firm for all stakeholders as part of the social section in the sustainability perspective were also recommended to have a spot in the strategy map. The objectives that were added to the strategy map with regard to the expert recommendations are presented in a different color in Figure 4.5.

After finalizing the strategy map, the result became mainly like the fully integrated and semi-hierarchical architecture of [5] (See Chapter 2). In this form, the financial perspective of the generic BSC is expanded to include the principles of the triple bottom line-namely economic, environmental, and social matters. Additionally, sustainability matters are also integrated within the other categories of the SBSC. Since this model is a semi-hierarchical form, most but not all of the cause-and-effect links are pointed upwards. It is worth mentioning that determining these causal relations between the strategic objectives is a very challenging task, and it does not represent a concrete relationship; thus, these relationships can be in a different form according to the objectives of each organization.

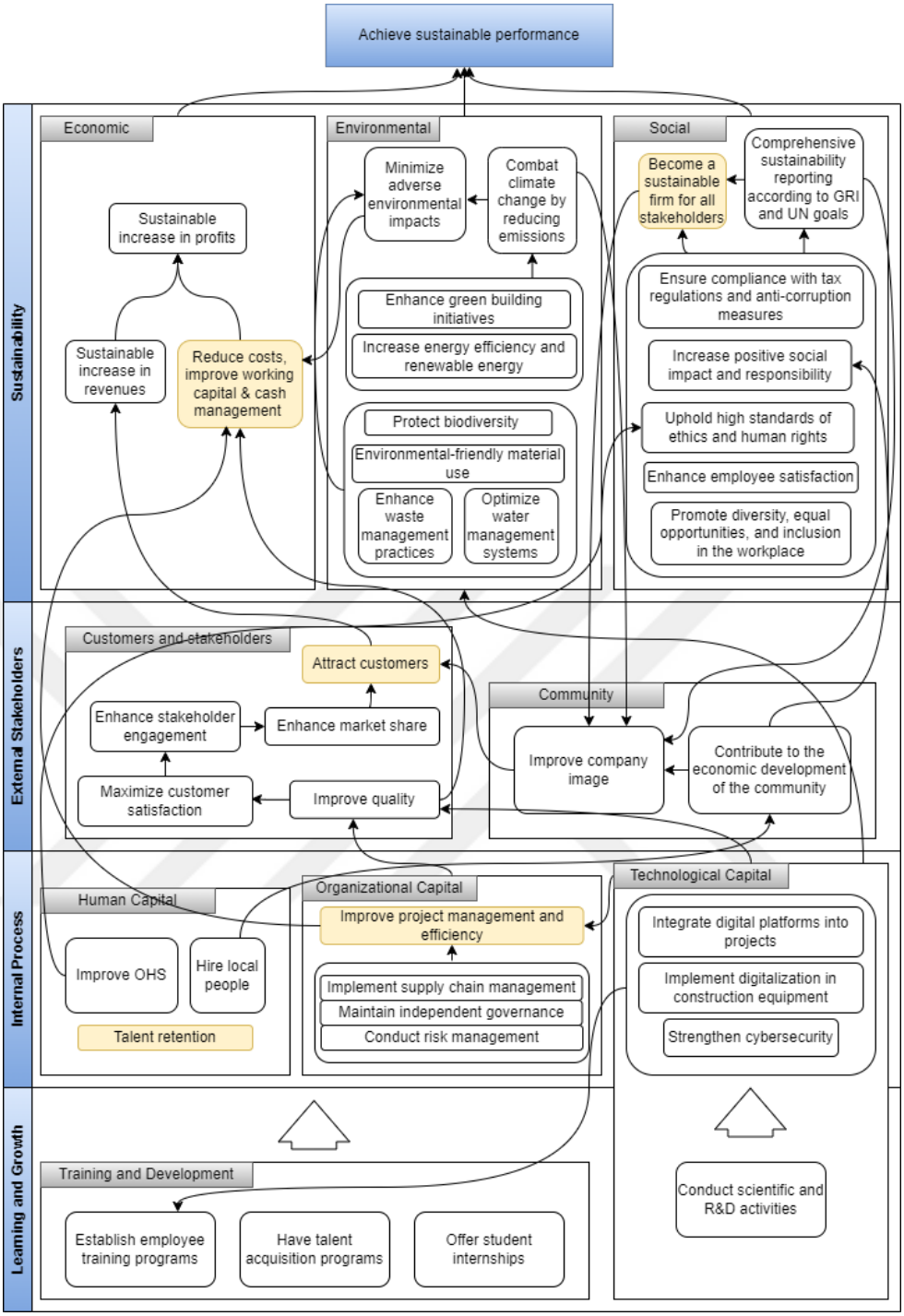

Figure 4.5: The proposed strategy map for achieving sustainable performance

4.2.3 Suggesting Key Performance Indicators (KPIs) for the SBSC

The most representative keywords that were retrieved for each issue based on their c-TF-IDF scores were combined with expert insights and were used as a starting point to propose some key performance indicators (KPIs) for the strategic objectives in order to complete the SBSC template for the construction industry.

Words like "training," "course," "employee," and "learning," for instance, can be used to generate a KPI like the number of employee training programs. Similar to this, terms like "women," "senior," "diversity," and "gender" can be used to infer measures like the diversity index and the percentage of women in senior positions. The existence of a term like "iso" is used to develop a metric, like the ISO compliance rate, for a strategic objective of improving quality, which measures the industry's level of quality. Words like "scope," "carbon," "emission," "machinery," and "reduction" can be utilized to produce metrics like scope 1, 2, and 3 carbon emission reduction rates for the strategic purpose of combating climate change and reducing carbon emissions. Terms like "water," "treatment," and "technology" can be combined to generate a metric like "advanced water treatment technology implementations." Words like "protected," "conservation," and "area" might be utilized to build a KPI like the protected area coverage rate for the strategic purpose of protecting biodiversity. Words like "machine," "equipment," "technology," and "intelligent" can be utilized to establish a metric like technology integration rate in equipment for the strategic objective of implementing digitalization in construction equipment. Similarly, terms like "vibration," "noise," "dust," and "pollution" can be utilized to establish a metric for these pollutants' reduction rates for the strategic objective of minimizing adverse environmental impacts.

In accordance, further KPIs were developed for various strategic goals. Tables 4.3, 4.4, 4.5, and 4.6 show the recommended KPIs for the SBSC template. For some objectives, it was not feasible to derive a KPI from the most representative keywords. Additionally, as mentioned in the previous section, a few objectives were added to the strategy map based on expert judgment. Therefore, the expert group was requested to recommend suitable KPIs that would represent the relative strategic objectives in the construction sector. These objectives and KPIs are marked with an asterisk (*) in the tables. For instance, as a metric for the improve quality objective, in addition to

the ISO compliance rate, it was also recommended to add the Non-Conformity report closure rate, as it is a popular metric in the construction sector. This KPI is a metric used to assess how well an organization resolves non-conformities found during quality control or audits. Additionally, it was recommended to assign gross profit KPI for evaluating sustainable increases in profits. The attracting customers and talent retention objectives can also be assessed by measuring customer increase rate and talent retention rate, respectively. A metric to evaluate how much a company has improved in project management and efficiency is measuring the project efficiency ratio in different fields, such as being within a budget and schedule. It is important to note that these metrics are only recommendations and that each company may modify its representative metrics and strategic objectives to fit its unique goals and capacities for measurement. In order to achieve sustainable performance in the construction sector, the tool that is being given provides a starting point for construction businesses to connect their strategic objectives with the way in which leading global construction organizations communicate their sustainability actions.

Table 4.3: The Sustainability category of the SBSC and suggested KPIs

Aspect	Strategic Objective	Key Performance Indicator (KPI)
Financial	Sustainable increase in revenues	Revenue Growth Rate
	Reduce costs, improve working capital and cash management*	Cost Reduction Percentage*
	Sustainable increase in profits	
		Gross Profit*
Environmental	Optimize water management systems	Advanced Water Treatment Technology Implementations
	Enhance waste management practices	Recycling Rate, Hazardous Waste Generation Rate
	Protect biodiversity	Protected Area Coverage
	Enhance green building initiatives	Green Building Certification Rate
	Environmental-friendly material use	Sustainable Materials Usage Rate
	Increase energy efficiency and renewable energy	Renewable Energy Usage Rate in Projects
	Minimize adverse environmental impacts	Noise and Dust Pollution Reduction Rate
	Combat climate change by reducing emissions	Scope 1, 2, and 3 Emissions Reduction Rate
Social	Promote diversity, equal opportunities, and inclusion	Percentage of Women in Senior Positions, Diversity Index
	Uphold high standards of ethics and human rights	Number of Reported Ethical Violations (e.g. Child Labor)
	Enhance employee satisfaction	Employee Satisfaction Rate
	Increase positive social impact and responsibility	Social Responsibility Projects, Volunteer Participation Rate
	Ensure compliance with tax regulations and anti-corruption measures	Number of Reported Corruption Incidents, Tax Compliance Rate
	Comprehensive sustainability reporting according to GRI and UN goals	GRI Disclosure Completeness Rate, UN Sustainable Development Goals (SDGs) Alignment Score
	Become a sustainable firm for all stakeholders*	N.A*

Table 4.4: The External Stakeholders category of the SBSC and suggested KPIs

Aspect	Strategic Objective	Key Performance Indicator (KPI)
Customers and Stakeholders	Improve quality	ISO Compliance Rate, Non-Conformity Report Closure Rate*
	Maximize customer satisfaction	Stakeholder/Customer Satisfaction Index
	Enhance stakeholder engagement	Stakeholder Engagement Rate
	Enhance market share	Market Share Percentage
	Attract customers*	Customer Increase Rate*
Community	Contribute to the economic development of the community	Poverty Alleviation Rate in Project Areas
	Improve company image	Implementing Activities to Improve Company Image

Table 4.5: The Internal Process category of the SBSC and suggested KPIs

Aspect	Strategic Objective	Key Performance Indicator (KPI)
Human Capital	Improve OHS	Total Incident Rate, OHS Training Completion Rate, Lost Time Incident Rate
		Local Employment Rate
	Hire local people	Talent Retention Rate*
	Talent Retention*	
Organizational Capital	Implement supply chain management	Supply chain management team implementation*
	Conduct risk management	Risk Mitigation Effectiveness
	Maintain independent governance	Board Independence Rate
	Improve project management and efficiency	Project Efficiency Ratio (budget, schedule, productivity)*
Technological Capital	Strengthen cybersecurity	Number of Cybersecurity Incidents
	Implement digitalization in construction equipment	Technology Integration Rate in Equipment
	Integrate digital platforms into projects	Number of BIM-Based Projects

Table 4.6: The Learning and Growth category of the SBSC and suggested KPIs

Aspect	Strategic Objective	Key Performance Indicator (KPI)
Training and Development	Establish employee training programs	Number of Employee Training Programs
	Have talent acquisition programs	Number of Talent Acquisition Events
	Offer student internships	Number of Internship Vacancies for Students
Technological Capital	Conduct scientific and R&D activities	Number of Scientific and R&D Projects

4.3 Case Study

After validating every framework, it should also be verified to prove that it serves the purpose for which it was created. The proposed SBSC framework was specifically designed to achieve sustainability within the construction sector, and it was created by analyzing the best practices of the leading global construction businesses. The proposed framework consisted of an SBSC with a strategy map foundation, which was enhanced by the integration of the triple bottom line principles and designed by defining strategic objectives with regard to the most dominant topics discussed within the sustainability reports of leading organizations. A suggested KPI list was also developed based on expert recommendations and the most representative keywords of the topics. The strategy map was designed with the aim of being a reference made according to the sustainability practice trends of the industry leaders for creating an SBSC customized to the objectives and performance indicators of each company. Therefore, the objectives presented in the strategy map were general statements such as combating climate change, promoting diversity, and increasing positive social impact.

Every business establishes its own strategies and unique measurements to achieve them, according to [28]. As a result, the strategy map offers a broad framework for attaining sustainability in the particular construction industry, but various businesses can set their own objectives within it. In the broad area of diversity, for instance, a business in the Middle East may strive to have a larger proportion of female employees in the management team, whilst a business in the USA may want to employ people from other ethnicities.

All in all, the framework that has been offered is merely meant to be used as a guide; it gives businesses ideas about the kinds of topics they should include in their objectives and initiatives in order to be regarded as sustainable organizations within their industry. Feedback from the company's management was sought to evaluate the framework's relevance and effectiveness in enhancing their sustainability efforts. Additionally, they were invited to suggest any improvements or modifications to improve the framework further.

4.3.1 Customizing the SBSC for a Turkish Construction Company

The proposed SBSC was customized according to the objectives, metrics, and priorities of a Turkish construction company to confirm the framework's applicability. In order to make the SBSC suitable to be used in their strategic planning, the necessary modifications were applied on the framework and the SBSC was customized according to the vision and priorities of the company. In this regard, the management of the company recommended modifying several items in the architecture of the strategy map; just like the way that technological capital was expanded through the learning and growth perspective and the internal process perspective, they preferred to transfer the social section of the sustainability perspective in the suggested strategy map to be extended between the internal process and the external stakeholders perspectives. They also preferred to transfer the environmental perspective to the internal process perspective and leave the sustainability category merely with the economic section. They stated that this model would better represent and visualize how sustainable initiatives contribute to the improvement in the financial performance of the company. The strategy map deemed appropriate by the company is depicted in Figure 4.6.
After arranging the strategy map, specific metrics of the company for assessing the strategic objectives were determined according to the company's evaluation capabilities. The customized SBSC for the sample Turkish company is shown in Figure 4.7. The strategic objectives and their relevant metrics in bold are items that the company already had in their strategic plan and they preferred to measure them with their own relative KPIs. The objectives and the metrics that are underlined are the new items that the company found to be useful and stated that they could be considered while making their strategic plan, as it reflects the priorities of the sector.

