Using machine learning to correlate ESG-related news articles to the stock market performance and detect corporate greenwashing

A dissertation submitted in partial fulfilment of the requirements for the MSc in Data Science by Luca Montalto

SCHOOL OF COMPUTING AND MATHEMATICAL SCIENCES

BIRKBECK, UNIVERSITY OF LONDON

September 2024

Academic Declaration

This report is the result of my own work except where explicity indicated in the text. I give my permission for it to be submitted to the TURNITIN Plagiarism Detection Service. I have read and understood the sections on plagiarism in the College website and the Policy and Procedures on academic integrity. The report may be freely copied and distributed provided the source is explicitly acknowledged.

Using machine learning to correlate ESG-related news articles
to the stock market performance and detect corporate greenwashing

Abstract

The integration of machine learning into finance has played an important role in the growth of the Financial Technology (Fintech) sector. A recognised approach is the application of Natural Language Processing (NLP) to leverage sentiment analysis models and analyse corporate news in relation to financial markets. In recent times, the financial sector has been experiencing significant changes driven by the emergence of Environmental, Social and Governance (ESG) factors. The purpose of sustainable investing is to combine the ESG rating system - which incorporates non-financial indicators - with the more traditional financial factors in the investment process.

The main aim of this project is to leverage the potential of machine learning techniques in the context of providing investors with useful insights to make informed investment decisions. To contextualise the research in a real world scenario and study variations between different market sectors, this project explored 10 stocks of companies from the automotive and physical retail sectors listed across different stock exchanges (i.e., London Stock Exchange (LSE), New York Stock Exchange (NYSE), Nasdaq Stock Market, Milano Stock Exchange (Borsa Italiana)).

This project comprised of two sub-projects. The first part studied the application of sentiment analysis to assess the strength of the correlation between the composite score specifically for ESG-related news articles collected over 90 days, and the variation of the return of the stock price during the same historical period (i.e., objective one). The results revealed varying behaviours across the companies on how the stock prices have been impacted by the ESG-related news. In addition, this research attempted to predict the stock market volatility using sentiment analysis of the ESG-related news (i.e., objective two). By performing Granger's causality testing between the ESG-related news sentiment time series and the daily return time series, the influence of news articles on stock performance was detected for two companies, and market volatility was predicted. A similar approach was applied between ESG-related news and news articles containing keywords associated with greenwashing (i.e., objective three). By performing Granger's causality testing, the influence of ESG-related news on news articles containing keywords associated with greenwashing was detected for two companies.

In the second part of this project, corporate greenwashing was investigated by analysing annual reports (i.e., objective four). Greenwashing can occur in multiple forms. The most common stems from misleading messages communicated through marketing or when a company tries to emphasise the sustainable aspects of a product by publishing a large quantity of ESG data with the intent to hide environmentally harmful practices. By performing sentence similarity and sentiment analysis between the annual corporate reports and Sustainable Development Goals (SDGs) across 10 years, an

anomalous trend of the content of thematic content across the time series for a particular company was detected. Validation was successfully completed by fact-checking news articles published immediately preceding the occurrence of the irregularity which demonstrated that this methodology is suitable for detecting greenwashing.

This project demonstrated that sentiment analysis represents a good approach to assess the correlation between ESG-related news articles and the variation of the return of the stock price. In addition, the approach developed in the second part of the project has the potential to significantly improve the detection of greenwashing.

Keywords: sentiment analysis, BERT, news analytics, Reuters, composite score, Granger's causality testing, sentence similarity, volatility, ESG, greenwashing.

Problem Statement

Can current opinion-minining algorithms process financial news around 'sustainable investing' to find signals that help in predicting stock performace, volatility and -finally- detect corporate greenwashing?
This graduation project will address the following detailed research questions:

- RQ1: How are ESG-related news correlated to stock prices and returns?
- RQ2: Can stock market volatility be predicted using ESG-related news?
- RQ3: Can ESG-related news articles influence news articles containing keywords associated with greenwashing?
- RQ4: Can greenwashing be detected?

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The world is undergoing significant change that is inevitably affecting the financial sector. One of the main drivers of this change is the emergence of Environmental, Social and Governance (ESG) factors. The purpose of sustainable investing is to combine the ESG rating system - which incorporates non-financial indicators - with the more traditional financial factors in the investment process. This allows companies to reduce capital risk by minimising impact on the planet and people as well as giving support to sustainable development projects. Reliable, relevant and updated ESG data is required to implement specific ESG investment strategies. ESG data can also play a significant role in describing how an organisation is capable of dealing with non-financial factors that could potentially affect the health of its business and its licence to operate in the long term. The Environmental factor for instance could examine non-financial factors such as the magnitude of Greenhouse Gas (GHG) emissions generated by a company. The Social factor could take on board further non-financial elements such as the management of the supply chain. Finally, the Governance factor could assess the robustness of the level of cybersecurity of an organisation.

The share price of a company and the stock market are influenced by several factors, including corporate news (Bapat et al. 2022). This type of news can come from a broad variety of sources such as press releases, newspaper articles, and social media. Therefore, being able to assess the strength of the correlation between the sentiment of corporate news, specifically for ESG factors, and the performance of a company's stock, could potentially assist in predicting how the stock market could react, assessing the market volatility and ultimately supporting investors in making informed investment decisions.

An additional aspect to consider is corporate greenwashing. Krafft et al. (2014) states that more than 90% of all final goods in North America are involved in greenwashing. This is due to misleading messages communicated through branding, mar-

keting or even packaging to convince consumers and investors that either the products manufactured by an organisation or its operations are environmentally friendly. On the other hand, greenwashing can also occur when a company tries to emphasise the sustainable aspects of a product by publishing a large quantity of ESG data with the intent to hide environmentally harmful practices. These few examples of how greenwashing can take place are sufficient to prove the negative consequences that greenwashing can have for consumers, investors and the environment.

Detecting greenwashing and its perpetrators is particularly relevant as global financial markets increase investment into apparent sustainable initiatives. This risk led researchers, to develop innovative deep neural network models aimed at identifying patterns across financial documents with the use of artificial intelligence (AI) for corporate disclosures, and specifically on sustainability and climate change topics (Bingler et al. 2022).

This project is of interest to current developments in the financial sector as it applies machine learning techniques to finance use case that involve sustainability; a crucial area for the future of investment. Also, this project bridges the gap between sentiment analysis and financial forecasting, specifically targeting the opportunity for company evaluation in sustainable investing. The integration of machine learning into finance plays an important role in Financial Technology (Fintech). Companies in the finance industry that apply Fintech have experienced benefits such as expansion in the market as well as a reduction of operational costs (Zhang et al. 2022), providing valuable tools for both investors and regulators.

Sentiment analysis has been widely applied to financial markets to analyse corporate news, particularly to analyse how news sentiment affects stock prices. One example is Bloomberg Terminal, which provides sentiment scores based on financial news. The system helps traders and investors anticipate market movements and volatility and make informed decisions by analysing the sentiment of news articles. Another example is represented by ESG rating agencies, such as Morgan Stanley Capital International (MSCI) and Sustainalytics, which rate the sustainability of listed companies based on their performance across ESG factors. Whilst these ratings are based on broad datasets, they often include sentiment analysis of corporate disclosures, news, and reports. Further applications and research in this topic will be expanded on in Chapter 3.

One of the key challenges of this project was the limited amount of news content during the observation period. Another aspect was the complexity of analysing corporate reports and distinguishing between genuine sustainable practices and greenwashing.

The thesis work is organised as follows: the project aim and objectives are described in Chapter 2, the literature review is presented in Chapter 3, the methodology is dis-

cussed in Chapter 4, the data is described in Chapter 5, Chapter 6 presents the analysis and design, the findings are described in Chapter 7 and in Chapter 8 conclusions will be discussed.

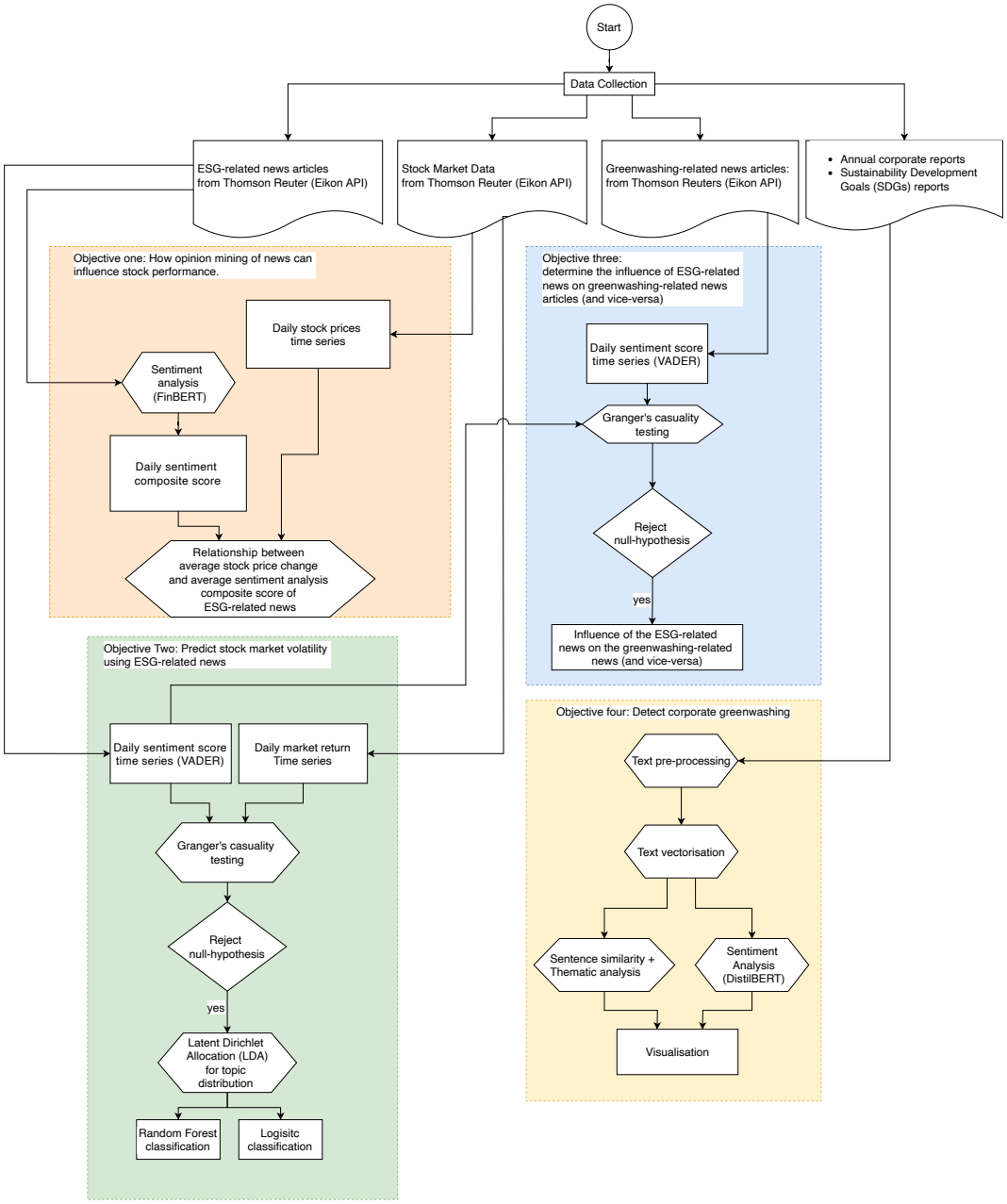

Figure 1.1: Proposed workflow and data pipeline

Chapter 2

Project Aim and Objectives

The main aim of this project is to leverage the potential of machine learning techniques in the context of providing investors with useful insights to make informed investment decisions. To address the research questions, the following four objectives were pursued and Figure 1.1 describes the process.

2.1 Objective One: How Opinion Mining of News Can Influence Stock Performance

Assess how opinion mining of news by sentiment analysis specifically related to sustainable investing disclosures can influence stock prices and returns. To achieve this objective, a similar approach to Bapat et al. (2022)'s project was implemented. The Python code created by Bapat et al. (2022) was used as guideline and it was useful to clarify the original intent. Bapat et al. (2022)'s code was implemented in accordance to the requirements of this objective. The following steps were performed:

- Download of news related to the selected companies based on keywords related to ESG topics over 90 days (Chapter 5.3);
- Text Preprocessing to convert raw data into a clean and structured format to optimise following analysis;
- Text Summarisation to reduce the number of words fed into the sentiment analysis model and further optimise the sentiment analysis classification process was adopted (Chapter 4.3);
- FinBERT sentiment classifier was used to perform the sentiment analysis and assess the sentiment score of each ESG-related news article (Chapter 5.7);

- Composite score was calculated to normalise the sentiment of the sentences across all the news articles (Equation 6.1);
- Daily stock market data of 10 selected companies were downloaded within the same period of observation (Chapter 5.1 and Chapter 5.6), and
- Aggregation of the average sentiment score and average opening price variation was performed to assess the correlations between the news and stock prices for each company (Chapter 7.1).

2.2 Objective Two: Predict Stock Market Volatility Using ESG-Related News

Assess how opinion mining of news by sentiment analysis specifically related to sustainable investing disclosures can predict stock market volatility. To achieve this objective, a similar approach to Deveikyte et al. (2022)'s project was implemented and the following steps were performed:

- Daily closing stock price of the selected companies was used to calculate the daily market return and 90-days volatility (Chapter 4.5 and Chapter 4.6);
- A binary market direction list was created to indicate the movement of the daily market return (i.e., +1 when the price goes up the value is +1, 0 when it goes down);
- VADER was used to perform sentiment classification of the ESG-related news articles and calculate the daily sentiment score (Chapter 4.2.2);
- Granger's causality test was performed for each company to determine whether the daily market return time series was influenced by the daily sentiment score time series (Chapter 4.9). If the null hypothesis was rejected then correlation was identified and topics were identified by the use of Latent Dirichlet Allocation (LDA) model (Chapter 4.8);
- ESG-related news are transformed into vectors and fed into a Latent Dirichlet Allocation (LDA) model for unsupervised classification in order to find group of clusters (topics). As a result, the news articles dataset was converted into a n x m matrix with n articles and m topics identified in the text corpus (Chapter 4.10);
- Topics are then converted into vectors to allow the machine to interpret the information, and

- A Random Forest and a Logisitc Regression model were used to assess the performance of the classification models in predicting the next day market movement based on the available topics (Chapter 4.11 and Chapter 4.12).

2.3 Objective Three: Determine the Influence of ESG-Related News on Greenwashing-Related News Articles

Assess the correlation between ESG-related news and news articles containing keywords associated with greenwashing.

To achieve this objective the following steps were performed:

- Download of news related to the selected companies based on keywords related to greenwashing topics over 90 days (Chapter 5.3);
- Text Preprocessing to convert raw data into a clean and structured format to optimise following analysis;
- VADER was used to perform sentiment classification of the greenwashing-related news articles and calculate the daily sentiment score (Chapter 4.2.2);
- The daily sentiment score of the ESG-related news articles determined in object two was used for this analysis, and
- Granger's causality test was performed for each company to determine influences between sentiment of ESG-related news time series and sentiment of greenwashing-related news articles time series (Chapter 4.9). If the null hypothesis was rejected then correlation was identified (Chapter 4.9).

2.4 Objective Four: Detect Corporate Greenwashing

Assess how sentiment analysis of annual corporate reports can detect greenwashing. To achieve this objective, a similar approach to Kang & Kim (2022a)'s project was implemented. The Python code created by Kang & Kim (2022a) was used as guideline and was useful to clarify the original intent. Kang & Kim (2022a)'s code was implemented in accordance to the requirements of this objective (Kang & Kim 2022b). The following steps were performed:

- 84 annual corporate reports were downloaded from the companies' websites (Chapter 5.4);

- 17 Sustainable Development Goals (SDG) reports were downloaded (Chapter 5.5);
- Text was extracted from every reports in PDF format (Chapter 4.15);
- Text Preprocessing was performed to convert raw data into a clean and structured format;
- Text Vectorisation using a pre-trained model to convert words into vectors was carried out to allow machines to interpret the data (Chapter 4.4);
- Two NLP methodologies were used in order to have multiple metrics to detect anomalies:
  - "Sentence Similarity": the UN's 17 Sustainable Development Goals (SDG) were used as benchmark to calculate "Cosine Similarity Measurement" in order to detect the similarities of the sentences across all the corporate reports (Chapter 4.14), and
  - Sentiment Analysis: DistilBERT (Chapter 4.2) was used to calculate the sentiment score of each sentence from the corporate reports to assess the trend of sentiment score over time.
- Both Sentence Similarity and Sentiment Analysis results were plotted to identify the tendency of companies to selectively report positive information. Anomalies and inconsistencies in the patterns (e.g., spikes) suggested greenwashing and required further investigation;
- Validation of the results was carried out by fact-checking news articles published immediately preceding the occurrence of the irregularity.

