Norwegian School of Economics
Bergen, Autumn 2024

The Reliability of ESG Disclosures as Indicators of Environmental Performance

A Machine Learning Approach to the Analysis of ESG Reports and Carbon Intensity in the Steel Industry

Birk Elias Nybø & Kåre Habbestad Skimmeland

Supervisor: Giacomo Benini

Master thesis, Economics and Business Administration

Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

Acknowledgements

This master's thesis was written as part of our master's degree in Economics and Business Administration at the Norwegian School of Economics (NHH).

We would like thank our supervisor Giacomo Benini for his guidance and constructive feedback throughout the process of writing this thesis. We first met him in his course at NHH "Pricing Analytics and Revenue Management" and his encouragement since then has been a big motivator for us.

We would also like to thank Marta Campagnoli for her help in gathering data, which saved us a lot of tedious work.

Norwegian School of Economics

Bergen, December 2024

Your name here

Your name here

Abstract

This thesis examines the relationship between ESG reporting and actual environmental performance in the steel industry, focusing on Scope 1 carbon intensity. Using natural language processing, we developed an Environmental Index to quantify the content of ESG reports and performed sentiment analysis to evaluate their tone. The Environmental Index did not significantly correlate with carbon intensity. This lack of a significant correlation may stem from three possible factors: limitations in the methodology used to construct the index, inaccuracies or misrepresentations in corporate ESG disclosures, or errors in third-party carbon intensity data. However, sentiment analysis revealed that a positive tone in reports was associated with lower carbon intensity, particularly in annual and integrated reports.

The study also finds that companies publishing standalone ESG reports often have higher carbon intensity. Similarly, firms emphasizing renewable energy were also found to have higher carbon intensity. This suggests that some companies may strategically use ESG reporting to influence stakeholder perceptions. These findings suggest a gap between ESG disclosures and actual emissions, raising concerns about the credibility of such reports.

Keywords – ESG, Machine Learning, NLP, Steel Industry, Environmental Index, Regression, Textual Analysis, Sentiment Analysis, Carbon Intensity, Greenwashing

Contents

List of Figures

List of Tables

1 Introduction

The steel industry is an important part of the global economy, as it provides essential materials used in construction, transportation, and manufacturing. However, this sector is also a major contributor to climate change, producing about 7% of the world's carbon dioxide (CO2) emissions (Mona, 2024). This large environmental impact comes from energy-intensive production processes, such as the combustion of coal in blast furnaces, as well as a high reliance on energy sources that release a lot of carbon (Edmond, 2024). As efforts to fight climate change increase, industries like steel face growing pressure to reduce their emissions and report their environmental impact clearly (DNV, 2024).

Governments, regulators, and private investors are paying closer attention to industries with large carbon footprints. Many companies have turned to environmental, social and governance (ESG) frameworks to show that they are committed to sustainability and meet the expectations of stakeholders. ESG reporting, which is often found in annual reports, integrated reports, and dedicated ESG reports, has become an important tool for companies to communicate how they address environmental challenges, social responsibilities, and governance practices.

Although it is becoming more common, ESG reporting continues to raise questions about its reliability and accuracy. Since ESG data are often self-reported by companies, concerns arise about whether these disclosures truly reflect the company's environmental performance. These doubts are especially relevant in energy-intensive industries such as steel, where there could potentially be huge differences between what companies report and their actual emissions. The accuracy of ESG reports is a critical issue for policymakers, who use these data to guide regulations, and for investors seeking to align their portfolios with environmental and long-term sustainability goals (IMD, 2023).

Several previous studies have focused on ESG ratings from third-party agencies like Bloomberg, Refinitiv and FTSE Russel, but these ratings often combine many factors from the environmental, social, and governance categories (iRIS Carbon, 2023). This makes it harder to directly assess the real environmental impact of a company. Furthermore, the methods used to create ESG ratings are not always clear, making it difficult to evaluate the connection between what companies report and their actual results. This thesis addresses these gaps by focusing only on the environmental (E) part of ESG reporting and how it relates to real emissions data.

This thesis investigates whether ESG disclosures from steel companies are reliable indicators of their actual environmental impact, specifically in terms of Scope 1 carbon intensity. Scope 1 emissions, which refer to direct greenhouse gas emissions from companies or controlled sources, represent a fundamental measure of the environmental responsibility of a company (Anthesis, 2024). Given the significant contribution of the steel industry to global emissions, understanding whether ESG reports correspond to real emissions data is important to evaluate whether their environmental disclosures can be seen as trustworthy and useful for understanding the environmental performance of companies.

The analysis is motivated by the need to provide clear and evidence-based insights into how effective ESG reporting is in capturing real environmental performance in industries with high emissions. By comparing raw emissions data with ESG information from annual reports, integrated reports, and dedicated ESG reports, the thesis seeks to uncover whether companies' self-reported environmental disclosures align with their actual emissions outcomes.

The core research objectives of this thesis can be split into three main parts: first, to analyze ESG disclosures from steel companies using machine learning techniques, particularly focusing on the environmental aspect; second, to create an Environmental Index and a Sentiment Score derived from the machine learning assisted text analysis; and third, to evaluate how well these metrics correlates with actual carbon intensity data.

1.1 Thesis Structure

This thesis is structured into eight chapters. Following the introduction, Chapter 2 provides the theoretical background and literature review, covering key concepts related to ESG reporting, Scope 1 emissions, and previous studies. Chapter 3 presents the data, including sources and preprocessing steps. Chapter 4 outlines the methodology, focusing on the machine learning models and the metrics we used. Chapter 5 contains analysis of unstratified data, highlighting the development of the Environmental Index and its correlation with Scope 1 emissions. Chapter 6 contains analysis of stratified data to uncover potential relationships that may not appear in the unstratified analysis. Finally, Chapter 7 concludes with a discussion of the results, limitations, and implications, followed by the conclusion in Chapter 8.

2 Theoretical Background and Literature Review

2.1 Theoretical Background

This section explains some key concepts that are important as background information for this thesis. First, we define what ESG reporting is and its role in communicating a company's sustainability practices. Then, we explain carbon intensity and its connection to Scope 1 emissions. Finally, we discuss the issue of greenwashing, where companies may exaggerate or mislead stakeholders about their environmental efforts.

2.1.1 ESG Reporting

ESG reporting refers to the disclosure of the environmental, social and governance performance of a company to stakeholders, including investors, customers, and regulators (PricewaterhouseCoopers, 2024). This practice has gained significant traction in recent years, with 63% of organizations disclosing ESG information in 2023, up from 56% the previous year, according to the Wall Street Journal yearly survey (Breg, 2023). ESG reporting aims to provide transparency on a company's sustainability efforts and impact on people and the environment.

ESG reporting has gained importance due to regulatory frameworks such as the European Union's Sustainable Finance Disclosure Regulation and the Corporate Sustainability Reporting Directive (CSRD) (Khazin, 2024). These initiatives aim to standardize sustainability disclosures, ensuring stakeholders have access to transparent information needed to evaluate a company's environmental and social impact. For high-emission industries like steel, ESG reporting plays a central role in dealing with sustainability challenges.

The steel industry, which is responsible for approximately 7% of global CO2 emissions (Mona, 2024), is under increasing pressure from regulators, customers, and investors to prioritize sustainability and provide clear ESG disclosures (DNV, 2024). However, the lack of standardization across reporting frameworks and the risk of greenwashing present significant challenges. In response, initiatives such as the Global Steel Climate Council (GSCC) are working to establish unified standards, focusing on reliable and consistent reporting of carbon emissions (Herrera, 2023).

2.1.2 Carbon intensity and Scope 1 Direct Emissions

Carbon intensity measures how much carbon dioxide (CO2) is produced for every ton of steel made. The steel industry is responsible for about 3.7 billion tons of CO2 emissions per year, which is more than all emissions from passenger cars worldwide (Lempriere, 2023). Scope 1 direct emissions are the greenhouse gases released from sources that a company owns or controls directly, such as emissions from furnaces and production equipment (National Grid, 2024). In this thesis, we will use the term Carbon Intensity (CI) to refer specifically to scope 1 carbon emissions.

Reducing carbon intensity and Scope 1 emissions is crucial for the steel industry as it transitions toward a low-carbon future. Companies are exploring new technologies, such as Electric Arc Furnaces (EAF), which use electricity instead of coal to produce steel. As of early 2023, about 43% of the planned steel production capacity will use EAF technology, an increase from 33% the previous year (Lempriere, 2023).

Accurate measurement and reporting of carbon intensity and Scope 1 emissions are necessary. From a stakeholder perspective, these figures help to demonstrate how a company impacts the environment and whether it is meeting sustainability goals. From the company's perspective, this is relevant because investors and customers increasingly rely on this information to make decisions about which companies to support (World Economic Forum, 2023). Furthermore, a McKinsey report finds that an estimated 14% of steel companies' potential value is at risk if they fail to reduce Scope 1 emissions and environmental impact (Christian Hoffmann et al., 2020).

However, measuring these emissions can be challenging. The steel industry's CO2 emission intensity has remained relatively stable in recent years but must decrease significantly to align with net-zero scenarios (Martin Kueppers, 2023). The complexity and variability of steel production methods globally make it difficult to collect consistent and comparable data between producers.

2.1.3 Greenwashing

Greenwashing involves companies making misleading claims about their environmental benefits or sustainability efforts to appear more environmentally conscious than they are. This practice has become increasingly common, with 25% of climate-related risk incidents from 2022 to 2023 related to greenwashing, up from 20% the previous year (Johnson, 2023). In the steel sector, greenwashing includes vague claims about "green" or "eco-friendly" products without clear criteria or significant emission reductions (Antto Jokinen, 2023).

The lack of standardized definitions and regulations for "green steel" has raised concerns about greenwashing, potentially misleading consumers and investors (Birand, 2024). For example, certain steel producers have been accused of overstating their environmental impact, with carbon reduction efforts representing less than 1% of overall emissions while marketing a significant portion of their steel as "low carbon" (Birand, 2024).

To combat greenwashing, the industry is exploring solutions such as third-party assurance and the development of unified standards for carbon emissions reporting (Antto Jokinen, 2023). For example, DNV offers green steel verification services to ensure the credibility of environmental claims (Antto Jokinen, 2023). Some companies, such as SSAB, are striving for transparency by clearly distinguishing their fossil-free steel made without fossil fuels from other "green" labels (Birand, 2024).

2.2 Literature Review

There are some studies that explore the relationship between ESG reporting and environmental performance, but they are usually focused on utilizing ESG ratings as a benchmark. A notable study comparing ESG reports and ESG ratings is “Bridging the Gap in ESG Measurement: Using NLP to Quantify Environmental, Social and Governance Communication” by Schimanski et al. (Schimanski et al., 2023). This paper introduces a new approach to quantifying ESG communication using Natural Language Processing (NLP) techniques, specifically a custom-made BERT model. The authors used a data set of more than 13.8 million text samples to analyze ESG communication, supplemented by expert-annotated datasets for the environmental, social, and governance dimensions. Their models demonstrate strong classification performance and reveal a significant positive relationship between ESG communication and ESG ratings while controlling for other variables (Schimanski et al., 2023).

Furthermore, there are several studies that examine how advanced tools can be used to analyze ESG data. The paper "Proposing an Integrated Approach to Analyzing ESG Data via Machine Learning and Deep Learning Algorithms" by Lee et al. (2022) explores methods for evaluating ESG performance using machine learning models like Light Gradient Boosting Machine (LGBM) and Long Short-Term Memory (LSTM). It constructs an ESG index by applying regression models to financial data, demonstrating how data-driven methods can improve the assessment of ESG performance. Although this study focuses mainly on financial metrics, this thesis uses NLP to analyze ESG reports. The work of Lee et al. highlights how advanced methodologies can improve ESG measurement, particularly in high-emission industries such as steel (Ook Lee et al., 2022).

Finally, there is a range of studies that investigate the relationship between ESG reporting and actual ESG performance. Barkemeyer, Comyns, Figge, and Napolitano (2014) conducted sentiment analysis on CEO statements from corporate ESG disclosures and found that sustainability reporting does not reliably reflect accountability. Their study suggests that CEO statements in ESG reports often reflect impression management rather than actual environmental performance (Barkemeyer et al., 2014).

Another study that found a lack of alignment between ESG reporting and actual emissions is “Cheap Talk and Cherry-Picking: What ClimateBert Has to Say on Corporate Climate Risk Disclosures” (Bingler et al., 2022). In this paper, Bingler et al. used natural language processing techniques to analyze climate-related disclosures and found evidence of "cheap talk" and "cherry-picking." Companies’ climate claims in ESG reports did not align with their actual emissions and climate initiatives. The study concludes that climate risk disclosures often do not provide a reliable indication of a company’s true climate impact, emphasizing the need for standardized and regulated reporting to filter out misleading information (Bingler et al., 2022).

3 Data

3.1 Introduction to Data

In this chapter, we present the two main data sources used in our analysis: carbon emissions data as computed by LCA engineering tools and the companies' Annual Reports, ESG Reports, and Integrated Reports. These sources provide the basis for exploring the relationship between reported sustainability efforts and actual environmental performance.

We explain how the data were collected and describe the scope of the dataset, including the companies analyzed, the time periods covered, and other features. A descriptive summary of the data is included to highlight the main characteristics and variations. Finally, we address the limitations and potential biases in the data, which are important for interpreting the results of the analysis.

3.2 Data Overview

Since this thesis is focused on the steel industry, only companies operating within this sector were included in the dataset. The data used for the analysis can be categorized into two main parts.

The first part consists of emissions data as computed by LCA engineering tools, which includes numerical values representing companies' actual environmental performance. These data serve as the response variable in the regression models described later in the methodology chapter. Specifically, we focus on Scope 1 emissions, which account for direct greenhouse gas emissions from processes that companies have direct control over (National Grid, 2024). Emissions from other parts of the supply chain, such as upstream or downstream activities, are excluded from the scope of this study.

The second part consists of textual data, used to construct the Environmental Index and create sentiment scores. For this, we rely on ESG Reports, Annual Reports, and Integrated Reports. While annual reports primarily focus on financial performance and regulatory compliance, integrated reports offer a more "holistic view", and incorporate aspects of sustainability and other forms of value creation alongside financial metrics (Dearnell, 2022). These textual sources are analyzed to quantify the environmental narratives of the companies and assess how they communicate sustainability efforts.

To include a temporal dimension in the analysis through panel data, the dataset is limited to companies that have at least four years of publicly available ESG data, specifically for the years 2019 to 2022. This ensures consistency and allows for comparisons across the same time period.

3.3 Data Gathering

3.3.1 Carbon Intensity (CI)

The CI data was obtained from a database compiled by Steelstat, a statistics tracking company specializing in the steel industry. Steelstat calculates Scope 1 emissions using plant-specific data, including technical specifications, material usage, and energy consumption. Standardized conversion factors from the World Steel Association are applied to ensure consistency and comparability across facilities, which allows for reliable emissions comparisons between companies (Steelstat, 2024). From this database, we extracted all relevant data on Scope 1 emissions, along with additional information about each company's regional location. This dataset serves as the basis for understanding direct emissions under the companies' operational control.

3.3.2 ESG and Company Reports

For the textual data, we manually gathered public reports in which ESG-related information was disclosed for each company. Some companies provided standalone ESG reports, while others included ESG information within their annual or integrated reports, depending on their reporting practices and what was publicly available for each company. The reports were sourced directly from company websites to ensure accuracy.

Each document was inspected to identify and resolve potential issues, such as:

- **Non-textual data:** Some reports included scanned images, locked PDFs or other types of data which are incompatible with text-scraping tools. These documents were excluded from the dataset.

- **Corrupt or empty files:** Any unusable files were removed to maintain data quality.
- **Language Barriers:** Reports that were not published in English were excluded due to limitations in text-processing tools

The process of extracting text from these reports, cleaning the data, and preparing them for analysis is described in detail in the Methodology section.

3.4 Scope and Descriptive Statistics

3.4.1 Scope

After gathering all the reports, we filtered out any non-English documents. A significant number of reports were in languages such as Korean, Chinese, or Russian. However, since our analysis relies primarily on the use of large language models (LLMs) to quantify textual data, it was necessary to focus on a single language. This decision was driven by the fact that different languages are tokenized differently and often have different baseline sentiment levels (Jun, 2023), which could introduce inconsistencies in the analysis.