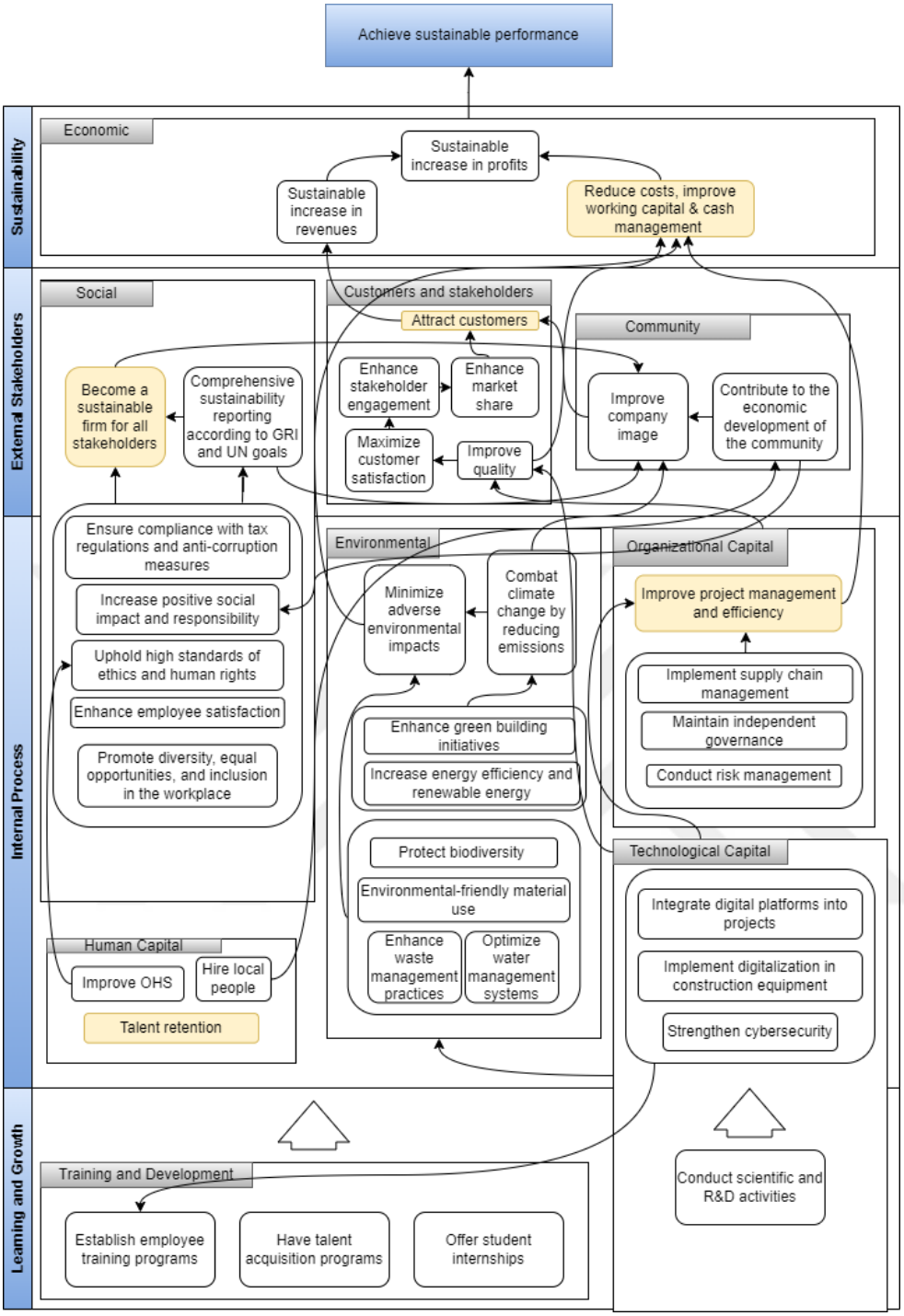

Figure 4.6: The proposed strategy map for the sample company

The sustainability report of this company in Turkey was also included in the input dataset for the topic modeling and creating the general SBSC in the first place; therefore, many factors stated in the SBSC were already in the strategy plan of this company. However, as this tool was created according to the combination of documents from leading construction companies all over the world, there were still some objectives that the company could consider adding to its profile. Additionally, some topics were addressed in their sustainability report but the actual performance of the organization in those topics was not necessarily measured and evaluated. The suggested KPI list, as a reference, assisted the company in this regard as well. To be more specific, in the sustainability report of the company, the strategic objective of combating climate change by emission reduction was addressed with the energy and emission management terminology. The company evaluated its performance by measuring its Scope 1 and Scope 2 carbon emissions. However, it is suggested in the KPI list to measure the indirect carbon emissions of the company as well (Scope 3). The company found it useful to include this metric as well in its future strategic plans, as this would make its initiatives more aligned with the sustainability trends and practices of the leading contractors on the global stage. In contrast, for this specific company, part of reducing its adverse environmental impacts was its noise, dust, and vibration impact on the project sites; with that being said, while in the KPI list, it was suggested to measure the reduction rate in these matters, the managers preferred to declare their actions that reduce the noise, dust, and the vibration from their projects, declaring that measuring this metric would also be one of their initiatives in the following years when they have enough resources. Similarly, for the objective of biodiversity protection, instead of choosing a metric to declare the protected area coverage, the company preferred to declare its relative plans for biodiversity protection within its projects.
Some objectives such as offering student internships, implementing digitalization in the construction equipment, conducting risk management, relations with local people, Eco-friendly material use, reporting comprehensiveness, anti-corruption, and employee rights and engagement were also addressed comprehensively in the sustainability report of the company but they were not clearly evaluated to compare the company's sustainability performance in consecutive years. The KPI list gave the management an initial reference list to incorporate some metrics into their program and communicate their sustainable impact more effectively by using the KPI list as

an initial reference list. The SBSC would be further completed when the company sets its targets and relevant initiatives and programs for reaching them within the following year.

Lastly, the managers offered their opinions regarding the suitability of the suggested SBSC for this study's purpose. According to them, this tool could be considered a very helpful resource that construction companies can use as a reference to align their strategies with the best sustainable practices carried out by the top global contractors. By doing this, they will be able to improve their sustainability impact and, ideally, secure a spot among the top construction companies with minimum adverse global impacts. Furthermore, the most trending topics in the field of sustainable constructions that were found during the topic modeling phase and the strategy map developed by the experts could be used as a foundation for construction managers and the sustainability teams within an organization to publish more effective and comprehensive sustainability reports.

They also added that the fact that this tool was designed with regard to a vast number of documents from companies of diverse continents yielded a general tool that can be customized according to the priorities and evaluation capabilities of each company. This is advantageous because different businesses in various locations and circumstances may have different goals, and they can use this framework to evaluate how well they are doing in achieving these objectives. Even the same company may decide after two or three years to shift its focus and set some specific goals for the upcoming year. Because of its adaptability and flexibility, these businesses can use this tool as a useful reference to adjust and synchronize its components with their own priorities. Nevertheless, it also covers most of the trending broad topics that need to be covered in their strategy plan in order for them to be regarded as a sustainable construction firm. As a result, the SBSC provided in this study is a useful resource for businesses looking to keep up with the sustainability trends used by top construction companies worldwide.

Achieve Sustainable Performance

	Aspect	Strategic Objective	Key Performance Indicator (KPI)
Sustainability	Financial	Sustainable increase in revenues Reduce costs, improve working capital & cash management Sustainable increase in profits	Revenue Growth Rate Cost Reduction Percentage Gross Profit
External Stakeholders	Community	Contribute to the economic development of the community **Improve company image**	Poverty Alleviation Rate in Project Areas **Activities by the Reputation Management Working Group**
	Customers and Stakeholders	**Improve quality** **Maximize customer satisfaction** **Enhance stakeholder engagement** Enhance market share Attract customers	**ISO Certifications, Non-Conformity Report Closure Rate** **Stakeholder/Customer Satisfaction Rate (Survey)** **Stakeholder Engagement Rate** Market Share Percentage Customer Increase Rate
	Social	**Diversity and Equal Opportunity** **Ethics and Compliance** **Employee Rights and Engagement** **Enhance employee satisfaction** **Social Impact and Responsibility** **Ensure compliance with tax regulations and anti-corruption measures** **Comprehensive sustainability reporting according to GRI and UN goals**	**Percentage of Women in Senior Positions**, Diversity Index **Number of Reported Ethical Violations (e.g. Child Labor)** Employee Engagement Rate **Employee Satisfaction Rate (Survey)** **Social Responsibility Projects, Hours of Volunteering Activities** Number of Reported Corruption Incidents, Tax Compliance Rate GRI Disclosure Completeness Rate, UN Sustainable Development Goals (SDGs) Alignment Score
Internal Process	Environmental	**Water Management** **Waste Management** **Protecting Biodiversity** **Green Building Certifications** **Eco-Friendly Material Use** **Energy and Emission Management** **Reducing Noise, Dust, and Vibration**	**Total Water Consumption and Water Discharge** **Total Amount of Waste Produced** **Biodiversity Management Plan in Projects** **BREEAM and LEED Certified Buildings** Sustainable Materials Usage Rate Renewable Energy Usage Rate in Projects, **Scope 1, 2,** and 3 **Emissions Reduction Rate** **Actions to Reduce Noise, Vibration, and Dust**
	Human Capital	**Occupational Health and Safety** **Relations with local people** **Talent Retention**	**Total Incident Rate, OHS Training Completion Rate, Lost Time Incident Rate** Local Employment Rate **Talent Retention Rate**
	Organizational Capital	**Responsible Supply Chain** **Conduct risk management** Maintain independent governance Improve project management and efficiency	**Conducting Supply Chain Management** Risk Mitigation Effectiveness Board Independence Rate Project Efficiency Ratio (budget, schedule, productivity)
Learning and Growth	Technological Capital	Strengthen cybersecurity **Implement digitalization in construction equipment** Integrate digital platforms into projects **Conduct scientific and R&D activities**	Number of Cybersecurity Incidents Technology Integration Rate in Equipment Number of BIM-Based Projects **Number of Activities by the R&D center**
	Training and Development	**Talent Management and Training** **Offer student internships**	**Number of Employee Training Programs, Number of Talent Acquisition Events** Number of Internship Vacancies for Students

Figure 4.7: SBSC for the sample company

CHAPTER 5

CONCLUSIONS

This chapter provides an overview of the research conducted for this thesis. It outlines the key objectives and the methodological approach employed to address the research questions. Subsequently, the contributions made by this study toward achieving sustainability within construction practices are stated. In this part, the practical contributions of the proposed framework for the construction managers and sustainability experts are highlighted, and its contributions to the current body of knowledge around the sustainability subjects for achieving sustainable constructions are emphasized. Finally, this chapter also acknowledges the limitations of the research. It also identifies areas where further investigation of the study could be beneficial and makes suggestions for future research directions.

5.1 Overview

Nowadays, the overall performance of construction companies is not merely judged based on their financial performance. To become a leader in the industry, these companies should not only excel in their projects but also should integrate sustainability issues into their strategic plans. Hence, construction businesses have integrated sustainable practices into their business plan in order to reduce their negative impact on the world, and they have disclosed their contributions to sustainable development in their sustainability reports. These initiatives should be in line with the strategic objectives of the companies, and they should be continually evaluated in order to have a positive impact on the performance of the company. However, a major barrier has been the absence of a comprehensive framework for assessing and contrasting the sustainability performance of construction enterprises. This study proposed the

sustainability balanced scorecard (SBSC), a novel framework for evaluating sustainability performance in construction-related sectors.

This framework was created by extending the aspects of the generic BSC in accordance with the triple bottom line principles. Due to the gap in the literature and the lack of generally accepted indicators, it was a challenging task to find the related indicators for the categories of this framework. Therefore, in order to fill these gaps, sustainability reports were analyzed to investigate the most significant sustainability topics in the construction industry. These reports lacked a consistent format; thus, in order to improve their analysis, artificial intelligence (AI) techniques had to be integrated. The present work used BERTopic, an unsupervised machine learning methodology that leverages natural language processing (NLP), to conduct topic modeling and identify the most frequently addressed subjects and practices related to corporate social responsibility (CSR) in the sustainability reports of leading worldwide contractors.

As a result, 90 topics were identified by BERTopic as the most important sustainability subjects in construction-specific industries. These findings were then given to a focus group of experts consisting of construction managers, civil engineers, sustainability specialists, and university professors in the construction management field to categorize similar topics and define the most prevalent subjects in achieving sustainability within construction practices. The categorized topics were then used to define strategic objectives for achieving sustainable performance in the construction industry. Since every BSC/SBSC is linked to a strategy map with strategic objectives and causal relationships between those objectives, a strategy map for achieving sustainable performance in the construction industry was created by the expert group. Subsequently, the most representative keywords of the topics were used along with the insights and recommendations of the specialists to suggest related KPIs for evaluating each strategic objective in the SBSC. The proposed framework was considered only a base that each construction company could customize according to its needs, vision, and available metrics.

In order to verify the applicability of this tool, the proposed SBSC was customized for a Turkish construction company, and the top management of the company was asked to give feedback about the relevance and applicability of this tool in making them be-

come more sustainable. They verified this framework as a useful tool that includes the main sustainability trends within the construction sector and would help construction companies plan and organize their sustainability initiatives in accordance with the way that the leaders of the industry communicate their sustainability actions. They also indicated that the best aspect of this framework was that it could be modified according to different scenarios and different needs and objectives of the companies. Furthermore, the most trending topics in the field of sustainable constructions that were found during the topic modeling phase and the strategy map developed by the experts could be used as a foundation for construction managers and the sustainability teams within an organization to publish more effective and comprehensive sustainability reports.

5.2 Contributions

The contributions of this research are twofold: First of all, the most prevalent topics that are discussed within the best sustainable practices of the leading global construction companies were found in this research using BERTopic, which is a topic modeling approach that leverages NLP techniques. This study adds to the body of knowledge on text classification and NLP applications related to the analysis of sustainability reports in the unique construction sector. The findings of this thesis can assist construction firms in publishing more comprehensive annual sustainability reports. These results can serve as a reference for matching their sustainability initiatives and strategic objectives with the latest worldwide trends. Furthermore, the disclosed topics provide the companies with a series of subjects through which they can communicate their sustainability initiatives in the construction industry.

Second, the suggested SBSC framework for the construction industry is built upon the strategy map created with the assistance of sustainability and construction management specialists, as well as a proposed KPI list according to the expert recommendations and the most representative keywords of each categorized subject. It should be noted that this tool is merely a reference, and each company can implement the strategy map along with its relative metrics to customize them according to their specific needs, objectives, and available metrics. Therefore, the proposed framework provides

managers with a novel and innovative performance evaluation tool that is specifically designed for the construction industry and is adaptable enough to be utilized on a global scale and adjusted when the construction companies' sustainability priorities shift over time.

5.3 Limitations and Future Research Directions

Of course, like any other research, this study was not free of limitations. First, among 250 top construction companies, approximately half of them published a sustainability report or an annual report with some sustainability-related sections. Therefore, future research can analyze a bigger corpus of datasets for an even more detailed text classification result. Additionally, while BERTopic has many advantages as one of the most updated topic modeling methods, future studies can perform topic modeling using other techniques as well to conduct a comparative study and examine their advantages and disadvantages in performing a comprehensive text classification in the construction field.

