

Sustainable Profitability: Investigating the Relationship between ESG Initiatives and Financial Performance

Author: Hermanis Ozols
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,

The purpose of this study is to determine whether an organization's Environmental, Social, and Governance (ESG) initiatives and its financial performance are related. ESG concerns have emerged as a noteworthy component for both investors and consumers as a result of the rise in environmental consciousness among consumers, which has fuelled a growth in sustainable investing. However, stakeholders frequently struggle to understand the sustainability promises made by businesses.

We analyse the ESG criteria that firms prioritize and find emerging themes by applying text mining algorithms to their annual reports. The implemented empirical approach in this study extracts useful data from these reports, such as the frequency and context of terms related to ESG principles. The outcomes measure a business' adherence to ESG values and demonstrate the breadth of its commitment to sustainability and ethical business conduct.

In the current era of ESG investing and the rising customer preference for sustainable products, it is especially important to distinguish genuine commitment to ESG principles. The developed framework will make it easier for stakeholders to distinguish whether businesses are really devoted to ESG principles from those who are simply "greenwashing" or exaggerating their ESG efforts. By assisting investors and stakeholders in making better decisions, the text mining analysis's findings will ultimately help create a better understanding of sustainable practices and their impact on businesses financial performance.

Graduation Committee members:

DR. I. SKUTE (IGORS)

DR. M. GOETHNER (MAXIMILIAN)

Keywords

Sustainable investing, ESG, Sustainability, Financial performance, Text mining, Market trends, Greenwashing

1. INTRODUCTION

Consumers are increasingly becoming more and more environmentally conscious, often choosing brands and products offering a greener alternative to their usual choice. This shift in consumer behaviour has also led to the rise of sustainable investing, where investors consider ESG factors alongside traditional financial metrics when making investment decisions (PricewaterhouseCoopers, 2021).

This trend has become increasingly popular in recent years, according to Statista.com (2022) the assets allocated to ESG ETF funds have grown steadily every year, showing that ESG is an increasingly more relevant factor for companies to focus on. As such, in recent years, there has been a growing push towards incorporating environmental, social, and governance (ESG) factors into business practices and decision-making processes, to align with stakeholder interests and values. According to a survey by (PricewaterhouseCoopers, 2021) investors expect ESG to be an integral part of corporate strategy.

One of the most noteworthy events that signalled a shift in how organisations and regulatory bodies think about ESG happened in 2015, when the United Nations adopted 17 sustainability goals (United Nations, 2015) that are now affecting how businesses operate and will most likely become increasingly more important in the future as the importance of sustainability does not show signs of slowing down. Sustainable Development Goals give organisations a valuable guide for trying to align their business strategies with global sustainability objectives, leading to profound economical and societal impact. Technological advancements can play a role in accelerating transition in businesses becoming truly sustainable (Schwartz, 2021).

According to research by Kearney (2022) consumers want to buy more-sustainable products and live more sustainably overall. But when bombarded by countless claims from “net zero emissions” and “recyclable packaging” to “locally sourced” and “environmentally friendly,” consumers are sometimes overwhelmed and confused about how to be sustainable and what actions are the most impactful.

Businesses can greatly benefit from understanding the possible risks and opportunities by developing their business strategies while being aware of the environmental sustainability goals set out by the UN. Not only can companies contribute to sustainable development, but they can also find new market opportunities, enhance their brand image, and attract socially and environmentally responsible investors. As governments and regulators increasingly integrate the UN sustainability goals into their policy frameworks, businesses may face new compliance requirements, reporting standards, and regulatory pressure.

This growing trend has the potential to influence the success or failure rate of business ventures. Businesses that

effectively integrate ESG considerations into their operations may be better positioned for long-term growth and resilience regarding regulatory pressure. Company executives need to effectively evaluate and integrate ESG initiatives into the decision-making process, as well as in research and development processes, to develop increasingly more sustainable products that are valued by consumers.