To ensure comparability over time, we also limited the dataset to companies that met the following criteria:

- Reports were available for all four years between 2019 and 2022.
- Reporting type (e.g., ESG reports, annual reports, integrated reports) remained consistent across all years to allow for meaningful year-over-year comparisons. Companies that switched reporting format in one of the four years we analyzed were dropped from the dataset.

After applying these filters, we were left with a balanced a dataset comprised of 37 steel companies, each with four years of observations. This resulted in a total of 148 data points, which form the foundation for the analysis in subsequent chapters.

3.4.2 Descriptive Statistics for Carbon Emissions Data

For the numerical data, the emission values of carbon intensity range from approximately 0.1 to 2.4. This value represents the ratio of tons of CO_2 or CO_2-equivalents emitted to tons of steel produced. For example, a carbon intensity of 1 indicates that a company emits 1 ton of CO_2 for every ton of steel it manufactures. Summary statistics for the average carbon intensity data are presented in Table 3.1.

Table 3.1: Summary Statistics for average Scope 1 carbon intensity Data

Statistic	Value
Count	148
Mean	1.26
Standard Deviation	0.77
Minimum	0.11
25th Percentile	0.35
Median (50th Percentile)	1.56
75th Percentile	1.94
Maximum	2.39

The carbon emissions data used in this study is sourced from steel companies and steel mills operating in various countries. However, since this analysis relies exclusively on reports in English, the dataset may not fully represent the global distribution of steel companies and mills (World Steel Association, 2022). Despite this limitation, as shown in Figure 3.1, the data include companies from multiple countries in all major regions, with the aim of providing a representative sample of the global steel industry.

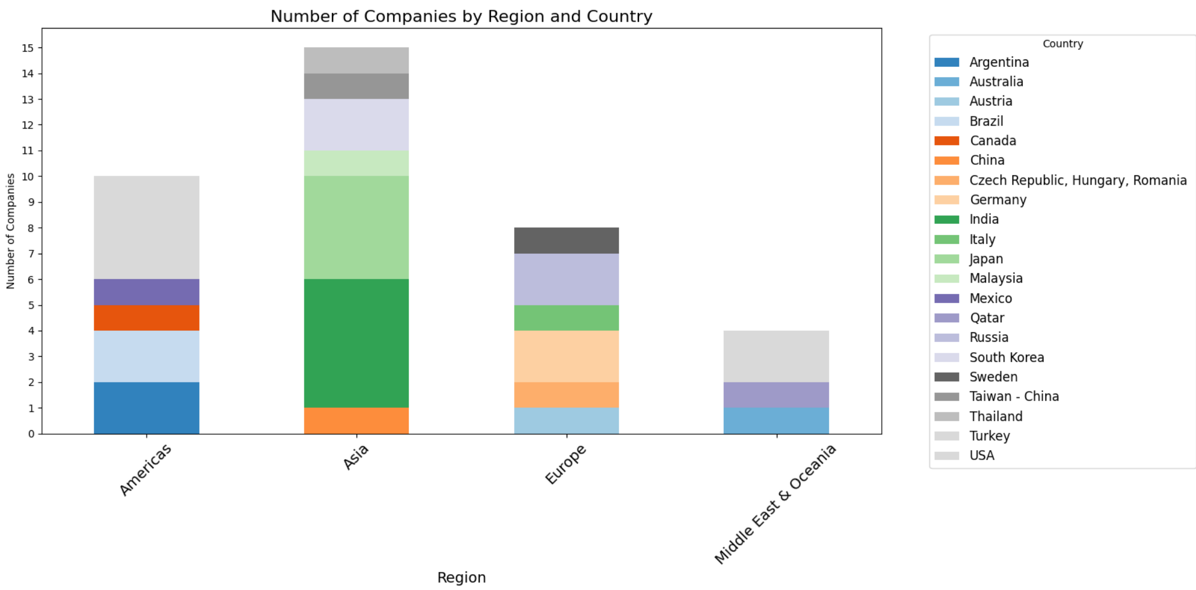

Figure 3.1: Geographic distribution of steel mills in the dataset

Figure 3.2 illustrates the trend lines for the Scope 1 emissions data, with each line representing the carbon intensity of an individual company over the four years included in our analysis. The thick dashed line represents the overall aggregated trend. Most companies exhibit relatively stable carbon intensity levels, with only minor fluctuations and occasional decreases/increases observed. The overall trend line suggests a slight downward trajectory, indicating a modest reduction in carbon intensity in the steel industry as a whole. This appears to align with findings from the International Energy Agency, which reports a similar trend for the global steel industry (Martin Kueppers, 2023).

From a visual inspection of the figure, we can also observe an indication of clustering among the companies. For instance, there appears to be a concentration of companies below the 0.5 line and another grouping around the 2.0 line. The division into groups of high and low emitters will be discussed in more detail in the analysis chapter.

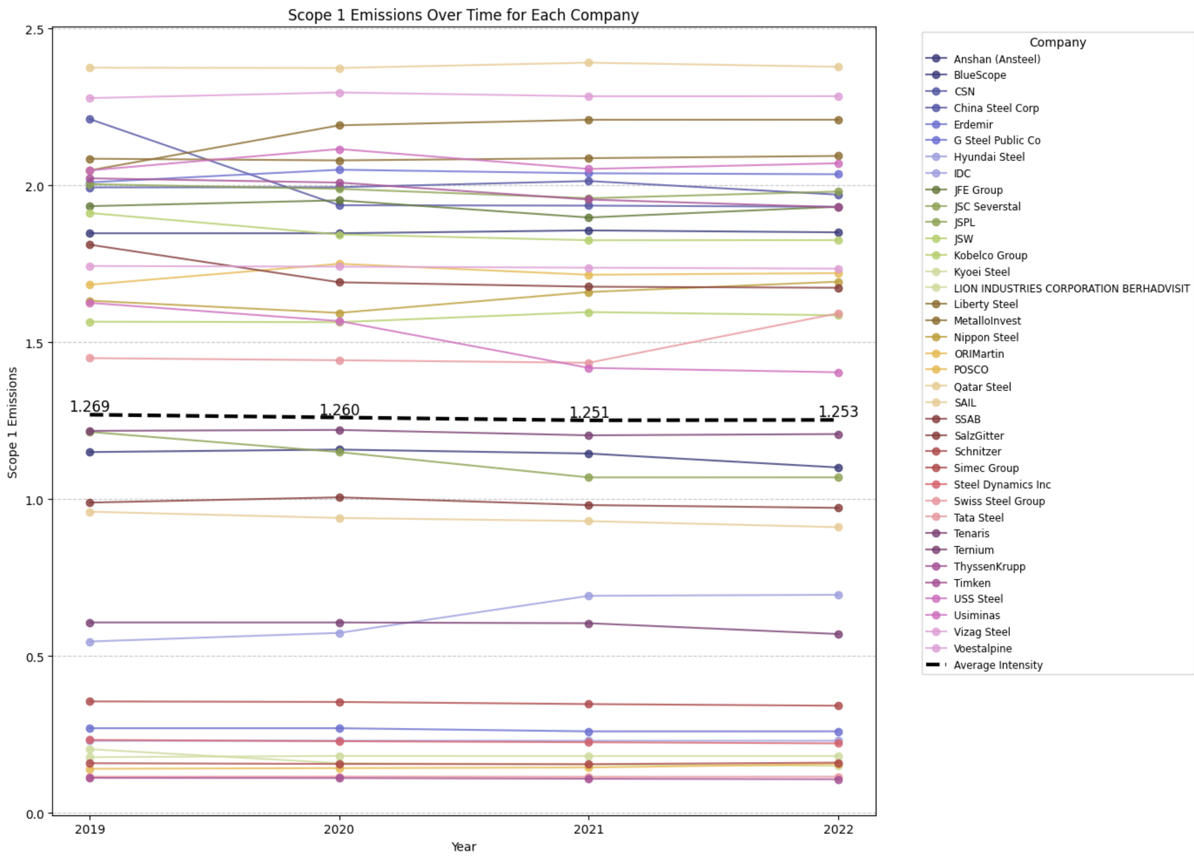

Figure 3.2: Scope 1 emissions trend over time for each company

3.4.3 Descriptive Statistics for ESG reporting data

Since the sample of companies is the same for both numerical and textual data, the geographical distribution aligns with what was described in the previous subsection. During the data collection process, we observed that not all companies produce standalone ESG reports. Instead, many companies disclose their ESG performance and goals within annual or integrated reports. Given that these two types of reports are often similar and can substitute for each other, they have been grouped into a single category.

In Figure 3.3, the left side illustrates the breakdown of report types within the data set, showing the proportion of companies that use standalone ESG reports compared to those that rely on annual or integrated reports. On the right side, the figure presents the average Scope 1 carbon emissions for the two report groups. From a visual inspection, there appears to be a noticeable discrepancy between the report types, with companies publishing ESG reports exhibiting an average carbon intensity that is approximately 30% higher than those relying on annual or integrated reports. In the analysis chapter, we investigate whether this observed difference is statistically significant.

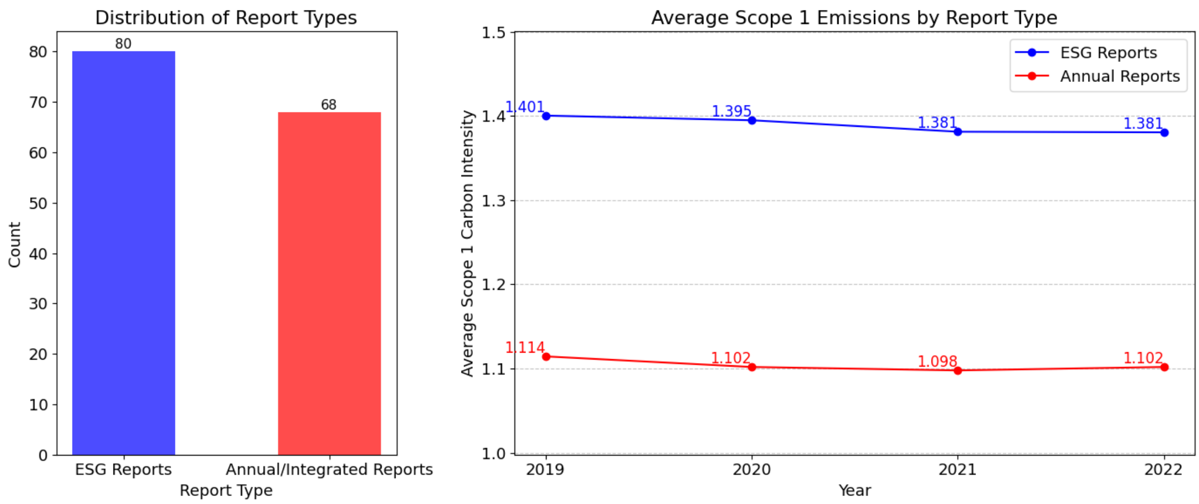

Figure 3.3: Report type distribution and scope emissions per group

The observed difference may be due to companies with higher emissions facing greater public pressure to address their impact, causing them to use standalone ESG reports to improve their reputation. Research suggests that high ESG ratings do not always correlate with lower emissions, as the good publicity from the rating reduces incentives for improvement (Treepongkaruna et al., 2024). Additionally, ESG reports can prioritize the image over the actual performance, which we defined earlier as the concept of greenwashing (Johnson, 2023). In contrast, companies with lower emissions may be content with describing their ESG efforts in their annual or integrated reports, as their environmental performance already meets expectations.

3.5 Limitations and Biases in the Data

The dataset used in this study is subject to several limitations and potential biases that could affect the interpretation and generalizability of the results:

- **Language Restriction:** Only English-language reports were included in the dataset. This exclusion of non-English reports, such as those in Chinese, Korean, or Russian, may lead to a regional bias, as companies from certain regions are less likely to produce reports in English. This limitation could under-represent key steel-producing countries, particularly in Asia.

- **Lack of Standardization in ESG Reports:** ESG reporting is not standardized globally (International Federation of Accountants, 2024), resulting in significant variability in the content and structure of the report. This lack of uniformity requires extra data processing and limits the choice of index methodology. For example, we could not weigh environmental scores based on a fixed set of standardized sections, which could otherwise improve the accuracy and comparability of the results.

- **Exclusion of Non-Reporting Companies:** Some companies did not provide ESG disclosures, either in standalone ESG reports or integrated/annual reports. These companies were excluded from the analysis. This introduces a potential bias, as the dataset may over-represent companies with better ESG practices or those more willing to voluntary report their environmental statistics.

- **Regional Imbalance:** According to a McKinsey report, the steel industry's highest-emitting region is China (Christian Hoffmann et al., 2020), which is underrepresented in our dataset relative to its global production capacity (Sustainable Ships, 2024). This regional imbalance could limit the generalizability of the findings, as the dataset may not fully capture emissions or reporting trends in the most significant steel-producing regions.

These limitations underscore the challenges of analyzing ESG data in the steel industry and must be taken into account when interpreting the findings.

4 Methodology

4.1 Introduction

In this segment, we describe the methodology used in the thesis. First, a combination of different Python libraries such as NLTK, Apache Tika, and RegEx are used to extract sentences from the reports, all of which are in PDF format. For sentence classification, we use two approaches: Method #1: a zero-shot learning model based on Google's DeBERTa model. Method #2: specialized BERT models fine-tuned to detect environmental sentences, and more specific subtopics. For example, the EnvironmentalBERT model, a classifier pre-trained with large amounts of ESG-related data, to classify sentences as either environmental or non-environmental.

The proportions of sentences related to the different environmental topics classified by the different models are then calculated for all companies from 2019 to 2022, and these metrics are stored in a separate data frame. In addition to creating this, we also perform sentiment analysis on environmental-related sentences to generate sentiment scores for each company and year.

Finally, the relationship between the calculated environmental metrics (the proportions) and the actual scope 1 carbon intensity is examined, with the environmental metrics as the explanatory variable and scope 1 carbon intensity as the response variable.

Using environmental metrics as the explanatory variable and scope 1 carbon intensity as the response variable is reasonable, especially considering the self-selection bias in ESG reporting. Self-selection means, in this case, that companies can choose which environmental topics to emphasize, often highlighting areas where they perform well while downplaying less favorable aspects (Contreras-Pacheco & Claasen, 2017). This creates subjectivity in the environmental metrics, as they may reflect strategic communication rather than communication representative of the company's environmental performance.

By comparing these metrics with the response variable scope 1 carbon intensity, which is data verified by a third-party and thus less prone to self-selection bias (Steelstat, 2024), we can evaluate whether companies' emphasis on environmental issues in their reporting aligns with their actual environmental performance. In this way we can examine whether the reported environmental communication is representative of actual environmental performance or if it is more strategic.

Similarly, a regression is performed using sentiment scores as the explanatory variable. Given the panel data structure, with 38 companies over 4 years, Fixed Effects and Random Effects regression models are used.

4.2 Data Preprocessing - Extraction of text from PDFs

Sentences are extracted from the PDF reports using a rule-based text parsing method based on an approach developed by Morgan (2022). This method utilizes several Python libraries, including Apache Tika for text extraction (The Apache Software Foundation, 2024), Natural Language Toolkit (NLTK) for sentence tokenization (Bird et al., 2009), and Regular Expressions (RegEx) for text cleaning and processing (Friedl, 1997). In the first step, raw text is parsed from the PDF report using Apache Tika. This is then cleaned by removing all non-ASCII characters, replacing all tabs with spaces, and the text is then split into lines. Lines where every letter is in uppercase are removed, as these are typically headers not relevant for sentence-level analysis. Also, lines starting with a whitespace, lowercase letter, or following a previous line not ending with a period, are combined into a sentence fragment, to handle sentences stretching over multiple lines. An example of a sentence stretching over multiple lines, taken from JFE Group's 2021 CSR report, is shown below:

'*JFE Shoji has established a stable global supply chain that sources high-quality electrical steel sheets which are*'

'*essential for improving the efficiency of motors and transformers from JFE Steel and other manufacturers and processes*'

'*the products for meeting customer needs. Customers who require high-quality electrical steel sheets, such as motor*'

'*manufacturers and transformer manufacturers, typically operate manufacturing facilities across the globe. To align itself*'

'*to this trend, the company has been expanding its electrical steel sheets supply chain based in a global quad-polar*' '*organization that includes Japan, America, China, and ASEAN. By further expanding its supply chain and processing*'

'*capabilities and collaborations with alliance companies, the company is striving for significant improvements in the*'

'*distribution and processing of electrical steel sheets, as described in the Seventh Medium-term Business Plan, and more*'

'*thoroughly responding to customer needs.*'

These lines are then gathered into a sentence fragment:

'*JFE Shoji has established a stable global supply chain that sources high-quality electrical steel sheets which are essential for improving the efficiency of motors and transformers from JFE Steel and other manufacturers and processes the products for meeting customer needs. Customers who require high-quality electrical steel sheets, such as motor manufacturers and transformer manufacturers, typically operate manufacturing facilities across the globe. To align itself to this trend, the company has been expanding its electrical steel sheets supply chain based in a global quad-polar organization that includes Japan, America, China, and ASEAN. By further expanding its supply chain and processing capabilities and collaborations with alliance companies, the company is striving for significant improvements in the distribution and processing of electrical steel sheets, as described in the Seventh Medium-term Business Plan, and more thoroughly responding to customer needs.*'

Each sentence fragment is then subjected to more cleaning steps: URLs, headers starting with a number, and sequences of numbers with more than 5 digits are all dropped. Following this, each sentence fragment is split into individual sentences using the NLTK sentence tokenizer, and sentences containing "table of contents" or shorter than 5 characters are removed. Additionally, extremely long sentences (with more than 512 tokens) are split into shorter sentence chunks, to ensure that the sentences are not too long for the BERT-classification models used later (Brzozowski, 2023).