Second, for this research, the "all-MiniLM-L6-v2" pre-trained sentence transformer model was used for the transformation of the input documents to numerical representations. Future researchers can employ different pre-trained models for word embeddings in order to do more detailed and comprehensive text analysis. Similarly, BERTopic offers alternatives for each of the five steps through its algorithm. Future studies can perform topic modeling with different model combinations to achieve the best outcomes. Moreover, some of the reports of the above-mentioned dataset were in a language other than English, and since the pre-trained model used for the word embeddings was specific to the English language, they had to be excluded from the corpus. Future research can focus on the analysis of datasets with more diverse languages by using more comprehensive word embeddings during the topic modeling phase.

Third, the input dataset consisted of many unrelated words, like the names of the cities, people, countries, organizations, etc., that adversely affected the topic modeling results. These words had to be manually added to the extended stop words list after each topic modeling attempt to get more clean results. Although it was aimed

to reduce these noises as much as possible, minor unrelated words between the final results were inevitable. Future researchers can attempt to automatize this manual and time-consuming process. Additionally, it is worth mentioning that BERTopic offers continuous topic modeling rather than discrete topic modeling [101]. Consequently, when modeling is repeated again, the stochastic nature of the model produces different results [84]. Nevertheless, the research's findings showed that although the number of topics and their representations varied after each run, they were remarkably similar, and the categorization of the determined topics yielded almost the same final topic series defined by the expert group. Finally, by using different text classification methods, the results obtained in this study can be used to analyze the different reporting trends in the construction industries of different continents. This will highlight how political, social, and economic issues indirectly affect construction companies' reporting practices and the critical areas of focus in various global regions.

REFERENCES

[1] R. S. Kaplan and P. David, "Norton, the balanced scorecard: Translating strategy into action," *Harvard Business Review Press*, vol. 1, 1996.

[2] D. Katic and L. Bevanda, "Balanced scorecard in construction," in *International Scientific Conference People Buildings and Environment-Czech Republic*, 2010.

[3] P. R. Niven, *Balanced scorecard: Step-by-step for government and nonprofit agencies*. John Wiley & Sons, 2008.

[4] F. Figge, T. Hahn, S. Schaltegger, and M. Wagner, "The sustainability balanced scorecard–linking sustainability management to business strategy," *Business strategy and the Environment*, vol. 11, no. 5, pp. 269–284, 2002.

[5] E. G. Hansen and S. Schaltegger, "The sustainability balanced scorecard: A systematic review of architectures," *Journal of Business Ethics*, vol. 133, pp. 193–221, 2016.

[6] A. Shamshiri, K. R. Ryu, and J. Y. Park, "Text mining and natural language processing in construction," *Automation in Construction*, vol. 158, p. 105200, 2024.

[7] P. Ahi and C. Searcy, "An analysis of metrics used to measure performance in green and sustainable supply chains," *Journal of cleaner production*, vol. 86, pp. 360–377, 2015.

[8] K. Kabirifar, M. Mojtahedi, C. Wang, and V. W. Tam, "Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review," *Journal of cleaner production*, vol. 263, p. 121265, 2020.

[9] Y. C. f. E. . A. United Nations Environment Programme, "Building materials and the climate: Constructing a new future," 2023-09.

[10] H. Zhao and N. Li, "Evaluating the performance of thermal power enterprises using sustainability balanced scorecard, fuzzy delphic and hybrid multi-criteria decision making approaches for sustainability," *Journal of Cleaner Production*, vol. 108, pp. 569–582, 2015.

[11] H. Al-Mawali, "Proposing a strategy map based on sustainability balanced scorecard and dematel for manufacturing companies," *Sustainability Accounting, Management and Policy Journal*, vol. 14, no. 3, pp. 565–590, 2023.

[12] P. Norton David, "The balanced scorecard: translating strategy into action," *Harvard Business School Press, Boston*, 1996.

[13] R. S. Kaplan and D. P. Norton, "Putting the balanced scorecard to work," in *The economic impact of knowledge*, pp. 315–324, Routledge, 2009.

[14] M. J. Epstein and P. S. Wisner, "Using a balanced scorecard to implement sustainability," *Environmental quality management*, vol. 11, no. 2, pp. 1–10, 2001.

[15] B. Torgautov, A. Zhanabayev, A. Tleuken, A. Turkyilmaz, C. Borucki, and F. Karaca, "Performance assessment of construction companies for the circular economy: A balanced scorecard approach," *Sustainable Production and Consumption*, vol. 33, pp. 991–1004, 2022.

[16] F.-H. Chen, T.-S. Hsu, and G.-H. Tzeng, "A balanced scorecard approach to establish a performance evaluation and relationship model for hot spring hotels based on a hybrid mcdm model combining dematel and anp," *International Journal of Hospitality Management*, vol. 30, no. 4, pp. 908–932, 2011.

[17] C. Mio, A. Costantini, and S. Panfilo, "Performance measurement tools for sustainable business: A systematic literature review on the sustainability balanced scorecard use," *Corporate social responsibility and environmental management*, vol. 29, no. 2, pp. 367–384, 2022.

[18] M. Gunduz and N. H. Al-Naimi, "Construction projects delay mitigation using integrated balanced scorecard and quality function deployment," *Engineering, Construction and Architectural Management*, vol. 29, no. 5, pp. 2073–2105, 2022.

[19] Global Reporting Initiative, "Global reporting initiative standards," n.d. Accessed: 2024-07-22.

[20] K. Chowdhary and K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.

[21] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[24] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[25] D. Angelov, "Top2vec: Distributed representations of topics," *arXiv preprint arXiv:2008.09470*, 2020.

[26] P. Gupta, B. Ding, C. Guan, and D. Ding, "Generative ai: A systematic review using topic modelling techniques," *Data and Information Management*, p. 100066, 2024.

[27] Z. Wang, J. Chen, J. Chen, and H. Chen, "Identifying interdisciplinary topics and their evolution based on bertopic," *Scientometrics*, pp. 1–26, 2023.

[28] B. D. Dağıdır and B. Özkan, "A comprehensive evaluation of a company performance using sustainability balanced scorecard based on picture fuzzy ahp," *Journal of Cleaner Production*, vol. 435, p. 140519, 2024.

[29] P. R. Niven, *Balanced scorecard step-by-step: Maximizing performance and maintaining results*. John Wiley & Sons, 2002.

[30] J. Dobrovič, M. Urbański, P. Gallo, E. Benková, and V. Čabinová, "Balanced scorecard concept as a tool of strategic management and its usage in the construction industry," *Polish Journal of Management Studies*, vol. 18, no. 2, pp. 59–72, 2018.

[31] S. Koprivica, J. Škondrić, and M. Bendić, "Balanced scorecard implementation in construction industry," in *8th INTERNATIONAL CONFERENCE Contemporary achievements in civil engineering, Subotica, Serbia, 2021, 569-578*, 2021.

[32] M. Kagioglou, R. Cooper, and G. Aouad, "Performance management in construction: a conceptual framework," *Construction management and economics*, vol. 19, no. 1, pp. 85–95, 2001.

[33] R. A. Stewart and S. Mohamed, "Utilizing the balanced scorecard for it/is performance evaluation in construction," *Construction innovation*, vol. 1, no. 3, pp. 147–163, 2001.

[34] T. K. Chan and P. T. Hiap, "A balanced scorecard approach to measuring industry performance," *Journal of Construction in Developing Countries*, vol. 1, no. 2012, pp. 23–41, 2012.

[35] T.-V. Luu, S.-Y. Kim, H.-L. Cao, and Y.-M. Park, "Performance measurement of construction firms in developing countries," *Construction Management and Economics*, vol. 26, no. 4, pp. 373–386, 2008.

[36] L. O. Oyewobi, A. O. Windapo, and J. O. B. Rotimi, "Measuring strategic performance in construction companies: a proposed integrated model," *Journal of Facilities Management*, vol. 13, no. 2, pp. 109–132, 2015.

[37] L. Koo, "Building balanced scorecard on the house of quality," in *The 1st Industrial Engineering and Management (IEM) Symposium, Transformational Strategy towards the 21st Century*, pp. 20–21, 1998.

[38] A. Price, "The strategy process within large construction orgnisations," *Engineering, Construction and Architectural Management*, vol. 10, no. 4, pp. 283–296, 2003.

[39] M. Vukomanovic and M. Radujkovic, "The balanced scorecard and efqm working together in a performance management framework in construction industry," *Journal of Civil Engineering and Management*, vol. 19, no. 5, pp. 683–695, 2013.

[40] S. Agrawal, R. K. Singh, and Q. Murtaza, "Outsourcing decisions in reverse logistics: Sustainable balanced scorecard and graph theoretic approach," *Resources, Conservation and Recycling*, vol. 108, pp. 41–53, 2016.

[41] A. Gunasekaran and A. Spalanzani, "Sustainability of manufacturing and services: Investigations for research and applications," *International journal of production economics*, vol. 140, no. 1, pp. 35–47, 2012.

[42] A. Eifert and C. Julmi, "Challenges and how to overcome them in the formulation and implementation process of a sustainability balanced scorecard (sbsc)," *Sustainability*, vol. 14, no. 22, p. 14816, 2022.

[43] T. Tsalis, I. Nikolaou, E. Grigoroudis, and K. Tsagarakis, "A framework development to evaluate the needs of smes in order to adopt a sustainability-balanced scorecard," *Journal of Integrative Environmental Sciences*, vol. 10, no. 3-4, pp. 179–197, 2013.

[44] L. Araújo, H. Oliveira, and L. Gomes, "Sustainability balanced scorecard for a brazilian agricultural and livestock company," in *E3S Web of Conferences*, vol. 159, p. 04005, EDP Sciences, 2020.

[45] I. E. Nikolaou and T. A. Tsalis, "Development of a sustainable balanced scorecard framework," *Ecological Indicators*, vol. 34, pp. 76–86, 2013.

[46] I. Dias-Sardinha, L. Reijnders, and P. Antunes, "Developing sustainability balanced scorecards for environmental services: A study of three large portuguese companies.," *Environmental Quality Management*, vol. 16, no. 4, 2007.

[47] I. D. Sardinha, L. Reijnders, and P. Antunes, "Using corporate social responsibility benchmarking framework to identify and assess corporate social responsibility trends of real estate companies owning and developing shopping centres," *Journal of Cleaner Production*, vol. 19, no. 13, pp. 1486–1493, 2011.

[48] S. Falle, R. Rauter, S. Engert, and R. J. Baumgartner, "Sustainability management with the sustainability balanced scorecard in smes: Findings from an austrian case study," *Sustainability*, vol. 8, no. 6, p. 545, 2016.

[49] L. J. Córdova-Aguirre and J. M. Ramón-Jerónimo, "Designing a sustainability assessment framework for peruvian manufacturing small and medium enterprises applying the stakeholder theory approach," *Sustainability*, vol. 16, no. 5, p. 1853, 2024.

[50] A. Rabbani, M. Zamani, A. Yazdani-Chamzini, and E. K. Zavadskas, "Proposing a new integrated model based on sustainability balanced scorecard (sbsc) and mcdm approaches by using linguistic variables for the performance evaluation of oil producing companies," *Expert systems with applications*, vol. 41, no. 16, pp. 7316–7327, 2014.

[51] J.-S. Kang, C.-F. Chiang, K. Huangthanapan, and S. Downing, "Corporate social responsibility and sustainability balanced scorecard: The case study of family-owned hotels," *International Journal of Hospitality Management*, vol. 48, pp. 124–134, 2015.

[52] T. Fatima and S. Elbanna, "Advancing sustainable performance management in the hospitality industry: A novel framework based on a health-inclusive balanced scorecard," *Tourism Management Perspectives*, vol. 48, p. 101141, 2023.

[53] C. De Villiers, P. Rouse, and J. Kerr, "A new conceptual model of influences driving sustainability based on case evidence of the integration of corporate sustainability management control and reporting," *Journal of cleaner production*, vol. 136, pp. 78–85, 2016.

[54] F. M. Tsai, T.-D. Bui, M.-L. Tseng, K.-J. Wu, and A. S. Chiu, "A performance assessment approach for integrated solid waste management using a sustainable balanced scorecard approach," *Journal of cleaner production*, vol. 251, p. 119740, 2020.

[55] G. H. Brundtland, "World commission on environment and development," *Environmental policy and law*, vol. 14, no. 1, pp. 26–30, 1985.

[56] J. Elkington, "The triple bottom line for 21st century business," *Journal of Experimental Psychology: General*, vol. 136, 1997.

[57] R. Kiani Mavi, D. Gengatharen, N. Kiani Mavi, R. Hughes, A. Campbell, and R. Yates, "Sustainability in construction projects: a systematic literature review," *Sustainability*, vol. 13, no. 4, p. 1932, 2021.

[58] H. Kang and J. Kim, "Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods," *Applied Sciences*, vol. 12, no. 11, p. 5614, 2022.

[59] M. Gutierrez-Bustamante and L. Espinosa-Leal, "Natural language processing methods for scoring sustainability reports—a study of nordic listed companies," *Sustainability*, vol. 14, no. 15, p. 9165, 2022.

[60] E. Park, Y. Kim, A. Lee, J. Kim, and H. Kong, "Study on the global sustainability of the korean construction industry based on the gri standards," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, p. 4231, 2023.

[61] E. Commission, "A renewed eu strategy 2011-14 for corporate social responsibility," 2011.

[62] G. Halkos and S. Nomikos, "Corporate social responsibility: Trends in global reporting initiative standards," *Economic Analysis and Policy*, vol. 69, pp. 106–117, 2021.

[63] United Nations Global Compact, "The ten principles of the un global compact," n.d. Accessed: 2024-07-22.

[64] M. Angin, B. Taşdemir, C. A. Yılmaz, G. Demiralp, M. Atay, P. Angin, and G. Dikmener, "A roberta approach for automated processing of sustainability reports," *Sustainability*, vol. 14, no. 23, p. 16139, 2022.

[65] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[66] L. Hillebrand, M. Pielka, D. Leonhard, T. Deußer, T. Dilmaghani, B. Kliem, R. Loitz, M. Morad, C. Temath, T. Bell, *et al.*, "sustain. ai: a recommender system to analyze sustainability reports," in *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pp. 412–416, 2023.

[67] A.-I. Moreno and T. Caminero, "Application of text mining to the analysis of climate-related disclosures," *International Review of Financial Analysis*, vol. 83, p. 102307, 2022.