Many business ventures often operate in highly specialised fields, making it hard for stakeholders to evaluate their business operations. According to Kenan Institute of Private Enterprise (2022) evaluating firms ESG focus is often difficult, this can potentially cause financial losses, lead to misalignment of interests and erosion of stakeholder trust. Effectively evaluating ESG performance increases trust in the ESG activities undertaken by organisation and allows for a greater transparency. Moreover, the ESG trends can change quickly, making it difficult for investors to make decisions based on long-term trends, that could rapidly change. These possible new trends could also be influenced by various factors, such as new regulations, developments in new technology or other market trends. With the increasing digitalisation and the rapid advancements of text analysis tools, decision-makers now have the access to larger data samples and more efficient data evaluation techniques. Leveraging these advanced capabilities, we can utilize data analytics and text mining techniques to effectively determine the extent of an organization's Environmental, Social, and Governance (ESG) activities. Data analytics and text mining techniques could be used by investors and other stakeholders as a tool to make informed decisions about their potential investment opportunities.

As ESG investing has only become a growing trend in the last few years, the topic has not been widely researched yet. The aim of the research is to look at the relationship of companies increasing their ESG activities and the subsequent change in stock price. This study seeks to address the research question: "***To what extent can the growth and financial prosperity made by organisations be explained by their efforts on sustainable performance?***".

By developing a method that assesses the emphasis placed by companies on each of the Environmental, Social, and Governance (ESG) variables within their annual reports, a quantifiable analysis between the selected companies is possible. This analysis aims to continue research of identifying and measuring the sustainability efforts of organisations, and to examine genuine ESG practices from greenwashing activities. Furthermore, the connection between ESG variables present in annual reports will be analysed in regard to their financial prosperity. Text mining will allow to extract insights from unstructured data, allowing it to transform it in analysable information. The use of text mining will allow for analysis of larger sample size of annual reports than more traditional methods.

By analysing the connection between high-tech venture sustainability efforts and their financial performance, this research is expected to reveal insights that can help potential investors make better decisions.

Potential findings of the research may discover a range of technological, geographical, sustainability, and

2. THEORETICAL FRAMEWORK

2.1 The role of ESG activities on organizations

In today's rapidly evolving business landscape, the development and management of emerging technologies have become critical for organizations seeking to maintain a competitive advantage. The volatile, uncertain, complex, and ambiguous (VUCA) framework offers a valuable lens through which to examine the challenges and opportunities that organizations face in their research and development (R&D) efforts (Bennett & Lemoine, 2014). The VUCA framework is helpful to better understand and analyse complex situations that might occur in the business world. Volatility aspects of the framework refers to the amount of change that could happen in terms of rate and also importance of the change. Complexity refers to how many parts or variables are included in the situation. Uncertainty aspect emphasizes how predictable is the situation and to what possibility is it to predict mitigating strategies. While ambiguity talks about how many possible interpretations of a situation are there for a situation. The VUCA framework helps to recognize and understand the various challenges they face in today's business world and helps develop strategies to understand these complex situations better. By analysing a certain situation using the elements of VUCA, leaders can develop better strategies and approaches to making decisions.

Another key framework of process of strategic management of R&D project portfolios, is highlighted by McMillan and McGrath (2002).

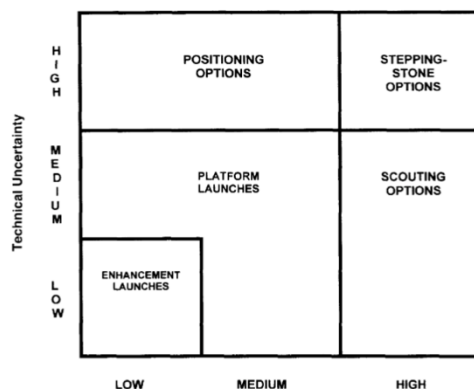


Figure 1: Technical uncertainty matrix (MacMillan and McGrath, 2002).

This framework helps organizations understand some of the risks associated with their projects, allowing them to make more thought out decisions when allocating resources and selecting projects for their R&D portfolios.

environmental protection factors, along with other yet undiscovered elements that can contribute to the success of a high-tech venture. Analysing the role and effect of sustainable practices and environmental thinking in the development of new products and ongoing operations of organisations can reveal how these factors influence firms' financial performance.