Chapter 3

Literature Review

Recent literature shows an emerging consensus, on the observation that the lack of a unique methodology and an integrated reporting framework for ESG often results in unrelated and inconsistent reporting and absence of transparency (Vitto et al. (2023), FinancialTimesAdviser (2021), Bapat et al. (2022) and Whelan et al. (2015)). This can make it difficult and time consuming to compare ESG efforts across multiple firms. In addition, ESG data is often assessed by companies themselves, which might affect the quality of the disclosed data and eventually the decision-making process. For this reason, when ESG ratings are examined, they should be considered as opinions rather than clear and standardised ratings. Therefore, investors should exercise caution when using these ratings to assess the sustainability performance of a company.

With regard to the correlation between ESG factors and financial returns, there are different opinions in literature that suggest uncertainty and unreliability. Bapat et al. (2022) demonstrate that whilst ESG-related news articles can impact positively on the stock prices for companies within certain sectors such as automotive, online retail and finance, the banking sector is inversely correlated. La Torre et al. (2020) suggest that investments in ESG can generate positive returns in specific sectors such as energy and utilities as long as the sustainable investment strategies are communicated. Whelan et al. (2015) state that most studies undertaken have reported positive or neutral results for sustainable investments. However, a few studies have found a negative correlation between ESG and financial performance. Brammer et al. (2006) discover a negative correlation amongst the UK firms with high social performance scores. As a result, those organisations had lower returns than those with lower social performance scores. Another study correlates greenwashing to a companies' returns to reveal a positive medium-term correlation (Kornreich 2022). Nevertheless, companies that did not engage in greenwashing had better long-term market performance. In line with this, Dumitrescu et al. (2023) also indicate that companies that engage in greenwashing are

more likely to underperform. They also confirm that the lack of a clear definition of greenwashing makes impossible to evaluate whether claims of greenwashing are legitimate.

In the recent work by Kang & Kim (2022a) the authors highlighted the shortcomings of "word frequency-based text" method because it doesn't take into account the semantic context of the sentences. Brookes & McEnery (2018) demonstrated that this approach treats words independently and without a correlation to the overall context. Also, a model that counts the frequency of specific words within a written document, is particularly limited in performing quantitative analysis of the content. To overcome this limitations, Wang et al. (2020) attempted to manually classify and assign each paragraph to one of the Sustainable Development Goals (SDGs). However, this kind of analysis, which involves manual classification across a vast amount of reports, is not sustainable and an automatic classification methodology is preferable. Nevertheless, a manual labelling task would be intrinsically affected by personal opinion and likely introduce cognitive bias and as a result the outcome would be affected (Székely & vom Brocke 2017). The "keyword matching method" (Lee & Kim 2021) is another word-frequency method with some fundamental flows as this model triggers only when the words compared are exactly the same regardless of the context of the documents compared. On the other hand, the use of sustainability reports could raise reservations, as organisations may use the reports to convey only positive aspects of particular topics (Liew et al. 2014) with lack of transparency. Therefore a sentiment analysis performed over 10 years was proposed to identify any anomalies in the sentiment scores patterns related to sustainability. With regard to the accuracy of data, whilst the use of social media content for sentiment analysis is used in several works (Bollen et al. 2010) (Vu et al. 2012), studies conducted by Peramunetilleke & Wong (2001), Readshaw & Giani (2021) and Deveikyte et al. (2022) suggest that social media content could contain more noise hence less accurate data compared to news articles.

Chapter 4

Methodology and Methods

4.1 Natural Language Processing

The expression “natural language” refers to the mean of communication either spoken or written used by humans. On the other hand “Natural Language Processing” (NLP) indicates a branch of Artificial Intelligence (AI) that gives the machines the ability to read, interpret and manipulate natural language. The field of NLP could span from tasks as simple as counting words within a text to the more complex tasks which include the comprehension of spoken conversation or assessing the sentiment of text (Steven Bird 2009).

4.2 Sentiment Classification Technique

“Sentiment analysis”, also known as “opinion mining”, is an application of NLP that leverages machine learning and NLP techniques to distinguish and classify the emotional tone and opinions expressed in messages and texts as “positive”, “negative”, or “neutral”. Such classification is also known as polarity.

4.2.1 Bidirectional Encoder Representations from Transformers (BERT)

In their recent work, Alaparthi & Mishra (2021) compared four different sentiment analysis methods: (i) unsupervised lexicon-based model using SentiWordNet, (ii) traditional supervised machine learning model using logistic regression, (iii) supervised deep learning model using Long Short-Term Memory (LSTM), and (iv) advanced supervised deep learning model using Bidirectional Encoder Representations from Transformers (BERT). And the superiority of BERT for sentiment classification purposes was demonstrated.

BERT, a Google neural network-based Language model, was designed by Devlin et al. (2019b) and distributed by HuggingFace (Devlin et al. 2019a). Unlike traditional unidirectional language models (i.e., left-to-right), BERT is capable of being pre-trained from an unlabelled corpus of text with a bidirectional approach. Bidirectional approach refers to the ability of considering both the left (preceding) and right (succeeding) contexts of a single word within a sentence when processing its meaning. Standard language models - such as OpenAI GPT - are unidirectional models and every token of a sentence is restricted in dealing with the previous token only. In contrast, BERT analyses the whole sentence and every single word simultaneously in order to determine how words are mutually influenced regardless of their order in the sentence. For example, in the sentence "My favourite season is summer.", BERT will look at "season" while considering both "My favourite" (preceding) and "is summer" (succeeding) when creating an embedding for "season". Embeddings are high-dimensional vectors representing words in their specific context. The conversion from words into vectors is essential to allow machines interpret words and language, and return accurate results. Therefore in the sentence "The recipe says to season the dish with plenty of chilly", "season" would have a different embedding because "season" here refers to the action of adding further flavour to a dish rather than the period of the year.

For the above considerations it makes sense to believe that a language model that looks at words from both directions is more robust than one that only reads from left to right or combines reading left-to-right and right-to-left. However, if the model reads both directions at once, it might cheat by "seeing" the word that it is supposed to predict. To avoid this, the bidirectional feature of BERT is enabled by the use of a Masked Language Model (MLM) also known as Cloze. MLM randomly masks some percentage of the words in a sentence and then trains the language model to predict the masked words based on the context. For example, in the sentence "My favourite *[MASK]* is summer.", BERT can predict the masked word "season" by considering the surrounding context defined by the preceding words "My favourite" and the succeeding words "is summer". Finally, the "Transformer encoder architecture" enables the bidirectional relationship between every pair of words in a sentence, regardless of their position in a sentence.

The combination of those features makes BERT "bidirectional" in contrast to standard language model that process text sequentially. For the reasons highlighted above, BERT was chosen in this project to carry out sentiment analysis of text. Following the launch of BERT, further pre-trained NLP models based on the same approach, were trained for specific domains. For this project two sentiment analysis models founded upon BERT were used: FinBERT and DistilBERT. The first model was pre-trained on the finance domain. Similarly to the work done by Bapat et al. (2022), FinBERT was used to

address objective one as news articles contained high volume of financial information and correlation with stock market prices was required (huggingface.co 2024b). DistilBERT was trained on more generic context and it was used to achieve objective four as corporate reports that have been analysed were not only related to sustainability topics but also to financial matters huggingface.co (2024a). The same model was also used by Kang & Kim (2022a) in their study for a similar application.

4.2.2 Valence Aware Dictionary and sEntiment Reasoner (VADER)

VADER is a ruled-based sentiment analysis tool part of Python Natural Language Toolkit (NLTK) library and it is optimised for social media text content (Hutto 2020). Ruled-based means that VADER uses a dictionary called Lexicon to determine the sentiment of a given text. Lexicon contains a set of words and phrases with their sentiment ratings being already assigned. VADER breaks down the sentences into individual words and assigns to each word a sentiment numeric score based on the Lexicon dictionary and this contribution will determine the overall sentiment score. As a result, VADER returns individual "positive", "negative" and "neutral" sentiment scores to reflect the proportion of words with that sentiment within the sentence, as well as the "compound" score that reflects the overall intensity and polarity of the sentiment of the text. The compound score is a weighted sum of the sentiment scores normalised to fall between -1 and +1 where -1 denotes very negative sentiment and +1 denotes very positive sentiment. The main advantage of VADER is that it doesn't require an additional pre-trained NLP model making it more transparent and faster. However, this advantage implies a downside as this tool has limited versatility and flexibility. VADER was used to address objectives two and three. The same model was also used by Deveikyte et al. (2022) in their study for a similar application.

4.3 Text Summarisation Using BART

To further optimise the sentiment analysis classification process in objective one, text summarisation of each individual news article was required in order to reduce the number of words fed into the sentiment analysis model. For this purpose the "Facebook/bart-large-cnn" model was used. "bart-large-cnn" was first introduced by Facebook AI in 2019, is pre-trained on English language model, fine-tuned on CNN Daily Mail and distributed by Hugging-Face (Lewis et al. 2019b). Besides Text summarisation, Bidirectional and Auto-Regressive Transformers (BART) include further other functions such as natural language generation, translation, and comprehension (Lewis et al. 2019a). A BART model is based on a pre-training method that consists of two phases: (i) noise is

introduced and as a result, the original text is corrupted, and (ii) a sequence-to-sequence model is able to reconstruct the original text. Its architecture can be considered a combination of (i) BERT (Chapter 4.2.1) as it inherits the bidirectional function, and (ii) Generative Pre-training Transformer (GPT) as it uses the left-to-right decoder function for prediction and generation of text. The contribution of BERT allows the model to select randomly tokens in the sentence, introduce some noise by replacing those tokens with mask symbols, and finally encode the document bidirectionally (left-to-right and right-to-left). The information encoded by BERT is fed into a GPT which allows the model to make auto-regressive decoding (i.e., predictions) which means that it can be used for the generation of the masked elements of the text.

4.4 Text Vectorisation

As mentioned in Chapter 4.2.1, machines are unable to interpret words, unless they are converted into a numerical structure such as vectors (i.e., embeddings) that computers can handle for further analysis. The dedicated "all-MiniLM-L6-v2" sentence transformer distributed by Hugging Face was used to convert the sentences contained in the corporate reports into vectors before feeding the cosine similarity model (Aarsen 2021). This particular type of sentence-encoder was designed for short paragraphs, making it ideal for this scenario since the reports contain brief sections.

4.5 Daily Market Return

The daily market return on day t was calculated as follows:

$$r_t = ln\left(\frac{CLOSE_t}{CLOSE_{t-1}}\right) \tag{4.1}$$

where $CLOSE_t$ is the closing price on day t and $CLOSE_{t-1}$ is the previous day closing price (Deveikyte et al. 2022). Natural logarithm is used to deal with returns over time (e.g., compound returns). In addition, logarithms allows for simplified calculations, such as using sums rather than multiplications. The "closing price" is considered a better indicator than "opening price" - for daily market returns - as it captures the final consensus amongst traders after a full day of negotiations allowing the true value of a stock at that point in time to emerge (Gratton 2024).

4.6 Volatility

Volatility, in finance, expresses the tendency of a price of a stock to fluctuate over a given period. High volatility denotes rapid and unpredictable changes of the price of a stock and it can also be associated to the propensity to generate high profits as well as expose the investors to high risks. Whereas low volatility is associated to a more stable trend, hence a safer context, which is associated to low risks. From a statistical point of view, the volatility can be expressed as the deviation standard of the daily market return values, in other words it measures how much the daily market returns are spread out from the average. The annualised volatility of a stock market index based on daily returns within a defined time window was calculated as follows:

$$Vol = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(r_t - \bar{r})^2} \cdot \sqrt{252} \tag{4.2}$$

where N is the total number of days during a window time of observations (e.g., previous 90 days), and 252 is the total number of trading days in a single year (Deveikyte et al. 2022).

4.7 Daily Sentiment Score

The daily sentiment score was calculated as follows:

$$Sent_d = \frac{N_d(pos) - N_d(neg)}{N_d(pos) + N_d(neut) + N_d(neg) + 3} \tag{4.3}$$

where $N_d(neg)$, $N_d(neut)$, and $N_d(pos)$ were calculated by sentiment analysis and denote the volume of negative, neutral, and positive news headlines or news articles on day t by aggregating the sentiment scores from the daily news articles and headlines (Deveikyte et al. 2022).

4.8 Hypothesis Testing

Hypothesis testing refers to the process of testing whether or not a sample of data confirms a particular hypothesis. In testing a hypothesis, the first step consists in defining a null hypothesis and an alternative hypothesis. Either the null hypothesis will be rejected and the alternative hypothesis will be considered true, or the null hypothesis will fail to be rejected. The latter doesn't automatically imply that the null hypothesis is true but rather that there is not enough evidence to reject it.

For this project the null hypothesis will be used to support the hypothesis in which no relationship exists between two sets of data. One set of data could be the sentiment score time series and the second one the daily market returns time series to address objective two, for instance. Therefore the null hypothesis claims that a relationship between the two sets of data doesn't exist. If the null hypothesis fails to be rejected then there is not enough evidence to prove the relationship between the two given sets of data. On the contrary, if the null hypothesis is rejected, then a relationship between the two sets of data exists. If this is the case, then a particular test statistic such as the "Granger's causality testing" (Chapter 4.9) will determine the corresponding p-value to assess the magnitude of influence between the two time series.

4.9 Granger's Causality Testing

The impact of the sentiment score on day t to the stock market performance in the following day $t+1$ was assessed by performing "Granger's causality testing" which is used to determine causality between two variables and whether one time series variable is relevant in forecasting the other over a given lagged period. In this particular case, Granger's causality testing was used to test whether the sentiment obtained from financial news (variable X) was relevant in forecasting the stock market performance and volatility (variable Y). The generic Granger's causality testing equation can be expressed as follows (Deveikyte et al. 2022):

$$y_t = \alpha + \sum_{j=1}^{k} \beta_j Y_{t-1} + \sum_{j=1}^{k} \lambda_j X_{t-j} + \epsilon_t \tag{4.4}$$

The function provides several test statistics and corresponding p-values that will reveal the influence of the financial news over the stock performance. If the null hypothesis is rejected and the p-value is smaller than 0.05, it is possible to conclude that variable X (sentiment) influences significantly stock market changes and volatility (variable Y). Interestingly, the test can be also carried out reversely to verify if the volatility could potentially affect the sentiment of news in reverse by inverting the order of the variables. An important variable of the Granger's causality test is "lags". As the test is run across two time series, lags represent the number of past values in time that may predict future values. The lags value depends on the scenario in which the model is applied. In a finance application where the market reaction to economic events is relatively quick, a small lag (i.e., up to five) is enough to detect a correlation. The Granger's causality testing (Equation 4.4) was performed using the dedicated Python library "grangercausalitytests" (statsmodels.org 2024).

4.10 Latent Dirichlet Allocation (LDA)

In the research conducted by Deveikyte et al. (2022) a Latent Dirichlet Allocation (LDA) model was implemented to categorise sets of news stories into specific topics. The LDA is an unsupervised model applied to NLP aimed at categorising clusters of words by topics. As a result, the LDA model generated a K-dimensional vector where K is the number of topics for every news article. The generic LDA equation can be expressed as follows (Blei 2012):

$$p(\beta_{1:k}, \theta_{1:\mathrm{D}}, z_{1:\mathrm{D}}, w_{1:\mathrm{D}}) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(z_{\mathrm{d,n}} \mid \theta_d) p(w_{\mathrm{d,n}} \mid \beta_{1:\mathrm{K}}, z_{\mathrm{d,n}}) \right) \tag{4.5}$$

where D is the total number of documents and N is the total number of words within documents. $\beta_{1:\mathrm{K}}$ are all the topics in document d, $\theta_{\mathrm{d:K}}$ is the weight of topic k in document d, $z_{\mathrm{d,n}}$ is the classification of the topic for the nth word in document d, w_d are the observed words in document d and $w_{\mathrm{d,n}}$ is the nth word in document d, which is an element from the fixed vocabulary.

Coherence score ranges from 0 to 1 and it measures similarity between words. The highest coherence score determines the optimal number of clusters (i.e., topics). "Word cloud" is a visual representation of the topics that helps determine if the number of topics proposed by the model is coherent. Additional preprocess of the text was required until word cloud results don't seem consistent. Finally a topic number was assigned to each news article and a vector that describes the distribution of the topics was passed to the classifier models, such as random forest and logistic regression, which predicted the movement of the volatility for the day after. The LDA model (Equation 4.5) was deployed using the "gensim/LdaModel" along with "gensim/CoherenceModel" libraries on Python (Rehurek 2024).