Using this rule-based text parsing method, we mostly obtain cleaned, meaningful sentences which can later be used in the classification models. Continuing the example above, we obtain these sentences from the sentence fragment:

'*jfe shoji has established a stable global supply chain that sources high-quality electrical steel sheets which are essential for improving the efficiency of motors and transformers from jfe steel and other manufacturers and processes the products for meeting customer needs.*'

'*customers who require high-quality electrical steel sheets, such as motor manufacturers and transformer manufacturers, typically operate manufacturing facilities across the globe.*'

'*to align itself to this trend, the company has been expanding its electrical steel sheets supply chain based in a global quad-polar organization that includes japan, america, china, and asean.*'

'*by further expanding its supply chain and processing capabilities and collaborations with alliance companies, the company is striving for significant improvements in the distribution and processing of electrical steel sheets, as described in the seventh medium-term business plan, and more thoroughly responding to customer needs.*'

However, the method also has its flaws, as not all sentences are meaningful. "more specifically." is an example of a sentence that does not provide enough context to be informative, extracted from ORIMartin 2022. This is a consequence of varying formatting in the PDF reports, as it is hard to take into account all possible formats with a rule-based approach. Despite these flaws, the method generally works for the purpose of this thesis, as the majority of extracted sentences seem to be meaningful and appropriate for analysis.

While the method is relatively flexible and can be used on several different PDF structures, it does not account for all structures, particularly image-based PDFs. Optical character recognition (OCR) is required to extract text from such PDFs (Nätt, 2024). OCR can also be used on regular text PDFs and is therefore more flexible than our approach.

For computational purposes, however, we have chosen not to include OCR and rather focus on pure text parsing. As OCR is computationally heavy and time-intensive, our rule-based text parsing method is faster (Paudel et al., 2024), which is advantageous when analyzing a large number of PDF files. To improve the method's flexibility and enable it to handle a wider range of PDF formats, future research should consider also including more flexible methods of text extraction, for example OCR.

4.3 BERT-based Classification Models

After the sentences have been extracted and cleaned, they act as input in different classification models that are based on the large language model BERT. The classification models we have used are of two types: BERT-based Zero-Shot Text Classification and specialized BERT models fine-trained on specific classification tasks.

4.3.1 NLP and the BERT-model

Natural Language Processing, or NLP, is a field within artificial intelligence where machine learning is utilized to make it possible for computers to understand and interact with human language (Stryker & Holdsworth, 2024). Large language models are prominent within the NLP-field, and these are deep learning models trained on vast amounts of text data making them able to understand and generate natural (human) language (IBM, 2024). Such models are usually built on a transformer architecture. Transformer models are a type of neural network designed to understand context and meaning by analyzing relationships in sequential data, for example the words of a sentence (Merritt, 2022). BERT is built on a multi-layered transformer architecture, and is an example of a large language model (IBM, 2024).

BERT, or Bidirectional Encoder Representations from Transformers, is a self-supervised language model that does not require pre-training with labeled data, meaning that it can learn from any type of text input (NVIDIA, 2023). It uses both Masked Language Modeling (MLM) and Next-Sentence Prediction (NSP) for learning (Devlin et al., 2019). Masked Language Modeling means that some of the words in a sentence are randomly hidden and the model must predict the hidden words based on the context, meaning that sentences are processed in both directions at the same time, thereby "bidirectional". Compared to older models, the inclusion of Masked Language Modeling, in addition to Next-Sentence Prediction, makes BERT better at understanding the meaning of words by looking at their context.

BERT converts, or encodes, text to numerical representations by assigning each word to a fixed vector based on a pre-trained vocabulary (token embedding), by assigning all the words of a sentence to a fixed vector so that the model understands which words belong to a certain sentence (segment embedding), and by assigning the position of the words to a fixed vector (position embedding) (Devlin et al., 2019). Then these three embeddings are summed to one vector, representing the words and their context in numerical terms. This representation of the text then goes through the system of transformers and the relationships between the words in the input text are captured.

BERT has been trained on large quantities of text and can be fine-tuned to perform different tasks like answering questions, sentiment analysis, and text classification (NVIDIA, 2023). In this thesis, we use several different adjusted versions of BERT to classify text.

To provide an overview of the process of how BERT classification is actually used, 4.1 illustrates the flowchart of the classification pipeline we use in this thesis. The process begins with raw ESG, annual, and integrated reports in PDF format collected from company websites. These documents undergo text extraction and cleaning to ensure compatibility with the classification models. The cleaned text is tokenized and classified using BERT-based models to identify whether sentences are environmental or non-environmental. The final output is structured data ready for analysis, including the calculation of the environmental index and sentiment scores.

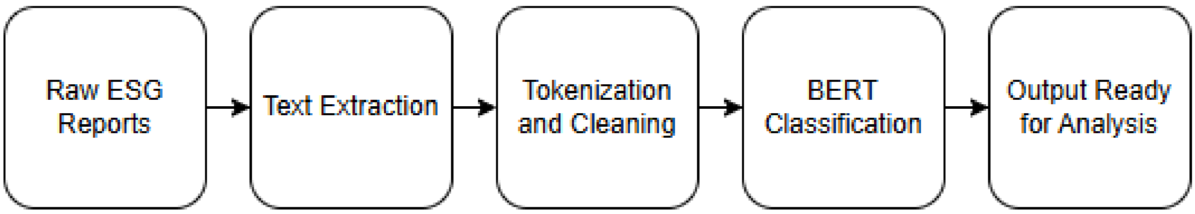

Figure 4.1: Flowchart of the classification pipeline

4.3.2 Zero-Shot Text Classification with the DeBERTa Model

Zero-Shot Classification is a method where a model is trained to categorize data into classes that it has not seen during training (Bergmann, 2024). Rather than learning from labeled examples of every possible class, the model utilizes supplementary information like semantic descriptions, attributes, or textual explanations about the unseen classes. In this way, the model can relate new classes to information it has already learned. This enables the model to categorize data into unseen classes without explicitly being trained on them. Zero-Shot Classification, therefore, makes classification tasks possible even when there is a lack of labeled data.

The type of Zero-Shot Classification we are interested in is Zero-Shot Text Classification on a sentence-level, and the model we use is based on Microsoft's language model DeBERTa (Decoding-enhanced BERT with disentangled attention). DeBERTa is an improved version of BERT that includes two improvements for better language understanding: disentangled attention and enhanced mask decoders (He et al., 2021). These improvements consist of embedding the meaning of words and their positions in two separate vectors, rather than one like in BERT. The word meanings and word positions embedded in vectors by DeBERTa function as the supplementary information which the Zero-Shot Classifier utilizes to categorize data into pre-defined unseen classes.

We have constructed a Zero-Shot Text Classifier based on DeBERTa that detects whether sentences belong to certain pre-defined classes or not. The pre-defined classes include "environmental", "carbon emissions", "pollution", "waste management", "renewable energy", "water usage", "energy consumption".

These classes are chosen because they cover several important environmental challenges in the steel industry (Conejo et al., 2020). For example, one of its biggest challenges, is that the industry is responsible for around 7% of total global carbon emissions (Mona, 2024). It also causes significant air pollution by releasing sulfur dioxide, nitrogen oxides, particulate matter, and other harmful compounds (TechnoServe, 2024). These emissions can cause health problems, like respiratory and cardiovascular diseases, for people living nearby, and they also contribute to acid rain. Additionally, steel production requires large amounts of water for cooling and cleaning processes (TechnoServe, 2024). The wastewater often contains heavy metals and oils, which can harm aquatic life and disturb the food chain. The industry also uses energy-intensive processes that mainly rely on coal, which increases CO2 emissions further (Edmond, 2024). Another problem is solid waste, such as slag, dust, and sludge (TechnoServe, 2024). If these are not handled properly, they can pollute the soil and create other environmental issues. However, there are solutions, like renewable energy technologies, including electric arc furnaces and green hydrogen, which can help steel producers lower their emissions and make production more sustainable (Ellerbeck, 2022).

We thus believe that the pre-defined classes we have chosen capture the most important environmental challenges faced by the steel industry. The pre-defined classes are illustrated in Table 4.1.

Classifier/Model	Pre-defined class (or topic)
DeBERTa-based Zero-Shot Classifier	“environmental”
	“carbon emissions”
	“pollution”
	“waste management”
	“renewable energy”
	“water usage”
	“energy consumption”

Table 4.1: Overview of topics detected by the Zero-Shot Classifier

For each class, the classifier returns a score, on the scale 0-1, for the analyzed sentence. This score indicates how likely it is that the sentence in question belongs to the class. A threshold is then defined, and the sentences with a score higher than the threshold are regarded as belonging to the class, while the rest are not. This is done for each of the seven pre-defined classes, and in practice, we therefore have seven binary Zero-Shot Classifiers. We will in section 5 test different values for the threshold, ranging from 0.1 to 0.9, to examine which score is best for the panel data regressions.

The fact that Zero-Shot Classification models do not need to be specifically trained on labeled data makes it a reasonable choice when classifying a vast amount of text (Hugging Face, 2023), as manually labeling sentences is very time-consuming.

4.3.3 Text Classification with EnvironmentalBERT

EnvironmentalBERT is a pre-trained BERT-based classifier specifically designed for identifying sentences related to the "E" (Environmental) aspect of ESG (Schimanski et al., 2023). EnvironmentalBERT is a binary classifier that categorizes sentences as either "environmental" or "non-environmental."

This model has been optimized using domain-specific pretraining on a large collection of ESG-related text, including corporate reports, news articles, and other sustainability-related disclosures (Schimanski et al., 2023). The pretraining process uses over 2.1 million text samples to optimize the model for tasks specific to the environmental domain, which in turn ensures a higher accuracy and reliability in classification tasks compared to general language models (Schimanski et al., 2023). An overview comparing the pre-training dataset with the base data is presented in Table 4.2.

Table 4.2: Overview of Pre-Training Dataset and Base Data. Source: (Schimanski et al., 2023)

Domain	Num. of Sentences	Avg. Num. of Words		
		Q1	Mean	Q3
Base Data	13,846,000	16	25.12	30
Environmental	2,100,586	17	27.03	32

EnvironmentalBERT was chosen for this thesis as the second primary classification method due to its accuracy in detecting environmental themes in textual data. The output of this model directly feeds into the construction of the Environmental Index and is the main textual foundation for creating the sentiment scores.

4.3.4 Text Classification with Specialized BERT Models

In addition to EnvironmentalBERT, we have also used specialized BERT models fine-trained for classifying specific environmental topics in text. More specifically, the models detect whether sentences are related to these topics: emission reductions, environmental claims (claims the company comes with about their environmental efforts), water (usage and contamination), and renewable energy. The reason we have included the topics emission reductions, water, and renewable energy, is because they are relevant when it comes to the environmental challenges of the steel industry (Conejo et al., 2020). The environmental claims topic is included because with this, we can examine the consistency between the company's claimed environmental efforts and its actual emissions.

The specialized BERT models used for this classification task and the topics they detect are illustrated in table 4.3.

Classifier/Model	Class (or topic)
climatebert/netzero-reduction	Emission reductions (and net-zero targets)
climatebert/environmental-claims	Environmental claims
ESGBERT/EnvironmentalBERT-water	Water usage and contamination
climatebert/renewable	Renewable energy

Table 4.3: Overview of specialized classifiers and which topics they detect

Each of these models classify sentences based on a score of how likely it is that the sentence in question belongs to the specific environmental topic. The sentences with a score over a certain threshold is then classified as belonging to the environmental topic in question, and for each of these models, the default threshold is 0.5, which we will use in Section 5.

The model detecting sentences related to net zero reduction targets have three classes: “net-zero”, “reduction” and “none”. As all the other models are binary with two classes, indicating whether the sentence is related to the topic in question or not, we choose to combine “net-zero” and “reduction” to one class, so that this model also becomes binary.

These BERT models are fine-tuned versions of EnvironmentalBERT-base and ClimateBERT. Both are versions of DistilRoBERTa that are further pre-trained on environmental data (Schimanski et al., 2023; Webersinke et al., 2022). EnvironmentalBERT-base is pre-trained on a large quantity of annual reports, ESG reports, and environmental-related news (Schimanski et al., 2023). ClimateBERT, while also pre-trained on annual and ESG reports and news, additionally includes research paper abstracts in its pre-training (Webersinke et al., 2022).

DistilRoBERTa, in turn, is a distilled version of the RoBERTa (Robustly Optimized BERT Approach) base model. DistilRoBERTa uses the same training process as DistilBERT (Sanh et al., 2019), by being pre-trained on the same text data as its mother-model RoBERTa, and using RoBERTa as its teacher. RoBERTa is an improved version of BERT, pre-trained on over 160 GB of text (Zhuang et al., 2021). Compared to BERT, RoBERTa is trained on more text with bigger batch sizes. Also, RoBERTa does not focus on next-sentence prediction like BERT, but solely on masked word prediction. Additionally, RoBERTa uses dynamic masking instead of static masking of words.

In summary, the specialized models we have used for classification are all fundamentally based on a distilled version of RoBERTa, and have subsequently been fine-tuned in several steps to become specialized models capable of detecting specific environmental topics in text. As these models are pre-trained and specialized for specific classification tasks, they can be used when labeled data is lacking. To avoid time-consuming labeling of the vast amount of text we have gathered, we find it reasonable to use these specialized models.

4.4 Sentiment Analysis

4.4.1 Understanding Sentiment Analysis in ESG Reporting

Sentiment analysis is a method within Natural Language Processing (NLP) that focuses on identifying the emotional tone in text (AWS, 2024). This technique is often used to study opinions, attitudes and emotions in written content, and has become important in areas such as analyzing social media posts, customer reviews, and ESG reports. In the context of ESG disclosures, sentiment analysis helps us understand whether statements made by companies are positive, negative, or neutral (Symanto, 2022). For example, a company may express positive sentiment when highlighting achievements such as meeting CO_2 reduction targets or implementing successful renewable energy projects. In contrast, negative sentiment might be evident in disclosures about incidents such as fuel spillage or failing to achieve sustainability goals. Both positive and negative events are classified as environment-related by the BERT models if they are connected to the company's environmental impact or actions (Schimanski, 2023). By using sentiment analysis, we can go beyond factual information and gain insight into the emotional tone or emphasis that companies use when discussing environmental issues.

4.4.2 Sentiment Analysis with BERT

For the sentiment analysis, we use the pre-trained BERT sentiment model. This model is designed to classify text into five sentiment categories: 1 (very negative), 2 (negative), 3 (neutral), 4 (positive), and 5 (very positive). Using this model, we analyze the emotional tone of the environmental-related sentences extracted from the company reports.

While EnvironmentalBERT helps identify whether a sentence is environmental or not, the sentiment BERT model focuses on evaluating the tone of these sentences. This allows us to study not only the content and quantity of the environmental communication, but also the emotional framing used by the companies.

Both positive and negative events are analyzed if they are related to the company's environmental actions or impact. For example, a statement about achieving CO_2 reduction goals might be classified as positive, while a report about an environmental accident, such as a fuel spill, would likely be classified as negative.