The VUCA framework ties together with the Technical Uncertainty Matrix by providing a comprehensive understanding of the environment in which organizations operate. The VUCA world framework proposes that organizations might face various different uncertainties, but there is a need to prioritize these risks based on the relevance to the organization. The Technical Uncertainty Matrix can be seen as a practical tool that complements the VUCA framework by helping organizations categorize and prioritize uncertainties.

By integrating the VUCA framework and the Technical Uncertainty Matrix, organizations can better anticipate and adapt to the challenges and opportunities that arise in a complex and rapidly changing world. With ESG initiatives, this approach is particularly crucial due to the evolving landscape of technology, regulations, and stakeholder expectations. These initiatives inherently contain elements of technical and market uncertainty - from rapidly emerging clean technologies to unpredictable regulatory changes and stakeholder reactions. ESG initiatives can be scored based on the level of technical uncertainty and potential impact on financial performance. By systematically evaluating ESG initiatives within this matrix, organizations can identify those with high potential for positive financial returns while also considering the technical risks involved allowing for a systematic view on a complex and unpredictable problems in the VUCA world.

According to Friede et al. (2015) there is a positive correlation between ESG impact and corporate financial performance (CFP). Vote-count studies, meta-analyses, and a best-evidence synthesis approaches were used. The research revealed a strong case for ESG investing, with ESG impact having a stable impact on positive financial returns. The analysis offered detailed insights and conclusions that can be applied for portfolio and non-portfolio studies, as well as various asset classes, regions, and categories of ESG. Nonequity asset classes both for bonds and real estate displayed a considerably higher share of positive findings over equities. The strongest correlation was found to be in the emerging markets and North America. The observed correlation was determined to be positive, albeit with a relatively modest magnitude. The inherent limitations of this study are that only articles and studies up to 2014 were analysed. The research does not look at more recent ESG initiatives that have occurred after 2014, which may significantly affect the current landscape of ESG and its relationship to financial performance.

According to another study (Bauer & Hann, 2010), poor environmental practices influenced the credit standing of borrowing firms through the legal, reputational, and regulatory risks associated with environmental incidents. The study examined corporations between the years 1995 and 2006, suggesting that firms with environmental concerns pay a premium on their cost of debt financing and are assigned lower credit ratings. In contrast, firms with

proactive environmental engagement benefit from a lower cost of debt financing.

As consumers, investors and company employees become increasingly more environmentally conscious, some businesses and brands have promoted their products as ESG conscious, when their impact is negligible. Greenwashing poses a challenge to businesses trying to achieve true sustainability, greenwashing has potential to erode customer trust, leading to destroyed brand image. Regulatory bodies have also increasingly put pressure on companies engaging in greenwashing activities, by creating new reporting standards and increasingly prosecuting companies engaging in environmentally irresponsible activities. According to research (Gatti et al., 2019), several non-governmental organizations (NGOs), such as Greenpeace or TerraChoice, assume today the roles of market monitors or “watchdogs.” The research also sets out idea that voluntary dimension is also not enough to prevent greenwashing and new mandatory dimension needs to be added. Mandatory dimension would add new reporting standards and accountability frameworks that would be mandatory to follow by businesses.

One of the most notable frameworks is developed by TerraChoice, it sets out seven ways that companies engage in greenwashing activities.. According to (Green Business Bureau, 2021), the seven sins of greenwashing as set out by TerraChoice are:

1. Sin of the Hidden Trade-off: a product is marketed as being green based on only one factor, while hiding other factors.
2. Sin of No Proof: providing no evidence about the environmental impact.
3. Sin of Vagueness: providing no specific information about environmental claims.
4. Sin of Worshipping False Labels: using false labels on their products, misleading consumers.
5. Sin of Irrelevance: Advertising obvious environmental features that are not applicable
6. Sin of Lesser of Two Evils: advertising product as sustainable when it inherently is not.
7. Sin of Fibbing: outright lying about environmental impact

By recognizing the seven sins of greenwashing, consumers, investors, and regulatory bodies can make more analytical decisions about the environmental impact of the products and services, and companies can work towards fostering genuine sustainability in their practices.