4.11 Random Forests

To overcome the risk of overfitting, random forest models are used as they gather multiple decision tress (ensemble). Random forest algorithms are supervised models that are used for both classification and regression purpose. Whilst in a classification problem the final prediction is given by the majority voting, in a regression problem the output is obtained by averaging all predictions across the trees in the forest. Random forests are characterised by being trained with the bagging method in which multiple subsets of the original training dataset are not correlated as they are different from

each other and it repeatedly fits decision trees to these subsets. With bagging the variance is reduced remarkably when compared to the variance of the individual decision trees. The Random Forest model was deployed using the dedicated Python library "sklearn/RandomForestClassifier" ((scikit learn.org 2024b).

4.12 Logistic Regression

Logistic regression models use the sigmoid function to convert linear outputs into probability functions from 0 to 1.

$$S(x) = \frac{1}{1 + e^{-X\beta}} \tag{4.6}$$

where X is the set of predictor features and β is the corresponding vector of weights. If the probability generated by S(x) is higher than the threshold 0.5 then the observation is classified as a "1", and a "0" otherwise. Logisitc regression is simple to use, however under performs when the decision boundary is not linear due to the fact it is characterised from being a low-variance and high-bias model. The Random Forest model was deployed using the dedicated Python library "sklearn/LogisticRegression" ((scikit learn.org 2024a)).

4.13 Confusion Matrix

With a classifier model the purpose is to minimise the number of misclassified observations and increase the number of successful classifications. A confusion matrix is a table that helps organise and visualise the performance of a classification model. Besides the true positive (TP) and true negative (TN) observations, also the misclassified observations called false positive (FP) and false negative (FN) are displayed in the confusion matrix. A false positive occurs when the model mistakenly predicts that an instance belongs to the positive class. Whereas in a false negative, the model produces a negative class, when in reality, is positive.

The accuracy of a classification model is a metric that indicates the probability of the model to make correct predictions overall. The accuracy is the ratio of the correct predictions (TP+TN) to the total number of samples and is calculate as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.7}$$

The precision of a classification model is a metric that indicates the probability of a model to predict a specific target class. The precision is the ratio of the true positives (TP) to the total positive predictive samples (TP+FP) and is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4.8}$$

4.14 Sentence Similarity

"Cosine similarity" is a robust metric for determining the similarity between two strings as similarity is measured by the cosine of the angle between two non-zero vectors and determines whether two vectors are pointing in approximately the same direction and if this occurs similarity between two sentences is established (Tata & Patel 2007). Also, cosine similarity method is based on the orientation of the vectors rather than the magnitude. This aspect is particular relevant as after converting the sentences into vectors, a high-dimensional sparse matrix is generated and in this circumstance a model based on the orientation of the vectors is more accurate than a model based on the magnitude.

After converting the sentences into vectors (Chapter 4.4), it became possible to assess the similarity between each sentence in the corporate reports against those in the SDG reports. To achieve this the "cosine similarity" method was used. The similarity for each SDG throughout the report was calculated to determine the ratio of how frequently representative words appeared in the sentences.

$$Cosine\ \ Similarity\ \ Score = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_i^n A_i \times B_i}{\sqrt{\sum_i^n (A_i)^2} \times \sqrt{\sum_i^n (B_i)^2}} \tag{4.9}$$

where A is a sentence vector from corporate reports, B is a representative sentence vector from the SDGs reports, and n is the dimension of vectors (Kang & Kim 2022a).

To compare the similarity scores across the corporate reports, "min-max scaler" (Equation 4.10) was used to normalise the scores to the same scale between 0 and 100:

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \times 100 \tag{4.10}$$

where x is the similarity score of sentence i, and z is the scaled similarity score (Kang & Kim 2022a).

The assessment of the similarity score was carried out as follows:

- Let n_i be the total number of sentences extracted from the i-th corporate report (i.e., n_1 is the total number of sentences from Report 1, n_2 is the total number of sentences from Report 2, and so on);
- Let i be the total number of corporate reports;
- Let n_k be the total number of sentences extracted from the k-th SDG report, and

- Let k be the total number of SDG reports;

where:

k = 1, 2, ... , 17 (k is the index associated to the SDG reports);

j = 1, 2, ..., n_k (j is the index associated to a sentence extracted from the SDG report k);

i = 1, 2, ..., n_p (i is the series of sentences from 1 to n_p within corporate Report r);

Then, the similarity score $S_{ij}^{k}|_r$ is calculated between the i-th sentence in Report r and the $j-th$ representative sentence from k-th SDG report.

The total of sentences across the corporate reports and the SDG reports were 168,277 and 641 respectively. As a result, the total of similarity scores assessed was 107,865,557 (168,277 x 641). The average and standard deviation of the similarity scores were 43.08 and 10.60 respectively.

The similarity score of each sentence 'i' extracted from the corporate reports 'r' against each SDG report 'k' were obtained by averaging all of the similarity scores as follows:

$$S_{i}^{k}|_r = \frac{1}{n_k}\sum_{j=1}^{n_k} S_{ij}^{k}|_r \tag{4.11}$$

The similarity scores for the r-th corporate report across the 17 SDG statements were calculated as follows:

$$S^{k}|_r = \frac{1}{n_r}\sum_{i=1}^{n_r} S_{i}^{k}|_r \tag{4.12}$$

Finally, the similarity score for the r-th report across the six SDG sub-categories (Economic and Technological development, Environments, Equity, Life, Resources, Social Development) was assessed by averaging the scores across a specific sub-category (Kang & Kim 2022a). Cosine similarity was performed using the "Sentence Transformer" framework for semantic embedding distributed by Hugging Face (HuggingFace 2021). Figure 6.1 summarises the sentence similarity assessment process.

4.15 Extraction of Text from the PDF Files

The PyMuPDF Python library (McKie 2016) was used to extract blocks of text from PDF documents as it is capable of isolating and distinguishing text from images.

Chapter 5

Data

5.1 Choice of Companies

A prerequisite for the selection is that the companies must be listed on a stock market (i.e., London Stock Exchange (LSE), New York Stock Exchange (NYSE), Nasdaq Stock Market, Milano Stock Exchange (Borsa Italiana)). Also, to diversify the perspective of the correlations, the companies belong to significant sectors in the economy. The companies selected for this study are the following:

- Automotive industry:
  - Ford;
  - Stellantis;
  - Toyota;
  - Polestar (Owned by Volvo Cars), and
  - Tesla.

  The choice of the first three companies reflects the Kwok et al. (2023)'s report where those three companies rank in the highest, middle and bottom position respectively. The last two are manufacturers of all-electric vehicles, with their head quarter based in Europe and North America respectively.

- Physical retail industry:
  - Asda (owned by Walmart);
  - Marks and Spencer;
  - Ocado;

- Sainsbury's, and
- Tesco.

The above retail chains are the predominant leaders in the United Kingdom's supermarket industry.

5.2 News Articles

For this project two main datasets of news articles were used. The first one is the set of "ESG-related news articles" which includes news containing keywords associated with ESG content. The second set of news includes news articles containing keywords associated with greenwashing content also indicated as "greenwashing-related news articles". The choice of the keywords is explained in Chapter 5.3. In addition, to address the unreliability of ESG rating systems (Chapter 3), the correlation of the ESG disclosures in relation to the stock performance relied on sentiment analysis of ESG-related news articles rather than ESG reports. For these reasons, this paper relied on the information extracted from news articles in English language. To avoid duplication of pieces of information, a single and reliable source of news was used. Thomson Reuters on Eikon API was the preferred option as it provides access to accurate real-time financial information and news. The raw ESG-related news articles dataset comprised of 5,166 rows, whilst the raw greenwashing-related news articles dataset comprised of 2,649 rows. Both datasets went through cleaning reducing the first dataset to 990 rows and the latter to 1,295 rows. By leveraging Python, the news articles published between 1.05.2024 and 31.7.2024 were downloaded from the Eikon-Reuters Code-Book web platform.

5.3 Choice of Keywords for the Selection of News Articles

Table 5.1 and 5.2 show the keywords that have been used to select the ESG-related news and greenwashing-related news respectively. Those keywords reflect the list proposed by Dumitrescu et al. (2023) in their work where they have defined a dictionary of words that appear more frequently in investment objectives of funds categorised under ESG principles. The keywords used for the ESG-related news are the words contained in the initials of the word "ESG". The more generic "sustainability" was also added. The rest of the words in Dumitrescu et al. (2023)'s list have been used as a benchmark for the greenwashing-related news. In particular for each of the companies, the period chosen for the download of the news articles spans from 1.05.2024 to 31.07.2024.

List of keywords for the selection of ESG-related news	
esg	social
environment	sustainability
governance	

Table 5.1: List of five keywords used for the selection of ESG-related news articles. This list reflect the benchmark proposed by (Dumitrescu et al. 2023).

List of keywords for the selection of greenwashing-related news	
alternative energy	low carbon
carbon-neutral	natural
clean energy	organic
climate	renew
fair	transition
fossil free	waste
green	zero waste

Table 5.2: List of 14 keywords for the selection of greenwashing-related news articles. This list reflect the benchmark proposed by (Dumitrescu et al. 2023).

5.4 Annual Corporate Reports

To perform the greenwashing detection model, a total of 84 annual corporate reports across 10 companies spanning from 2014 to 2023 were downloaded in PDF format from the respective companies websites based on availability. For the purpose of this project the term “corporate reports” indicates either ”corporate sustainability reports” and “corporate financial reports”. Across the companies observed, only seven of them issued dedicated sustainability reports, and of those, only four issued reports for the entire period of observation. In particular, Polestar and Tesla published their reports later due to be newer companies. Asda released its first sustainability report in 2020, and before that year, not even financial reports were publicly available. Ocado and Tesco issued only annual financial statements which include sustainability chapters in their reports. Sainsbury’s has issued sustainability reports consistently since 2018. It is worth noticing that all the automotive companies covered in this study issued dedicated annual sustainability reports. The list of the reports for each company is summarised in Tables 5.3 and in more detail in Table B.1.

		2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Automotive	Ford	●	●	●	●	●	●	●	●	●	●
	Polestar							●	●	●	●
	Stellantis	●	●	●	●	●	●	●	●	●	●
	Tesla						●	●	●	●	●
	Toyota	●	●	●	●	●	●	●	●	●	●
Retail	Asda							●	●	●	●
	Marks & Spencer	●	●	●	●	●	●	●	●	●	●
	Ocado	○	○	○	○	○	○	○	○	○	○
	Sainsbury's	●	○	○	○	●	●	●	●	●	●
	Tesco	○	○	○	○	○	○	○	○	○	○

● = dedicated corporate annual sustainability report

○ = corporate annual report or financial statement with sustainability claims incorporated

Table 5.3: Summary that shows the available type of annual corporate report from 2014 to 2023. Contrary to the supermarkets chains analysed, all the five automotive companies have published dedicated sustainability reports. This is due to the fact that the automotive industry is a sector that is heavily scrutinised for its direct environmental implications (Wagemans 2023). Supermarkets are less keen to issue dedicated sustainability reports. In fact, two companies out of five have issued only financial statements over the last 10 years.

5.5 Sustainable Development Goal (SDG) Reports

To assess the sentence similarity, the United Nations 17 Sustainable Development Goals (SDG) (2016-2030) (United Nations 2016) were used as benchmark for the similarity score across the sentences extracted from the corporate reports. A total of 17 reports in PDF format were downloaded from the Compass website (Charlotte Portier 2015). To simplify the similarity score assessment, the 17 SDGs were grouped in six subcategories (Economic and Technological development, Environments, Equity, Life, Resources, Social Development) (Wu et al. 2018). Table 5.4 summarises the six categories.

Equity	
Goal 4	Quality education
Goal 5	Gender equality
Goal 10	Reduced inequalities
Social Development	
Goal 11	Sustainable cities and communities
Goal 16	Peace justice and strong institutions
Goal 17	Partnerships for the goals
Life	
Goal 1	No poverty
Goal 2	Zero hunger
Goal 3	Good health and well-being
Economic and Technological development	
Goal 8	Decent work and economic growth
Goal 9	Industry, innovation, and infrastructure
Resources	
Goal 6	Clean water and sanitation
Goal 7	Affordable and clean energy
Goal 12	Responsible consumption and production
Goal 14	Life below water
Environments	
Goal 13	Climate action
Goal 15	Life on land

Table 5.4: The 17 SDGs were grouped into six sub-categories (Wu et al. 2018). Data were organised in sub-groups to improve the analysis and reduce complexity. The corporate reports were compared against those six categories using sentence similarity method.

5.6 Stock Market Data

The stock market dataset comprised of 636 rows (week days only) corresponding to trading days across the 10 companies analysed. The dataset includes data which spans from 1.5.2024 to 31.7.2024, including closing and opening price on that date as well as the highest and lowest price the stock reached during the day. Similar to the source of news articles, Thomson Reuters on Eikon API was the preferred option to gather the stock market data for the 10 companies (Table 5.5). The dataset was downloaded from

the Eikon Code-Book web platform using a Python script that was written to allow the connection and identification via API.

No	Company	Symbol	Date	Close	Open	High	Low
0	Asda	WMT	2024-05-01	58.85	59.31	59.4100	58.7200
1	Asda	WMT	2024-05-02	59.71	58.94	59.8850	58.5800
2	Asda	WMT	2024-05-03	59.82	59.62	59.9800	59.1400
...	...	...	...	...	...	...	...
633	Toyota	TM	2024-07-29	192.48	193.00	193.2000	191.8067
634	Toyota	TM	2024-07-30	193.11	194.96	195.4800	192.2650
635	Toyota	TM	2024-07-31	193.55	194.18	194.8900	192.9000

Table 5.5: Daily stock market dataset for Asda, Ford, Marks & Spencer, Ocado, Polestar, Sainsbury's, Stellantis, Tesco, Tesla and Toyota for a total of 636 rows (week days only) downloaded from Thomson Reuters Eikon API from 1.5.2024 to 31.7.2024, including closing and opening price on that date as well as the highest and lowest price the stock reached during the day.

5.7 Software Development Framework

Based on the nature of the data and the type of analysis required, Python was the programming language preferred for this project. Python offers a broad range of free open-source libraries and machine learning models relatively easy to implement to serve multiple purposes including import and preprocess data from APIs, import and preprocess text from PDF documents, perform sentiment analysis, determine correlations, and visualise the results.

Chapter 6

Analysis and Design

6.1 Objective One: How Opinion Mining of News Can Influence Stock Performance

Assess how opinion mining of news by sentiment analysis specifically related to sustainable investing disclosures can influence stock prices and returns. To achieve this objective, a similar approach to Bapat et al. (2022)'s project was implemented and the following steps were performed:

6.1.1 Text Preprocessing

The model is based on real-time newspaper articles downloaded using the Thomson Reuters API (Eikon) pipeline. By leveraging Python, the ESG-related news articles published between 1.05.2024 and 31.7.2024 were downloaded and organised with the full article content, timestamp, storyID aggregated by companies and appropriate keywords for a total of 5,166 rows (Chapter 5.2). Text preprocessing was required to format the date, drop empty cells, remove HTML tags as well and finally convert the text into small caps to optimise further data manipulation. As a result, the number of rows was reduced and a total of 990 stories were obtained (Table C.1).

Before passing the dataset to the sentiment classifier model, "Facebook/bart-large-cnna" was used to summarise the text content. A Bidirectional and Auto-Regressive Transformer (BART) was required to further condense the text to a manageable length, reduce the data set size and increase the performance of the sentiment analysis classifier (Chapter 4.3).

6.1.2 Sentiment Analysis of ESG-Related News

The dataset was fed into "ProsusAI/FinBERT" for sentiment classification (Chapter 4.2.1). The model returned a "sentiment label" for each news story (i.e., positive, neutral and negative) and a "confidence score" spanning from 0 to 1 indicating the accuracy of each classification. A "sentiment value" was determined by converting the sentiment label into numbers by assigning (i.e., mapping) -1 to negative sentiment, 0 to neutral sentiment and +1 to positive sentiment.

$$Composite\ \ Score = Sentiment\ \ Value \times Confidence\ \ Score \tag{6.1}$$

The "composite score" normalised the results between -1 and +1, representing the sentiment of a sentence based on the weight assigned by the "confidence score". Because the correlation is done against the variation of the price of stocks, ideally a neutral sentiment corresponds to a very small or nil variation of the price, therefore is accepted that the composite score ignores the sentences with neutral sentiment (Bapat et al. 2022). Table D.1 shows an example of the sentiment analysis of the ESG-related news using FinBERT.

6.1.3 Stock Prices

Stock prices were downloaded with a daily granularity and for this study opening prices were considered in order to assess the variation of price from the previous day (Bapat et al. 2022). Rather than closing prices, opening prices are considered as the direction of an entire session is often ruled by the early trading of the same day (Investopedia.com 2023). Table 5.5 shows an example of the stock prices dataset.