The results of the sentiment analysis for each sentence are stored and will be used to calculate sentiment scores, which are explained in Section 4.6.

4.5 Defining Environmental Communication Metrics

To measure the quantity of environmental communication in ESG and annual reports, we use the proportion of sentences detected in a specific environmental topic, as done in Schimanski et al. (2023). The primary metric, the Environmental Index, which is based on EnvironmentalBERT, measures the degree of general environmental communication by dividing the number of sentences classified as "environmental" by the total number of sentences extracted from the report. Additionally, we calculate supplementary metrics for more specific environmental topics, such as renewable energy, pollution, water and energy consumption, claims, and emission reductions, using the same method. These calculations are performed for each company and year and are illustrated in equation 4.1, where the subscript i represents company and t represents year:

$$\text{Environmental Metric}_{it} = \frac{\text{Number of sentences related to topic}_{it}}{\text{Total number of sentences extracted from the report}_{it}} \tag{4.1}$$

4.5.1 Descriptive Statistics for the Environmental Index

Summary statistics for the computed Environmental Index across all companies and years are presented in Table 4.4. In this table we can see that the Environmental Index ranges from 0.01 to 0.70. The mean value is equal to 0.23 with a standard deviation of 0.15, suggesting a relatively low average Environmental Index with moderate variation. The median is equal to 0.25, indicating a slight skewness in the data. As the 25th percentile is 0.08 and the 75th percentile is 0.33, most companies have an Environmental Index between these points.

Table 4.4: Summary Statistics for the Environmental Index

Statistic	**Value**
Count	152
Mean	0.23
Standard Deviation	0.15
Minimum	0.01
25th Percentile	0.08
Median (50th Percentile)	0.25
75th Percentile	0.33
Maximum	0.70

In Figure 4.2, the trend lines for the Environmental Index are illustrated, with each line representing the Environmental Index of an individual company over the period 2019 to 2022. The thick dashed line represents the overall aggregated trend, in the form of the yearly median value. Over the whole period, we observe an upwards trajectory in the overall trend line, suggesting an increase in environmental communication in reports in the steel industry. This aligns with the rising focus on sustainability from regulators, investors, and industrial customers (Paragamian et al., 2021), which can incentivize companies to dedicate more space for environmental concerns in their reports to please these stakeholders.

When we examine the trend line more closely, we observe that in the periods 2019 to 2020 and 2021 to 2022, the median Environmental Index is relatively stable, while there is a moderate increase between 2020 and 2021. This might be the effects of COVID-19, as this global pandemic lead investors and regulators to demand better transparency, comparability and consistency in ESG-reporting (Adams & Abhayawansa, 2022). However, as the focus of this thesis is not to find the drivers of environmental communication, we recommend examining this increase in future research.

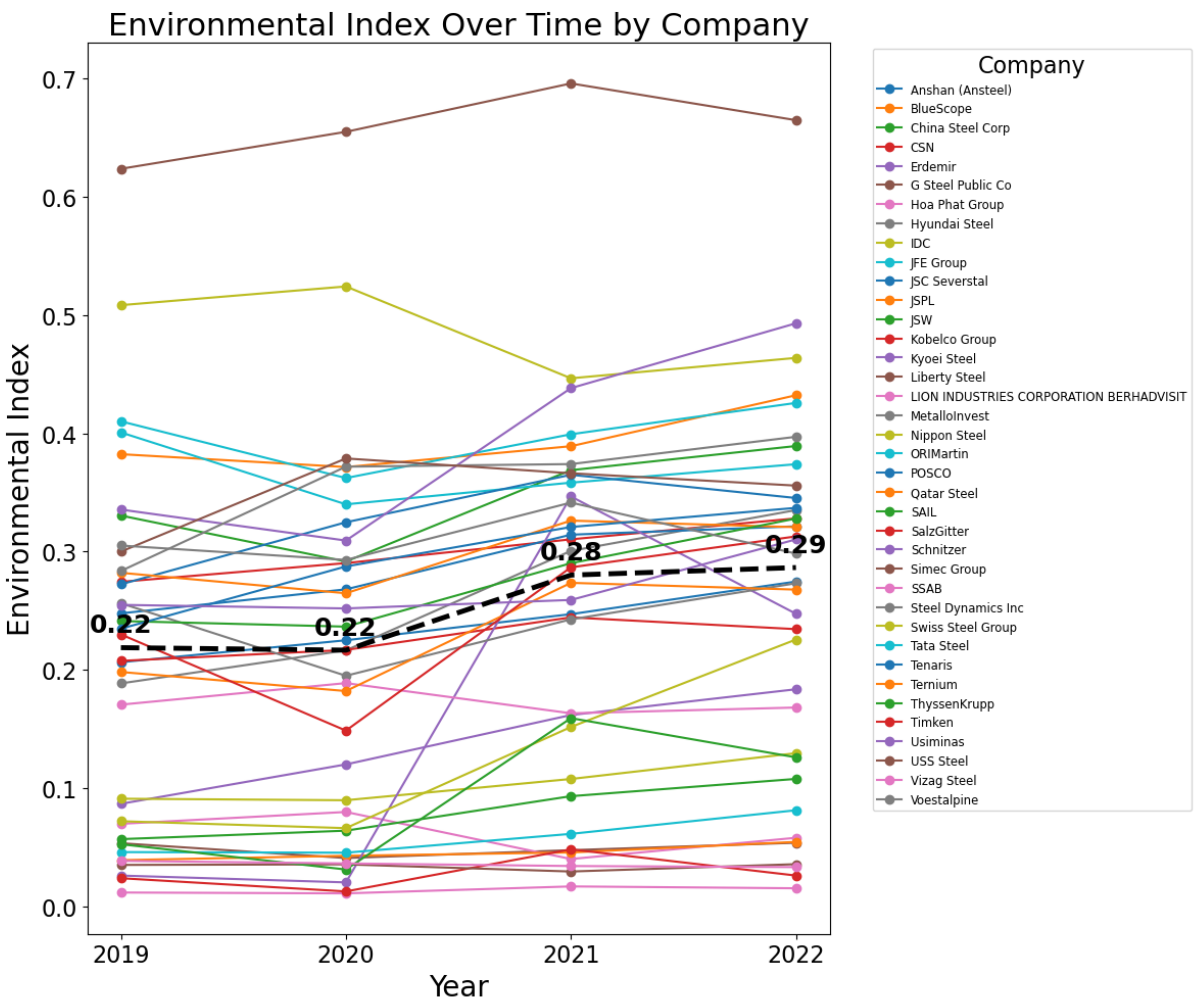

Figure 4.2: Environmental Index trend over time for each company

4.6 Defining the Sentiment Score

In this subsection, we describe the method we use to calculate sentiment scores for each company and year. The sentiment score is a metric that quantifies the emotional tone in environmental sentences extracted from company reports. This score is created through a two-step process: computing the logit probabilities for sentiment classification and aggregating these probabilities into an annual sentiment score for each company.

The BERT sentiment analysis model, which uses a multinomial logit framework, classifies sentences into five sentiment categories, where 1 represents highly negative sentiment and 5 represents highly positive sentiment (Gao et al., 2019). The model outputs logit probabilities for each category, which indicate the likelihood of a sentence belonging to a specific sentiment class. These probabilities are derived from a multinomial logit model and are calculated by applying the softmax function to the model's logits. Specifically, for a given sentence j, the probability of it being classified into sentiment class k is given by:

$$P_{ijk} = \frac{\exp(z_{ijk})}{\sum_{k=1}^{K} \exp(z_{ijk})} \tag{4.2}$$

Where:

- P_{ijk}: Probability of sentence j from company i in year t belonging to sentiment class k,
- z_{ijk}: Raw logit value for sentiment class k,
- K: Total number of sentiment classes (5 in this case).

We then use he probabilities P_{ijk} for each sentiment class to compute the weighted sentiment score for individual sentences. The sentiment score for a company in a given year aggregates the sentiment scores of all environmental sentences by taking a weighted average of the sentiment scores for each sentence. The formula for calculating the annual sentiment score can be shown mathematically as follows:

$$\text{SentimentScore}_{it} = \frac{\sum_{j=1}^{N_{it}} \left(\sum_{k=1}^{5} k \cdot P_{ijk} \right)}{N_{it}} \tag{4.3}$$

Where:

- $\text{SentimentScore}_{it}$: The average sentiment score for company i in year t,
- P_{ijk}: Probability of sentence j from company i in year t belonging to sentiment class k,
- k: Sentiment class (1 to 5),
- N_{it}: Total number of environmental sentences classified for company i in year t.

The sentiment score provides a numerical representation of the emotional tone in a company's environmental communication, and will be used as an explanatory variable in the regression analysis to explore its relationship with the actual carbon intensity. A higher score indicates a more positive sentiment, while a lower score suggests a more negative sentiment. By using logit probabilities instead of choosing the maximum likelihood category, this metric accounts for more of the uncertainty in sentence classification and aims to give a realistic view of the sentiment in ESG-related disclosures.

4.7 Comparing the Distributions of the Metrics

The distribution plot in Figure 4.3 shows the differences in the Environmental Index and Sentiment Score across the dataset. The Sentiment Score (green) follows a smoother, unimodal distribution that is close to normal. This aligns with expectations based on the Central Limit Theorem (CLT) (Ganti, 2024), as the score is an average of sentiment probabilities across many sentences, as defined in Equation 4.3. When the number of sentences (N_{it}) is sufficiently large, the aggregation of sentiment scores naturally tends toward a normal distribution.

In contrast, the Environmental Index (blue) displays a bimodal distribution with two peaks, indicating structural differences within the dataset. This can be explained by the presence of two distinct report types: ESG reports and annual/integrated reports. ESG reports focus heavily on environmental topics, leading to higher Environmental Index values, while annual reports contain a broader range of content, including financial and operational information, resulting in lower Environmental Index values. This segmentation causes the bimodality observed in the plot (Feldman, 2024).

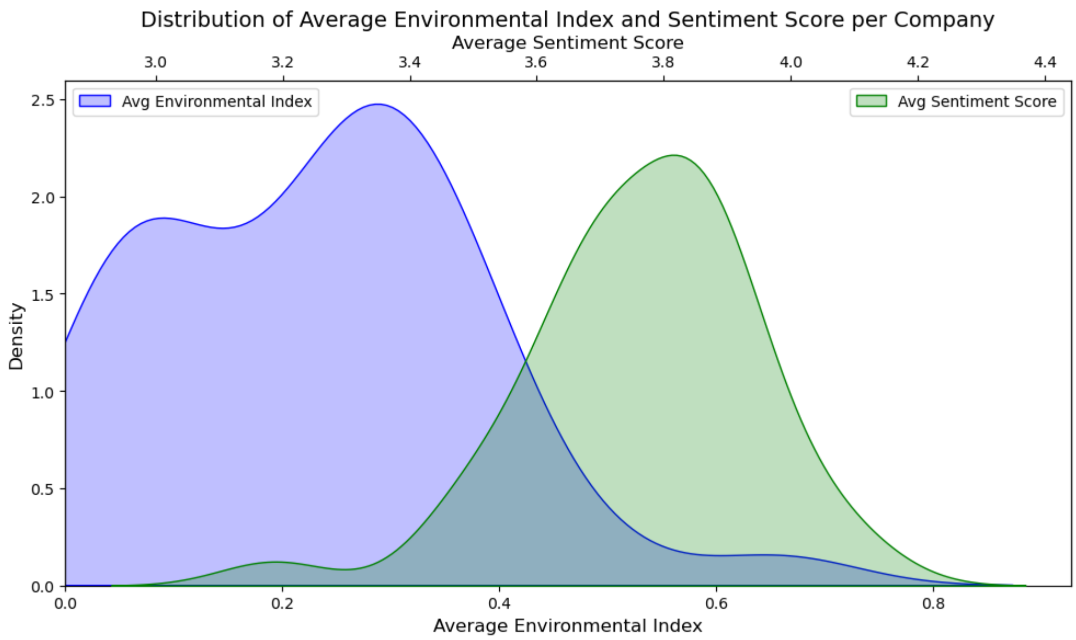

Figure 4.3: Comparison of the two metrics' distributions

4.8 Statistical Thresholding with K-means Clustering

In the descriptive analysis of the data, we observed a grouping pattern in the distribution of companies' Scope 1 carbon intensity. Some companies were clustered around lower carbon intensity values, while others were concentrated at higher levels. To quantify these observations and establish a data-driven threshold for analysis, we used the K-means clustering algorithm.

K-means clustering is a method used to divide data into k distinct groups, called clusters. It works by assigning each data point to the closest cluster center, also known as a centroid. The centroids are adjusted step by step, and the data points are reassigned until the groups become stable (Eda Kavlakoglu & Vanna Winland, 2024). The goal of the algorithm is to create clusters where the data points in the same group are as similar as possible, while the points in different groups are as different as possible (Javatpoint, 2024). By applying k-means, we divide companies into low, medium, and high carbon emitters based on their Scope 1 carbon intensity.

Figure 4.4 provides an illustration of how k-means clustering organizes data into clear groups.

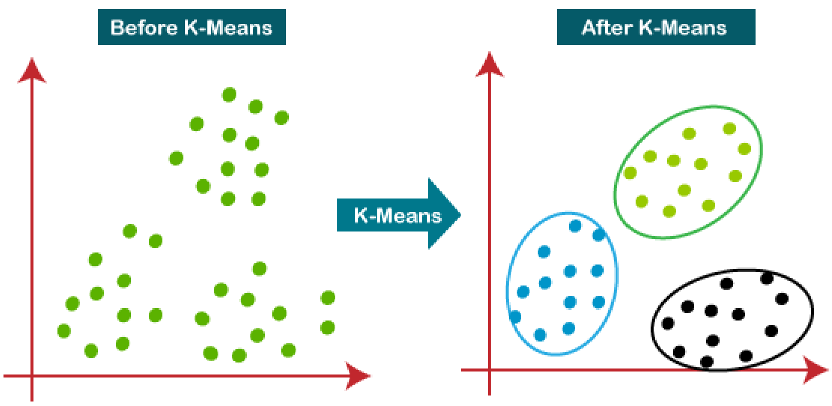

Figure 4.4: Report type distribution and scope emissions per group

5 Analysis - Unstratified

In this section, we analyze the entire dataset, performing regressions without distinguishing between company types. The goal is to evaluate the overall relationship between the environmental index, sentiment score and carbon intensity across all companies, regardless of report type or emission level.

5.1 Analysis of Metrics based on BERT Classification Models

In this section we will analyze the relationship between the calculated environmental metrics and scope 1 carbon emissions intensity using Fixed Effects and Random Effects models (Batalgi, 2005).

Each regression uses a single environmental metric as the explanatory variable, and each metric is analyzed separately. All calculated metrics are multiplied by 100 to obtain them in percentage values, making coefficients easier to interpret.

For Zero-Shot Classifier-based metrics, we test classification score thresholds ranging from 0.1 to 0.9, and only the regression results from the thresholds giving the lowest P-values are included. As the threshold testing for the Zero-Shot Classifier did not give any significant results, and since the specialized BERT models require considerably more computational resources, we use the default threshold of 0.5 for these.

5.1.1 Fixed Effects Regression Results

5.1.1.1 FE-Regression Results for the Zero-Shot Classifier

For the Zero-Shot Classifier none of the environmental metrics have a significant relationship with scope 1 carbon emission intensity for any threshold of the classification score with the Fixed Effects model. The *renewable energy* metric has the coefficient $\hat{\beta}$ with the lowest p-value when accounting for all score thresholds, and is thus the most significant. However, this p-value is 0.4571, and is therefore far from significant. The lowest P-value obtained for each metric, for all thresholds, and other relevant regression statistics are illustrated in 5.1.

Environmental Metric	Threshold	$\hat{\beta}$	Std. Err.	P-value	R^2	Adj. R^2
Environmental Share	0.9	-0.0025	0.0045	0.5794	0.0027	-0.3326
Carbon Emissions Share	0.9	-0.0059	0.0168	0.7261	0.0011	-0.3348
Pollution Share	0.2	-0.0024	0.0078	0.7637	0.0008	-0.3352
Energy Consumption Share	0.1	0.0029	0.0042	0.4981	0.0041	-0.3308
Water Usage Share	0.7	-0.0105	0.0195	0.5907	0.0026	-0.3329
Waste Management Share	0.5	0.0033	0.0095	0.7260	0.0011	-0.3348
Renewable Energy Share	0.6	-0.0076	0.0102	0.4571	0.0049	-0.3297

Table 5.1: FE-regression results for the Zero-Shot Classifier

5.1.1.2 FE-Regression Results for the EnvironmentalBERT Classifier

The EnvironmentalBERT classifier uses a binary classification approach to identify whether sentences are environmental or not. The proportion of environmental sentences, what we call the "Environmental Index", is used as the explanatory variable in the regression.