In summary, integrating the VUCA framework and the Technical Uncertainty Matrix can provide organizations with a comprehensive understanding of the complex environment in which they operate, especially regarding ESG initiatives. As the studies by Friede et al. (2015) and Bauer and Hann (2010) have established a positive correlation between ESG initiatives and financial performance, the organizations can greatly benefit from being proactive with ESG initiatives either by avoiding large environmental scandals or by enhancing their brand reputation. Organizations however need to be aware of the risks of doing activities that could result into greenwashing scandals.

As organizations continue to adapt to evolving technologies, regulations, and stakeholder expectations, further research is needed to examine the specific ESG related variables and their impact on financial performance.

2.2 The role of Green Image on firm performance

The role of brand reputation and its image has an increasingly become more important variable on the organization’s overall performance. A positive brand image can enhance a company's brand overall competitiveness and give unique value to customers by signalling to consumers, investors, and other stakeholders that the company is committed to sustainability and responsible business practices. Integrating ESG initiatives into R&D management can also help companies attract and retain top talent who are passionate about sustainability and social responsibility.

Incorporating ESG perspectives into R&D management can contribute to a more sustainable and responsible innovation process. A strong brand that incorporates ESG values can not only attract environmentally and socially conscious customers but also motivate companies to focus on sustainable product development. This may involve investing in R&D projects that prioritize eco-friendly materials, energy-efficient technologies, or socially responsible production processes.

According to Xie et al. (2019), there are two primary strategies of green technology innovation: green process innovation and green product innovation. The researchers suggest a conceptual model, in which the relationship between green technology innovation and financial performance is mitigated by the improved green image of a brand as well as the possible green subsidies organization can receive.

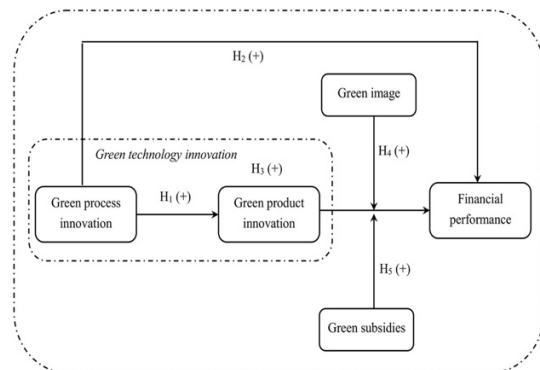


Figure 3: Conceptual model (Xie et al., 2019).

The study results show that green process innovation is significantly correlated with green product innovation. The results also show that a firm's green image and green subsidies are significantly related to its green process innovation, as well as to its green product innovation (Xie et al., 2019).

2.3 The role of green innovation on firm performance

According to (Davila, 2014) bigger more established companies beat start-ups in more complex breakthroughs – those that called for creations or transformations of whole markets and industries. This however depends on the industry. The author talks about four distinct types of innovation, incremental innovation driven from the top, incremental innovations driven from the bottom, incremental innovations driven from the bottom and lastly incremental innovations driven from the top. Corporate bureaucracy can slow down these breakthrough innovations due to the lack of flexibility and long approval processes.

As such smaller more flexible companies can develop more rapidly.

In the context of R&D management, the Schumpeterian theory of innovation can have several implications. It emphasizes how crucial it is to foster an innovative workplace environment where staff members are encouraged to think creatively and take initiative in taking calculated risks in order to develop new concepts and products.

The Schumpeterian theory of innovation suggests that R&D activities and management should be viewed as a long-term investment, rather than a short-term cost. The benefits of the increased R&D activities are not instant and might take several years to realize.