6.1.4 Correlations

Table 7.1 shows the resulting correlations between the weekly averages of "composite score" (Equation 6.1) and average stock prices variation.

6.2 Objective Two: Predict Stock Market Volatility Using ESG-Related News

To predict the movement of the market on the next day and its volatility by leveraging sentiment analysis of ESG-related news, the methodology proposed by (Deveikyte et al. 2022) was followed.

6.2.1 Daily Market Returns Time Series and Volatility

The daily market returns time series was calculated for every stock (Chapter 4.5) and Table 7.2 shows the results. The volatility within the period of observation was also calculated for each stock (Chapter 4.6) and Table 7.3 shows the results.

6.2.2 Sentiment Analysis of ESG-Related news

The sentiment analysis of the ESG-related news was performed using the VADER sentiment classification model described in Chapter 4.2.2 and Table D.3 shows the results.

6.2.3 Granger's Causality Testing

Granger's causality testing was performed to assess whether one time series affects the other (Chapter 4.9). The statistical hypothesis testing was rejected for Marks & Spencer and Tesla as their probability value (p-value) was smaller than 0.05. Table 7.4 shows the results. This suggested that ESG-related news for Marks & Spencer and Tesla influenced their respectively stock market performance. Therefore further analysis for the prediction of volatility was conducted and restricted to those two companies.

6.2.4 Latent Dirichlet Allocation (LDA) for Topic Distribution

Two dedicated LDA models were trained, one for each company (i.e., Marks & Spencer and Tesla), to obtain the optimal number of topics. To train the LDA models, the ESG-related news stories for M&S and Tesla were converted into a list, then each story in the list was converted into tokens. The dictionary with the tokenised news was converted into a bag of words corpus. Finally, the bag of words corpus was fed into the LDA model which generated a pre-trained topic model which was passed, along with the dictionary and the tokenised list, to a Coherence Model which returned coherence scores as a function of the number of topics (Chapter 4.10). Figure 7.1 shows the coherence score plot for Marks & Spencer. Figure 7.2 shows the coherence score plot for Tesla. For both of them the number of topics with the highest coherence score was seven.

Word clouds were used to visualise the most relevant words for each topic in order to establish whether the results from the LDA model were consistent and appropriate. Figure 7.3 shows the word clouds visualisation for Marks & Spencer. Figure 7.4 shows the word clouds visualisation for Tesla.

The number of topics was required to create a distribution of the news stories in relation to the topics which was used as feature matrix to train the classifier models. On the other hand, the daily returns list were converted into a binary format where

+1 identified a positive daily return from the previous day, whereas 0 identified a loss. The daily returns list was used as a target vector to train the classifier models. Unlike the research by (Deveikyte et al. 2022), that used a logistic regression model to predict the next day return, this work explored the use of two supervised classification models, random forest and logistic regression model.

6.2.5 Random Forest Classifier

Two random forest classifiers were trained (Chapter 4.11). One for Marks & Spencer with a total of 24 data points, and another one for Tesla with a total of 35 data points. For both models a splitting ratio of 70:30 (i.e., 70% of the data is for training and 30% for testing) was applied. Table 7.5 shows accuracy and precision metrics (Chapter 4.13). Figure 7.5 shows the confusion matrix for Marks & Spencer and Figure 7.7 shows the confusion matrix for Tesla.

6.2.6 Logistic Regression Classifier

Two logistic regression classifiers (Chapter 4.12) were trained. For comparison purpose, the data points used and the splitting ratio were the same as those for the random forest classifiers (i.e., 70:30). Table 7.5 shows accuracy and precision metrics (Chapter 4.13). Figure 7.6 shows the confusion matrix for Marks & Spencer and Figure 7.8 shows the confusion matrix for Tesla (Chapter 4.13).

6.3 Objective Three: Determine the Influence of ESG-Related News on Greenwashing-Related News Articles

6.3.1 Sentiment Analysis of Greenwashing-Related News

The sentiment analysis of the greenwashing-related news was performed using the VADER sentiment classification model described in Chapter 4.2.2.

6.3.2 Granger's Causality Testing

Granger's causality testing was performed to assess whether one time series affects the other (Chapter 4.9). The statistical hypothesis testing was rejected for Marks & Spencer and Ford as their probability value (p-value) was smaller than 0.05. Table 6.1 shows the results. In particular, for Ford the sentiment of ESG-related news influences sentiment of greenwashing-related news articles. On the other hand, for Marks & Spencer the sentiment of greenwashing-related news influences sentiment of ESG-related news.

	p-value	
	ESG-related news articles influence on Greenwashing-related news	**Greenwashing-related news influence on ESG-related news articles**
Ford	0.035	-
M&S	-	0.0018

Table 6.1: Granger's testing significant p-values (¡0.05). For Ford the sentiment of ESG-related news influences sentiment of greenwashing-related news articles. For Marks & Spencer the sentiment of greenwashing-related news influences sentiment of ESG-related news.

6.4 Objective Four: Detect Corporate Greenwashing

To achieve this objective and address the limitations described in Chapter 3, the methodology proposed by Kang & Kim (2022a) was followed. This study used two natural language processing methods:

- "Sentence similarity" method to perform "thematic analysis" (Chapter 4.14), and
- "Sentiment analysis" (Chapter 4.2).

For this study "sentence similarity" method was applied to overcome the limitations of word frequency-based text methods. Kang & Kim (2022a) demonstrated that sentence similarity methods - in contexts such as corporate reports - have performed better than word frequency-based methods.

"Thematic analysis" helped in understanding the semantic relationships between sentences and assessing the content of the reports. To identify the thematic structures of the corporate reports in a sustainable context, it was required to assess the weight of information linked to "sustainability" for each of the corporate reports. In order to achieve that, the model assigned one of the SDGs categories to each sentence extracted from the corporate reports (Chapter 5.5). The SDG framework was adopted by the United Nations in 2015 to specify sustainability goals and metrics for nations and organisations. The SDGs are 17 objectives that have the purpose of enhancing the environment, improve the welfare of human beings, guarantee peace and fight discrimination.

Finally, sentiment analysis classification (Chapter 4.2) of the corporate reports was applied to assess the sentiment scores across the years when the reports were available in order to determine any changes in the balance of positive and negative sentiment

scores across the period of observation as this would point out anomalies that would potentially suggest greenwashing. The following steps were performed:

6.4.1 Text Preprocessing

Corporate and SDG reports were downloaded in PDF format (Sections 5.4 and 5.5 which both required appropriate preprocess as follows:

- **Extract the text from the PDF files**: The PyMuPDF Python library was used to extract blocks of text from PDF documents (Chapter 4.15). As suggested by Bapat et al. (2022) a further filter was applied to ignore sentences composed by less than 10 words to optimise the dataset by removing titles or slogans with little information content;
- **Break down the text into sentences**: Each block of text was broken down into a list of sentences by a pre-trained tokenizer module for Python included in the Natural Language Toolkit (NLTK) for English (Loper & Bird 2002);
- **Clean the sentences**: The last step consisted in cleaning each sentences by removing any characters other than letters, numbers and special characters that are not on standard keyboards - which would not provide any useful information for the analysis - and replacing them with blank spaces.

As a result of preprocessing a total of 168,277 sentences across 84 reports were obtained from corporate annual reports (Table 6.2), and 641 sentences from 17 SDGs (Table 6.3).

6.4.2 Text Vectorisation

To convert sentences into vectors which are comprehensible to machines, the text vectorisation method described in Chapter 4.4 was used.

6.4.3 Sentence Similarity

After converting the sentences into vectors (Chapter 4.4), it became possible to assess the similarity between each sentence in the corporate reports against those in the SDG reports. To achieve this the "cosine similarity" method was used (Chapter 4.14). Figure 6.1 shows the proposed workflow.

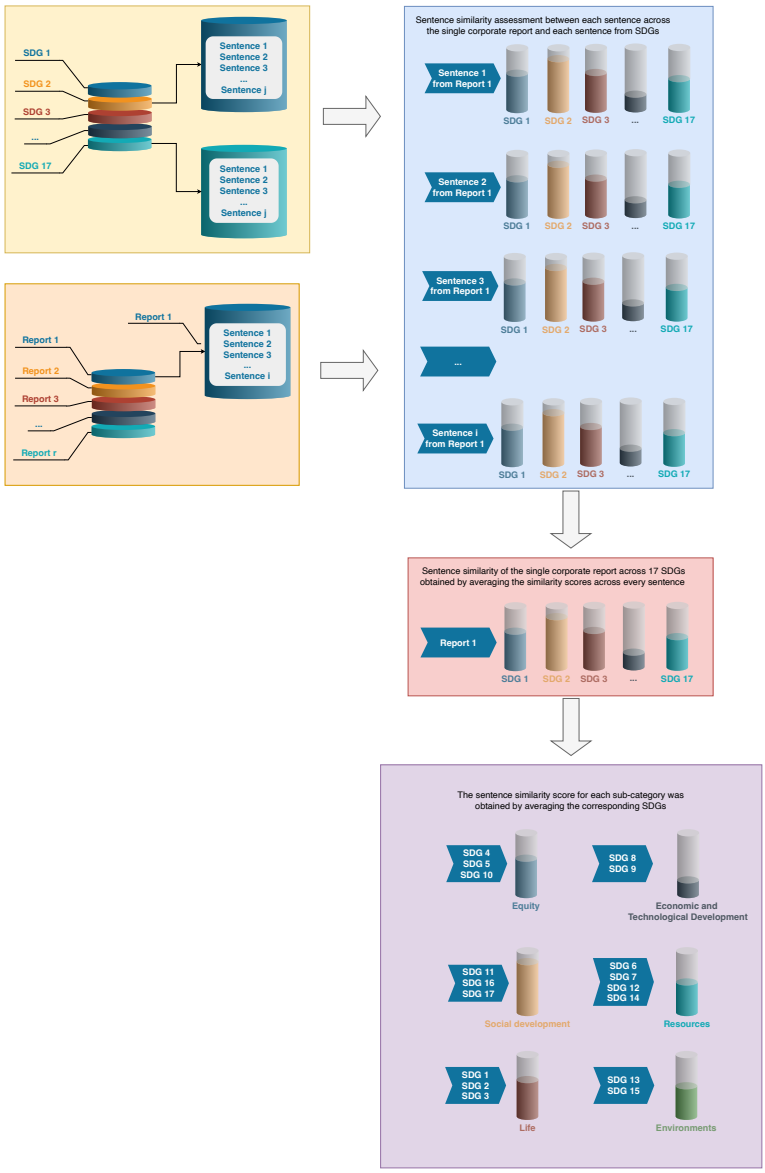

Figure 6.1: Proposed workflow and data pipeline for sentence similarity assessment. SDG reports and corporate reports were split into sentences. Each sentence was compared to the sentences aggregated across the 17 SDG reports. Sentence similarity of each single corporate report was obtained by averaging the similarity scores across every sentence. Finally the sentence similarity score for each SDG sub-category was obtained by aggregating the SDG sub-category (i.e., Equity, Economic and Technical Development, Social Development, Resources, Life, Environment) and averaging the corresponding sentence similarity scores.

No	DocID	File Name	Sentence
0	1	Asda_2020.pdf	Our action on sustainability supports the...
1	1	Asda_2020.pdf	In particular, our efforts are contributing to...
2	1	Asda_2020.pdf	For example, our work to tackle food poverty...
3	1	Asda_2020.pdf	Our CCFB strategy covers every aspect of...
4	1	Asda_2020.pdf	It also covers International Procurement...
...	...	...	...
168272	84	Toyota_2023.pdf	Environmental Data [O] Remanufactured...
168273	84	Toyota_2023.pdf	Management of significant waste...
168274	84	Toyota_2023.pdf	Operations and suppliers in which...
168275	84	Toyota_2023.pdf	Assessment of the health and safety...
168276	84	Toyota_2023.pdf	...To be updated throughout the year as necessary...

Table 6.2: A total of 168,277 sentences across 84 corporate reports for the 10 companies were obtained from corporate annual reports following data pre-processing.

6.4.4 Sentiment Analysis

Sentiment analysis was carried out to assess the ratio between the positive and negative sentiment scores balance and determine the inclination of organisations to prefer reporting positive rather than negative content (Chapter 4.2). To achieve this objective the *DistilBERT* sentiment classification model was used to perform sentiment analysis of the corporate reports (Chapter 4.2.1). The sentiment score spans between 0 and 1. The closer the score is to 1, the more positive the sentiment is, and the closer the score is to 0, the more negative the sentiment is. A classification of the sentiment score was required and sentences were labelled as either *Positive* or *Negative* depending on whether their sentiment score was above or below the threshold of 0.5 respectively. The ratio of *Positive* labels and *Negative* labels was calculated for each report. Finally, the yearly ratio for each company was plotted to identify any change in the trends which could suggests anomalous behaviour worth it further investigation.

Figure 6.2 summarises the sentiment analysis process.

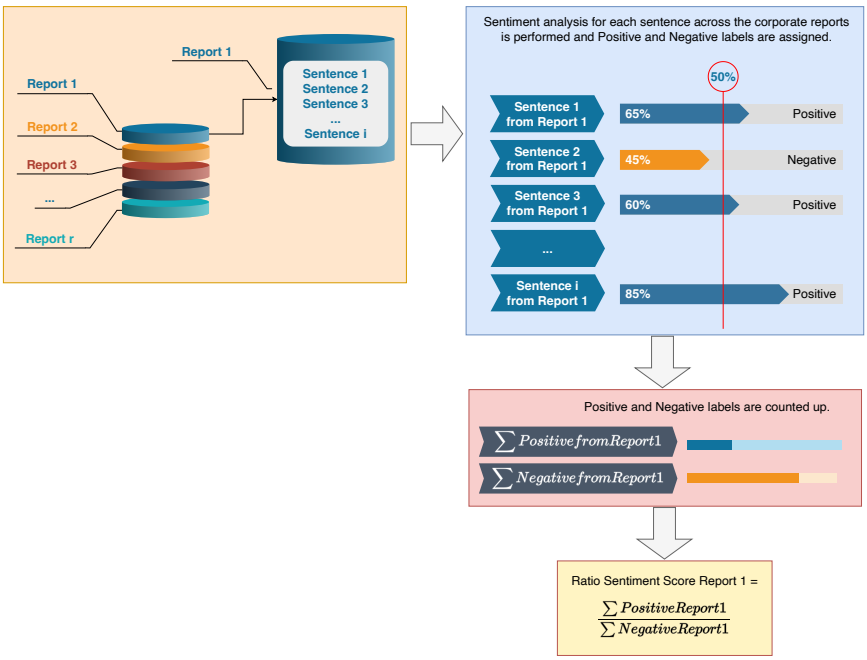

Figure 6.2: Proposed workflow and data pipeline for sentiment analysis assessment for objective four using DistilBERT. The ratio sentiment score was calculated through every sentences across the same report for a specific year. This methodology generated a time series that allowed the identification of suspected behaviour.

No	Category	Goal	Sentence
0	Life	Goal 1	End poverty in all its forms everywhere
1	Life	Goal 1	Despite progress under the MDGs, approx...
2	Life	Goal 1	Over the past decade, markets in devel...
3	Life	Goal 1	Certain groups are disproportionately...
4	Life	Goal 1	These include women, persons with disabil...
...	...	...	...
636	Social development	Goal 17	Enhance the global partnership for...
637	Social development	Goal 17	Encourage and promote effective...
638	Social development	Goal 17	Data, monitoring and accountability...
639	Social development	Goal 17	... enhance capacity-building...
640	Social development	Goal 17	... build on existing initiatives...

Table 6.3: A total of 641 sentences across 17 Sustainable Development Goals were obtained from corporate annual reports following data pre-processing.

Chapter 7

Findings and Discussion

7.1 Objective One: How Opinion Mining of News can Influence Stock Performance

Unlike Bapat et al. (2022) who carried out their research across four companies representing specific industrial sectors, this project explored 10 companies split across two industrial domain sectors.

Company	Average Sentiment Analysis Composite Score	Average Stock price change (in percentage)
Asda	0.376	0.434
Ford	-0.178	-0.023
Marks & Spencer	0.185	0.379
Ocado	-0.132	1.725
Polestar	-0.183	0.752
Sainsbury's	0.366	-0.275
Stellantis	-0.281	-0.677
Tesco	-0.126	0.204
Tesla	-0.100	0.432
Toyota	0.096	-0.430

Table 7.1: Correlations between the average sentiment analysis composite score for ESG-related news and the average stock prices. The news articles and stock prices cover a 90-day period from 1.05.2024 to 31.07.2024.