The results of the Fixed Effects regression for the EnvironmentalBERT classifier are presented in Table 5.2. The coefficient estimate ($\hat{\beta} = -0.0812$) suggests a negative relationship between the Environmental Index and Scope 1 carbon intensity, but this relationship is not statistically significant (p-value = 0.2748). Furthermore, the R^2 value of 0.0105 indicates that the Environmental Index explains only a small portion of the variation in Scope 1 carbon intensity.

Environmental Metric	$\hat{\beta}$	Std. Err.	P-value	R^2	Adj. R^2
Environmental Index	-0.0812	0.0740	0.2748	0.0105	-0.0070

Table 5.2: FE-regression results for the EnvironmentalBERT Classifier

In conclusion, while the Environmental Index we created using the EnvironmentalBERT classifier provides a quantitative metric for environmental communication, it does not show a significant relationship with the Scope 1 carbon intensity in this dataset.

5.1.1.3 FE-Regression Results for the Specialized BERT Classifiers

When it comes to the environmental metrics based on the specialized BERT classifiers, only one of these have a significant coefficient: The proportion of renewable energy sentences, classified by *climatebert/renewable*, has a p-value of 0.0206. As this is lower than 0.05, this indicates that the relationship between the proportion of renewable energy sentences and scope 1 carbon emission intensity is statistically significant on a 95% confidence-level. The relationship is as follows:

$$CI_{it} = 0.0325 \cdot renewable\ share_{i,t-1} + \mu_i + v_{it} \tag{5.1}$$

where CI_{it} is the scope 1 carbon emission intensity for company i in year t, $\hat{\beta} = 0.0325$, the share of renewable sentences for company i in year *t-1* is the explanatory variable, μ_i is the fixed effects of company i, and v_{it} is the noise component. The fixed effects of the model are shown in appendix A.

The way to interpret this relationship is that an increase of 1 percentage point in the proportion of renewable energy sentences in reports is associated with an increase in scope 1 carbon emission intensity of 0.0325. The R-squared is 0.0465, and adjusted R-squared is -0.2741, suggesting that the variation in the renewable energy-metric can explain 4.65% of the variation within groups (entity or time period) in the scope 1 carbon intensity. This is a modest fit, indicating that other variables, not accounted for, may explain the variation in scope 1 carbon intensity better.

The Fixed Effects regression results for all the environmental metrics based on the specialized BERT models are showcased in table 5.3.

Environmental Metric	**Threshold**	$\hat{\beta}$	**Std. Err.**	**P-value**	$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$
Emissions Reduction Share	0.5	-0.0003	0.0025	0.9110	0.0001	-0.3361
Environmental Claims Share	0.5	-0.0062	0.0105	0.5586	0.0030	-0.3322
Water-Related Share	0.5	0.0028	0.0223	0.8992	0.0001	-0.3361
Renewable Energy Share	0.5	0.0325	0.0138	0.0206	0.0465	-0.2741

Table 5.3: FE-regression results for the specialized BERT Classifiers

5.1.2 Random Effects Regression Results

5.1.2.1 RE-Regression Results for the Zero-Shot Classifier

None of the measured environmental metrics based on the Zero-Shot Classifier have a statistically significant relationship with scope 1 carbon emission intensity when performing random effects regressions, as their p-values are well above 0.05. The share of sentences related to water is the metric with the lowest p-value for its coefficient, with a p-value equal to 0.3653, but the relationship is not statistically significant. Table 5.4 shows the RE-regression results for the lowest p-values across all thresholds for the different environmental metrics.

Environmental Metric	**Threshold**	$\hat{\beta}$	**Std. Err.**	**P-value**	**Intercept**	**Int. P-value**	$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$
Environmental Share	0.8	0.0009	0.0037	0.8051	1.2350	3.84E-17	0.0004	-0.0063
Carbon Emissions Share	0.6	0.0041	0.0101	0.6821	1.2370	3.94E-21	0.0011	-0.0055
Pollution Share	0.5	0.0034	0.0083	0.6803	1.2300	4.43E-19	0.0011	-0.0055
Energy Consumption Share	0.1	0.0033	0.0041	0.4141	1.1850	5.25E-15	0.0044	-0.0022
Water Usage Share	0.2	0.0103	0.0114	0.3653	1.2020	2.97E-20	0.0054	-0.0012
Waste Management Share	0.4	0.0029	0.0082	0.7209	1.2354	1.78E-19	0.0009	-0.0058
Renewable Energy Share	0.6	-0.0059	0.0100	0.5543	1.2882	8.05E-21	0.0023	-0.0043

Table 5.4: RE-regression results for the Zero-Shot Classifier

5.1.2.2 RE-Regression Results for the EnvironmentalBERT Classifier

The Random Effects (RE) regression results for the EnvironmentalBERT classifier are summarized in Table 5.5. The coefficient ($\hat{\beta}$) for the Environmental Index is -0.0733, with a p-value of 0.3215, indicating that the relationship between the Environmental Index and Scope 1 carbon intensity is not statistically significant. The R^2 value is 0.0066, showing that the Environmental Index explains only a very small proportion of the variance in Scope 1 carbon intensity.

These findings are consistent with the results from the Fixed Effects regression, where no significant relationship was observed between the Environmental Index and Scope 1 carbon intensity.

Environmental Metric	$\hat{\beta}$	**Std. Err.**	**P-value**	$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$
Environmental Index	-0.0733	0.0737	0.3215	0.0066	-0.0054

Table 5.5: RE-regression results for the EnvironmentalBERT Classifier

5.1.2.3 RE-Regression Results for the Specialized BERT Classifiers

The proportion of sentences related to renewable energy, classified by *climatebert/renewble*, has a p-value below 0.05 (0.0104) also in the Random Effects regression, while none of the other metrics are significant. This suggests that the relationship between the proportion of renewable energy sentences and scope 1 carbon intensity is statistically significant. The relationship is illustrated in 5.2:

$$CI_{it} = 1.1271 + 0.0348 \cdot renewable\ share_{i,t-1} + \mu_i + v_{it} \tag{5.2}$$

where CI_{it} is the scope 1 carbon emission intensity for company i in year t, the overall intercept $\alpha = 1.1271$, $\hat{\beta} = 0.0348$, the share of renewable sentences for company i in year *t-1* is the explanatory variable, μ_i is the random effects of company i, and v_{it} is the noise component.

This relationship indicates that an increase of 1 percentage point in the proportion of renewable energy sentences in reports is associated with an increase in scope 1 carbon intensity equal to 0.0348, all else equal. R-squared equals 0.0420, and adjusted R-squared equals 0.0356. Thus, this suggests that the total variation in the renewable energy-metric can explain 4.2% of the variation in the scope 1 carbon intensity. This is a modest explanatory power, indicating that there are other variables that have not been accounted for that may explain the variation in scope 1 carbon intensity better.

The proportion of water-related sentences has the second lowest p-value (0.2251). The rest of the metrics based on the specialized BERT classifiers are insignificant, and their RE-regression results are showed in table 5.6.

Environmental Metric	**Threshold**	$\hat{\beta}$	**Std. Err.**	**P-value**	**Intercept**	**Int. P-value**	$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$
Emissions Reduction Share	0.5	0.0001	0.0025	0.9574	1.2542	0.0000	0.0000	-0.0066
Environmental Claims Share	0.5	-0.0032	0.0097	0.7436	1.2764	0.0000	0.0007	-0.0059
Water-Related Share	0.5	0.0247	0.0204	0.2251	1.1849	0.0000	0.0097	0.0031
Renewable Energy Share	0.5	0.0348	0.0136	0.0104	1.1271	0.0000	0.0420	0.0356

Table 5.6: RE-regression results for the specialized BERT Classifiers

5.1.3 Model Evaluation

We will now evaluate the regression models where statistically significant relationships have been found for the BERT Classifiers, using different statistical tests and tools. We use the Hausman Test to find out which model is preferred between Fixed Effects and Random Effects, and we also test the model for heteroscedasticity and autocorrelation in the residuals. Finally, we examine the distribution of the residuals.

5.1.3.1 Share of Renewable Energy Sentences

The share of renewable energy sentences based on *climatebert/renewble* had a significant coefficient both with Fixed Effects and with Random Effects, and this will therefore be evaluated.

We begin by performing a Hausman Test (Baltagi, 2014), to check which model we should choose, between the Fixed Effects and Random Effects model. When we perform the Hausman Test, we obtain a p-value of 0.3743, which means that we fail to reject the null-hypothesis that the individual effects are not correlated with the explanatory variable Renewable Energy Share (Amini et al., 2012). Under this assumption, the Random Effects model provides estimates that are more efficient than those of the Fixed Effects model, making Random Effects the preferred model. We therefore choose the Random Effects model. This means that we have a common intercept that is used across companies with a random individual component that accounts for company-specific effects, and this is different from Fixed Effects where each company gets its own intercept in the form of the company-specific effect (Greene, 2019).

After this, we analyze the residuals and thus the noise component of the Random Effects model with share of renewable energy sentences as explanatory variable. First, the model is tested for heteroscedasticity with the Breusch-Pagan Test (Breusch & Pagan, 1979). This gives a p-value of 0.6362, and we cannot reject the null-hypothesis that the residuals are homoscedastic. This suggests that the residuals are homoscedastic, and thus the assumption that the noise component is homoscedastic (Batalgi, 2005) seems to hold.

Then, the model is tested for serial correlation with the Breusch–Godfrey Test for panel models (Breusch, 1978; Godfrey, 1978). The p-value of this test is 6.059e-10, « 0.05, and we therefore reject the null-hypothesis about no autocorrelation in the noise component. This indicates that there is autocorrelation in the noise component which can lead to biased estimates and less efficient results (Drukker, 2003).

To address this issue, one should consider using robust standard errors or alternative model specifications that account for autocorrelation (Batalgi, 2005), to ensure that inference can be done reliably. We therefore use Arellano's (1987) method for robust standard errors on our model. These robust standard errors are consistent in the presence of heteroskedasticity and autocorrelation (Arellano, 1987). Using this method, allows us to observe how the standard errors and the p-values of the coefficient estimates change when accounting for heteroskedasticity and autocorrelation.

When using this method on the model, both the standard error and the p-value of *renewable share*'s coefficient increase. The standard error increases from 0.0136 to 0.0312, and the p-value increases from 0.0104 to 0.2775 as shown in table 5.7. This means that the relationship is no longer statistically significant when accounting for autocorrelation in the residuals.

Type of Std. Err.	$\hat{\beta}$	**Std. Err.**	**P-value**
Regular	0.0348	0.0136	0.0104
Robust (Arellano)	0.0348	0.0312	0.2775

Table 5.7: Standard error and p-value for the Renewable Energy Share before and after using robust standard errors

Next, we examine the assumption about the residuals being independently distributed with a mean of 0 (Batalgi, 2005), using the histogram and QQ-plot shown in figure 5.1:

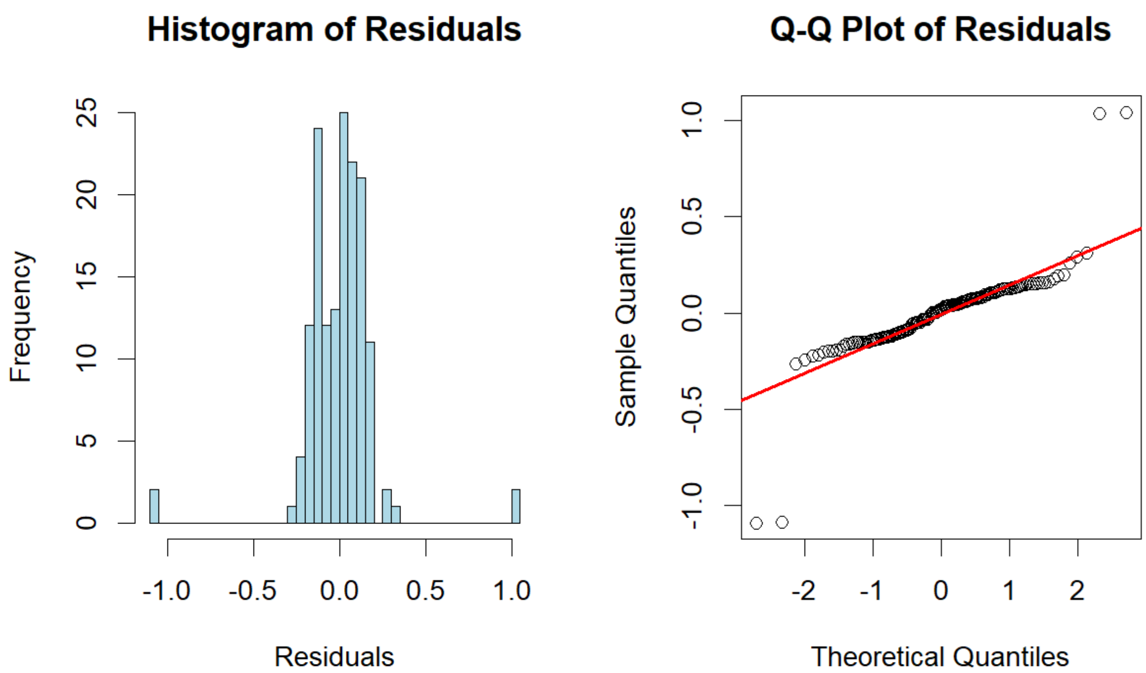

Figure 5.1: Histogram and QQ-plot of Residuals

In the histogram we can see that the residuals seem to be fairly normally distributed around 0, but with a few extreme outliers both in the high and low end. The QQ-plot also indicates that the residuals are normally distributed, as most residuals are distributed closely around the QQ line. However, here we also see four extreme outliers, two in the high end and two in the low end. The residuals thus seem to have a normal distribution with mean around 0, with some slight deviations.

5.2 Sentiment Analysis

5.2.1 Descriptive Analysis of Sentiment

The heatmap in Figure 5.2 provides a visual representation of the sentiment scores across companies and years. The color gradient shows variations in sentiment, with higher scores (red) representing more positive sentiment and lower scores (blue) representing more negative sentiment. From the heatmap, we observe that sentiment scores vary both across companies and over time. Some companies consistently maintain higher sentiment scores, while others fluctuate or trend downward.

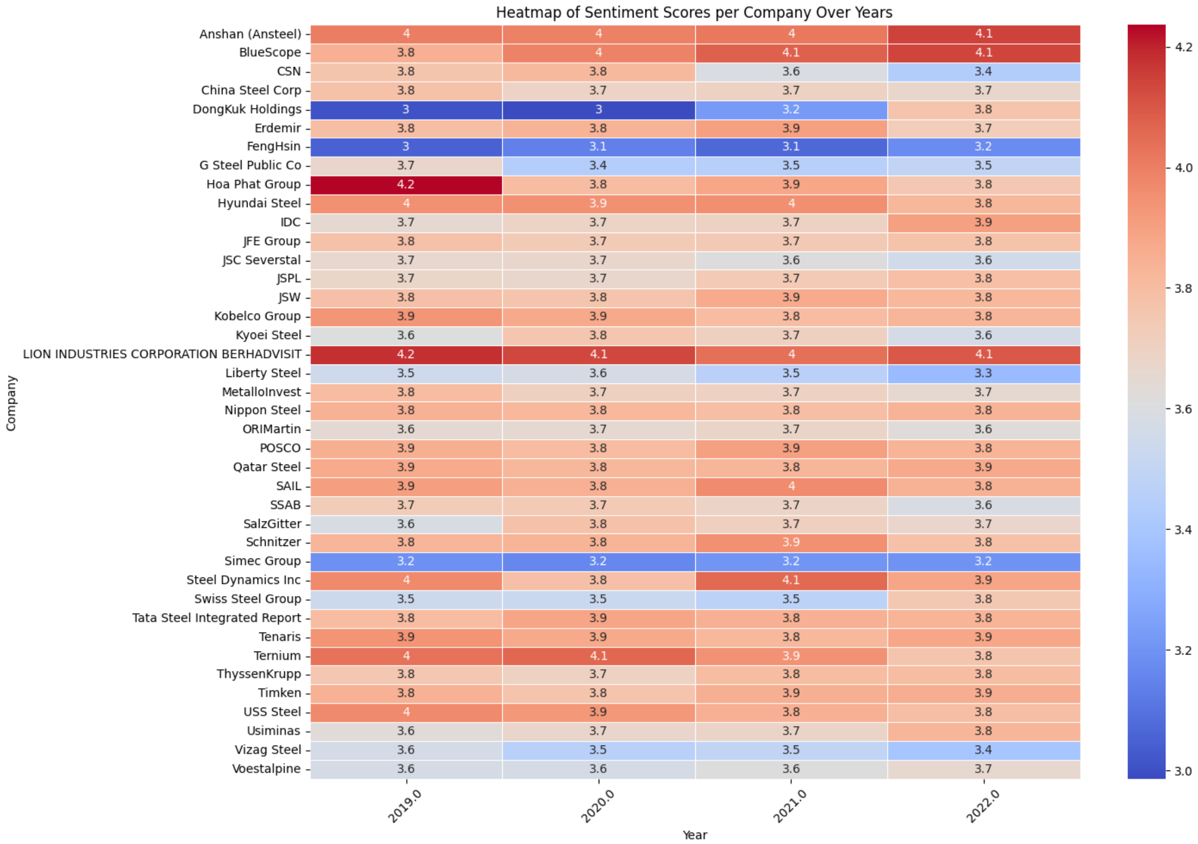

Figure 5.2: Heatmap of Sentiment Scores per Company Over Time

5.2.2 Regression Analysis on Sentiment Score

To investigate the relationship between sentiment scores and Scope 1 carbon intensity, Fixed Effects (FE) and Random Effects (RE) regression models were performed. The results are summarized in Tables 5.8 and 5.9.