The Schumpeterian theory of innovation can be extended to encompass Environmental, Social, and Governance (ESG) initiatives. This can be done in various ways such as investing in research and development of new products and technologies that are environmentally friendly or selling products that use environmentally friendly packaging, as well as implementing sustainable practices and reducing the organization's environmental impact. The connection between Schumpeterian theory and ESG initiatives could be observed from different perspectives:

1. Gain a competitive advantage by incorporating ESG initiatives in the R&D strategy. This differentiation strategy can give an advantage when compared to the competitors and lead to a better brand awareness, increased market share and profitability.

2. Resilience to increased regulatory changes regarding ESG initiatives. By proactively addressing environmental issues, organizations can enhance their resilience to external shocks and disruptions, ensuring their long-term viability and competitiveness.

3. New value proposition and new market creation, companies can offer innovative solutions that cater to the growing demand for sustainable products and services. This can result in the expansion of existing markets or the creation of entirely new ones, contributing to economic growth and development. These new environmental products could contribute to the destruction and replacement process of existing markets and products.

Encouraging R&D managers to focus on or to seek out market opportunities for sustainable solutions, and to disrupt existing industries and business models with sustainable alternatives will give a company and its product portfolio unique advantage on companies that are staying true to their old ways of innovating.

2.4 Text mining application in management

Text mining offers significant advantages in business decision-making by enabling organizations to efficiently gather and analyse vast amounts of data from various sources, such as news articles, research papers, social media posts, and business reports. The use of text mining methods can benefit businesses and researchers in multiple ways, they allow to analyse data much more effectively when compared to manual analysis. This in turn means that predicting trends, finding relevant information is more efficient. This allows organizations to make more informed decisions and gain a competitive edge in the market.

In the context of our problem statement, the use of text mining techniques can provide valuable insights into the relationship between ESG initiatives and financial

performance. By applying text mining techniques to the problem of distinguishing genuine ESG initiative incorporating companies to those using greenwashing, we can systematically analyse large volumes of data to identify the factors that contribute to successful ESG integration.

A study developed an algorithm and hierarchical approach to help improve project selection (Rathore et al.). The developed algorithm outperformed the existing technique of research project selection using KNN. The k-nearest neighbours algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point (IBM, 2023).

A paper by Hou and Li (2010) developed a multiple regression model and VAR model, that looks to examine the relationship between R&D investments and financial performance of publicly listed companies from Shanghai Stock Exchange. VAR is a statistical model that aims to capture interdependencies among multiple variables over time. VAR models are often used in finance and economics. The study found that there is a positive correlation between R&D expenditure and companies' financial success. The impact was most significant after 3 years of an R&D expenditure. Main variables used were R&D expenditure, companies' assets and operating profit.

3. RESEARCH DESIGN

3.1 Data collection

To gather the requisite data for this study, the Eikon Refinitiv database will be used. It is a comprehensive and extensive and reliable sources of public company information. Eikon Refinitiv gives access to financial data, industry classifications, audited ESG statements and other data points that were used in the company selection process.

Eikon Refinitiv was used to find the largest publicly traded US companies by their market cap in food and beverage industry. The 3 largest companies included in the analysis were, Coca-Cola Co, PepsiCo Inc, Philip Morris International Inc. In the final analysis 47 companies were chosen with the largest market cap being 264.75B USD and smallest being 39.46M USD. A sample size of 47 was chosen as this amount sufficiently captured the largest players in the chosen industries, but also allowed for efficiency during the data analysis process. The selection process was focused on the largest players in the industry, as they might typically be considered the trend setters and market movers.

To obtain the annual reports of these selected companies, NASDAQ website was used to retrieve the companies' SEC filings, 10-K annual reports for the year 2018 were chosen as the data for analysis. The annual reports are analysed to calculate each company a certain ESG score.

The obtained ESG scores are compared against the average stock price difference between January of 2018 and January of 2023. Stock price information is retrieved through Yahoo finance python API. This allows for stock price information to be analysed automatically when processing the annual reports.

3.2 Text pre-processing

After data is collected from companies operating in the specific industry. Pre-processing includes addressing missing or incomplete data points, standardizing data, and aggregating the chosen variables for analysis. The pre-processing phase ensures that the datasets from all the

companies are consistent and have all the necessary information for analysis.