Following the correlations obtained in Table 7.1, the following clustering is pro-

posed and identified as follows: Ford and Stellantis had both a positive correlation and are known for being traditional automakers which are transitioning towards the electric vehicles (EV) technology. Whereas Polestar, Tesla and Toyota, with negative correlation, are more focused on sustainability and technology which have already made significant progress in the EV field. Contrary to the common belief that Tesla is the pioneer of the modern highway-capable full electric vehicles, Toyota has been playing a big part in the EV evolution as well. For example the RAV4 EV was the first generation of modern full battery vehicles manufactured by Toyota from 1996 until 2003 (Callahan 2024). Asda and Marks & Spencer had both a positive correlation and are both companies that have made significant efforts towards sustainability. Asda and Marks & Spencer, along with Sainsbury's, stand out as the only supermarket chains in the group of this study with dedicated sustainability reports and initiatives, focusing on reducing environmental impact and improving ethical sourcing, stated on "Plan A Reports" and "Environmental, Social & Governance Report" respectively for Asda and Marks & Spencer. On the other hand, Tesco, Sainsbury's and Ocado rank first for having a significant presence on online grocery delivery by leading the net sales in the UK in 2023 (Statista.com 2024). These results, that might appear in contradiction, should not dissuade corporations from taking part in sustainable investing and responsible initiatives but instead to emphasise the complexity of the subject and promote a constructive collaboration between investors and organisations aimed to reinforce the significance of sustainable initiatives.

7.2 Objective Two: Predict Stock Market Volatility Using ESG-Related News

Whilst the study by (Deveikyte et al. 2022) focused on a single index (i.e., FTSE100) as the sole research subject, this work explored 10 individual stocks of companies. Following the results obtained from the Granger's causality testing the analysis focused on Marks & Spencer and Tesla. Table 7.5 shows the accuracy and precision for random forest and logistic regression.

For Marks & Spencer the random forest model performed slightly better than logistic regression in terms of accuracy (0.75 against 0.71), and significantly better in terms of precision (1.00 against 0.71). This suggests that whilst random forest tended to identify more true positive leading to less false positive, logistic regression was slightly more inclined to misclassification. The random forest is an ensemble classification model in which the model tends to reduce the variance (Chapter 4.11). In this particular case, the precision was equal to 1.00 and suggested that the model was able to correctly classify

positive cases. However, the high probability value could also suggest a possible overfitting (i.e., low bias, high variance) where the models are not able to generalise out of sample. On the other hand, the logistic regression model, which is a generalised linear model hence simpler than a random forest, might be prone to higher bias, as it is not flexible enough in capturing complex relationships, and lower bias. This would explain the under performance in precision of the logistic regression but with similar accuracy to random forest.

For Tesla the logistic regression model outperformed in both accuracy (0.54 against 0.27) and precision (0.54 against 0.33). This suggests that for Tesla, logistic regression was able to capture the patterns more effectively, while random forest struggled to make accurate predictions. The low performance of the random forest model could be justified by high variance (i.e., overfitting to the training data but the model fails to generalise to new data) despite random forest models tend to reduce the variance. The fact that logistic regression performed better than random forest, it might due to its simpler nature and it might imply that the underlying relationships in Tesla data might be more linear.

In a real world correlation, Marks & Spencer is a traditional retail company, characterised by a volatility lower than Tesla (Table 7.3), and with stable and predictable business operations that random forest can capture effectively. On the other hand, the lower performance of logistic regression might reflect its inability to fully capture the complicated dynamics of this sector.

In contrast to Marks & Spencer, Tesla operates in a more dynamic and fast paces industry, characterised by a volatility higher than Marks & Spencer (Table 7.3), and with factors like innovation and regulatory changes that might influence the underlying relationships. Logistic regression, which technically performs well with linear and simpler relationship, outperformed random forest, and this behaviour suggests that random forest struggles with the high volatility of Tesla.

In conclusion, the results demonstrated that random forest classifiers are more suitable for companies with low volatility and in an established industry. Whereas, logistic regression classifiers are more suitable for companies with high volatility and in a more dynamic industry.

No	Company	Symbol	Date	Close	Daily return
0	Asda	WMT	2024-05-01	58.85	NaN
1	Asda	WMT	2024-05-02	59.71	0.015
2	Asda	WMT	2024-05-03	59.82	0.0018
...	...	...	...	...	...
633	Toyota	TM	2024-07-29	192.48	-0.00021
634	Toyota	TM	2024-07-30	193.11	0.0033
635	Toyota	TM	2024-07-31	193.55	0.0023

Table 7.2: Daily stock market returns calculated for the ten companies using daily closing stock price from 1.05.2024 to 31.07.2024 (Chapter 4.5)

Company	Volatility %	Quartile
Polestar	1.047	4
Ocado	0.745	4
Tesla	0.566	4
Ford	0.477	3
Stellantis	0.295	3
Marks & Spencer	0.250	2
Toyota	0.224	2
Sainsbury's	0.195	1
Asda	0.187	1
Tesco	0.121	1

Table 7.3: Volatility distribution calculated for each stock for 90 days from 1.05.2024 to 31.07.2024 (Chapter 4.6)

7.3 Objective Three: Determine the Influence of ESG-Related News on Greenwashing-Related News Articles

Granger's causality testing suggested that influence of ESG-related news on news articles containing keywords associated with greenwashing was detected for Ford in the period of observation. Whereas the opposite relationship was detected for Marks & Spencer.

Company	p-value
M&S	0.00353
Tesla	0.0209

Table 7.4: Significant probability values obtained with Granger's causality testing. This suggested that ESG-related news for Marks & Spencer and Tesla influenced the respectively stock market performance.

Figure 7.3: Word clouds for different topics for Marks & Spencer. The clouds suggested the following topics: 1. fragrance, 2. fashion, 3. clothing, 4. animals, 5. financial investment, 6. fashion collection, 7. miscellaneous.

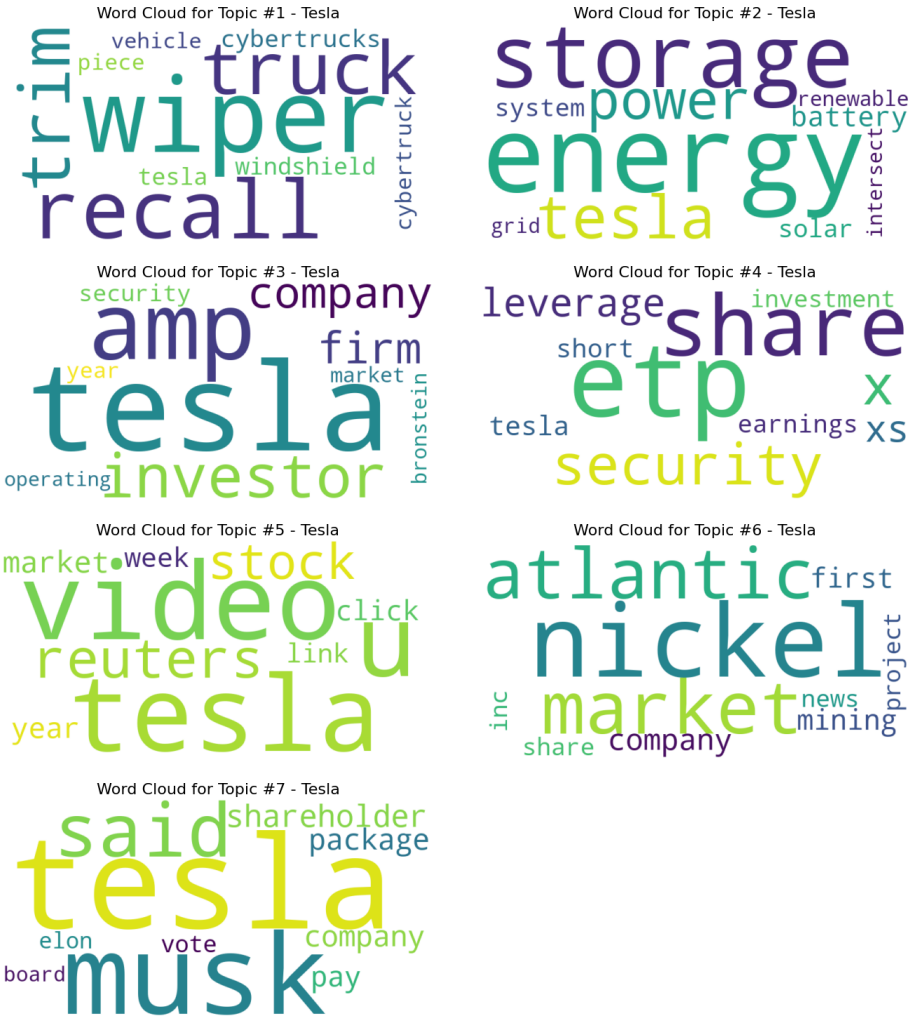

Figure 7.4: Word clouds for different topics for Tesla. The clouds suggested the following topics: 1. vehicle, 2. electric energy, 3. market, 4. financial investment, 5. media, 6. miscellaneous, 7. company identity.

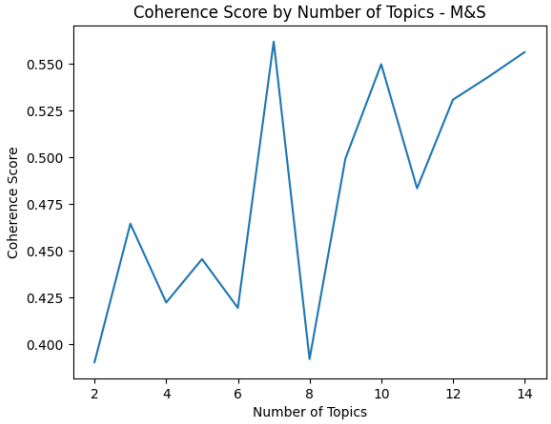

Figure 7.1: The coherence score for M&S suggested that an appropriate number of topics for the corpus provided to the model was seven

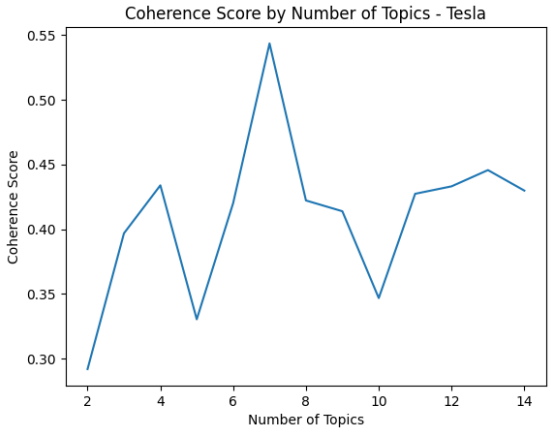

Figure 7.2: The coherence score for Tesla suggested that an appropriate number of topics for the corpus provided to the model was seven.

A proposed interpretation of the results follows. The automotive industry is a sector that is heavily scrutinised for its direct environmental implications (Wagemans 2023). Ford is a leader in this domain (Carlier 2024) and any ESG initiative is crucial in shaping the expectation and the perceptions of the public around sustainability and greenwashing implications. In a real world scenario, if Ford is seen genuinely working towards sustainable solutions, such as developing new electric vehicles and public sentiment is positive, this could reduce the likelihood of Ford for being accused of greenwashing. On the other hand, Marks & Spencer is one of the major retailer in the UK, and it is acknowledged for its sustainability initiatives such as "Plan A" program aimed at

	Random forest		Logistic regression	
	Accuracy	**Precision**	**Accuracy**	**Precision**
M&S	0.75	1.00	0.71	0.71
Tesla	0.27	0.33	0.54	0.54

Table 7.5: After training random forest and logistic regression models for M&S and Tesla, accuracy and precision on the test subset of data were obtained. Random forest performs slightly better with dataset from Marks & Spencer. Whereas logistic regression performs better with the dataset from Tesla.

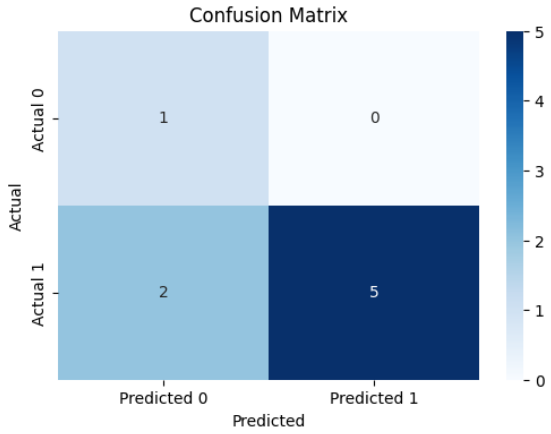

Figure 7.5: Confusion matrix that shows the performance of the Random Forest classifier used on the test subset of data for Marks & Spencer.

reducing its environmental impact. Nevertheless, consumers are becoming more conscious of the impact on the environment of the products they buy. In a real world scenario, if Marks & Spencer is found guilty of greenwashing due to misleading sustainability claims about its products, this could lead to skepticism towards Marks & Spencer sustainability goals and its entire ESG initiatives.

7.4 Objective Four: Detect Corporate Greenwashing

7.4.1 Sentence Similarity and Thematic Analysis

The similarity score is a metric that determines the thematic structure of each corporate reports in order to measure how much each corporate report aligns with the six SDG sub-categories. The plots in Figure 7.11 were created using the Python Matplotlib library and show the trends of the thematic structure across the period of observation

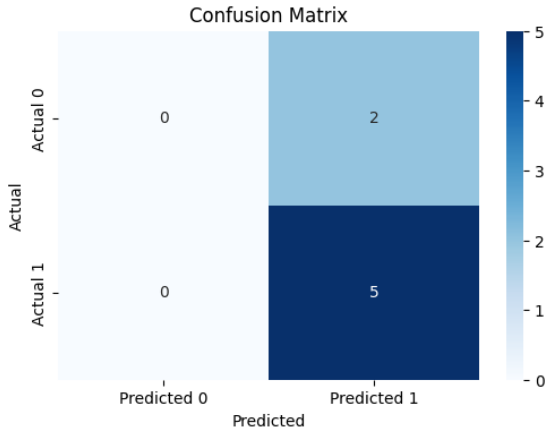

Figure 7.6: Confusion matrix that shows the performance of the Logistic Regression classifier used on the test subset of data for Marks & Spencer.

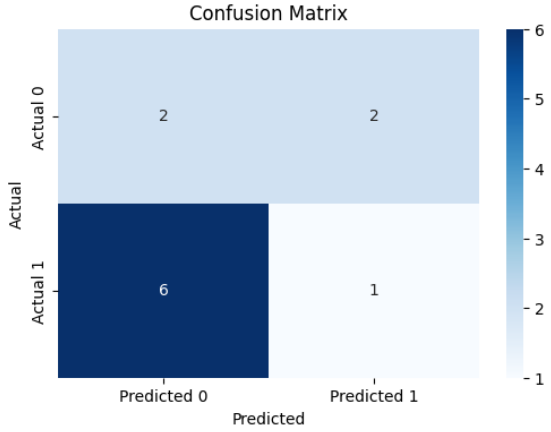

Figure 7.7: Confusion matrix that shows the performance of the Random Forest classifier used on the test subset of data for Tesla.

(matplotlib 2024).

To validate the results, a closer examination was carried out on companies with similarities. Polestar and Tesla for the automotive sector, as well as Ocado and Tesco for the retail industry were chosen.

Polestar and Tesla are both primarily focused on manufacturing fully electric vehicles, driven by the commitment to promoting an alternative way of transportation that contributes to reduce the impact on the environment. Therefore the expectation is for the *"Environments"* sub-category to be predominant against the other categories. This expectation is fulfilled and Figure 7.9 demonstrates that the 'Environments' topic is

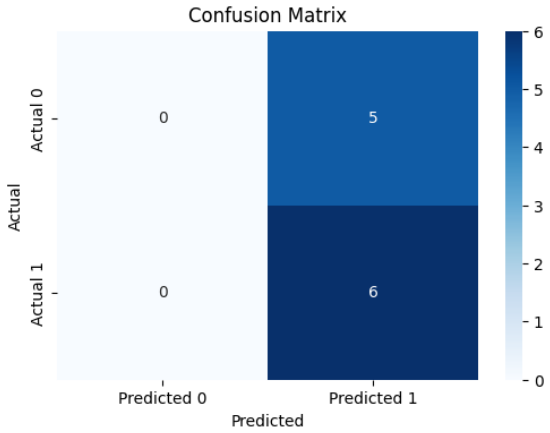

Figure 7.8: Confusion matrix that shows the performance of the Logistic Regression classifier used on the test subset of data for Tesla.

consistently on top. Interestingly, also the overall thematic structure is the same across Polestar and Tesla with a few exceptions.