For both the FE and RE models, the sentiment score showed a statistically significant negative relationship with Scope 1 carbon intensity. The FE model yielded a coefficient ($\hat{\beta}$) of -0.5348 ($p\text{-}value = 0.0286$), while the RE model produced $\hat{\beta} = -0.4832$ ($p\text{-}value = 0.0361$). These results suggest that higher sentiment scores are associated with lower Scope 1 carbon intensity, indicating a potential link between positive sentiment and improved environmental performance. However, the R^2 values for both models were relatively low, indicating limited explanatory power.

Variable	$\hat{\beta}$	**Std. Err.**	***P-value***	R^2	**[95% CI]**
Intercept	3.2467	0.9068	0.0005	0.0436	[1.4492, 5.0442]
Sentiment Score	-0.5348	0.2411	0.0286	0.0436	[-1.0127, -0.0570]

Table 5.8: FE-regression results for Sentiment Analysis

Variable	$\hat{\beta}$	**Std. Err.**	***P-value***	R^2	**[95% CI]**
Intercept	3.0653	0.8687	0.0006	0.0374	[1.3483, 4.7823]
Sentiment Score	-0.4832	0.2285	0.0361	0.0374	[-0.9348, -0.0317]

Table 5.9: RE-regression results for Sentiment Analysis

The regression analysis highlights a significant negative relationship between sentiment scores and Scope 1 carbon intensity, suggesting that companies with more positive sentiment scores tend to have a lower carbon intensity.

5.3 Summary of Results for Unstratified Analysis

5.3.1 Zero-Shot Classifer

No significant relationships have been found between any of the Zero-Shot Classifier-based metrics and scope 1 carbon intensity, for either the Fixed Effects regression or the Random Effects regression.

5.3.2 EnvironmentalBERT Classifier

No significant relationship was found between the environmental index created using the classified sentences from the EnvironmentalBERT classifer when regressing on the entire dataset.

5.3.3 Specialized BERT Classifiers

For the specialized BERT models all the computed environmental metrics are insignificant, except for the renewable energy metric: The proportion of renewable energy sentences, based on the *climatebert/renewable* classifier, has a statistical positive relationship with scope 1 carbon emission intensity in both the Fixed Effects and Random Effects regressions. Based on the Hausman Test, we find that the Random Effects model is most suitable. The positive relationship suggests that the more a company talks about renewable energy in their reports, the higher are their scope 1 emission intensity.

The test results confirm that the assumption of homoscedasticity is satisfied, but there is significant autocorrelation in the residuals. The residuals are fairly normally distributed with a mean around zero, with few extreme outliers are present. We use robust standard errors to account for the autocorrelation, and find that the coefficient is no longer statistically signficant. This suggests that there is no significant relationship between the share of sentences related to renewable energy and *CI*. Therefore, the statistical significance of the relationship is ambiguous, and inference should be done with caution.

5.3.4 Sentiment Analysis

Sentiment analysis revealed a significant negative relationship between sentiment scores and Scope 1 carbon intensity. Both Fixed Effects and Random Effects models demonstrated that companies with higher sentiment scores had lower carbon intensity, suggesting a possible link between positive sentiment and better environmental performance.

6 Analysis - Stratified

In this section, we stratify the dataset to investigate specific subgroups of companies. By breaking the data into smaller segments, we try to uncover potential relationships that may not appear in the unstratified analysis.

6.1 Report Type Analysis

6.1.1 Introduction to report type analysis

The type of report (ESG or annual/integrated) may affect how companies disclose their environmental performance. Companies with higher Scope 1 emissions might prefer ESG reports to highlight their environmental efforts. In this section, we compare Scope 1 carbon intensity between report types and analyze the relationship further using Fixed Effects and Random Effects models for each group.

6.1.2 Statistical Difference in Scope 1 Carbon Intensity Between Report Types

The Figure 3.3 from the data section of the thesis illustrates the average Scope 1 carbon intensity over time, separated by report type. From a visual inspection, it looks like the companies that publish ESG reports have consistently higher carbon intensity in their production. We will now use statistical tools to see if that is the case.

To formalize the observed differences, two statistical tests were performed: a t-test to compare the means of Scope 1 carbon intensity between report types and an OLS regression to further quantify the relationship.

6.1.2.1 T-Test Results:

A t-test revealed a statistically significant difference in Scope 1 carbon intensity between the two report types ($t\text{-}statistic = 2.2626$, $p\text{-}value = 0.0252$). This indicates that companies using ESG reports tend to have higher carbon intensity than those relying on annual reports.

6.1.2.2 OLS Regression Results:

An OLS regression model was also used to look at the relationship between report type and Scope 1 carbon intensity. The results are summarized in Table 6.1. The negative coefficient for the report type variable (-0.2854) confirms that annual report users have, on average, significantly lower carbon intensity compared to ESG report users ($p\text{-}value =$ 0.0243).

Variable	$\hat{\beta}$	**Std. Err.**	**t-statistic**	**P-value**	**[95% CI]**
Intercept	1.6747	0.193	8.661	0.000	[1.293, 2.057]
Report Type	-0.2854	0.125	-2.276	0.024	[-0.533, -0.038]

Table 6.1: OLS Regression Results: Scope 1 Carbon Intensity by Report Type

6.1.3 Panel Data Regression Results by Report Type

To explore the relationship between the Environmental Index and Scope 1 carbon intensity across different report types, Fixed Effects (FE) and Random Effects (RE) models were run separately for companies using ESG reports and those using annual or integrated reports. The results are presented below.

6.1.3.1 Fixed Effects Regression Results

The Fixed Effects regression results indicate that the Environmental Index does not have a statistically significant relationship with Scope 1 carbon intensity for either report type. For companies using ESG reports, the coefficient for the Environmental Index is $\hat{\beta} = -0.1648$ with a p-value of 0.1888, as shown in Table 6.2. Similarly, for companies using annual reports, the coefficient is $\hat{\beta} = -0.0288$ with a p-value of 0.7664, providing no evidence of a significant relationship.

Report Type	$\hat{\beta}$	**P-value**	**Std. Err.**	R^2	**[95% CI]**
ESG (FE)	-0.1648	0.1888	0.1240	0.0291	[-0.4129, 0.0833]
Annual (FE)	-0.0288	0.7664	0.0964	0.0018	[-0.2225, 0.1649]

Table 6.2: FE-regression results for ESG and Annual Reports

6.1.3.2 Random Effects Regression Results

The Random Effects regression results also show no statistically significant relationship between the Environmental Index and Scope 1 carbon intensity for either report type. For ESG reports, the coefficient is $\hat{\beta} = -0.1641$ with a p-value of 0.1854. For annual reports, the coefficient is $\hat{\beta} = -0.0249$ with a p-value of 0.7962. These results are summarized in Table 6.3.

Report Type	$\hat{\beta}$	**P-value**	**Std. Err.**	R^2	**[95% CI]**
ESG (RE)	-0.1641	0.1854	0.1228	0.0224	[-0.4087, 0.0804]
Annual (RE)	-0.0249	0.7962	0.0961	0.0010	[-0.2168, 0.1670]

Table 6.3: RE-regression results for ESG and Annual Reports

6.1.4 Sentiment Analysis of Report Types

In this section, we look at the relationship between sentiment scores and CI for companies that publish ESG reports and Annual/Integrated reports. The purpose is to see if the sentiment in different report types is linked to carbon intensity.

6.1.4.1 Random Effects and Fixed Effects Regression Results

We used both Random Effects (RE) and Fixed Effects (FE) regression models to study the two report types. Table 6.4 presents the results for ESG reports, while Table 6.5 shows the results for Annual/Integrated reports.

Report Type	$\hat{\beta}$	**P-value**	**Std. Err.**	R^2	**[95% CI]**
ESG (RE)	0.0748	0.3887	0.0863	0.0259	[-0.0970, 0.2466]
ESG (FE)	0.0832	0.3378	0.0861	0.0161	[-0.0892, 0.2556]

Table 6.4: Sentiment Analysis Results for ESG Reports

The results for ESG reports show no significant relationship between sentiment scores and carbon intensity. The coefficients for both models are positive but not statistically significant. This means that sentiment in ESG reports does not appear to explain differences in carbon intensity.

Report Type	$\hat{\beta}$	P-value	Std. Err.	R^2	[95% CI]
Annual/Integrated (RE)	-0.6895	0.0898	0.4005	0.0430	[-1.4891, 0.1101]
Annual/Integrated (FE)	-0.9130	0.0411	0.4355	0.0808	[-1.7876, -0.0383]

Table 6.5: Sentiment Analysis Results for Annual/Integrated Reports

For Annual/Integrated reports, the Fixed Effects model shows a significant negative relationship between sentiment scores and Scope 1 carbon intensity ($\hat{\beta} = -0.9130$, $p = 0.0411$). This result suggests that companies with more positive sentiment in their Annual/Integrated reports tend to have lower carbon intensity. The Random Effects model gives similar results but with slightly weaker significance.

6.1.5 Summary of the Report Type Analysis

Statistical tests revealed that companies publishing standalone ESG reports have significantly higher carbon intensity compared to those that rely on annual or integrated reports. Fixed Effects and Random Effects models were applied separately to analyze the relationship between the Environmental Index and carbon intensity, but no statistically significant relationship was found in either group. In contrast, sentiment analysis showed that sentiment in Annual/Integrated reports has a significant negative relationship with carbon intensity, while sentiment in ESG reports does not show any significant relationship.

6.2 High vs. Low Emitters

6.2.1 Introduction to High vs. Low Emitters Analysis

From the descriptive analysis, we observed a wide range of Scope 1 carbon emissions between companies. In this section, we use statistical thresholding with the K-means clustering algorithm to define thresholds for separating companies into low and high emitters. Based on these thresholds, we analyze differences in their environmental communication and its relationship to Scope 1 carbon intensity.

6.2.2 Determining the Thresholds Using K-Means Clustering

To find the thresholds for analyzing low vs. high emitters, we must first use K-Means Clustering to group the companies based on scope 1 emissions. To determine the optimal number of clusters for classifying companies into emission groups, we applied the elbow method to the standardized average Scope 1 carbon intensity values. The elbow diagram (Figure 6.1) indicates that the inertia, which measures the sum of squared distances to the nearest cluster center, decreases sharply up to k=3. After this point, the reduction in inertia becomes minimal, suggesting that k = 3 is the optimal choice for clustering.

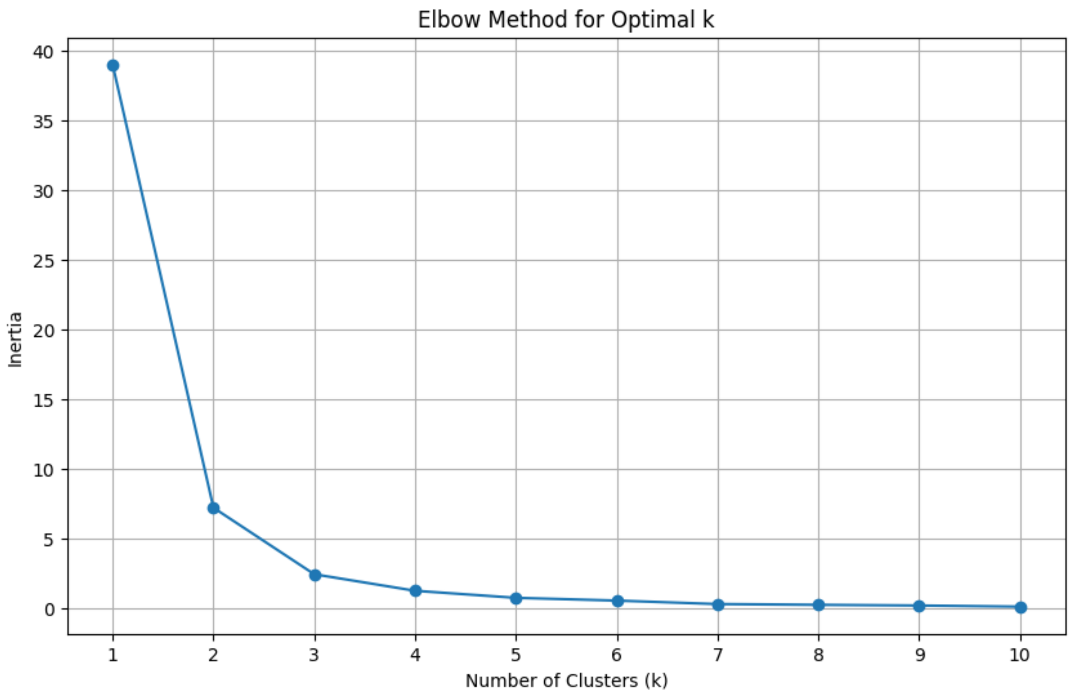

Figure 6.1: Elbow method to determine optimal k

Using k=3, the K-means algorithm grouped the companies into three distinct clusters, representing low, medium, and high emitters. Figure 6.2 shows the clustering results, with each company's average Scope 1 carbon intensity assigned to a specific emission group.

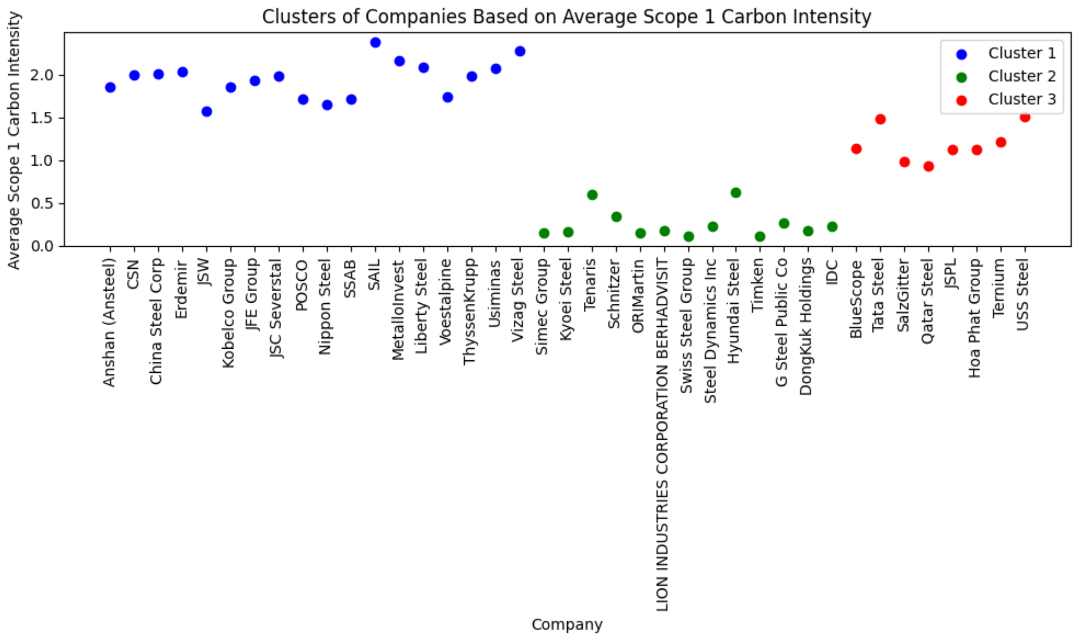

Figure 6.2: Clustering of companies based on carbon intensity

Based on the clustering results, we identify two thresholds at approximately 0.5 and 1.5 for average Scope 1 carbon intensity. These thresholds will be used for the regression analyses to explore the differences between emission groups.