The first step in the data pre-processing phase contains identifying and addressing missing or incomplete data.

After the annual reports are gathered from the NASDAQ website, they serve as input for our Python program. We are using the pdfplumber library for this task. pdfplumber is a Python library that makes extracting text, images, and metadata from PDF files easy. It provides a simple API for converting the text found in PDF file to analysable text information.

The gathered text information from each of the annual reports is subjected to pre-processing. This is necessary step to ensure that data is cleaned, normalized and ready for analysis.

The first step in the pre-processing pipeline is the removal of non-word characters from the text. These might include punctuation and special characters found in the text. These characters are removed from the text as they do not contribute to the semantic meaning of the text.

Next, we handle the issue of multiple spaces between words. Extra spaces can occur when we remove punctuation or special characters and can interfere with the tokenization process.

The third step is to convert all the text to lowercase. This is done to ensure that the same word, appearing in different cases, is not considered as different words. For example, 'Environment', 'environment', and 'ENVIRONMENT' should all be considered as the same word. Once the text is cleaned, we proceed to tokenize it. Tokenization splits the text into individual words or tokens. This is achieved using the `word_tokenize(text)` function from the Natural Language Toolkit (NLTK).

The next step is to remove stop words from our tokens. Stop words are common words like 'is', 'the', 'and', etc., that do not carry much meaningful information and can be safely removed. We use the `stopwords.words('english')` function from NLTK to remove all English stop words from our data.

As the last step of text pre-processing, we perform lemmatization on our tokens. Lemmatization is the process of reducing a word to its base or root form. For example, "environment", "environmentally", and "environmental" are all variations of the word "environment" and lemmatizing these words would reduce them to "environment". We use the `WordNetLemmatizer().lemmatize(word)` function from NLTK for this purpose.

A well-executed pre-processing phase will allow for reliable and effective analysis of the data, reducing potential biases, outliers and errors, enabling to draw reliable conclusions from the processed text data analysis.

3.3 Data analysis

After the data is collected from NASDAQ and subjected to the data pre-processing, the data will be analysed to gain insights in the textual content.

4. RESULTS

Detailed analysis was conducted on a select group of companies within the food and beverage industry. The selection process was done through Refinitiv Eikon software. The parameters set for the search were specifically

A word frequency analysis will be conducted on the pre-processed annual reports, leading to insights of how frequently keywords related to environment, social issues and corporate governance are used in each of the company's annual reports.

The word frequency analysis will be performed by analysing how frequently keywords related to each of the three pillars of ESG (Environmental, Social, and Governance) variables occur within the annual reports. Each report will be given its own score for the environmental, social and governance aspect. The score is calculated by summing up the frequency of each keyword in the report, dividing by the total count of words in the report, and multiplying by 100,000 to give us a score of how frequently a keyword related to each of the variables occur per 100,000 words. These individual ESG scores for each report will provide a quantitative measurement of a company's ESG performance and priorities as conveyed in their annual reports.

In addition to calculating the scores, a word cloud will also be generated for each of the three pillars, producing a visualization of the most common keywords used. For this task python library "wordcloud" will be used.

3.4 Clustering

For cluster analysis, one of the most employed techniques is k-means clustering. K-means clustering aims to group data points (in this case, companies) based on similarity of the results of ESG keyword frequency analysis. The clusters will allow for more objective analysis of company financial performance, as the companies will be analysed as a group.

For instance, one cluster might consist of companies with high Environmental scores. These companies could be considered as putting a great focus in their annual reports on environmental issues. Another cluster might include companies with low Environmental scores. These companies could be seen as lagging in their Environmental practices.

To implement k-means clustering in this research, a python library "sklearn" in addition with "numpy" are used. The data containing ESG scores, will serve as input for the k-means algorithm.