[68] Y. Li and M. Rockinger, "Unfolding the transitions in sustainability reporting," *Sustainability*, vol. 16, no. 2, p. 809, 2024.

[69] Y. Yu, S. Scheidegger, J. Elliott, and Å. Löfgren, "climatebug: A data-driven framework for analyzing bank reporting through a climate lens," *Expert Systems with Applications*, vol. 239, p. 122162, 2024.

[70] S. Q. Jafari, S. Shokouhyar, and S. Shokoohyar, "Producer-consumer sustainability continuum: Mutual understanding to implement extended producer responsibility," *Journal of Cleaner Production*, vol. 374, p. 133880, 2022.

[71] N. Raman, G. Bang, and A. Nourbakhsh, "Mapping esg trends by distant supervision of neural language models," *Machine Learning and Knowledge Extraction*, vol. 2, no. 4, pp. 453–468, 2020.

[72] I. Nechaev and D. S. Hain, "Social impacts reflected in csr reports: Method of extraction and link to firms innovation capacity," *Journal of Cleaner Production*, vol. 429, p. 139256, 2023.

[73] A. Gupta, A. Chadha, and V. Tewari, "A natural language processing model on bert and yake technique for keyword extraction on sustainability reports," *IEEE Access*, 2024.

[74] R. Niveditha, N. K. NS, M. R. Parimi, A. Raam, and S. Babu, "Develop csr themes using text-mining and topic modelling techniques," in *2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, pp. 67–71, IEEE, 2020.

[75] N. Székely and J. Vom Brocke, "What can we learn from corporate sustainability reporting? deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique," *PloS one*, vol. 12, no. 4, p. e0174807, 2017.

[76] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism management*, vol. 59, pp. 467–483, 2017.

[77] R. Churchill and L. Singh, "The evolution of topic modeling," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–35, 2022.

[78] M. Selvi, K. Thangaramya, M. Saranya, K. Kulothungan, S. Ganapathy, and A. Kannan, "Classification of medical dataset along with topic modeling using lda," in *Nanoelectronics, Circuits and Communication Systems: Proceeding of NCCS 2017*, pp. 1–11, Springer, 2019.

[79] H. Kim, I. Cho, and M. Park, "Analyzing genderless fashion trends of consumers' perceptions on social media: using unstructured big data analysis through latent dirichlet allocation-based topic modeling," *Fashion and Textiles*, vol. 9, no. 1, p. 6, 2022.

[80] H. Gadekar and N. Bugalia, "Automatic classification of construction safety reports using semi-supervised yake-guided lda approach," *Advanced Engineering Informatics*, vol. 56, p. 101929, 2023.

[81] Z. Wu, P. Xie, J. Zhang, B. Zhan, and Q. He, "Tracing the trends of general construction and demolition waste research using lda modeling combined with topic intensity," *Frontiers in Public Health*, vol. 10, p. 899705, 2022.

[82] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin, "Experimental explorations on short text topic mining between lda and nmf based schemes," *Knowledge-Based Systems*, vol. 163, pp. 1–13, 2019.

[83] R. Egger and J. Yu, "Identifying hidden semantic structures in instagram data: a topic modelling comparison," *Tourism Review*, vol. 77, no. 4, pp. 1234–1246, 2021.

[84] R. Egger and J. Yu, "A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts," *Frontiers in sociology*, vol. 7, p. 886498, 2022.

[85] G. Yazıcı and T. Ozansoy Çadırcı, "Creating meaningful insights from customer reviews: a methodological comparison of topic modeling algorithms

and their use in marketing research," *Journal of Marketing Analytics*, pp. 1–23, 2023.

[86] L. Gan, T. Yang, Y. Huang, B. Yang, Y. Y. Luo, L. W. C. Richard, and D. Guo, "Experimental comparison of three topic modeling methods with lda, top2vec and bertopic," in *International Symposium on Artificial Intelligence and Robotics*, pp. 376–391, Springer, 2023.

[87] S. C. Turan, K. Yildiz, and B. Büyüktanir, "Comparison of lda, nmf and bertopic topic modeling techniques on amazon product review dataset: A case study," in *International Conference on Computing, Intelligence and Data Analytics*, pp. 23–31, Springer, 2023.

[88] Y. Liu, A. H. B. Alias, N. A. Haron, N. A. Bakar, and H. Wang, "Exploring three pillars of construction robotics via dual-track quantitative analysis," *Automation in Construction*, vol. 162, p. 105391, 2024.

[89] Engineering News-Record, "Enr top 250 international contractors 2021." `https://www.enr.com/toplists/2021-Top-250-International-Contractors-Preview`, 2021. Accessed: 2022-10-05.

[90] I. Scarpino, C. Zucco, R. Vallelunga, F. Luzza, and M. Cannataro, "Investigating topic modeling techniques to extract meaningful insights in italian long covid narration," *BioTech*, vol. 11, no. 3, p. 41, 2022.

[91] H. Walsh, "Csr report nlp walkthrough." `https://github.com/hannahawalsh/HTTF4-ESG-and-NLP/blob/main/CSR_Report_NLP_Walkthrough.ipynb`, 2024. Accessed: 2024-08-11.

[92] "Apache tika." `https://tika.apache.org/`. Accessed: 2024-08-11.

[93] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

[94] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," 2019.

[95] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[96] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*, pp. 160–172, Springer, 2013.

[97] A. Kousis and C. Tjortjis, "Investigating the key aspects of a smart city through topic modeling and thematic analysis," *Future Internet*, vol. 16, no. 1, p. 3, 2023.

[98] T. Joachims *et al.*, "A probabilistic analysis of the rocchio algorithm with tfidf for text categorization," in *ICML*, vol. 97, pp. 143–151, Citeseer, 1997.

[99] Skanska, "Environmental report 1998." `https://group.skanska.com/4907e6/siteassets/sustainability/reporting-publications/sustainability-reporting/environ_report98.pdf`, 1998. Accessed: 2024-08-07.

[100] Corel Corporation, "Mindmanager." `https://www.mindmanager.com`, 2023. Version 21.

[101] A. Alcoforado, T. P. Ferraz, R. Gerber, E. Bustos, A. S. Oliveira, B. M. Veloso, F. L. Siqueira, and A. H. R. Costa, "Zeroberto: Leveraging zero-shot text classification by topic modeling," in *International Conference on Computational Processing of the Portuguese Language*, pp. 125–136, Springer, 2022.

APPENDICES

A. ENR's top 250 international contractors

RANK 2021	RANK 2020	FIRM	2020 REVENUE $ MIL. INT'L	2020 REVENUE $ MIL. TOTAL	2020 NEW CONTRACTS $ MIL	GENERAL BUILDING	MANUFACTURING	POWER	WATER SUPPLY	SEWER / WASTE	INDUS. / PETROLEUM	TRANSPORTATION	HAZARDOUS WASTE	TELECOM
1	1	**ACS ACTIVIDADES DE CONSTRUCCIÓN Y SERVICIOS SA,** Madrid, Spain†	36,687.0	42,357.1	36,685.6	36	1	8	2	0	6	30	0	4
2	2	**HOCHTIEF AG,** Essen, Germany†	27,536.0	28,612.0	26,461.0	45	2	1	1	1	4	25	0	5
3	3	**VINCI,** Rueil-Malmaison, France†	23,463.0	50,141.0	49,619.0	6	0	21	2	0	4	44	1	12
4	4	**CHINA COMMUNICATIONS CONSTRUCTION GROUP LTD.,** Beijing, China†	21,348.4	100,811.6	226,906.8	12	1	0	1	2	0	84	0	0
5	5	**BOUYGUES,** Paris, France†	17,284.0	32,160.0	30,214.0	28	1	6	0	0	1	60	0	2
6	6	**STRABAG SE,** Vienna, Austria†	15,936.1	18,954.5	20,538.3	30	0	1	4	2	5	57	0	0
7	7	**POWER CONSTRUCTION CORP. OF CHINA,** Beijing, China†	13,007.9	65,717.6	124,133.8	5	0	67	4	1	0	23	0	0
8	9	**SKANSKA AB,** Stockholm, Sweden†	11,342.0	14,523.1	16,262.3	42	5	5	1	1	3	42	0	1
9	8	**CHINA STATE CONSTRUCTION ENGINEERING CORP. LTD.,** Beijing, China†	10,746.2	195,658.7	389,167.8	74	2	3	0	0	2	19	0	0
10	11	**FERROVIAL,** Madrid, Spain†	10,154.3	13,505.3	16,851.6	21	0	6	5	6	0	60	0	0
11	12	**CHINA RAILWAY CONSTRUCTION CORP. LTD.,** Beijing, China†	8,375.0	134,745.0	370,680.0	21	0	7	3	0	0	63	0	1
12	**	**FLUOR,** Irving, Texas, U.S.A.†	7,517.4	11,672.4	9,004.9	9	0	3	0	0	75	1	0	12
13	13	**CHINA RAILWAY GROUP LTD.,** Beijing, China†	7,419.9	141,852.7	377,774.2	14	0	0	0	0	3	76	0	0
14	**	**SAIPEM SPA,** San Donato Milanese, Italy†	7,132.0	7,574.0	9,488.0	0	0	8	0	0	89	2	0	0
15	17	**EIFFAGE,** Velizy-Villacoublay, France†	5,240.0	17,221.0	19,000.0	28	10	23	4	0	8	26	1	1
16	14	**HYUNDAI ENGINEERING & CONSTRUCTION CO. LTD.,** Seoul, South Korea	5,075.6	14,518.4	NA	9	13	22	0	0	42	14	0	0
17	21	**ROYAL BAM GROUP NV,** Bunnik, Netherlands†	4,517.0	7,790.0	NA	56	0	0	0	0	0	44	0	0
18	19	**WEBUILD SPA,** Milan, Italy†	4,324.5	6,070.3	6,093.4	10	0	0	19	9	0	56	0	0
19	22	**CHINA NATIONAL CHEMICAL ENG'G GROUP CORP. LTD.,** Beijing, China†	4,221.8	18,325.7	44,654.5	6	0	11	0	0	74	7	0	0
20	18	**TECNICAS REUNIDAS SA,** Madrid, Spain†	4,195.3	4,299.7	2,200.0	0	0	6	0	0	94	0	0	0
21	15	**CHINA ENERGY ENGINEERING CORP. LTD.,** Beijing, China†	4,177.4	28,469.7	83,774.8	5	0	68	12	6	4	2	0	0
22	**	**EXYTE GMBH,** Stuttgart, Germany†	3,928.7	4,322.3	4,197.2	0	90	0	0	0	3	0	0	6
23	29	**KAJIMA CORP.,** Tokyo, Japan†	3,894.1	15,062.1	15,502.7	72	15	1	1	1	6	2	0	2
24	16	**BECHTEL,** Reston, Va., U.S.A.†	3,778.0	12,239.0	3,722.0	0	0	0	3	0	51	46	0	0
25	27	**LARSEN & TOUBRO LTD.,** Chennai, India†	3,684.7	13,502.5	19,643.0	5	6	40	4	2	34	9	0	0
26	42	**SACYR,** Madrid, Spain†	3,682.1	5,425.6	1,379.0	10	0	3	3	1	2	81	0	0
27	20	**PETROFAC LTD.,** Jersey, U.K.†	3,664.7	3,664.7	1,384.5	0	0	4	0	0	96	0	0	0
28	23	**RENAISSANCE CONSTRUCTION,** Ankara, Turkey†	3,652.8	4,530.7	4,149.8	42	1	17	0	0	18	20	0	0
29	32	**PORR AG,** Vienna, Austria	3,461.0	6,316.0	NA	0	0	0	0	0	0	0	0	0
30	33	**SAMSUNG ENGINEERING CO. LTD.,** Seoul, South Korea†	3,423.3	5,365.5	8,136.0	0	12	0	1	10	77	0	0	0
31	24	**OBAYASHI CORP.,** Tokyo, Japan†	3,400.0	15,204.0	15,204.0	24	0	4	24	2	0	34	0	0
32	38	**ORASCOM CONSTRUCTION PLC,** Dubai, U.A.E.†	3,347.8	3,371.1	2,919.0	63	2	4	7	3	6	16	0	0
33	34	**CHINA PETROLEUM ENGINEERING CO. LTD.,** Beijing, China†	3,340.5	10,593.5	13,496.9	0	0	0	0	0	100	0	0	0
34	36	**SAMSUNG C&T CORP.,** Seoul, South Korea	3,254.1	9,844.6	7,908.9	12	20	22	14	0	0	32	0	0
35	25	**CHINA NATIONAL MACHINERY INDUSTRY CORP.,** Beijing, China†	3,113.0	6,927.6	9,562.5	8	4	62	9	0	8	5	0	2
36	37	**FCC SA,** Madrid, Spain†	3,049.5	7,556.5	3,162.2	5	0	0	17	47	0	22	0	0
37	55	**WOOD,** Aberdeen, Scotland, U.K.	3,011.3	3,662.3	NA	1	0	11	2	0	86	0	0	0
38	40	**HYUNDAI ENGINEERING CO. LTD.,** Jongro-gu, South Korea	2,800.5	6,091.6	7,798.0	1	16	19	1	1	63	0	0	0
39	31	**LENDLEASE,** Barangaroo, Australia†	2,724.2	5,856.5	5,857.0	98	1	0	0	0	0	0	0	1
40	47	**OHLA,** Madrid, Spain	2,554.3	3,084.4	2,468.5	0	0	0	0	0	0	0	0	0
41	39	**MAIRE TECNIMONT SPA,** Milan, Italy†	2,475.4	2,615.7	1,706.8	0	0	0	0	0	100	0	0	0
42	45	**JGC HOLDINGS CORP.,** Yokohama, Japan†	2,397.0	3,820.0	6,800.0	1	0	0	0	0	99	0	0	0
43	51	**JAN DE NUL GROUP (SOFIDRA SA),** Capellen, Luxembourg†	2,242.0	2,249.0	3,755.0	12	0	0	0	2	5	8	0	0
44	50	**ROYAL BOSKALIS WESTMINSTER NV,** Papendrecht, Netherlands†	2,221.0	2,885.0	3,741.0	0	0	23	0	0	37	30	0	0
45	52	**DANIELI & C. O.M. SPA,** Buttrio, Italy†	2,069.0	2,225.0	2,097.0	0	0	0	0	0	100	0	0	0
46	60	**WORLEY LTD.,** North Sydney, Australia	1,930.8	1,930.8	2,415.4	0	0	1	0	0	98	0	0	0
47	26	**CONSOLIDATED CONTRACTORS GROUP,** Athens, Greece†	1,872.4	1,872.4	1,152.9	22	0	2	1	1	37	34	0	0
48	48	**BESIX,** Brussels, Belgium†	1,838.5	3,151.3	2,594.8	67	0	2	2	6	0	23	0	0
49	43	**CIMIC GROUP LTD.,** Sydney, Australia†	1,803.1	10,491.6	5,996.2	18	0	1	0	4	2	19	0	0
50	44	**GS ENGINEERING & CONSTRUCTION,** Seoul, South Korea†	1,764.0	8,402.5	9,357.9	15	13	3	15	0	37	18	0	0