6.2.3 Regression Results With Thresholds

Based on the two thresholds derived from the cluster centroids, we performed regressions twice. First, the companies were split into two groups: those with average Scope 1 carbon intensity below 0.5 (low emitters) and those above 0.5. In the second analysis, the split was made at 1.5, dividing the companies into medium or lower emitters (below 1.5) and high emitters (above 1.5).

6.2.3.1 Threshold at 0.5

For companies with Scope 1 carbon intensity below 0.5 (Low Emitters), both Fixed Effects (FE) and Random Effects (RE) regression models revealed a significant negative relationship between the Environmental Index and Scope 1 carbon intensity. The FE model produced a coefficient ($\hat{\beta}$) of -0.0568 (p-value = 0.0021), while the RE model gave $\hat{\beta} = -0.0540$ (p-value = 0.0030). These results suggest that higher Environmental Index values are associated with lower carbon intensity for low emitters. The R-squared values were 0.2525 for FE and 0.1492 for RE, indicating moderate explanatory power.

For companies with Scope 1 carbon intensity above 0.5 (High Emitters), no significant relationship was observed. The FE model yielded $\hat{\beta} = -0.1081$ (p-value = 0.4068), and the RE model produced $\hat{\beta} = -0.1028$ (p-value = 0.4168).

Group	$\hat{\beta}$	**P-value**	**Std. Err.**	R^2	**[95% CI]**
Low Emitters (FE)	-0.0568	0.0021	0.0170	0.2525	[-0.0914, -0.0222]
Low Emitters (RE)	-0.0540	0.0030	0.0172	0.1492	[-0.0886, -0.0193]
High Emitters (FE)	-0.1081	0.4068	0.1296	0.0085	[-0.3659, 0.1498]
High Emitters (RE)	-0.1028	0.4168	0.1261	0.0343	[-0.3527, 0.1472]

Table 6.6: Regression Results for Threshold 0.5

6.2.3.2 Threshold at 1.5

When applying the 1.5 threshold, no significant relationships between the Environmental Index and Scope 1 carbon intensity were found for either category of emitters. For Low Emitters (below 1.5), the FE model gave $\hat{\beta} = -0.0453$ (p-value = 0.4044) and the RE model $\hat{\beta} = -0.0430$ (p-value = 0.4241). For High Emitters (above 1.5), the FE model produced $\hat{\beta} = -0.1140$ (p-value = 0.4620), and the RE model $\hat{\beta} = -0.1194$ (p-value = 0.3885).

Group	$\hat{\beta}$	**P-value**	**Std. Err.**	R^2	**[95% CI]**
Low Emitters (FE)	-0.0453	0.4044	0.0539	0.0122	[-0.1533, 0.0627]
Low Emitters (RE)	-0.0430	0.4241	0.0535	0.0231	[-0.1496, 0.0636]
High Emitters (FE)	-0.1140	0.4620	0.1538	0.0099	[-0.4223, 0.1944]
High Emitters (RE)	-0.1194	0.3885	0.1377	0.1572	[-0.3938, 0.1549]

Table 6.7: Regression Results for Threshold 1.5

6.2.4 Sentiment Analysis of High vs. Low Emitters

Here we want to look at whether sentiment scores are correlated with carbon intensity when companies are grouped based on their CI levels.

6.2.4.1 Threshold at 0.5

Group	$\hat{\beta}$	P-value	Std. Err.	R^2	[95% CI]
Low Emitters (RE)	-0.0048	0.7550	0.0153	-0.0448	[-0.0357, 0.0261]
Low Emitters (FE)	-0.0043	0.7835	0.0156	0.0026	[-0.0362, 0.0275]
High Emitters (RE)	0.0260	0.7327	0.0758	0.0574	[-0.1244, 0.1763]
High Emitters (FE)	0.0459	0.5408	0.0748	0.0049	[-0.1030, 0.1948]

Table 6.8: Sentiment Analysis Results for Emission Threshold 0.5

For the 0.5 threshold, no significant relationships were found between sentiment scores and Scope 1 carbon intensity for either group.

6.2.4.2 Threshold at 1.5

Group	$\hat{\beta}$	P-value	Std. Err.	R^2	[95% CI]
Low Emitters (RE)	-0.0445	0.2733	0.0403	0.0273	[-0.1248, 0.0359]
Low Emitters (FE)	-0.0481	0.2386	0.0403	0.0266	[-0.1291, 0.0329]
High Emitters (RE)	0.0327	0.7090	0.0873	0.2020	[-0.1414, 0.2068]
High Emitters (FE)	0.0653	0.4680	0.0893	0.0100	[-0.1138, 0.2444]

Table 6.9: Sentiment Analysis Results for Emission Threshold 1.5

The results for the 1.5 threshold also show no significant relationships between sentiment scores and carbon intensity.

6.2.5 Summary of High vs. Low Emitters

The results indicate a significant negative relationship between the Environmental Index and carbon intensity for low emitters (below 0.5), but no significant relationships were observed for higher emitters (above 0.5 or 1.5). This suggests that the Environmental Index has more explanatory power for companies with lower carbon intensity. In contrast, sentiment scores do not show any significant relationship with carbon intensity for either threshold (0.5 or 1.5), indicating that sentiment does not explain variations in CI among low or high emitters.

6.3 Summary of Results for Stratified Analysis

6.3.1 Report Type Analysis

The analysis revealed a statistically significant difference in carbon intensity (CI) between companies that publish ESG reports and those that publish annual or integrated reports. Companies using ESG reports have, on average, a CI approximately 0.3 higher than those relying on annual or integrated reports.

Fixed Effects and Random Effects regression models were applied separately to each report type to examine the relationship between the Environmental Index and CI. No statistically significant relationships were found for either group. However, sentiment analysis revealed an important difference: while sentiment scores in ESG reports did not show any significant relationship with CI, sentiment in annual or integrated reports exhibited a significant negative relationship with CI. This suggests that companies with more positive sentiment in their annual or integrated reports tend to have lower carbon intensity.

6.3.2 High vs. Low Emitters

Companies were further stratified based on emission thresholds (0.5 and 1.5) identified through K-means clustering. For low emitters (below a CI of 0.5), both Fixed Effects and Random Effects regression models revealed a significant negative relationship between the Environmental Index and CI. This result indicates that higher Environmental Index scores are associated with lower carbon intensity for low emitters. However, no significant relationships were observed for medium or high emitters (above 0.5 or 1.5).

When analyzing sentiment scores across the same thresholds, no significant relationships were found between sentiment and carbon intensity for either low or high emitters. These results suggest that while the Environmental Index can explain variations in CI among low emitters, sentiment does not provide meaningful insights into carbon intensity within any of the emission groups.

7 Discussion

7.1 Discussion of BERT Models - Unstratified

In section 5 we found that there were no significant relationships between any of the Zero-Shot Classifier-based metrics and carbon intensity, in either the Fixed Effects regression or the Random Effects Regression. This suggests that none of the pre-defined classes are related to the carbon intensity metric.

An implication of this is that there could be a disconnect between environmental communication in corporate reports and actual environmental performance in the steel industry, and that corporate reports may not be reliable sources for steel companies' environmental performance. Based on this, ESG communications might require better reporting standards.

However, another explanation for these insignificant relationships could be that, unlike the specialized BERT models and EnvironmentalBERT, the Zero-Shot Classifier is not pre-trained on ESG data, nor fine-tuned for classification tasks, thereby making it worse at accurately detecting and categorizing specific ESG-related topics in the text.

EnvironmentalBERT, on the other hand, was designed specifically for ESG contexts and demonstrated greater relevance. Although the Environmental Index created from this model showed no significant overall relationship with carbon intensity in the regressions, it captured environmental themes more effectively than the Zero-Shot Classifier. This highlights the importance of using ESG-specific models when analyzing corporate reports. EnvironmentalBERT's specialized training may provide insights that general classifiers fail to detect, although this also suggests that alternative or more advanced metrics are needed to bridge the gap between textual analysis and actual emissions data.

In regards to specialized models, the *Renewable Energy Share*, based on the specialized BERT model *climatebert/renewable*, turned out to be a notable addition. This metric showed a positive statistically significant relationship with carbon intensity in both the Fixed Effects and Random Effects regressions. The Hausman test indicated that the Random Effects model was the preferred one. However, inference with this model should be taken with caution, as we observed that the relationship was no longer statistically significant when accounting for autocorrelation with robust standard errors.

Nonetheless, we interpret the relationship, and the positive coefficient suggests that companies with more talk about renewable energy in their reports tend to have higher emissions. Seen in the light of legitimacy theory, a reason for this can be that companies want to defend their legitimacy, or reputation, in the eyes of the public (Şeker & Şengür, 2021). In this case, companies with high emissions may feel the need to communicate their use of, or investment in, renewable energy to compensate for their high emissions, and thus maintain their reputation. As the statistical significance of this relationship is ambiguous, we recommend investigating it more closely in future research.

For the other specialized BERT-based metrics, the proportion of water-related sentences has the second lowest, but not significant, p-value (= 0.2251) in the Random Effects regression, and a positive coefficient (0.0247). Interestingly, water-related sentences obtained the lowest, but not significant, p-value (= 0.3653) in the Random Effects regression with the Zero-Shot Classifier, with a positive coefficient (0.0103). Although these relationships are not significant, they suggest a positive relationship between the proportion of water-related sentences and carbon intensity. This could be the result of noise in the text data, but we recommend examining it further in future research.

No significant relationship was found between the share of environmental claims in companies' reports and their scope 1 carbon intensity, nor between their share of emission reductions-related sentences and carbon intensity. An implication of this can be that there is no connection between what companies say they do for the environment and what they actually do. In this case, there may also be a need for better reporting standards.

It should be noted, though, that our proportion-based metrics are relatively simple, as only quantity is measured, and thus might not be the best at capturing the nuances in environmental communication in reports. Additionally, it could be that there are better methods to extract text from PDF files compared to what we have used, and we recommend exploring alternative methods in future research. Also, as we saw in the Model Evaluation section (5.1.3), the noise component seems to be autocorrelated in the Random Effects model with renewable energy share, based on climatebert/renewble, as the explanatory variable. If this also applies for the other models, their results might be biased.

7.2 Discussion of Report Types

The report type analysis showed that companies publishing ESG reports tend to have a higher carbon intensity compared to those relying on annual or integrated reports. This difference, seen both in descriptive statistics and statistical tests, may occur because companies with higher emissions favour ESG reports to comprehensively address and potentially justify their environmental impact. These companies might use detailed ESG reports strategically to mitigate stakeholder concerns, aligning with the concept of greenwashing (Treepongkaruna et al., 2024). In addition, the stratified sentiment analysis shows that the tone of ESG reports is generally more positive than that of the annual reports, further highlighting that the use of optimism in communication could be a strategy for reputation management rather than actual environmental action.

Despite these differences, no significant relationship was found between the Environmental Index and carbon intensity for either report type. This suggests that the type of report does not affect how environmental communication is related to actual emissions.

7.3 Discussion of High vs. Low Emitters

Our analysis based on thresholds for high and low emitters showed that the Environmental Index had a statistically significant negative relationship with carbon intensity for low emitters (below the 0.5 threshold). This indicates that companies with lower emissions are more likely to align their environmental communication with actual environmental performance. This could suggest that companies that are already on track with reducing their emissions are more likely to communicate genuine initiatives, where the environmental communication reflects actual and tangible actions rather than just rhetoric.

For high emitters, the lack of significance suggests that their environmental communication does not appear to be linked to their actual emissions. This implies that whether a company writes extensively about ESG topics or says very little, there is no clear indication of their emission levels either way. Instead, the environmental communication may be influenced by external expectations, such as political or stakeholder demands, or efforts to protect their legitimacy. This finding also seems to be supported in the literature, especially if the company is large (Lagasio, 2023).

7.4 Discussion of Sentiment Analysis

The sentiment analysis showed a statistically significant negative relationship between sentiment scores and carbon intensity. Companies with more positive language in their environmental communication often had lower carbon intensity, while those with more negative tones tended to have higher emissions. This could mean that low-emission companies use positive language to highlight their achievements, while high-emissions companies acknowledge their challenges or environmental issues more openly. The negative sentiment might come from acknowledging environmental challenges, such as high emissions, accidents, or operational difficulties.

In general, corporate reports use positive language, which aligns with the goal of maintaining investor confidence and avoid creating concerns among regulators and stakeholders. Using negative language could damage relationships or damage the company's reputation. However, sentiment analysis has limitations. While the method captures general tone, it may not fully reflect the complexity or strategic intent behind the communication. For example, positive sentiment could be used as a strategic tool, regardless of the company's actual environmental performance.

When stratified by report type, the analysis revealed that sentiment in annual/integrated reports showed a significant negative relationship with carbon intensity, whereas no such relationship was observed in ESG reports. This could suggest that the tone of annual/integrated reports may better reflect actual environmental performance compared to ESG reports, where communication might focus on managing perceptions rather than giving a realistic portrayal of progress. In stratified analyses by emission level, no significant relationships were found, indicating that sentiment scores do not effectively distinguish environmental communication between low and high emitters.

8 Conclusion

This thesis studies the relationship between environmental communication in ESG reports and the carbon intensity of production in the steel industry. The main goal was to see if these reports reflect actual environmental performance, specifically regarding Scope 1 carbon intensity. The study was motivated by concerns about greenwashing, where companies may mislead stakeholders by presenting a better image of their environmental efforts than reality.

Our results show that the Environmental Index we created does not co-move with carbon intensity. There are three possible reasons for this. First, the index might not be good enough to capture the full meaning of environmental communication. Second, companies might not give an accurate picture of their environmental performance in their ESG reports, intentionally or unintentionally. Third, there could be problems with the carbon intensity data from Steelstat, which we used in our analysis. These results suggest that self-published ESG information may be more useful as tools for stakeholder communication than as reliable indicators of environmental outcomes.

While the Environmental Index did not show a strong connection to carbon intensity, the sentiment analysis provided some interesting findings. In annual and integrated reports, a more positive tone in the language was linked to lower carbon intensity. However, this pattern was not found in standalone ESG reports. We also found that companies publishing ESG reports had much higher carbon intensity than those using only annual or integrated reports. This could mean that ESG reports are sometimes only used to highlight positive efforts, even though their actual emission statistics are less favorable.

In addition to these findings, we also looked at more specific environmental subtopics. Here we found that companies that have more emphasis on renewable energy in their reports tend to have higher carbon intensity. This can be a result of companies with high emissions compensating for this by strategically overemphasizing their communication about renewable energy to maintain their reputation.

Our findings are important for different stakeholders who rely on ESG reports. The lack of connection between the Environmental Index and carbon intensity suggests that ESG reports may not always provide an accurate picture of the environmental performance of a company. This means that investors, policymakers, and other stakeholders should be careful when using these reports, highlighting the importance of having more standardized reporting frameworks that better align disclosed information with a company's actual environmental profile.

With this in mind, we recommend considering other sources of data, such as verified emissions figures in addition to self-reported ESG information when making decisions. Future research could address this by incorporating multilingual NLP models to capture a broader dataset. Further methodological advancements, such as incorporating hybrid models that capture more than just the environmental communication ratio, may also enhance the robustness of an analysis.

This study has some limitations. We only used reports in English, which left out data from companies in non-English-speaking regions. The Steelstat data we relied on might also have inaccuracies. Additionally, the lack of standardization in ESG reporting introduces variability in how environmental topics are addressed, complicating comparisons across companies. The exclusion of companies that do not report ESG data also limits the representativeness of the findings, potentially skewing the sample toward firms with stronger reporting practices. Finally, the regional imbalance, particularly the underrepresentation of high-emission regions such as China, reduces the generalizability of the conclusions to the global steel industry.