The number of clusters will be set to 4, which means that the K-means algorithm will divide the data into 4 clusters. By setting the cluster count of 4 we aim, for four distinct clusters that are companies:

- Low – companies having a very low frequency of ESG keywords.
- Below average – companies have some ESG related keywords but still falling short of the average.
- Above average – companies have an ESG score of higher than average.
- Exceptional – companies that have ESG scores that are significantly higher than average.

tailored to ensure the relevance and specificity of the results. The country of exchange was set to the United States, as the focus was set on companies operating within this economic context to facilitate easier and more trustworthy data collection process.

Furthermore, the North American Industry Classification System (NAICS) was utilized to refine the search to companies within the food and beverage manufacturing sectors, as well as those in food manufacturing. This classification system allowed to sort of companies that only specialize in the chosen industry.

The final parameter employed during the selection procedure was the market capitalization of the organization. This measure was undertaken to guarantee the incorporation of the most prominent stakeholders in the domain. The analysis was able to concentrate on organizations that have a significant impact on the industry due to their size and reach by focusing on companies with the largest market capitalization. The approach ensured that the analysis was focused on the major contributors and trendsetters within the food and beverage industry.

By adopting this methodology, the used empirical approach helped to access data in reliable way, thereby giving credibility and dependability of the research outcomes. The utilization of the 10-K forms facilitated a standardized way of gathering the data and enabled cross-company comparisons, given that all publicly traded companies in the United States are mandated to submit these forms in a standardized format, which are made available in various file formats on the NASDAQ website.

During data collection, certain challenges were encountered that necessitated the exclusion of some companies from the analysis. Specifically, issues arose with processing the annual reports of some companies. In some instances, the SEC filings of certain companies were not available on the NASDAQ website or in other cases the stock price information was not available. In total 10-K annual reports for 47 companies were gathered and subjected to analysis.

4.1 Environment scores

The results of the word frequency analysis revealed significant findings in terms of the emphasis placed on different categories within the annual reports. Specifically, the 'Environment' category emerged as the most frequently referenced, indicating a heightened focus on environmental issues within these reports.

The scores vary significantly across the companies, ranging from as low as approximately 4.09 per 100'000 words to as high as approximately 128.65 per 100'000 words. This wide range suggests a considerable variation in the extent to which these companies emphasize environmental issues in their annual reports.

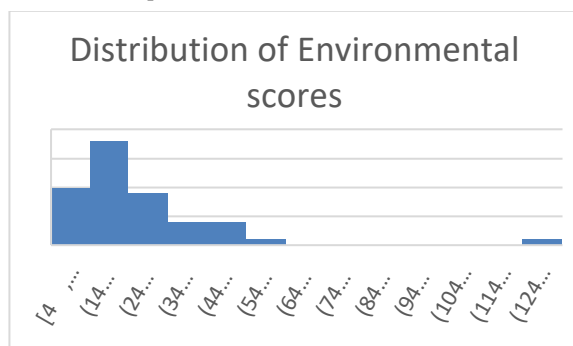


Figure 4: Distribution of environmental scores

Few of the selected companies have relatively high frequency of environmental related keywords (e.g., 128.65, 54.99, 52.99, 51.53), indicating a particularly strong emphasis on environmental issues in their annual reports.

On the other hand, many companies have relatively low scores (e.g., 4.09, 4.39, 4.54), suggesting a lesser emphasis on environmental issues in their annual reports.

Most of the companies have scores in the mid-range, indicating a moderate level of emphasis on environmental issues. This suggests that while environmental keywords are found in their annual reports, they may not be the primary focus. The average environment score is 24.67, meaning on average there 24,67 environment related keywords per 10'000 words.

The K-means clustering algorithm has been applied to the environmental scores of the companies, resulting in the formation of four distinct clusters. The clusters are differentiated based on the environmental score ('E Score').

Table 1 - K-means clustering results of environmental score

Cluster	Average E score	Stock change	Nr. Of companies
0	21,16	-1,78	27
1	128,64	6,12	1
2	44,63	20,93	9
3	5,80	11,40	10
Total	24,67	5,53	47

Cluster 0, which includes the majority of the companies (27 out of 47), has an average E Score of approximately 21.16, indicating a moderate level of emphasis on environmental issues. However, these companies have experienced an average stock price decrease of approximately 1.79% from 2018 to 2023.