Figure A.1: ENR's top 250 international contractors - Part 1

RANK 2021	RANK 2020	FIRM	2020 REVENUE $ MIL. INT'L	2020 REVENUE $ MIL. TOTAL	2020 NEW CONTRACTS $ MIL	GENERAL BUILDING	MANUFACTURING	POWER	WATER SUPPLY	SEWER / WASTE	INDUS. / PETROLEUM	TRANSPORTATION	HAZARDOUS WASTE	TELECOM
51	160	**SHANGHAI ELECTRIC GROUP CO. LTD.,** Shanghai, China	1,731.9	2,501.8	2,714.7	0	0	100	0	0	0	0	0	0
52	59	**DAEWOO ENGINEERING AND CONSTRUCTION CO. LTD.,** Seoul, South Korea†	1,721.8	7,478.6	12,787.3	18	0	2	2	0	39	19	6	0
53	41	**CHINA METALLURGICAL GROUP CORP.,** Beijing, China†	1,659.8	54,100.2	140,590.8	29	0	6	0	1	54	9	0	0
54	53	**SHAPOORJI PALLONJI & CO. PRIVATE LTD.,** Mumbai, India†	1,655.7	5,584.8	3,881.0	79	0	5	1	0	0	15	0	0
55	63	**CHINA ZHONGYUAN ENGINEERING CORP.,** Beijing, China	1,635.4	1,635.4	NA	0	0	100	0	0	0	0	0	0
56	61	**LIMAK INSAAT SANAYI VE TICARET AS,** Ankara, Turkey	1,632.0	3,985.0	1,660.0	11	0	0	9	0	0	77	0	0
57	80	**ANT YAPI INDUSTRY & TRADE INC.,** Istanbul, Turkey†	1,590.0	1,659.0	1,199.0	93	0	0	0	0	5	3	0	0
58	49	**CHIYODA CORP.,** Yokohama, Japan†	1,529.3	2,848.8	8,118.8	0	0	0	0	0	100	0	0	0
59	71	**KIEWIT CORP.,** Omaha, Neb., U.S.A.†	1,318.9	11,200.7	7,862.1	0	0	8	4	0	53	34	0	0
60	54	**SINOMA INTERNATIONAL ENGINEERING CO. LTD.,** Beijing, China†	1,297.8	2,449.4	4,956.9	0	85	3	0	0	8	3	0	0
61	64	**ABENGOA,** Seville, Spain†	1,270.6	1,425.3	910.4	4	0	68	27	1	0	1	0	0
62	68	**SK ECOPLANT,** Seoul, South Korea†	1,248.3	6,909.4	3,635.4	0	23	1	1	0	64	12	0	0
63	62	**CITIC CONSTRUCTION CO. LTD.,** Beijing, China†	1,242.1	2,518.9	5,238.1	37	0	8	0	0	38	16	0	0
64	69	**VAN OORD N.V.,** Rotterdam, Netherlands†	1,210.3	2,046.4	NA	0	0	19	0	0	18	63	0	0
65	57	**SNC-LAVALIN INC.,** Montreal, Quebec, Canada†	1,195.4	2,328.5	1,272.4	1	0	46	1	0	50	1	0	0
66	67	**PENTA-OCEAN CONSTRUCTION CO. LTD.,** Bunkyo-ku, Japan†	1,181.3	4,356.0	4,893.4	24	1	0	1	13	0	60	0	0
67	73	**CHINA GENERAL TECHNOLOGY (GROUP) HOLDING CO. LTD.,** Beijing, China†	1,151.7	3,906.1	5,514.7	0	0	80	1	0	3	10	0	6
68	78	**YAPI MERKEZI INSAAT VE SANAYI AS,** Istanbul, Turkey†	1,068.8	1,214.7	14.3	0	0	0	0	2	0	98	0	0
69	75	**CTCI CORP.,** Taipei, Taiwan†	1,059.5	1,835.4	4,171.7	0	3	11	0	3	83	1	0	0
70	86	**ENKA INSAAT VE SANAYI AS,** Istanbul, Turkey†	1,056.3	1,263.3	1,600.1	29	0	52	0	0	14	5	0	0
71	79	**TAKENAKA CORP.,** Osaka, Japan†	1,043.0	11,074.0	11,079.0	19	66	0	0	0	0	15	0	0
72	81	**CHINA JIANGXI INT'L ECON. AND TECH. COOP. CO. LTD.,** Nanchang, China†	1,023.6	1,086.8	1,712.5	25	0	26	24	0	0	25	0	0
73	111	**CHINA ELECTRIC POWER EQUIPMENT AND TECH. CO. LTD.,** Beijing, China†	1,019.4	1,019.4	190.2	0	0	100	0	0	0	0	0	0
74	88	**BONATTI GROUP,** Parma, Italy†	995.0	1,024.0	745.0	2	0	0	0	0	98	0	0	0
75	85	**ZHONGMEI ENGINEERING GROUP LTD.,** Nanchang, China	989.9	989.9	779.4	8	0	0	37	2	0	53	0	0
76	74	**ED. ZUBLIN AG,** Stuttgart, Germany†	972.8	4,921.3	NA	33	0	0	2	10	2	3	0	0
77	108	**THE ARAB CONTRACTORS CO. (OSMAN AHMED OSMAN),** Cairo, Egypt†	962.7	3,608.9	5,530.0	13	0	46	2	0	0	39	0	0
78	95	**HARBIN ELECTRIC INTERNATIONAL CO. LTD.,** Harbin, China	942.6	980.0	3,742.0	0	0	100	0	0	0	0	0	0
79	98	**ITINERA SPA,** Tortona, Italy†	923.6	1,290.2	1,116.4	26	0	0	0	0	0	74	0	0
80	65	**TEKFEN CONSTRUCTION AND INSTALLATION CO. INC.,** Istanbul, Turkey†	908.0	1,013.0	495.0	12	0	0	0	0	55	34	0	0
81	90	**NORINCO INTERNATIONAL COOPERATION LTD.,** Beijing, China†	894.9	1,165.8	1,571.3	6	2	45	0	0	2	4	0	0
82	66	**TOYO ENGINEERING CORP.,** Narashino-shi, Chiba, Japan†	891.5	1,662.0	1,110.1	0	0	3	0	0	93	0	0	0
83	89	**DL E&C CO. LTD.,** Seoul, South Korea†	880.6	6,147.7	6,545.8	0	0	6	3	0	59	31	0	0
84	82	**ZHEJIANG CONSTRUCTION INVESTMENT GROUP CO. LTD.,** Hangzhou, China†	871.6	14,377.9	20,300.8	86	0	1	0	0	10	3	0	0
85	76	**ATLAS GROUP,** Katy, Texas, U.S.A.†	824.0	849.0	614.0	100	0	0	0	0	0	0	0	0
86	70	**SINOPEC ENGINEERING (GROUP) CO. LTD.,** Beijing, China†	807.2	8,853.5	9,199.8	0	0	0	0	0	100	0	0	0
87	77	**SHIMIZU CORP.,** Tokyo, Japan†	779.6	13,698.1	11,259.2	33	25	5	0	0	15	19	0	1
88	**	**CBRE,** Dallas, Texas, U.S.A.†	773.0	806.0	NA	0	0	0	0	0	0	0	0	0
89	97	**CHINA INTERNATIONAL WATER & ELECTRIC CORP. (CWE),** Beijing, China	772.8	772.8	654.0	4	0	91	1	0	0	5	0	0
90	139	**SHANDONG HI-SPEED GROUP CO. LTD.,** Jinan, China†	736.1	736.1	381.6	41	1	0	0	0	23	35	0	0
91	102	**PER AARSLEFF HOLDING A/S,** Viby, Denmark†	694.7	2,088.6	2,561.0	2	46	4	0	22	0	27	0	0
92	116	**ONUR TAAHHUT TASIMACILIK INSAAT TIC. VE SANAYI AS,** Ankara, Turkey†	693.8	838.6	576.4	0	0	0	0	0	0	100	0	0
93	101	**SHANGHAI CONSTRUCTION GROUP CO. LTD.,** Shanghai, China	692.5	45,863.4	56,076.7	71	0	0	0	9	0	20	0	0
94	**	**QINGJIAN GROUP CO. LTD.,** Qingdao, Shandong, China†	685.3	9,660.5	11,772.6	92	0	0	1	0	1	6	0	0
95	103	**BL HARBERT INTERNATIONAL,** Birmingham, Ala., U.S.A.	645.9	1,199.5	1,085.5	94	0	0	0	0	0	0	0	0
96	121	**POSCO ENGINEERING & CONSTRUCTION,** Pohang-si, South Korea†	629.9	6,258.2	8,226.3	7	3	24	2	0	58	6	0	0
97	87	**IMPRESA PIZZAROTTI & C. SPA,** Parma, Italy†	629.4	1,112.0	148.5	52	0	0	0	1	0	47	0	0
98	115	**BAUER AG,** Schrobenhausen, Bavaria, Germany†	626.1	968.8	1,116.9	29	3	18	9	3	11	23	0	4
99	104	**CONTRACTING AND TRADING CO. (C.A.T.),** Beirut, Lebanon†	593.0	594.0	1,226.0	0	0	0	1	0	86	13	0	0
100	96	**CHINA GEO-ENGINEERING CORP.,** Beijing, China†	588.3	1,157.0	1,890.4	33	0	0	31	8	0	28	0	0