In summary, this thesis shows that ESG reports and the Environmental Index do not align with actual carbon intensity in the steel industry. However, the sentiment analysis shows a potential link between language tone and emissions, especially in annual and integrated reports. Additionally, companies highlighting renewable energy in reports tend to have higher carbon intensity. These findings show the importance of having more standardized reporting frameworks to ensure that disclosures are better matched with real environmental performance.

References

Adams, C. A., & Abhayawansa, S. (2022). Connecting the covid-19 pandemic, environmental, social and governance (esg) investing and calls for 'harmonisation' of sustainability reporting [Special Issue: Covid and the Environment in Crisis]. *Critical Perspectives on Accounting*, *82*, 102309. https://doi.org/https://doi.org/10.1016/j.cpa.2021.102309

Amini, S., Delgado, M. S., Henderson, D. J., & Parmeter, C. F. (2012, September). Fixed vs random: The hausman test four decades later. In B. H. Baltagi, R. Carter Hill, W. K. Newey, & H. L. White (Eds.), *Essays in honor of jerry hausman* (pp. 479–513, Vol. 29). Emerald Group Publishing Limited. https://doi.org/10.1108/S0731-9053(2012)0000029021

Anthesis. (2024). What Are Scope 1, 2 And 3 Emissions? | Anthesis. Retrieved November 30, 2024, from https://www.anthesisgroup.com/solutions/net-zero-decarbonisation/scope123/

Antto Jokinen. (2023, April). Genuine green steel or greenwashing? Common labeling rules are needed to benefit the climate and business. Retrieved November 30, 2024, from https://wwf.fi/en/uutiset/2023/04/genuine-green-steel-or-greenwashing-common-labeling-rules-are-needed-to-benefit-the-climate-and-business/

Arellano, M. (1987). Practitioners' corner: Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434. https://doi.org/https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x

AWS. (2024). What is Sentiment Analysis? - Sentiment Analysis Explained - AWS. Retrieved December 19, 2024, from https://aws.amazon.com/what-is/sentiment-analysis/

Baltagi, B. (2014). Panel data and difference-in-differences estimation. In A. J. Culyer (Ed.), *Encyclopedia of health economics* (pp. 425–433). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-375678-7.00720-3

Barkemeyer, R., Comyns, B., Figge, F., & Napolitano, G. (2014). CEO statements in sustainability reports: Substantive information or background noise? [Publisher: Elsevier]. *Accounting forum*, *38*(4), 241–257. Retrieved November 30, 2024, from https://ideas.repec.org//a/eee/accfor/v38y2014i4p241-257.html

Batalgi, B. H. (2005). Econometric analysis of panel data. John Wiley Sons, Ltd.

Bergmann, D. (2024). What is zero-shot learning? [Accessed: 27.11.24]. https://www.ibm.com/topics/zero-shot-learning

Bingler, J. A., Kraus, M., Leippold, M., & Webersinke, N. (2022). Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Research Letters*, *47*, 102776. https://doi.org/10.1016/j.frl.2022.102776

Birand, B. (2024, August). Steel's gambit: AI, transparency, and the battle against 'greenwashing'. Retrieved November 30, 2024, from https://www.smartindustry.com/benefits-of-transformation/environmental-health-safety/article/55129944/steels-gambit-ai-transparency-and-the-battle-against-greenwashing

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. O'Reilly Media Inc.

Breg, D. (2023). More Companies Are Disclosing Their ESG Data, but Confusion on How Persists. *Wall Street Journal*. Retrieved November 30, 2024, from https://www.wsj.com/articles/more-companies-are-disclosing-their-esg-data-but-confusion-on-how-persists-e667698c

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, *17*(31), 334–55. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287–1294. Retrieved December 16, 2024, from http://www.jstor.org/stable/1911963

Brzozowski, M. (2023, March). Fine-tuning BERT model for arbitrarily long texts, Part 1. Retrieved December 19, 2024, from https://www.mim.ai/fine-tuning-bert-model-for-arbitrarily-long-texts-part-1/

Christian Hoffmann, Michel Van Hoey, & Benedikt Zeumer. (2020, March). Decarbonization in steel | McKinsey. Retrieved November 30, 2024, from https://www.mckinsey.com/industries/metals-and-mining/our-insights/decarbonization-challenge-for-steel

Conejo, A. N., Birat, J.-P., & Dutta, A. (2020). A review of the current environmental challenges of the steel industry and its value chain. *Journal of Environmental Management*, *259*, 109782. https://doi.org/https://doi.org/10.1016/j.jenvman.2019.109782

Contreras-Pacheco, O. E., & Claasen, C. (2017, April). *Fuzzy reporting as a way for a company to greenwash: perspectives from the Colombian reality* (MPRA Paper No. 85472). University Library of Munich, Germany. https://ideas.repec.org/p/pra/mprapa/85472.html

Dearnell, A. (2022, December). Are You Ready For An Integrated Annual Report? [Section: Leadership Strategy]. Retrieved December 1, 2024, from https://www.forbes.com/sites/adriandearnell/2022/12/20/are-you-ready-for-an-integrated-annual-report/

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *North American Chapter of the Association for Computational Linguistics*. https://api.semanticscholar.org/CorpusID:52967399

DNV. (2024). Shaping the Future of Sustainable Steel: Download the Report. Retrieved November 30, 2024, from https://www.dnv.com/about/supplychain/shaping-the-future-of-sustainable-steel/

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal*, *3*(2), 168–177. https://journals.sagepub.com/doi/pdf/10.1177/1536867X0300300206

Eda Kavlakoglu & Vanna Winland. (2024, June). What is k-means clustering? | IBM. Retrieved December 2, 2024, from https://www.ibm.com/topics/k-means-clustering

Edmond. (2024, January). The CO2e impact of iron and steel production. Retrieved November 30, 2024, from https://edmondclimate.com/the-co2e-impact-of-iron-and-steel-production/

Ellerbeck, S. (2022, July). What is green steel and why does the world need more of it? [Accessed: 17.12.24]. https://www.weforum.org/stories/2022/07/green-steel-emissions-net-zero/

Feldman, K. (2024, November). Bimodal Distributions and Why They Matter. Retrieved December 18, 2024, from https://www.isixsigma.com/dictionary/bimodal-distribution/

Friedl, J. E. F. (1997). *Mastering regular expressions: Powerful techniques for perl and other tools*. O'Reilly & Associates Inc.

Ganti, A. (2024, August). Central Limit Theorem (CLT): Definition and Key Characteristics. Retrieved December 18, 2024, from https://www.investopedia.com/terms/c/central_limit_theorem.asp

Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-Dependent Sentiment Classification With BERT [Conference Name: IEEE Access]. *IEEE Access*, *7*, 154290–154299. https://doi.org/10.1109/ACCESS.2019.2946594

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, *46*(6), 1293–1301. Retrieved December 16, 2024, from http://www.jstor.org/stable/1913829

Greene, W. H. (2019). *Econometric analysis global edition* (8th ed.). Pearson-prentice Hall.

He, P., Liu, X., Gao, J., & Chen, W. Deberta: Decoding-enhanced bert with disentangled attention. In: In *Iclr 2021 papers*. 2021, October. https://arxiv.org/pdf/2006.03654

Herrera, Z. (2023, September). GSCC: Stressing the 'S' in ESG - Recycling Today. Retrieved November 30, 2024, from https://www.recyclingtoday.com/article/global-steel-climate-council-certification-stressing-the-s-in-esg/

Hugging Face. (2023, September). What is Zero-Shot Classification? - Hugging Face. Retrieved December 19, 2024, from https://huggingface.co/tasks/zero-shot-classification

IBM. (2024). What are llms? [Accessed: 29.11.24]. https://www.ibm.com/topics/large-language-models

IMD. (2023, August). ESG: Environmental, social, & governance investing explained. Retrieved November 30, 2024, from https://www.imd.org/blog/sustainability/esg-environmental-social-and-governance/

International Federation of Accountants. (2024, February). The state of play: Sustainability disclosure and assurance: 2019-2022 trends & analysis [Accessed: 18.12.24]. https://ifacweb.blob.core.windows.net/publicfiles/2024-02/IFAC-State-Play-Sustainability-Disclosure-Assurance-2019-2022_0.pdf

iRIS Carbon. (2023, January). A Beginner's Guide to ESG Rating Agencies and Methodologies. Retrieved November 30, 2024, from https://www.iriscarbon.com/a-beginners-guide-to-esg-rating-agencies-and-methodologies/

Javatpoint. (2024). K-Means Clustering Algorithm - Javatpoint. Retrieved December 19, 2024, from https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

Johnson, L. (2023, November). Greenwashing growing in frequency and complexity: Report. Retrieved November 30, 2024, from https://www.esgdive.com/news/greenwashing-rising-report-rep-risk-social-washing-sustainability/696289/

Jun, Y. (2023, December). All Languages Are NOT Created (Tokenized) Equal. Retrieved December 1, 2024, from https://towardsdatascience.com/all-languages-are-not-created-tokenized-equal-cd87694a97c1

Khazin, B. (2024, July). The Importance of Developing Accurate Automated Reporting Systems for ESG. Retrieved November 30, 2024, from https://www.environmentenergyleader.com/stories/the-importance-of-developing-accurate-automated-reporting-systems-for-esg,44856

Lagasio, V. (2023, September). Measuring Greenwashing: The Greenwashing Severity Index. https://doi.org/10.2139/ssrn.4582917

Lempriere, M. (2023, July). Steel industry makes 'pivotal' shift towards lower-carbon production. Retrieved November 30, 2024, from https://www.carbonbrief.org/steel-industry-makes-pivotal-shift-towards-lower-carbon-production/

Martin Kueppers. (2023, November). Iron & steel. Retrieved November 30, 2024, from https://www.iea.org/energy-system/industry/steel

Merritt, R. (2022). What is a transformer model? [Accessed: 29.11.24]. https://blogs.nvidia.com/blog/what-is-a-transformer-model/

Mona, G. (2024, September). New DNV report highlights path to a leaner, greener steel industry. Retrieved November 30, 2024, from https://www.dnv.com/news/sustainable-steel-report/

Morgan, H. (2022, March). Sustainability reports nlp [Accessed: 01.09.24]. https://github.com/hannahawalsh/HTTF4-ESG-and-NLP/blob/main/CSR_Report_NLP_Walkthrough.ipynb

National Grid. (2024, January). What are scope 1, 2 and 3 carbon emissions? | National Grid Group. Retrieved November 30, 2024, from https://www.nationalgrid.com/stories/energy-explained/what-are-scope-1-2-3-carbon-emissions

Nätt, T. H. (2024). Ocr. *Store Norske Leksikon*. https://snl.no/OCR

NVIDIA. (2023). Bert [Accessed: 29.11.24]. https://www.nvidia.com/en-us/glossary/bert/

Ook Lee, Hanseon Joo, Hayoung Choi, & Minjong Cheon. (2022). (PDF) Proposing an Integrated Approach to Analyzing ESG Data via Machine Learning and Deep Learning Algorithms. *ResearchGate*. https://doi.org/10.3390/su14148745

Paragamian, M., Anagnostou, J., & Schlorke, S. (2021, July). Strengthening sustainability in the steel industry [Accessed: 17.12.24]. https://www.ifc.org/content/dam/ifc/doc/2023/strengthening-sustainability-in-the-steel-industry-ifc-2023.pdf

Paudel, P., Khadka, S., G. C., R., & Shah, R. (2024). Optimizing nepali pdf extraction: A comparative study of parser and ocr technologies.

PricewaterhouseCoopers. (2024, September). ESG reporting and preparation of a Sustainability Report. Retrieved November 30, 2024, from https://www.pwc.com/sk/en/environmental-social-and-corporate-governance-esg/esg-reporting.html

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter. *ArXiv*, *abs/1910.01108*.

Schimanski, T. (2023, November). Analyzing ESG with AI and NLP (Tutorial#1): Report Analysis Towards ESG Risks and Opportunities. Retrieved December 19, 2024, from https://medium.com/@schimanski.tobi/analyzing-esg-with-ai-and-nlp-tutorial-1-report-analysis-towards-esg-risks-and-opportunities-8daa2695f6c5

Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2023). Bridiging the Gap in ESG Measurement: Using NLP to Quantify Environmental, Social, and Governance Communication. *Available on SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract*$_i d = 4622514$.

Şeker, Y., & Şengür, E. D. (2021). The impact of environmental, social, and governance (esg) performance on financial reporting quality: International evidence. *Ekonomika*, *100*(2). https://www.redalyc.org/journal/6922/692272891009/692272891009.pdf

Steelstat. (2024). Methodology. Retrieved December 1, 2024, from https://www.steelstat.com/methodology

Stryker, C., & Holdsworth, J. (2024). What is nlp? [Accessed: 29.11.24]. https://www.ibm.com/topics/natural-language-processing

Sustainable Ships. (2024). What is the carbon footprint of steel? Retrieved December 1, 2024, from https://www.sustainable-ships.org/stories/2022/carbon-footprint-steel

Symanto. (2022, August). ESG Sentiment Analysis: A Valuable Tool for Investors? | Symanto [Section: Blog]. Retrieved December 19, 2024, from https://www.symanto.com/blog/esg-sentiment-analysis-tracking-the-mood-of-stakeholders/

TechnoServe. (2024, August). Understanding the environmental impact of steel plant operations: Causes and solutions [Accessed: 17.12.24]. https://www.technoserveindustries.com/news/Environmental-Impact-of-Steel-Plant-Operations/194

The Apache Software Foundation. (2024). *Apache tika - a content analysis toolkit* [Accessed: 17.12.24]. https://tika.apache.org/

Treepongkaruna, S., Au Yong, H. H., Thomsen, S., & Kyaw, K. (2024). Greenwashing, carbon emission, and ESG [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bse.3929]. *Business Strategy and the Environment*, *33*(8), 8526–8539. https://doi.org/10.1002/bse.3929

Webersinke, N., Kraus, M., Bingler, J., & Leippold, M. (2022). ClimateBERT: A Pretrained Language Model for Climate-Related Text. *Proceedings of AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges*. https://doi.org/https://doi.org/10.48550/arXiv.2212.13631

World Economic Forum. (2023, November). Net-Zero Industry Tracker 2023 Edition. Retrieved November 30, 2024, from https://www.weforum.org/publications/net-zero-industry-tracker-2023/in-full/steel-industry-net-zero-tracker/

World Steel Association. (2022). World Steel in Figures 2022. Retrieved December 19, 2024, from https://worldsteel.org/data/world-steel-in-figures-2022/

Zhuang, L., Wayne, L., Ya, S., & Jun, Z. (2021, August). A robustly optimized BERT pre-training approach with post-training. In S. Li, M. Sun, Y. Liu, H. Wu, K. Liu, W. Che, S. He, & G. Rao (Eds.), *Proceedings of the 20th chinese national conference on computational linguistics* (pp. 1218–1227). Chinese Information Processing Society of China. https://aclanthology.org/2021.ccl-1.108

Appendices

A Fixed Effects: Renewable Energy-Share Regression

Company	Individual Effects
Anshan (Ansteel)	1.7677
BlueScope	1.0280
China Steel Corp	1.8286
CSN	1.9025
Erdemir	1.9425
G Steel Public Co	0.2576
Hoa Phat Group	0.9651
Hyundai Steel	0.4773
IDC	0.1169
JFE Group	1.8065
JSC Severstal	1.8664
JSPL	0.9555
JSW	1.4674
Kobelco Group	1.7087
Kyoei Steel	0.0737
Liberty Steel	1.7442
LION INDUSTRIES CORPORATION BERHADVISIT	0.1317
MetalloInvest	2.0607
Nippon Steel	1.5272
ORIMartin	-0.0081
POSCO	1.6343
Qatar Steel	0.8391
SAIL	2.2262
SalzGitter	0.9593
Schnitzer	0.2535
Simec Group	0.0756
SSAB	1.6322
Steel Dynamics Inc	0.1416
Swiss Steel Group	-0.0559
Tata Steel	1.3952
Tenaris	0.4994
Ternium	1.1163
ThyssenKrupp	1.7981
Timken	-0.0839
Usiminas	1.9609
USS Steel	1.3927
Vizag Steel	2.1809
Voestalpine	1.5699

Figure A.1: Company specific fixed effects from the regression with the explanatory variable Renewable Energy-Share based on climatebert/renewable