Both Ocado and Tesco provided annual financial reports over the 10 years of observation (Chapter 5.4). Their thematic structure is relatively similar with the "Economic" and the "Equity" topics resulting to be prominent compared to the rest of the topics. Interestingly, the same pattern can be seen between 2015 and 2017 for Sainsbury's as only annual financial reports were available. The spike at 2018 can be justified as the move from corporate financial reports to sustainability reports hence a change in the typology of reports and communication, resulting in a change in thematic structure.

7.4.2 Sentiment Analysis

The line graph in Figure 7.12 shows the trends of the ratio of positive content against negative content from the reports to help determine if changes in the trends occur across the period of 10 years of observation. Trends across companies are different. Ocado's trend was consistent all along the entire timeline which may suggest that greenwashing was unlikely. The line was pretty flat with a very little variation (average= 1.07, standard deviation= 0.14) which suggests that positive and negative sentences were well balanced. Also, Tesco's trend remained consistent from 2015 to 2023, with a significant drop of the ratio after 2014, reducing to nearly half of its previous level. However, it is worth noticing that Ocado and Tesco's annual reports were not dedicated sustainability reports but instead were financial reports with sustainability claims incorporated and this was probably the reasons why the trend was particular flat around the optimal

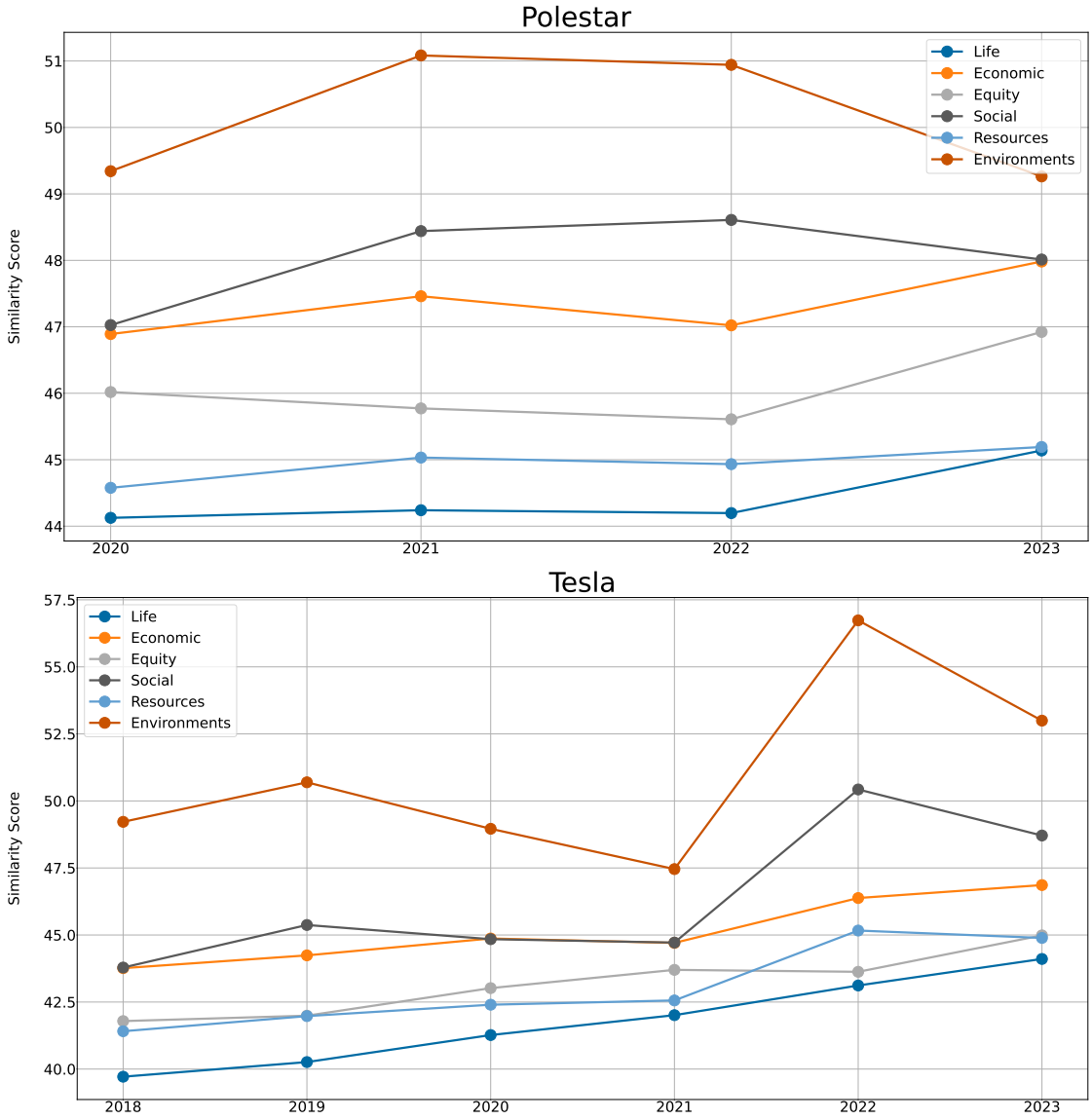

Figure 7.9: Thematic trends as result of sentence similarity for Polestar and Tesla. The sequence of the topics across the two automotive companies were similar and consistent along the period of observation. These result may suggest that the examined companies paid attention to sustainability issues that are relevant to their specific sector.

balance. Tesla's trend was also particularly consistent and balanced (average= 1.456, standard deviation= 0.27) with a slight increment of the ratio from 2022 to 2023. Marks & Spencer, Stellantis and Toyota's trends varied but overall they maintained an average ratio of positive comments that was nearly twice as high as that of negative comments.

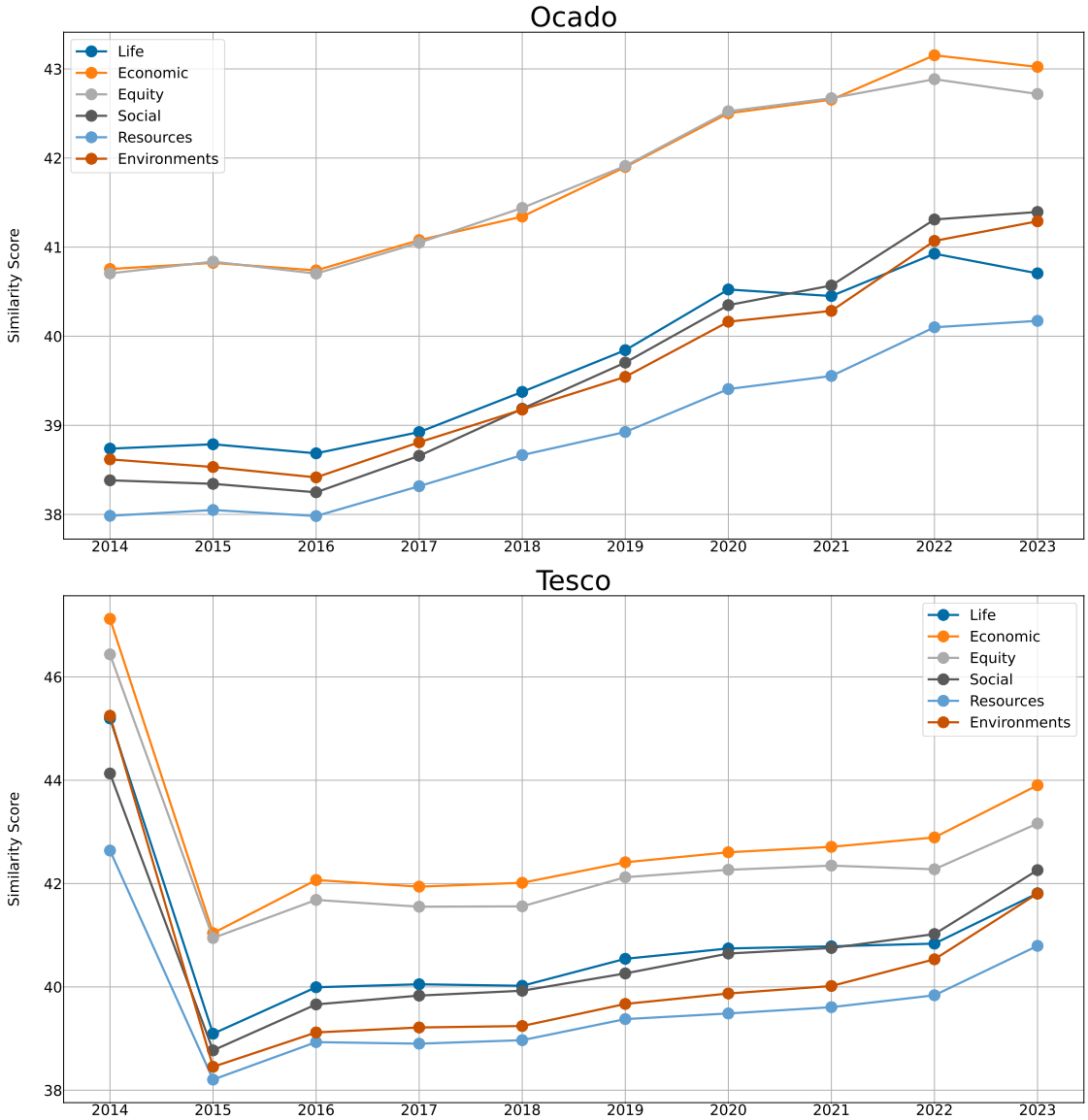

Figure 7.10: Thematic trends as result of sentence similarity for Ocado and Tesco. The sequence of the topics across the two retail companies were similar and consistent along the period of observation. These result may suggest that the examined companies paid attention to sustainability issues that are relevant to their specific sector.

Interestingly, Marks & Spencer, Stellantis and Toyota's had also in common the fact that their corporate annual reports were dedicated sustainability reports. Asda's first sustainability report issued in 2020 had a ratio of positive statements nearly five times as much as negative statements. From the following year the positive comments dropped

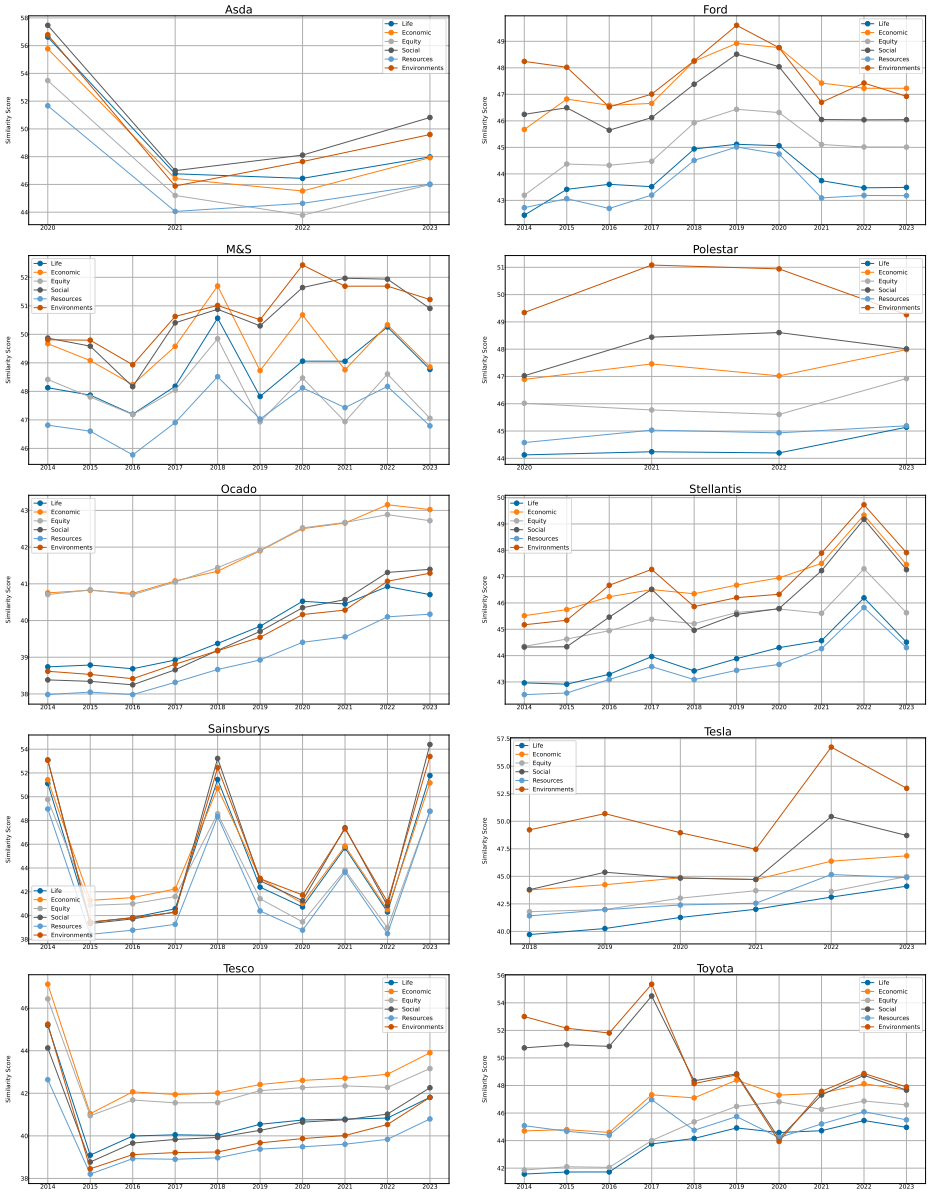

Figure 7.11: Thematic trends as result of sentence similarity for the 10 companies. The results show that the least sustainability themes across the retail and automotive sectors were "Resources" and "Life" respectively. While the top sustainability theme for automotive was "Environment". Whilst "Social" was the top theme for the retail companies that published dedicated corporate sustainability reports, "Economic" for the retail companies that published annual financial reports.

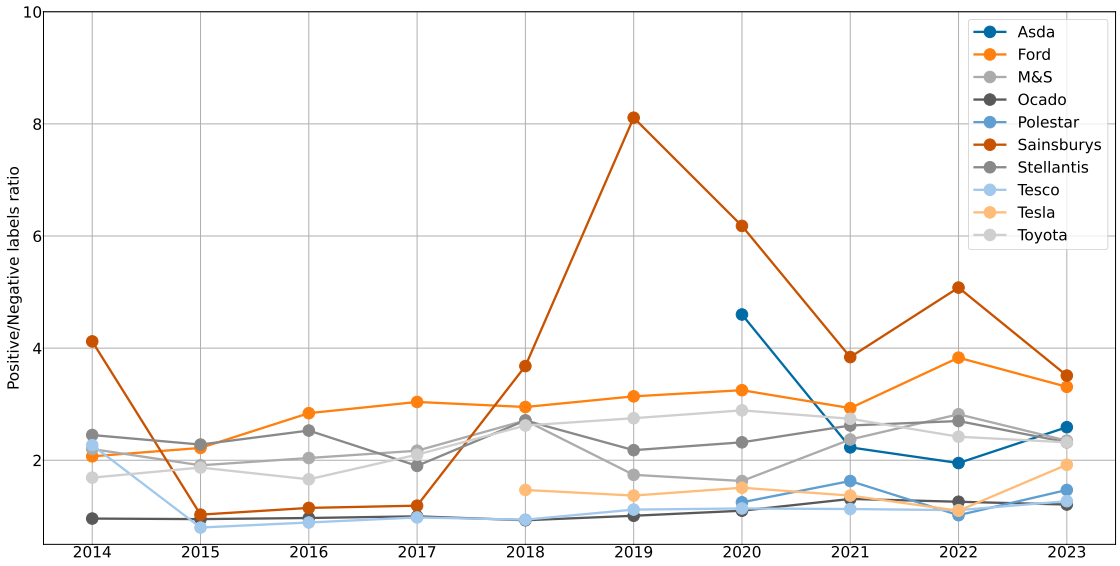

Figure 7.12: Sentiment score ratios across 10 companies from 2014 to 2023. Sainsbury's prominent peak from a ratio value of nearly four in 2018, and a ratio of eight in 2019, which referred to the 2019/2020 sustainability report, could potentially suggest greenwashing.

to remain around twice as high as negative comments. Ford's ratio of positive sentence was twice as much as negative statements in 2014 and increased across the entire period until a ratio of nearly three times. Sainsbury's prominent peak from a ratio value of nearly four in 2018, and a ratio of eight in 2019, which referred to the 2019/2020 sustainability report, could potentially suggest greenwashing. It is worth noticing that from 2015 and 2017, Sainsbury's reports were annual financial statements rather than sustainability, which explains the drop of sentiment ratio during this period. To validate the result fact-checking of news articles preceding the occurrence of the irregularity was conducted on the Internet. The year before the publication of the sustainability report, in March 2019 Sainsbury's was challenged by Greenpeace over its policy for reducing the plastic package. The environmental organisation defined the supermarket chain as "the worst in class of all major UK supermarkets for cutting plastic packaging" (Greenpeace 2019). It is possible that Sainsbury's wanted to emphasize the positive content in its 2019/2020 sustainability report to improve its reputation following the strong criticism occurred the year before.