Figure A.2: ENR's top 250 international contractors - Part 2

RANK 2021	2020	FIRM	2020 REVENUE $ MIL. INT'L	TOTAL	2020 NEW CONTRACTS $ MIL	GENERAL BUILDING	MANUFACTURING	POWER	WATER SUPPLY	SEWER / WASTE	INDUS. / PETROLEUM	TRANSPORTATION	HAZARDOUS WASTE	TELECOM
101	84	**TAV CONSTRUCTION,** Istanbul, Turkey	586.2	665.0	NA	20	0	0	0	0	0	80	0	0
102	109	**NUROL CONSTRUCTION AND TRADING CO. INC.,** Istanbul, Turkey†	565.9	904.6	448.0	39	0	0	29	33	0	0	0	0
103	134	**ALBERICI-FLINTCO,** St. Louis, Mo., U.S.A.†	559.3	2,622.5	2,949.5	0	0	0	2	4	93	0	0	0
104	130	**SICIM SPA,** Busseto, Italy†	537.0	541.3	1,200.0	0	0	0	0	0	100	0	0	0
105	110	**SINOPEC ZHONGYUAN PETROLEUM ENGINEERING LTD.,** Puyang City, China†	524.6	1,762.7	1,896.6	0	0	0	0	0	100	0	0	0
106	106	**YUNNAN CONST. AND INVEST. HOLDING GRP. CO. LTD.,** Kunming City, China†	516.8	516.8	454.3	28	0	0	0	0	0	72	0	0
107	99	**JIANGSU PROVINCIAL CONSTRUCTION GROUP CO. LTD.,** Nanjing, China†	515.1	11,012.9	17,543.1	76	0	1	1	3	2	15	0	0
108	122	**JIANGSU NANTONG SANJIAN CONSTR. GROUP CO. LTD.,** Haimen, China†	507.3	15,684.7	6,120.4	100	0	0	0	0	0	0	0	0
109	105	**BEIJING URBAN CONSTRUCTION GROUP CO. LTD.,** Beijing, China†	502.0	25,624.7	32,765.9	39	0	0	1	1	0	59	0	0
110	114	**DRA GLOBAL LTD.,** Perth, Australia†	497.8	718.7	234.4	0	0	0	0	0	0	0	0	0
111	93	**TBEA CO. LTD.,** Changji, China†	489.3	6,255.9	6,279.9	0	0	100	0	0	0	0	0	0
112	125	**STO BUILDING GROUP INC.,** New York, N.Y., U.S.A.†	483.0	8,080.0	10,600.0	87	0	0	0	0	0	0	0	13
113	168	**XPCC CONSTRUCTION & ENGINEERING (GROUP) CO. LTD.,** Urumqi, China†	476.3	4,177.9	5,837.4	28	0	0	0	13	3	28	0	0
114	149	**TAISEI CORP.,** Tokyo, Japan†	472.0	12,129.0	13,590.0	33	0	0	0	0	0	31	0	0
115	129	**ESTA INSAAT SANAYI LOJISTIK VE DIS TICARET AS,** Istanbul, Turkey	465.2	465.2	130.4	65	0	0	0	0	25	0	0	0
116	119	**AFCONS INFRASTRUCTURE LTD.,** Mumbai, India†	460.9	1,296.1	1,065.0	0	0	0	2	0	0	98	0	0
117	117	**BEIJING CONSTRUCTION ENGINEERING GROUP CO. LTD.,** Beijing, China†	457.4	16,234.2	30,353.5	76	0	0	1	5	0	18	0	0
118	135	**SSANGYONG ENGINEERING & CONSTRUCTION CO. LTD.,** Seoul, South Korea†	452.9	1,314.5	1,204.8	84	0	0	0	0	0	16	0	0
119	146	**YANJIAN GROUP CO. LTD.,** Yantai, China†	450.0	2,063.2	2,578.2	63	0	10	1	0	0	0	0	0
120	112	**GHELLA SPA,** Rome, Italy†	446.0	630.0	815.0	0	0	2	0	36	0	55	0	0
121	107	**CHINA HENAN INT'L COOPERATION GROUP CO. LTD.,** Zhengzhou, China	444.8	444.8	576.8	0	0	2	8	8	0	63	0	0
122	124	**GULERMAK,** Ankara, Turkey†	439.3	607.3	147.6	0	0	0	0	0	0	100	0	0
123	123	**DONGFANG ELECTRIC CORP.,** Chengdu, China†	427.9	5,534.3	7,369.3	0	0	100	0	0	0	0	0	0
124	120	**CHINA JIANGSU INT'L ECON. AND TECH. COOP. GROUP LTD.,** Nanjing, China†	427.0	1,904.7	1,810.6	68	0	0	1	0	2	29	0	0
125	**	**ASLAN YAPI VE TICARET AS,** Ankara, Turkey	416.3	422.8	131.3	100	0	0	0	0	0	0	0	0
126	152	**ELECTRA LTD.,** Ramat Gan, Israel†	410.5	2,325.0	2,890.8	63	0	0	5	0	0	27	0	4
127	126	**ANHUI SOGECOA FOREIGN ECON. CONSTR. (GROUP),** Hefei, Anhui, China†	410.2	414.9	5.5	59	0	6	4	0	0	3	0	0
128	128	**GRUPO EMPRESARIAL SAN JOSE SA,** Tres Cantos, Spain†	409.2	1,096.9	1,081.4	97	0	0	0	2	0	2	0	0
129	138	**CHINA WU YI CO. LTD.,** Fuzhou, China†	408.1	3,870.6	6,423.1	19	0	0	0	0	1	77	0	2
129A	**	**TREVI SPA,** Cesena, Italy	395.7	431.3	294.6	37	0	0	6	0	2	54	0	0
130	161	**PETRO. PROJECTS AND TECH. CONSULT. CO. - PETROJET,** New Cairo, Egypt	394.0	2,012.8	3,331.7	0	0	0	0	0	95	0	0	0
131	141	**CADDELL CONSTRUCTION CO. (DE) LLC,** Montgomery, Ala., U.S.A.	389.0	815.0	723.0	100	0	0	0	0	0	0	0	0
132	143	**JIANGXI WATER AND HYDROPOWER CONSTR. CO. LTD.,** Nanchang, China	388.7	571.1	290.8	25	0	11	43	4	0	18	0	0
133	**	**SEMBOL ULUSLARARASI YATIRIM TARIM PEYZAJ INSAAT,** Istanbul, Turkey	372.7	429.8	470.7	100	0	0	0	0	0	0	0	0
134	147	**TUTOR PERINI CORP.,** Sylmar, Calif., U.S.A.†	365.7	6,614.7	4,443.4	57	0	0	0	0	0	43	0	0
135	144	**ZHONGDING INTERNATIONAL ENGINEERING CO. LTD,** Nanchang, China	365.3	365.3	239.2	78	3	0	4	4	0	12	0	0
136	142	**KUZU GROUP,** Istanbul, Turkey†	365.0	426.0	430.0	75	0	0	0	25	0	0	0	0
137	94	**KOLIN INSAAT TURIZM SANAYI VE TICARET AS,** Ankara, Turkey	363.3	1,612.4	519.2	0	0	0	0	0	2	98	0	0
138	131	**RIZZANI DE ECCHER SPA,** Pozzuolo del Friuli, Italy†	360.4	463.2	671.6	53	8	0	0	0	0	21	0	0
139	162	**KINDEN CORP.,** Chiyoda-ku, Japan†	355.0	5,147.0	NA	5	85	0	0	0	0	0	0	0
140	163	**SUMITOMO MITSUI CONSTRUCTION CO. LTD.,** Chuo-ku, Japan†	347.9	2,908.5	2,857.9	9	7	0	0	0	5	80	0	0
141	211	**DPR CONSTRUCTION,** Redwood City, Calif., U.S.A.	347.6	6,458.0	7,897.0	26	0	0	0	0	0	0	0	73
142	155	**COMSA CORPORACIÓN,** Barcelona, Spain	343.0	889.0	NA	0	0	0	0	0	0	62	0	0
143	136	**CGCOC GROUP CO. LTD.,** Beijing, China†	331.7	387.4	719.7	4	0	3	34	0	0	59	0	0
144	118	**YUKSEL INSAAT CO. INC.,** Ankara, Turkey†	329.6	335.5	164.7	27	0	0	33	39	0	0	0	0
145	158	**ESER CONTRACTING,** Ankara, Turkey	324.2	324.9	NA	67	0	0	8	0	0	24	0	0
146	113	**BLACK & VEATCH,** Overland Park, Kan., U.S.A.†	321.9	1,575.1	1,265.3	0	0	37	14	13	37	0	0	0
147	185	**SHANGHAI URBAN CONSTRUCTION (GROUP) CORP.,** Shanghai, China†	321.3	9,444.0	14,858.4	4	0	0	0	0	0	96	0	0
147A	**	**CIMOLAI SPA,** Rome, Italy	321.0	529.0	198.0	47	0	0	0	0	20	33	0	0
148	145	**SINOSTEEL EQUIPMENT & ENGINEERING CO. LTD.,** Beijing, China	314.0	1,906.1	3,897.3	0	0	0	0	0	100	0	0	0

NOS. 129A AND 147A OMITTED FROM ORIGINAL LIST DUE TO DATA ERRORS. FIRMS SHOULD HAVE RANKED AT 129 AND 147 RESPECTIVELY.

Figure A.3: ENR's top 250 international contractors - Part 3

RANK 2021	RANK 2020	FIRM	2020 REVENUE $ MIL. INT'L	2020 REVENUE $ MIL. TOTAL	2020 NEW CONTRACTS $ MIL	GENERAL BUILDING	MANUFACTURING	POWER	WATER SUPPLY	SEWER / WASTE	INDUS. / PETROLEUM	TRANSPORTATION	HAZARDOUS WASTE	TELECOM
149	176	**COMBINED GROUP CONTRACTING CO. (K.S.C),** Kuwait, Kuwait	294.9	440.3	155.0	83	0	0	0	6	1	10	0	0
150	**	**DREES & SOMMER,** Stuttgart, Germany†	270.0	270.0	NA	57	18	3	0	0	7	9	0	5
151	157	**IC ICTAS INSAAT SANAYI VE TICARET AS,** Istanbul, Turkey†	268.3	883.3	NA	16	0	0	0	0	0	84	0	0
152	200	**SENER GRUPO DE INGENIERIA SA,** Getxo, Bizkaia, Spain†	260.6	285.2	197.6	0	28	45	0	0	24	0	0	2
153	**	**BLUE&P INFRASTRUCTURE GROUP,** Tehran, Iran†	257.5	280.8	42.8	0	0	68	32	0	0	0	0	0
154	151	**GILBANE BUILDING CO.,** Providence, R.I., U.S.A.†	256.9	6,404.6	6,800.1	36	0	0	0	0	62	1	1	0
155	132	**CHINA NONFERROUS METAL IND. FOR. ENG'G AND CONSTR.,** Beijing, China	244.3	245.1	420.5	0	0	0	10	0	90	0	0	0
156	172	**PARSONS,** Centreville, Va., U.S.A.†	241.7	607.6	1,393.4	75	0	0	0	0	0	24	0	0
157	171	**CALIK ENERJI SANAYI VE TICARET AS,** Istanbul, Turkey†	237.0	237.0	55.0	0	0	100	0	0	0	0	0	0
158	213	**ILK CONSTRUCTION,** Istanbul, Turkey	236.7	236.7	51.7	15	0	6	0	0	28	51	0	0
159	127	**CHINA NATIONAL AERO-TECHNOLOGY INT'L ENG'G CORP.,** Beijing, China†	231.5	533.4	746.8	53	0	0	7	0	0	40	0	0
160	159	**MWH CONSTRUCTORS INC.,** Broomfield, Colo., U.S.A.†	223.8	600.0	1,029.9	0	0	0	56	44	0	0	0	0
161	175	**MICHELS CORP.,** Brownsville, Wis., U.S.A.	223.4	2,979.2	3,704.2	0	0	1	0	10	88	0	0	0
162	216	**GAP INSAAT YATIRIM VE DIS TICARET AS,** Istanbul, Turkey†	220.9	260.8	49.6	63	37	0	0	0	0	0	0	0
163	203	**POLAT YOL YAPI SAN. VE TIC. AS,** Istanbul, Turkey†	218.1	297.4	36.9	0	0	0	0	0	0	100	0	0
164	223	**FORTIS CONSTRUCTION INC.,** Portland, Ore., U.S.A.†	218.0	1,188.0	1,675.0	0	0	0	0	0	0	0	0	100
165	189	**ALARKO CONTRACTING GROUP,** Istanbul, Turkey†	214.2	4,292.5	NA	0	0	0	0	0	0	100	0	0
166	169	**AEGION CORP.,** Chesterfield, Mo., U.S.A.†	213.9	999.8	990.0	4	0	0	0	45	48	3	0	0
167	**	**XIAN ELECTRIC ENGINEERING CO. LTD.,** Xi'an, Shaannxi, China	211.7	211.7	NA	0	0	100	0	0	0	0	0	0
168	137	**DEKINSAN GRUP INSAAT AS,** Ankara, Turkey	208.6	216.2	484.4	100	0	0	0	0	0	0	0	0
169	184	**RAILWORKS,** New York, N.Y., U.S.A.	206.8	892.3	NA	0	0	0	0	0	0	100	0	0
170	**	**ICM SPA,** Vicenza, Italy†	204.9	446.2	688.2	22	0	0	0	0	0	77	0	0
171	154	**SHENYANG YUANDA ALUMINUM INDUS. ENG'G CO. LTD.,** Shenyang, China†	197.3	375.5	355.9	99	0	0	0	0	0	1	0	0
172	148	**CHINA NATIONAL COMPLETE PLANT IMP. & EXP. GROUP,** Beijing, China†	197.0	197.0	219.2	71	7	3	0	9	10	0	0	0
173	186	**SHANXI CONSTRUCTION INVESTMENT GROUP CO. LTD.,** Taiyuan, China†	194.1	16,933.9	29,077.5	58	0	15	0	22	0	5	0	0
174	178	**ANHUI CONSTRUCTION ENGINEERING GROUP CO. LTD.,** Hefei City, China†	191.7	12,066.1	15,150.6	59	0	0	1	0	0	40	0	0
175	**	**SHANDONG DEJIAN GROUP CO. LTD.,** Dezhou City, Shandong, China	191.3	1,784.9	1,506.8	73	0	6	0	0	0	21	0	0
176	194	**LONGXIN CONSTRUCTION GROUP CO. LTD.,** Haimen, China	191.0	4,253.9	2,593.5	100	0	0	0	0	0	0	0	0
177	187	**SHANDONG ZIJIAN CONSTRUCTION GROUP CO. LTD.,** Zibo, China	189.6	1,902.3	1,470.3	97	0	0	0	0	0	0	0	3
178	214	**WALBRIDGE,** Detroit, Mich., U.S.A.	188.1	1,912.3	2,157.5	35	65	0	0	0	0	0	0	0
179	174	**ENG'G FOR THE PETROLEUM & PROCESS INDUSTRIES (ENPPI),** Cairo, Egypt†	187.5	585.5	2,700.0	0	0	0	0	0	100	0	0	0
180	191	**HUNAN CONSTRUCTION ENGINEERING GROUP CO. LTD.,** Changsha, China†	185.2	13,936.4	22,343.4	30	0	21	20	0	0	29	0	0
181	182	**GURBAG GROUP,** Ankara, Turkey†	183.8	189.5	246.1	61	0	0	0	0	0	39	0	0
182	181	**LOTTE ENGINEERING & CONSTRUCTION CO. LTD.,** Seoul, South Korea†	172.0	4,669.0	9,134.0	28	0	53	0	8	3	6	0	0
183	179	**TEPE INSAAT SANAYI AS,** Ankara, Turkey†	168.0	215.6	216.9	28	0	0	0	0	3	69	0	0
184	198	**ZHEJIANG DONGYANG THIRD CONSTR. ENG'G CO. LTD.,** Dongyang, China†	167.9	893.6	727.3	100	0	0	0	0	0	0	0	0
185	**	**MAKYOL INSAAT SANAYI TURIZM VE TICARET AS,** Istanbul, Turkey	166.2	702.0	112.0	0	0	0	1	0	0	99	0	0
186	**	**HEBEI CONSTRUCTION GROUP CO. LTD.,** Hebei, China†	162.7	7,893.0	9,129.9	0	0	70	5	2	23	0	0	0
187	164	**APTIM,** Baton Rouge, La., U.S.A.	162.3	1,053.4	1,339.6	23	0	0	0	0	72	0	4	0
188	180	**METAG GROUP,** Ankara, Turkey	162.1	174.4	180.0	24	0	0	34	9	0	32	0	0
189	205	**NANTONG CONSTRUCTION GROUP CO. LTD.,** Nantong, China†	161.6	2,079.7	2,919.7	61	0	11	8	0	20	0	0	0
190	201	**ZHEJIANG COMMUNICATIONS CONSTR. GROUP CO. LTD.,** Hangzhou, China†	160.4	5,017.3	9,968.0	23	0	0	2	0	0	75	0	0
191	170	**AECON GROUP INC.,** Toronto, Canada	157.0	3,643.0	3,308.0	0	0	7	0	0	6	3	0	0
192	221	**HUNAN ROAD & BRIDGE CONSTR. GROUP CO. LTD.,** Changsha, China†	156.3	2,209.4	3,469.1	0	0	0	0	0	0	100	0	0
193	**	**JIANGSU ZHONGNAN CONSTR. INDUSTRY GROUP CO. LTD.,** Haimen, China†	155.7	19,493.4	27,754.2	100	0	0	0	0	0	0	0	0
194	208	**JIANGXI CONSTRUCTION ENG'G (GROUP) CORP. LTD.,** Nanchang, China†	153.0	7,493.9	11,770.0	18	0	0	0	0	1	82	0	0
195	**	**RAMBOLL GROUP A/S,** Copenhagen, Denmark	147.5	147.5	114.7	0	5	0	3	4	51	0	36	0
196	**	**HASKELL,** Jacksonville, Fla., U.S.A.†	144.2	1,075.4	NA	0	4	0	0	0	94	2	0	0
197	140	**CHINA TRIUMPH INTERNATIONAL ENGINEERING CO. LTD.,** Shanghai, China†	143.0	1,704.1	1,875.2	0	0	56	0	0	44	0	0	0
198	**	**USTAY YAPI TAAHHUT VE TICARET AS,** Istanbul, Turkey†	139.7	162.9	157.0	0	0	0	0	0	100	0	0	0

Figure A.4: ENR's top 250 international contractors - Part 4

RANK 2021	RANK 2020	FIRM	2020 REVENUE $ MIL. INT'L	2020 REVENUE $ MIL. TOTAL	2020 NEW CONTRACTS $ MIL	GENERAL BUILDING	MANUFACTURING	POWER	WATER SUPPLY	SEWER / WASTE	INDUS. / PETROLEUM	TRANSPORTATION	HAZARDOUS WASTE	TELECOM
199	167	**TIANYUAN CONSTRUCTION GROUP CO. LTD.,** Linyi, China†	134.9	134.9	134.9	100	0	0	0	0	0	0	0	0
200	207	**CHONGQING INTERNATIONAL CONSTRUCTION CORP.,** Chongqing, China†	133.6	463.1	736.2	6	0	0	20	0	0	75	0	0
201	197	**YENIGUN INSAAT SANAYI VE TICARET AS,** Ankara, Turkey	130.1	138.3	142.1	64	0	2	0	0	0	34	0	0
202	204	**CHINA GANSU INT'L CORP. FOR ECON. AND TECH. COOP.,** Lanzhou, China†	125.9	198.5	400.7	49	0	0	6	0	21	24	0	0
203	166	**SUMMA TURIZM YATIRIMCILIGI AS,** Ankara, Turkey†	122.7	122.7	532.0	85	0	0	0	0	0	15	0	0
204	212	**GAMA,** Ankara, Turkey†	122.5	156.9	NA	0	0	89	0	0	11	0	0	0
205	217	**NATA INSAAT TURZ. TASIMACILIK TIC. VE SAN. AS,** Ankara, Turkey	116.0	523.6	292.0	0	0	0	0	0	0	100	0	0
206	219	**CENGIZ CONSTRUCTION INDUSTRY & TRADE CO. INC.,** Istanbul, Turkey	116.0	833.9	995.0	0	0	0	0	0	0	100	0	0
207	**	**GREENLAND INFRA. CONSTR. GROUP CO. LTD.,** Shanghai, China†	112.3	43,653.5	88,397.5	44	0	0	0	0	37	18	0	0
208	224	**PERNIX GROUP INC.,** Lombard, Ill., U.S.A.†	111.7	389.4	209.6	92	0	8	0	0	0	0	0	0
209	**	**TATA PROJECTS LTD.,** Secunderabad, India†	104.2	1,644.9	1,185.9	0	0	85	0	0	12	0	0	0
210	**	**ZHENGTAI GROUP CO. LTD.,** Taizhou, Jiangsu, China	100.5	1,103.5	1,828.5	54	0	0	46	0	0	0	0	0
211	**	**NANTONG SI JIAN CONSTRUCTION GROUP CO. LTD.,** Nantong City, China†	100.5	8,577.1	7,115.4	100	0	0	0	0	0	0	0	0
212	**	**MBD INSAAT SANAYI VE TICARET AS,** Ankara, Turkey†	95.5	134.6	134.6	66	0	4	10	10	0	10	0	0
213	210	**SICHUAN ROAD AND BRIDGE (GROUP) CO. LTD.,** Chengdu, China†	92.0	7,147.7	5,721.6	0	0	0	0	0	0	100	0	0
214	**	**THE BECK GROUP,** Dallas, Texas, U.S.A.	90.0	1,003.8	439.8	100	0	0	0	0	0	0	0	0
215	**	**FEKA INSAAT SANAYI TICARET AS,** Izmir, Turkey	89.0	92.4	125.6	16	0	39	2	0	0	41	0	0
216	**	**WSP GLOBAL INC.,** Montreal, Quebec, Canada†	84.8	84.8	NA	0	0	0	0	0	0	0	0	0
217	**	**CHINA DALIAN INT'L ECON. & TECH. COOP. GROUP CO. LTD.,** Dalian, China	84.7	115.7	386.4	100	0	0	0	0	0	0	0	0
218	**	**SOUTHLAND HOLDINGS LLC,** Roanoke, Texas, U.S.A.†	79.8	1,175.0	1,801.0	0	0	0	51	6	0	43	0	0
219	202	**SHANDONG KERUI PETROLEUM EQUIP. CO. LTD.,** Dongying, China	79.7	156.4	50.0	0	0	0	0	0	100	0	0	0
220	**	**ENERGOPROJEKT HOLDING PLC,** Belgrade, Serbia†	79.7	204.4	119.5	66	0	1	0	0	0	33	0	0
221	**	**CHINA ALUMINUM INTERNATIONAL ENG'G CORP. LTD.,** Beijing, China†	75.8	3,303.5	5,739.0	12	29	0	15	0	44	0	0	0
222	**	**IRIS INSAAT TURIZM SANAYI VE TICARET AS,** Istanbul, Turkey	75.5	79.9	106.1	0	0	0	0	0	0	96	0	0
223	**	**SMK GROUP,** Ankara, Turkey	75.4	79.1	97.0	19	0	24	24	0	32	0	0	0
224	**	**EGYPTIAN MAINTENANCE CO. (EMC),** New Cairo, Egypt	74.6	314.7	419.5	0	0	14	10	0	74	2	0	0
225	**	**MATRIX SERVICE CO.,** Tulsa, Okla., U.S.A.†	72.4	794.4	NA	0	0	0	0	0	100	0	0	0
226	225	**THE WALSH GROUP,** Chicago, Ill., U.S.A.	72.1	5,379.0	4,988.0	64	0	0	0	15	0	20	0	1
227	**	**VELESSTROY LLC,** Moscow, Russia	71.0	1,596.0	1,048.0	0	0	0	0	0	100	0	0	0
228	**	**CHINA BENGBU INT'L TECH. & ECON. COOP. LTD.,** Hefei, Anhui, China	70.1	70.1	78.0	0	0	0	1	0	0	99	0	0
229	**	**IRCON INTERNATIONAL LTD.,** New Delhi, India	68.4	609.1	540.1	0	0	0	0	0	0	100	0	0
230	192	**BARNARD CONSTRUCTION CO. INC.,** Bozeman, Mont., U.S.A.†	68.3	1,443.6	910.8	0	0	98	2	0	0	0	0	0
231	**	**ECC,** Burlingame, Calif., U.S.A.†	63.1	506.1	1,544.0	25	0	1	0	0	24	26	23	0
232	**	**JIANGSU NANTONG NO.2 CONSTRUCTION,** Qidong, Jiangsu, China†	61.7	13,232.3	9,473.2	83	0	0	0	11	0	6	0	0
233	**	**UNITED E&C INC.,** Piscataway, N.J., U.S.A.	60.8	509.1	494.7	0	0	100	0	0	0	0	0	0
234	**	**BOTHAR GROUP PTY LTD.,** Wakerley, Queensland, Australia†	60.6	84.2	90.2	0	0	0	0	91	9	0	0	0
235	**	**GOLDER,** Palm Beach Gardens, Fla., U.S.A.†	60.0	80.3	21.5	2	0	2	0	3	30	0	1	0
237	**	**MAEG COSTRUZIONI SPA,** Vazzola, Italy†	55.1	102.4	110.7	9	0	0	0	0	0	91	0	0
238	218	**STFA CONSTRUCTION GROUP,** Istanbul, Turkey†	51.7	54.4	2.0	0	0	0	0	0	0	85	0	0
239	165	**DOGUS INSAAT VE TICARET AS,** Istanbul, Turkey†	43.9	211.7	331.8	11	0	0	0	0	0	89	0	0
240	**	**BURNS & MCDONNELL,** Kansas City, Mo., U.S.A.†	43.3	1,871.9	3,596.5	0	0	55	0	0	45	0	0	0
241	**	**BARTON MALOW HOLDINGS LLC,** Southfield, Mich., U.S.A.†	43.0	2,336.5	3,879.9	0	10	38	0	0	5	48	0	0
242	177	**JIANGLIAN HEAVY INDUSTRY GROUP CO. LTD.,** Nanchang, China†	37.0	201.3	361.4	13	51	0	0	0	0	5	0	0
243	**	**WILLIAMS INDUSTRIAL SERVICES GROUP INC.,** Tucker, Ga., U.S.A.†	35.7	269.1	218.0	0	0	100	0	0	0	0	0	0
244	**	**SALFACORP SA,** Santiago, Chile†	34.0	562.0	1,899.0	100	0	0	0	0	0	0	0	0
245	72	**AECOM,** Los Angeles, Calif., U.S.A.	33.0	6,568.5	NA	80	0	0	0	0	0	20	0	0
246	**	**WEEKS MARINE INC.,** Cranford, N.J., U.S.A.†	32.5	779.0	697.7	0	0	0	0	31	12	57	0	0
247	206	**GRAY CONSTRUCTION,** Lexington, Ky., U.S.A.†	32.4	1,614.3	1,614.3	0	0	0	0	0	100	0	0	0
248	**	**CDM SMITH,** Boston, Mass., U.S.A.†	32.2	385.4	381.0	39	0	0	61	0	0	0	0	0
249	**	**ZACHRY GROUP,** San Antonio, Texas, U.S.A.†	31.5	1,973.2	1,256.3	0	0	0	0	0	100	0	0	0
250	**	**GREAT LAKES DREDGE & DOCK CO.,** Houston, Texas, U.S.A.	25.9	733.6	692.8	0	0	0	0	0	0	100	0	0

Figure A.5: ENR's top 250 international contractors - Part 5

B. The extended stop words list

hochtief, pgina, vanguard, China, th, co, ltd, better, readability, hereinafter, referred, us, quoted, comes, come, January, December, also, pre, rmb, nox, cod, nhx, ceec, ful, part, ic, ipat, ion, ointly, non, million, ka, jim, sc, op, em, si, kajima, ems, four, rl, fy, zeb, zeh, asbestos, kajimas, yeong, guan, taichung, ctci, ii, tcfd, led, chief, sbti, September, lotte, twice, year, bsc, April, take, sees, meets, actual, Shanghai, Shenyang, huanggu, zhuhai, shall, Hyundai, first, iii, vr, March, October, sacyrs, via, sacyr, go, one, thus, every, gumi, mill, bucheon, ty, ohsms, pyeongtaek, always, pagina, ferrovial, end, two, years, thereof, third, Mr, kim ju cheol, bod, February, August, compli, ance, ing, rm, June, dl, Korean, sinmungo, hansup, icfr, set, step, manag, ement, resp, onsible, toward, signing, half, dudrim, time, May, within, throughout, July, entire, ilo, make, krw, whenever, raw, sub, six, dssc, rba, smeta, whether, samsung, par, ticipation, could, manak, govt, either, ts, ayhan, ahenk, garanti, turkey, sinopec, etc, cro, taisei, bcp, abide, without, corp, Chinese, wbcsd, second, cop, zhenhai, yanshan, mansarover, fauna, since, yangtze, nbs, posco, ngos, tnfd, enerya, stfas, naat, Istanbul, kg, Turkmenistan, tonnes, far, eight, Beijing, Tianjin, gaoqiao, guangzhou, hainan, qilu, qingdao, xinjiang, Mongolia, Shandong, pem, cubic, metres, lng, dx, site, plm, savvy, roi, obayashi, raku, rfid, ict, sumitomo, mitsui, wowtalk, mamoru, biz, itb, epc, iepc, iku, lynchpin, hebei, rd, nd, ks, sqi, sns, daewoo, dec, five, cec, dongfang, eep, kom, iea, sds, mailto, seite, zurck, schaltflche, oct, inopec, ssrc, dr, strabag, ot, bam, nv, dendermonde, belgium, fm, konya, eiste, hadimi, viaduct, boukhroufa, souk, tleta, mraniye, ataehir, gztepe, veatch, black, dushanzi, urumqi, iv, agm, yen, ernst, young, shinnihon, llc, ksek, anders, danielsson, isas, sweden, c_gen_page c_gen_pagel c_gen, jersey, uk, amsterdam, llp van eimeren, dutch, kmr, kmrs, ngms, srv, aa, ap, dou, group, hong kong, sar, japan, mainland, seven, chiyodas, three, masakazu, sakakida, fuminori, hasegawa, koji tarutani, masao ishikawa, ota, ryo matsukawa, yutaka kunigo, shingo torii, mika narahashi, hisashi ito, tokyo, chiyoda, tamotsu iwamoto, xebio, nurol, algeria, eltarf, south bouteldja, umraniye, atasehir, goztepe, tizi ouzou bouira, ilsu, hepp, tigris, cengiz, hasankeyf, siirt, batman, yusufeli, gulsan, artvin, erzurum, silenkar, tekkale, hazuket, oord, van, netherlands, merweoord, consoord, ulran, sek, bn, usd, gbp, eur, nok, czk, pln, dkk, intra,

seop, united kingdom, ireland, belgium, freytag, ingenieurbau, ag, germany, bunnik, runnenburg az bunnik, petrofac, kuwaiti, arab emirates, november, ithaca, zhenhua, jsd, ias, comsa, chang, mmr, taiwan, shantou, mitsubishi, qinghai golmud, guangdong, jintan jiangsu, shnes yangyi, yangyi vi, gedaxiang damxung, lhasa, city tibet, yangyi, hanoi, ibst, vietnamese, seizogauchi miyazaki, tankaru, tcoe, seis, donmuimun, jeongseon condo, yeouido yongin han, daejeon, takenaka, kyoto, setagaya, chochikukyo, oyamazaki, cho, hokkaido, oce tohoku, yokohama, kanto, oce, nagoya, osaka, kobe, shikoku, hiroshima, kyushu, yumeshima, maki, nihon sekkei, japans, abeno harukas tenshiba, pelli clarke pelli, otemachi, cbd, umekita, nikken, jisho, nikken, minato mirai, yasuo kitayama kyoto higashiyama, kyoyamato, hyatt, pagoda hokanji, yasaka, limak, trkiyes, trkiye, ic holding, tatemono sogo kanri, paris, danieli, yiliang, yunnan, yanhe, guizhou, tongzi, huayuan, hunan, chengkou, chongqing, heishui, tongnan, wuchuan, shangdu, yu xubo, lu yimin, xiaohaizi, daqingshan, xi jinping, dongxiang, yopurga, yuexi, yingxiang, ponkan luxi, kiwi fenghuang, baingoin, zeku, nutreco, boon, boons, denise, koopmans, mrs, rotterdam, nijmegen, amstelland, heerema, cap gemini, oman, venterra, italy, bouw en vastgoed, nederland, nelis, laing orourke kier, wilkinson, per, significant, issuer, issuers, sydney, australian, sydneys, chatswood sydenham, glenrowan, wallgrove, new zealand, auckland, giulia parlato, nuove avventure sotterranee, st marys, bechtel, norway, dali yunnan, guoluo, qinghai, luzhou sichuan, henan, hubei, guangxi, luoyang, haidong haixi qinghai, linyi shandong, skanskas, swedens, anyang, seoul, amano, suzuki, ukraine, singapore, morinosuke, kasumigaseki, spain, spanish, mexico, yes, tsr, sti, lti, psp, daelim, saipem, kbr, fmc, alk, turkish, ahmet, denim, albtelecom, albanias, qatar, japanese, keidanren, secom, ssr, susta, ng, nab, es, korea, michiaki tanaka, rikkyo, hiroki fukasawa, jiro taniyama, genertec, sinomach, caeri, apr, jun, suntory, smcc, bams, cent, limkon, limakport, philippines, kazakhstan, aegon, ri, aodp, usa, aam, aifmd la banque postale, sa lbpam pelargos, pelargos saemor, chile, peru, czech, republic, slovakia, ohla, fortis, fortisbc, fortisbcs, hudsons, hudson, canada, npata, cccc, cccg, liulin, haihe, yanghekou, qinhuangdao, jinzhong, dongle, jiyun, yongding, shenzhen zhongshan, peljeac, croatia, karakoram, havelian thakot, changtai, guangdong hong, macao, xinjiang urumqi yuli, wufengshan yangtze, nanjing jiangxinzhou, jinan jinsha, huaping lijiang, yangbajing, beijing, jin, shenzhen, pingshan, wuhe, hulukeng, zhongcun, binheng, guangning, zhongshan, guishan, zhuhai qingzhou, haiyan, taishan, chaoyang,