Chapter 8

Conclusions and Recommendations

All of the objectives were successfully achieved. In particular objective one provided the correlations between sentiment scores of ESG-related news and the average variation of the stock prices for each company. Objective two was also achieved and it demonstrated how the ESG-related news can be used to predict the next day stock price for two companies and eventually the volatility. Objective three demonstrated that ESG-related news can influence greenwashing-related news and vice versa. Finally, objective four was surprisingly achieved as our analysis of the corporate reports led to the discover of a company that might be seen as conducting greenwashing practices.

8.1 Objective One: How Opinion Mining of News can Influence Stock Performance

As general consensus and regulatory pressures grow for public companies to integrate corporate responsibility practices and operate in a sustainable way, it seems important to assess how investors perceive the efforts of those companies in the environmental, social, and governance (ESG) areas and how these efforts impact their performance. To assess how ESG-related news correlated to stock prices of 10 public companies, the Thomson Reuters API (Eikon) was used to download ESG-related news stories as well as stock opening prices across 90 days. 990 news articled were gathered and to assess the polarity and composite score of them, sentiment analysis was carried out using the pre-trained NLP ProsusAI/FinBERT model trained specifically to analyse sentiment of financial text. The daily composite score and daily open price variation were determined for each sample stock and the averages across the period of observation were calculated. The results obtained helped determine either positive or inverse relation-

ships as follows: Asda, Marks & Spencer, Ford and Stellantis stock prices have been impacted in a positive correlation by the ESG news although the first two increased and the last two decreased their average stock price. Whereas Ocado, Tesco, Saisbury's, Polestar, Tesla and Toyota were inversely correlated to the ESG-related news.

8.2 Objective Two: Predict Stock Market Volatility Using ESG-Related News

Correlation analysis to compare sentiment scores from ESG-related news against changes in the stock returns was carried out to predict the next day stock market return and volatility. Thomson Reuters API (Eikon) was used to gather ESG-related news articles as well as stock closing prices across 90 days with a daily granularity. The Granger's causality testing confirmed the influence of ESG-related news articles on stock performance for Marks & Spencer and Tesla. Two supervised classification models (i.e., logistic regression and random forest) were trained and their performances were compared. The results obtained permitted to determine a relationship between the classification model and the industry where a specific company operates. The results suggested that random forest classifiers are more suitable for companies with low volatility and in an established industry. Whereas, logistic regression classifiers are more suitable for companies with high volatility and in a more dynamic industry.

8.3 Objective Three: Determine the Influence of ESG-Related News on Greenwashing-Related News Articles

Sentiment scores from both ESG-related news and news articles containing keywords associated to greenwashing was carried out to assess their reciprocal correlation. The daily sentiment score time series of ESG-related news stories calculated from the study of objective two was reused. To determine the keywords related to greenwashing the benchmark created by Dumitrescu et al. (2023) was used. Thomson Reuters API (Eikon) was required to gather the articles containing keywords associated to greenwashing over 90 days. VADER was used to calculate the daily sentiment score time series for the greenwashing-related news. Granger's causality testing suggested that influence of ESG-related news on greenwashing-related news articles was detected for Ford in the period of observation. Whereas the opposite relationship was detected for Marks & Spencer.

This type of analysis benefits organisations as it provides valuable insights regarding corporate strategy, relations with the public and risk management. It could help

companies like Ford and M&S understand how their sustainability initiatives and communications are perceived by consumers, manage risks on reputation, and enhance communication to improve trust of consumers and stakeholders. It is worth noticing how the type of industry and clients can influence Granger's testing results.

8.4 Objective Four: Detect Corporate Greenwashing

In the recent years, the general public and consumers have become more sensitive to thematic related to sustainability, and their decisions reflect this tendency when buying services and products. Companies and organisations have recognised this trend and have started publishing annual sustainability reports, along with their traditional annual financial statements, to outline and communicate their sustainability strategies, priorities and practices. To deal with the large amount of information contained in the reports, different NLP models have been deployed to identify the concepts contained in the documents. This project proposes an alternative approach to the most used "word frequency-based method". To detect greenwashing, 83 annual corporate reports from five companies amongst the automotive and other five companies amongst the retail sectors, were collected between 2014 and 2023 to perform sentiment analysis in combination with sentence similarity against the United Nations Sustainable Development Goals reports used as benchmark. Sentence similarity enabled thematic analysis to extract the patterns of the themes contained in the reports for each company. The results successfully demonstrates how themes had different ranking across the two industrial sectors although the themes patterns were consistent across the same sector except for little adjustments over the 10 years of observation. Also, this confirmed how the companies examined paid attention to sustainability issues that are relevant to their specific sector. In particular, the results showed that the least sustainability themes across the retail and automotive sectors were "Resources" and "Life" respectively. Whilst he top sustainability theme for automotive was "Environment", "Social" is the top theme for the retail companies that have published dedicated corporate sustainability reports, and "Economic" for the retail companies that published annual financial reports. These patterns confirmed previous findings from Kang & Kim (2022a). Additionally, sentiment analysis was carried out to determine the positive and negative sentiment score and the ratio of the sentiment assessed across the years of observation was plotted. Visualisation of the trends helped identify a spike suggesting the excessive use of positive content for *Sainsbury's* which could be flagged as potential greenwashing activity to cover up a scandal happened the year before the publication of the sustainability report. Validation of the results was conducted by fact-checking of news articles.

The findings of this research have successfully proved that this approach can be applied to detect automatically greenwashing by analysing the sustainability reports by sentiment analysis in combination with sentence similarity. This methodology could have multiple purposes and applications as it could be used by investigators to find out companies that are involved in greenwashing as well as by organisations keen to monitor the content of their sustainability reports in order to ensure the ratio of sentiment score is consistent over time and adjust the reports to provide more reliable content and avoid suggesting anomalous behaviours.

Nevertheless, by examining two industrial sectors and five companies in each sector, this research addressed the limitation highlighted by Kang & Kim (2022a) in their paper, which emphasised the need to increase the number of companies being examined within the same industry.

One challenge experienced in achieving this objective was the difficulty in collecting the annual sustainability reports for all companies involved in the study as some of them have published only financial reports. Where there was a combination of both financial and sustainability reports across the period of observation, this was reflected in the sentiment and similarity analysis results where discontinuities were visible due to the different nature of the reports. Also, the specialised topic chosen for this project did not allowed to gather as many news articles as required to obtain further correlations and additional material to analyse. For instance, Eikon API did not return any of the greenwashing-related news for Asda and Sainsbury's during the period of observation. However, despite the limited time series, this study successfully achieved meaningful results.

At the end of this graduation project, it is worth offering a summary of the so-called 'lessons learned': (i) handling numerous files and results generated by the implementation of different Python scripts across four different objectives was necessary. Being mindful of naming the files in a proper way and organising the work in folders optimised the execution of the work; (ii) it took a while to set up the Eikon-Reuters platform to download stock data and news articles. Taking time to approach a new tool was needed and engaging in conversations on appropriate forums to seek assistance by other users was beneficial.

Figures 7.1 and 7.2 summarise an important finding from of our work on objective two: the same quantity of topics was identified for two different companies despite they belonged to different sectors. It would be interesting for a future development to investigate whether the reason of that similarity was a coincidence or if it suggests that for the same type of mean of communication the number of topics converge. In general, extending the number of industries and the period of observation could also provide additional data useful to find further correlations.

Bibliography

Aarsen, T. (2021). Last accessed 14 September 2024.
URL: *https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2*

Alaparthi, S. & Mishra, M. (2021), 'Bert: a sentiment analysis odyssey', Journal of Marketing Analytics **9**(2), 118–126. Last accessed 14 September 2024.
URL: *https://doi.org/10.1057/s41270-021-00109-8*

Bapat, S. R., Kothari, S. & Bansal, R. (2022), 'Sentiment analysis of esg disclosures on stock market'. Last accessed 14 September 2024.
URL: *http://arxiv.org/abs/2210.00731*

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'. Last accessed 14 September 2024.
URL: *https://www.sciencedirect.com/science/article/pii/S1544612322000897*

Blei, D. M. (2012), 'Probabilistic topic models - surveying a suite of algorithms that offer a solution to managing large document archives.', Association for Computing Machinery . Last accessed 14 September 2024.
URL: *https://dl.acm.org/doi/pdf/10.1145/2133806.2133826*

Bollen, J., Mao, H. & Zeng, X.-J. (2010), 'Twitter mood predicts the stock market'. Last accessed 14 September 2024.
URL: *http://arxiv.org/abs/1010.3003*

Brammer, S., Brooks, C. & Pavelin, S. (2006), 'Corporate social performance and stock returns: Uk evidence from disaggregate measures', Management **35**, 97–116. Last accessed 14 September 2024.
URL: *https://www.jstor.org/stable/30137803*

Brookes, G. & McEnery, T. (2018), 'The utility of topic modelling for discourse studies: A critical evaluation', Discourse Studies **21**, 146144561881403. Last accessed 14 September 2024.
URL: *https://journals.sagepub.com/doi/10.1177/1461445618814032*

Callahan, R. (2024), 'History and mystery of toyota's first ev', www.topspeed.com . Last accessed 14 September 2024.
URL: *https://www.topspeed.com/toyota-first-ev-history/*

Carlier, M. (2024), 'Ford motor company in the united states - statistics & facts', statista.com . Last accessed 14 September 2024.
URL: *https://www.statista.com/topics/6154/car-brands-ford/*

Charlotte Portier, James Gomme, N. W. (2015), 'Sdg compass', SDG Compass . Last accessed 14 September 2024.
URL: *https://sdgcompass.org/*

Deveikyte, J., Geman, H., Piccari, C. & Provetti, A. (2022), 'A sentiment analysis approach to the prediction of market volatility', Frontiers . Last accessed 14 September 2024.
URL: *https://www.frontiersin.org/articles/10.3389/frai.2022.836809/full*

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019a), 'Bert', huggingface.co . Last accessed 14 September 2024.
URL: *https://huggingface.co/docs/transformers/en/model_doc/bert*

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019b), 'Bert: Pre-training of deep bidirectional transformers for language understanding'. Last accessed 14 September 2024.
URL: *https://arxiv.org/abs/1810.04805*

Dumitrescu, A., Gil-Bazo, J. & Zhou, F. (2023), 'Defining greenwashing'. Last accessed 14 September 2024.
URL: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4098411*

FinancialTimesAdviser (2021), 'How machine learning is being used to combat greenwashing'. Last accessed 14 September 2024.
URL: *https://www.ftadviser.com/investments/2021/05/17/how-machine-learning-is-being-used-to-combat-greenwashing/?page=1*

Gratton, P. (2024), 'Opening price: Definition, example, trading strategies', investopedia.com . Last accessed 14 September 2024.
URL: *https://www.investopedia.com/terms/o/openingprice.asp*

Greenpeace (2019), 'Greenpeace calls out sainsbury's as worst in class on plastics'. Last accessed 14 September 2024.
URL: *https://www.greenpeace.org.uk/news/greenpeace-calls-sainsburys-worst-class-p*

HuggingFace (2021), 'Sentence transformers', huggingface.co . Last accessed 14 September 2024.
URL: *https://huggingface.co/tasks/sentence-similarity*

huggingface.co (2024a), 'Distilbert', huggingface.co . Last accessed 14 September 2024.
URL: *https://huggingface.co/docs/transformers/en/model_doc/distilbert*

huggingface.co (2024b), 'Finbert', huggingface.co . Last accessed 14 September 2024.
URL: *https://huggingface.co/ProsusAI/finbert*

Hutto, C. (2020), 'vadersentiment', pypi.org . Last accessed 14 September 2024.
URL: *https://pypi.org/project/vaderSentiment/*

Investopedia.com (2023), 'What the market open tells you', www.investopedia.com . Last accessed 14 September 2024.
URL: *https://www.investopedia.com/articles/trading/11/analyzing-market-open.asp*

Kang, H. & Kim, J. (2022a), 'Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods', Applied Sciences (Switzerland) **12**. Last accessed 14 September 2024.
URL: *https://www.researchgate.net/publication/361014122_Analyzing_and_Visualizing_Text_Information_in_Corporate_Sustainability_Reports_Using_Natural_Language_Processing_Methods*

Kang, H. & Kim, J. (2022b), 'Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods', GitHub . Last accessed 14 September 2024.
URL: *https://github.com/llbtl/paper_ssm01/*

Kornreich, M. (2022), '"how does greenwashing affect firm value? empirical analysis from stock market reaction and companies' performance"'. Last accessed 14 September 2024.
URL: *http://hdl.handle.net/2078.1/thesis:36587*

Krafft, J., Saito, R., Hallberg, A. & Hauff, J. (2014), 'Greenwashing an experimental study about the effects of misleading and deceptive environmental claims in advertising'. Last accessed 14 September 2024.
URL: *https://core.ac.uk/download/pdf/43558099.pdf*

Kwok, J., Choi, H. H., Kong, A., Newport, E., Bao, H., Brigden, K., Choi, E., Cray, C., Dallos, G., Fletcher, B., Gehr, B., Hong, H., Huang, K., Liu, W., Miller, K., Read, D., Shiohata, M., Stephan, B., Zhang, Y. & Zheng, M. (2023), 'Automobile environmental guide (2023 edition) - a comparative analysis of decarbonisation efforts by global automakers'. Last accessed 14 September 2024.
URL: *https://www.greenpeace.org/static/planet4-eastasia-stateless/2023/10/327c0c30-auto-environmental-guide-2023_greenpeaceea.pdf*

La Torre, M., Mango, F., Cafaro, A. & Leo, S. (2020), 'Does the esg index affect stock return? evidence from the eurostoxx50', Sustainability **12**(16). Last accessed 14 September 2024.
URL: *https://www.mdpi.com/2071-1050/12/16/6387*

Lee, R. & Kim, J. (2021), 'Developing a social index for measuring the public opinion regarding the attainment of sustainable development goals', Social Indicators Research **156**, 1–21. Last accessed 14 September 2024.
URL: *https://www.researchgate.net/publication/349393085_Developing_a_Social_Index_for_Measuring_the_Public_Opinion_Regarding_the_Attainment_of_Sustainable_Development_Goals*

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. & Zettlemoyer, L. (2019a), 'Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension', Arxiv . Last accessed 14 September 2024.
URL: *http://arxiv.org/abs/1910.13461*

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. & Zettlemoyer, L. (2019b), 'facebook/bart-large-cnn', Facebook AI . Last accessed 14 September 2024.
URL: *https://huggingface.co/facebook/bart-large-cnn*

Liew, W. T., Adhitya, A. & Srinivasan, R. (2014), 'Sustainability trends in the process industries: A text mining-based analysis', Computers in Industry **65**(3), 393–400. Last accessed 14 September 2024.
URL: *https://www.sciencedirect.com/science/article/pii/S0166361514000207*

Loper, E. & Bird, S. (2002), 'NLTK: The natural language toolkit'. Last accessed 14 September 2024.
URL: *https://aclanthology.org/W02-0109*

matplotlib (2024), 'matplotlib', matplotlib . Last accessed 14 September 2024.
URL: *https://matplotlib.org/*

McKie, J. X. (2016). Last accessed 14 September 2024.
URL: *https://github.com/pymupdf/PyMuPDF*

Peramunetilleke, D. & Wong, R. (2001), 'Currency exchange rate forecasting from news headlines'. Last accessed 14 September 2024.
URL: *https://www.cse.unsw.edu.au/~wong/Papers/desh.pdf*

Readshaw, J. & Giani, S. (2021), 'Using company-specific headlines and convolutional neural networks to predict stock fluctuations', Neural Computing and Applications **33**, 17353–17367. Last accessed 14 September 2024.
URL: *https://link.springer.com/article/10.1007/s00521-021-06324-9*

Rehurek, R. (2024), 'Python framework for fast vector space modelling', pypi.org . Last accessed 14 September 2024.
URL: *https://pypi.org/project/gensim/*

scikit learn.org (2024a), 'Logisitc regression', scikit-learn.org . Last accessed 14 September 2024.
URL: *https://scikit-learn.org/1.5/modules/generated/sklearn.linear_model.LogisticRegression.html*

scikit learn.org (2024b), 'Random forest classifier', scikit-learn.org . Last accessed 14 September 2024.
URL: *https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html*

Statista.com (2024), 'Leading grocery online stores in the united kingdom in 2023, by e-commerce net sales', www.statista.com . Last accessed 14 September 2024.
URL: *https://www.statista.com/forecasts/870373/top-online-stores-food-beverages-united-kingdom-ecommercedb*

statsmodels.org (2024), 'grangercausalitytests', statsmodels.org . Last accessed 14 September 2024.
URL: *https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.grangercausalitytests.html*

Steven Bird, Ewan Klein, E. L. (2009), Natural Language Processing with Python, O'Reilly Media, Inc.