binhai, dagukou, jingu, gezhouba, suizhou xinyang, guannan, dingyuan, hualong, karachi, porr, ebt, romania, franki, porr oevermann, sdschnell, hanover porr, neckartal, horb, powidz, poland, porrs, bouygues, chuo, koga, seiko, fukae chikuyu, equans, gap, tarihi, naats, belgelendirme, dng, gediktepe, balkesir, lidya, ashgabat, taksim, kronsberg, austria, indonesia, india, scandinavia, finland, denmark, aarsleffl, div, takefumi saito, kojiro shimizu, nobuhiko sasaki, kazuhiko kato, osamu nakagawa, naruki ohashi, takayuki sakakima, hiroyuki fujimoto, hideo yokoyama, takuya, masakazu hyodo, shimizu america, sadao matsuhashi, katsumi sawahata, naomi onishi, shuji kakegawa, fumio asami, kenji tsuzuki, toshiyuki nakahara, aecon, ontario, aecons, daniele zanotti, toronto, alberta, eglinton, canadas, mn, vs, nte, nan, aberdeen, israel, soes, pattani narathiwat, thailand, vance, powerchina, dongtai, heilongjiang, gansu, shendiantong, guokong, puneandmumbai, baden wrttemberg, hessen, bavaria, hamburg, black veatch, veatchs, binaviktori, australia, melbourne, santiago, centro, mm, olivotto ferre, brazil, ku, tvtrk, lfl, miyagi, orascom, besix, shimizu, tcnicas, reunidas, de, la, soilmec, trevi, spera, sto, wsp, sommer, snc, fluor, watpac, yasunishi, gilbane, electra, otis, elevators, elevator, escalators, vo, qingliu, caa, dalaman, jgc, corporacin, toyos, toyo, soe, nsw, ugl, thai, kiewit, nebraska, erzhong, chinas, asahi, tak, cbre, victoria, gj, stfa, aarsleff, ags, eiffage, turin, foundat, vers, brah, ar, educat, madencilik, enerji, sergio, weve

C. Topic Word Scores

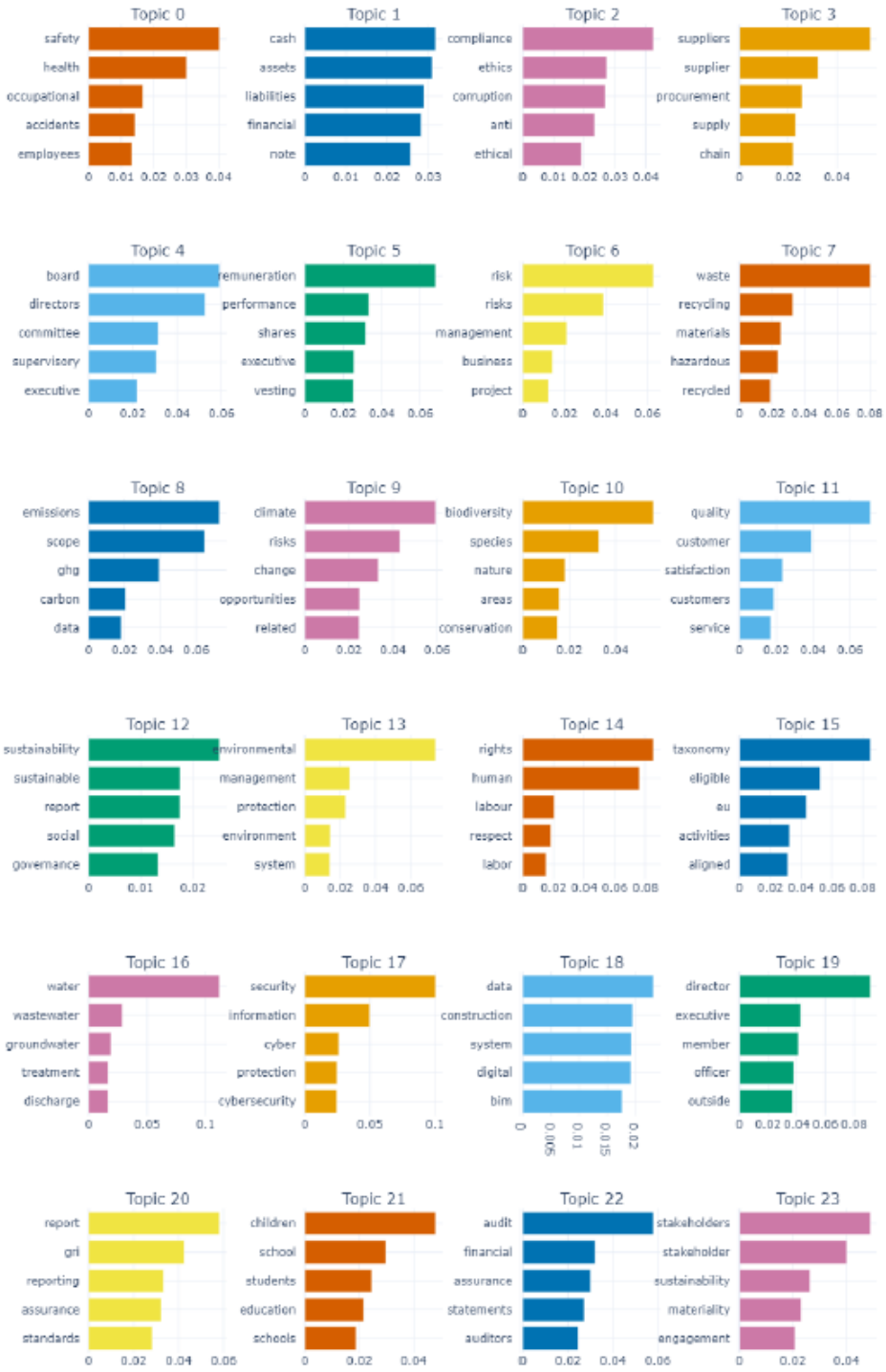

Figure C.6: Topic Word Scores Part 1

Figure C.7: Topic Word Scores Part 2

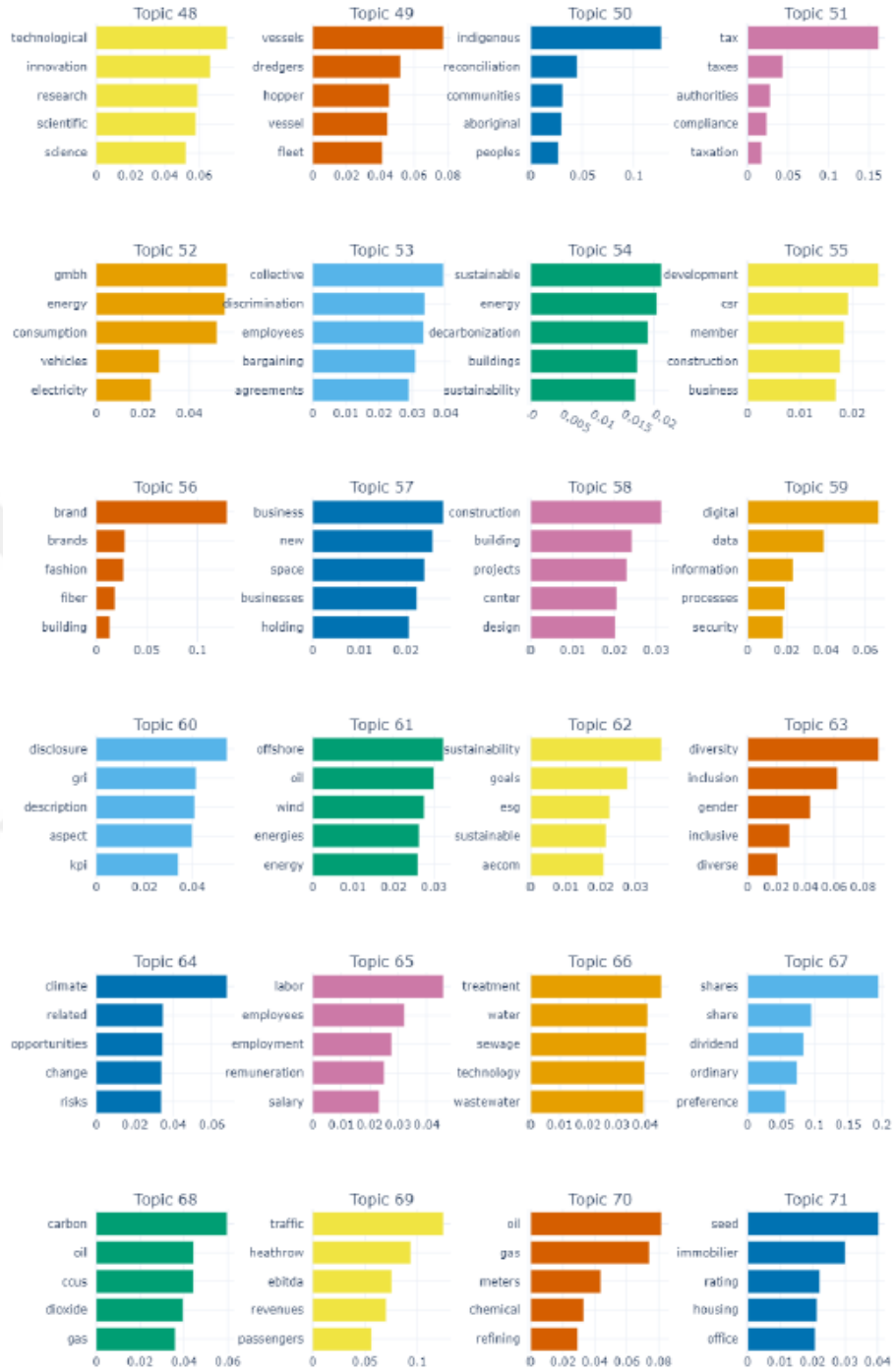

Figure C.8: Topic Word Scores Part 3

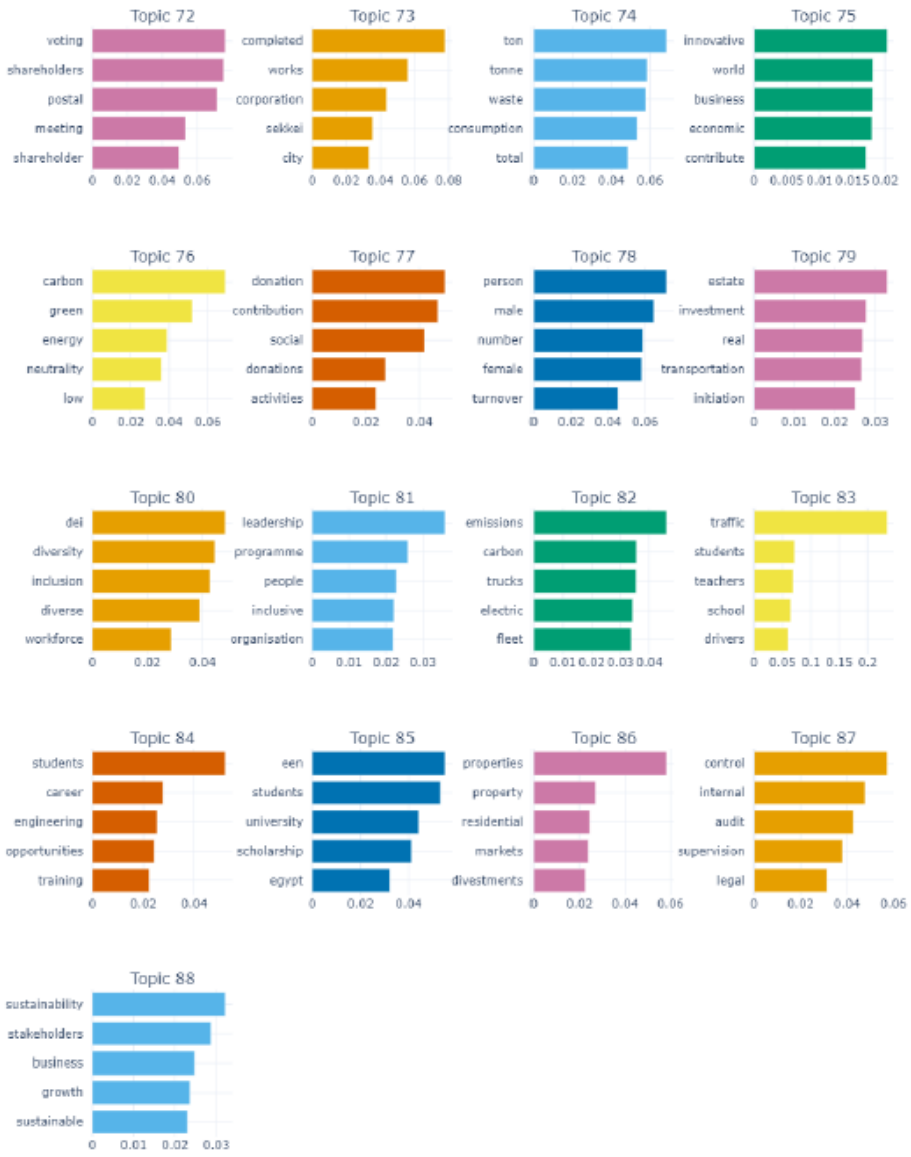

Figure C.9: Topic Word Scores Part 4