Székely, N. & vom Brocke, J. (2017), 'What can we learn from corporate sustainability reporting? deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique', PLOS ONE **12**, 1–27. Last accessed 14 September 2024.
URL: *https://doi.org/10.1371/journal.pone.0174807*

Tata, S. & Patel, J. (2007), 'Estimating the selectivity of tf-idf based cosine similarity predicates', Sigmod Record **36**.

Vitto, A. D., Marazzina, D. & Stocco, D. (2023), 'Esg ratings explainability through machine learning techniques', Annals of Operations Research . Last accessed 14 September 2024.
URL: *https://link.springer.com/article/10.1007/s10479-023-05514-z*

Vu, T. T., Chang, S., Ha, Q. T. & Collier, N. (2012), 'An experiment in integrating sentiment features for tech stock prediction in twitter'. Last accessed 14 September 2024.
URL: *https://dev.twitter.com/docs/streaming-apis/streams/public*

Wagemans, S. (2023), 'Decarbonisation strategies move beyond the tailpipe', Automotive World . Last accessed 14 September 2024.
URL: *https://www.automotiveworld.com/articles/decarbonisation-strategies-move-beyond-the-tailpipe/*

Wang, X., Yuen, K. F., Wong, Y. D. & Li, K. X. (2020), 'How can the maritime industry meet sustainable development goals? an analysis of sustainability reports from the social entrepreneurship perspective', Transportation Research Part D: Transport and Environment **78**, 102173. Last accessed 14 September 2024.
URL: *https://www.sciencedirect.com/science/article/pii/S1361920919309472*

Whelan, T., Atz, U. & Clark, C. (2015), 'Esg and financial performance: Uncovering the relationship by aggregating evidence from 1,000 plus studies'. Last accessed 14 September 2024.
URL: *https://sri360.com/wp-content/uploads/2022/10/NYU-RAM_ESG-Paper_2021-2.pdf*

Wu, J., Guo, S., Huang, H., Liu, W. & Xiang, Y. (2018), 'Information and communications technologies for sustainable development goals: State-of-the-art, needs and perspectives', IEEE Communications Surveys and Tutorials **20**(3), 2389–2406.

Zhang, X., Zhao, T., Wang, L. & Dong, Z. (2022), 'Does fintech benefit financial disintermediation? evidence based on provinces in china from 2013 to 2018', Journal of Asian Economics **82**, 101516. Last accessed 14 September 2024.
URL: *https://www.sciencedirect.com/science/article/pii/S1049007822000720*

Appendix A

Repository

The source code and corporate reports used for this study are available at:
`https://github.com/Birkbeck/msc-projects-2023-4-lucalto`
(Last accessed on 2 October 2024).

Appendix B

List of Corporate Reports

No	Sector	Company	Year	Title
1	Automotive	Ford	2023	Integrated Sustainability and Financial Report 2023
2			2022	Integrated Sustainability and Financial Report 2022
3			2021	Integrated Sustainability and Financial Report 2021
4			2020	Sustainability Report 2020
5			2019	Sustainability Report 2018/19
6			2018	Sustainability Report 2017/18
7			2017	Sustainability Report 2016/17
8			2016	Sustainability Report 2015/16
9			2015	Sustainability Report 2014/15
10			2014	Sustainability Report 2013/14
11	Automotive	Polestar	2023	Sustainability Report 2023
12			2022	Sustainability Report 2022
13			2021	Sustainability Report 2021
14			2020	Sustainability Report 2020
15	Automotive	Stellantis	2023	Corporate Social Responsibility Report 2023
16			2022	Corporate Social Responsibility Report 2022
17			2021	Corporate Social Responsibility Report 2021
18			2020	Sustainability Report 2020
19			2019	Sustainability Report 2019
20			2018	Sustainability Report 2018
21			2017	Sustainability Report 2017
22			2016	Sustainability Report 2016
23			2015	Sustainability Report 2015
24			2014	Sustainability Report 2014
25	Automotive	Tesla	2023	Impact Report 2023
26			2022	Impact Report 2022

Continued on next page

Table B.1 continued from previous page

No	Sector	Company	Year	Title
27			2021	Impact Report 2021
28			2020	Impact Report 2020
29			2019	Impact Report 2019
30			2018	Impact Report 2018
31	Automotive	Toyota	2023	Sustainability Data Book 2023
32			2022	Sustainability Data Book 2022
33			2021	Sustainability Data Book 2021
34			2020	Sustainability Data Book 2020
35			2019	Sustainability Data Book 2019
36			2018	Sustainability Data Book 2018
37			2017	Environmental Report 2017
38			2016	Environmental Report 2016
39			2015	Environmental Initiatives 2015
40			2014	Environmental Initiatives 2014
41	Retail	Asda	2023	Our Brighter Living Report 2023
42			2022	Environmental, Social & Governance Report 2022
43			2021	Environmental, Social & Governance Report 2021
44			2020	Asda Creating Change for Better 2020
45	Retail	M&S	2023	Sustainability Report 2023
46			2022	Sustainability Report 2022
47			2021	Plan A Report 2021
48			2020	Plan A Report 2020
49			2019	Plan A Performance Update 2019
50			2018	Plan A Report 2018
51			2017	Plan A Report 2017
52			2016	Plan A Report 2016
53			2015	Plan A Report 2015

Continued on next page

Table B.1 continued from previous page

No	Sector	Company	Year	Title
54			2014	Plan A Report 2014
55	Retail	Ocado	2023	Annual Report and Accounts 2023
56			2022	Annual Report and Accounts 2022
57			2021	Annual Report and Accounts 2021
58			2020	Annual Report and Accounts 2020
59			2019	Annual Report and Accounts 2019
60			2018	Annual Report and Accounts 2018
61			2017	Annual Report and Accounts 2017
62			2016	Annual Report and Accounts 2016
63			2015	Annual Report and Accounts 2015
64			2014	Annual Report and Accounts 2014
65	Retail	Sainsbury's	2023	Sustainability Update 2023/2024
66			2022	Sustainability Update 2022/2023
67			2021	Sustainability Update 2021/2022
68			2020	Sustainability Update 2020/2021
69			2019	Sustainability Update 2019/2020
70			2018	Sustainability Update 2018
71			2017	Annual Report and Financial Statements 2017
72			2016	Annual Report and Financial Statements 2016
73			2015	Annual Report and Financial Statements 2015

Continued on next page

Table B.1 continued from previous page

No	Sector	Company	Year	Title
74			2014	An update on our progress so far 2014
75	Retail	Tesco	2023	Annual Report and Financial Statements 2023
76			2022	Annual Report and Financial Statements 2022
77			2021	Annual Report and Financial Statements 2021
78			2020	Annual Report and Financial Statements 2020
79			2019	Annual Report and Financial Statements 2019
80			2018	Annual Report and Financial Statements 2018
81			2017	Annual Report and Financial Statements 2017
82			2016	Annual Report and Financial Statements 2016
83			2015	Annual Report and Financial Statements 2015
84			2014	Annual Report and Financial Statements 2014

Table B.1: List of available corporate sustainability and financial reports from 2014 to 2023. (1) FCA merged with the PSA Group to create Stellantis in 2019.

Appendix C

ESG-Related News Articles dataset

No	Story	Date	Company	Keyword
0	...litigation firm announces that it is investigating claims on behalf of investors of ford motor company...	2024-07-29	Ford	esg
...	...	...	...	...
472	...advisory firm has come out against reinstating a pay package for tesla ceo elon musk that was voided earlier this year...	2024-06-03	Tesla	governance
...	...	....	...	...
604	...have agreed to collaborate in the development of new engines tailored to electrification and the pursuit of carbon neutrality...	2024-05-29	Toyota	environment
...	...	....	...	...
859	...the retailer will offer alterations and repairs to customers from august amid increased demand for...	2024-06-29	M&S	social
...	...	....	...	...
989	...tesco has revealed plans to use a new more sustainable packaging format for its range of fresh mince meat products the plans follow a successful trial period...	2024-05-03	Tesco	sustainability

Table C.1: ESG-related news articles dataset with 990 articles following data cleaning and text preprocessing.

Appendix D

Sentiment Analysis Classification Results

No	Text summary	Date	Company	Keyword	Sentiment	Conf. Score	Sent. Value	Comp. Score
0	...litigation firm announces that it is investigating claims on behalf of investors of ford motor company...	2024-07-29	Ford	esg	negative	0.915	-1	-0.915
...	...	...	...	...	...	...	...	...
472	...advisory firm has come out against reinstating a pay package for tesla ceo elon musk that was voided earlier this year...	2024-06-03	Tesla	governance	...	...	...	...
...	...	...	...	...	...	...	...	...
604	...have agreed to collaborate in the development of new engines tailored to electrification and the pursuit of carbon neutrality...	2024-05-29	Toyota	environment	...	...	...	...
...	...	...	...	...	...	...	...	...
859	...the retailer will offer alterations and repairs to customers from august amid increased demand for...	2024-06-29	M&S	social	...	...	...	...
...	...	...	...	...	...	...	...	...
989	...tesco has revealed plans to use a new more sustainable packaging format for its range of fresh mince meat products the plans follow a successful trial period...	2024-05-03	Tesco	sustainability	...	...	...	...

Table D.1: Sentiment analysis classification of ESG-related news articles by FinBERT to implement objective one. Dataset with 990 articles as the result of text preprocessing.

Company	Sentence	Score	Label
Asda	We re working on a number of important projects to support water management and drought resilience in our supply chain, including commitments to multi-stakeholder projects in Spain and South Africa, as well as with UK suppliers.	1.00	Positive
	Similarly, we are working with the digital platform Too Good To Go, which offers any excess, but still edible, food to app users.	0.85	Positive
	Risk of lost revenue due to possible future developments in fuel regulation and green energy transition.	0.00	Negative
	In response to these trends, we have delayed our target date from 2025 and aim to engage with store colleagues to ensure self-drop boxes are present in-store and easily identifiable to support greater recycling.	0.17	Negative
Ford	Our climate change efforts are aligned with the United Nations Framework Convention on Climate Change (Paris Agreement).	0.99	Positive
	Significant changes will be required to decarbonize global energy and transport systems, and we expect these changes will occur in different product segments and regions at different times.	0.72	Positive
	Commercial customers turn over 10-15% of their vehicles per year on average, meaning the transition from gas to electric will take time.	0.02	Negative

Continued on next page

Table D.2 continued from previous page

Company	Sentence	Score	Label
	However, such a transition comes with challenges that must be addressed.	0.26	Negative
Marks & Spencer	As an own brand retailer, working closely with our supply partners is crucial to achieving our net zero ambitions.	0.61	Positive
	We have a clear line of sight, based on projects that are underway, to 62% of the 2.1m tonne emissions reduction we are committed to deliver in 2025/26.	1.00	Positive
	However, due to the modelling approach for supply chain carbon emissions, a number of programmes that we have in place to deliver emissions reductions cannot yet be seen in our disclosed emissions.	0.01	Negative
	This has led to an increase in our base year footprint of 0.45m tonnes of CO2e.	0.10	Negative
Ocado	Driving more sustainable and efficient ways of doing business responsibly We are committed to be carbon Net Zero in our own operations (Scope 1 and 2) by 2035 and in our value chain (Scope 3) by 2040.	0.87	Positive

Continued on next page

Table D.2 continued from previous page

Company	Sentence	Score	Label
	Ocado can play an important role in a sustainable future, where our products and customer proposition through our online grocery delivery model result in lower levels of food waste and reduce our partners energy consumption levels by removing millions of weekly shopping basket miles.	1.00	Positive
	In high-labourcost markets, fulfilling orders manually in store, whereby employees carry out the picking and packing, is not profitably scalable.	0.01	Negative
	Retail operating costs increased by 6.3% to 2,347.1m (FY22: 2,207.0m) largely driven by the growth in orders, continued inflation and incremental OSP fees year-on-year.	0.01	Negative
Polestar	With Polestar, having fun driving a sports car can be a way to build a more sustainable society.	1.00	Positive
	This translates to striving for higher carbon efficiency in new car programs as well as decreasing the footprint of running car programs in the long-term.	0.99	Positive
	The speed of the transition to electric mobility can't be taken for granted right now.	0.01	Negative

Continued on next page

Table D.2 continued from previous page

Company	Sentence	Score	Label
	Many reports shows that global inequality is on the rise, GHG emissions are still increasing, and we are currently experiencing severe biodiversity loss and species extinction.	0.04	Negative
Sainsbury's	To leave a measurable positive impact on the communities we serve and source from and address food poverty by providing good food for all of us.	0.77	Positive
	We want all our colleagues to feel that there are opportunities to grow and progress should they wish to.	0.78	Positive
	Agriculture is responsible for about 70 per cent of freshwater use globally and in some areas contributes to water scarcity.	0.17	Negative
	Performance impacted by the delay of unified recycling across the UK and the lack of alternative materials to replace some types of plastic packaging, such as PET film.	0.00	Negative
Stellantis	A clear and transparent acknowledgement of the risks and issues related to climate change is therefore vital for Stellantis to work towards sustainability.	0.98	Positive
	Stellantis evaluates the probability of future impacts due to earthquakes and to extreme weather events on its sites and on its supply chain taking into account climate change impact on risk occurrence.	1.00	Positive

Continued on next page

Table D.2 continued from previous page

Company	Sentence	Score	Label
	The Strategy Council meets monthly, to direct the strategy regarding vehicle CO2 emissions and to review on a quarterly basis the overall Carbon Net Zero roadmap with the Top Executive Team.	0.27	Negative
	Elimination of waste and losses will mitigate the impacts of against price volatility.	0.32	Negative
Tesco	The better we do our job for customers, the more we can reinvest.	1.00	Positive
	We are also committed to cutting food waste across the supply chain by 50% by 2030.	0.61	Positive
	Downstream activities represent 40% of our footprint including primarily emissions resulting from customers using our products.	0.06	Negative
	The reduction in tax this year reflects the lower retail operating profits and a one-off charge in the prior year related to the revaluation of deferred tax.	0.12	Negative
Tesla	As costs continue to decline, more customers will be able to financially benefit from turning to renewable energy.	1.00	Positive
	Electric vehicles and sustainable energy products have a far better environmental impact than fossil fuel alternatives.	0.99	Positive
	Air pollution from burning fossil fuels leads to premature deaths.	0.02	Negative
	Many people are unlikely to buy products just because they have a low lifetime carbon footprint.	0.01	Negative

Continued on next page

Table D.2 continued from previous page

Company	Sentence	Score	Label
Toyota	Toyota promotes open and fair business practices and is making constant progress with initiatives to promote sustainability.	1.00	Positive
	We are also working closely with suppliers and dealers to improve quality, as well as providing safety and peace-of-mind to our customers, to achieve a high level of customer satisfaction.	0.99	Positive
	If no improvements are made, business relationship may be reconsidered.	0.03	Negative
	Recognize that migrant workers are vulnerable to exploitation and forced labor.	0.03	Negative

Table D.2: DistilBERT results for sentiment analysis classification to implement objective four.

No	Story	Date	Company	Keyword	neg	neu	pos	compound	Sentiment
0	...litigation firm announces that it is investigating claims on behalf of investors of ford motor company...	2024-07-29	Ford	esg	0.100	0.799	0.121	0.5267	positive
...	...	...	...	...	...	...	...	...	...
472	...advisory firm has come out against reinstating a pay package for tesla ceo elon musk that was voided earlier this year...	2024-06-03	Tesla	governance	0.115	0.812	0.073	-0.7579	negative
...	...	...	...	...	...	...	...	...	...
604	...have agreed to collaborate in the development of new engines tailored to electrification and the pursuit of carbon neutrality...	2024-05-29	Toyota	environment	0.000	0.850	0.150	0.9897	positive
...	...	...	...	...	...	...	...	...	...
859	...the retailer will offer alterations and repairs to customers from august amid increased demand for...	2024-06-29	M&S	social	0.008	0.858	0.134	0.9590	positive
...	...	...	...	...	...	...	...	...	...
989	...tesco has revealed plans to use a new more sustainable packaging format for its range of fresh mince meat products the plans follow a successful trial period...	2024-05-03	Tesco	sustainability	0.009	0.797	0.194	0.9943	positive

Table D.3: Sentiment analysis classification of ESG-related news articles by VADER to implement objective two. Dataset with 990 articles as the result of text preprocessing.