Cluster 1, which consists of a single company, stands out due to its exceptionally high average E Score of approximately 128.65. This company has also experienced a stock price increase of approximately 6.12% over the same period. However, upon further manual analysis of the report this high score was due to the use of word environment in various contexts, some of the examples being “in an environment of increased regulation and taxation” and “The competitive environment and our competitive position”.

Cluster 2, comprising 9 companies, has a higher average E Score of approximately 44.64, indicating a stronger emphasis on environmental issues. Interestingly, these companies have seen a substantial average stock price increase of approximately 20.93% from 2018 to 2023.

Cluster 3 includes 10 companies with the lowest average E Score of approximately 5.81, suggesting a lesser emphasis on environmental issues. Despite this, these companies have experienced a significant average stock price increase of approximately 11.41% over the period.

The word cloud results for the Environmental (E) scores show a strong emphasis on the term 'environment', which appears in almost all the entries. This is expected, as it is a broad term that encompasses many aspects of a company's environmental impact.

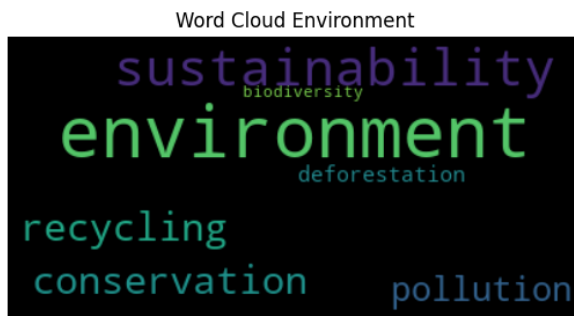


Figure 5: Word cloud results of environmental score

Additionally, the word "sustainability" is the second most frequently used keyword leading to believe that many businesses are taking initiatives to perform business activities in a sustainable manner. Initiatives to cut back on resource consumption as well as reduce waste, or adopt sustainable sourcing methods could all fall under this category.

Pollution, recycling, and conservation are less frequently used but still important terms. These terms suggest that some companies are discussing specific environmental issues in their reports. 'Pollution' could refer to efforts to reduce emissions or manage waste, 'recycling' could indicate initiatives to recycle materials or reduce waste, and 'conservation' could relate to efforts to conserve resources or protect biodiversity.

4.2 Social scores

The average social score of approximately 4.79 is significantly lower than the average environmental score of approximately 24.67. This suggests that, on average, the companies in the food and beverage industry place a greater emphasis on environmental issues compared to social issues in their annual reports, however the keywords related to social sustainability could be considered more specific than those of environment and play a role in the average score. The scores vary significantly across the companies, ranging from 0 to approximately 33.74.

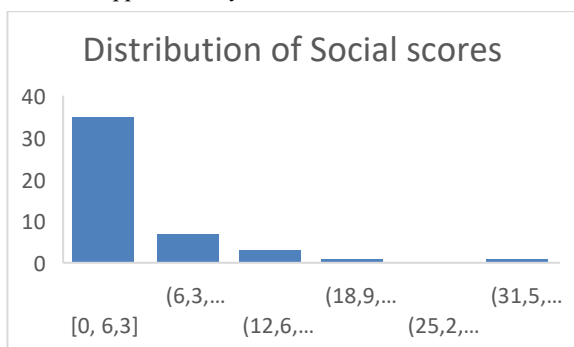


Figure 6: Distribution of social scores

The majority of the companies have scores in the low to mid-range, indicating a moderate to low level of emphasis on social issues. This suggests that while social considerations are a part of their corporate discourse, they may not be part of the focus in annual reports. Contrary, some companies have a score of 0, meaning that social issues were not addressed in their annual reports at all.

The K-means clustering algorithm has been applied to the social scores of the companies, resulting in the formation of four distinct clusters. The clusters are differentiated based on the social score ('S Score').

Table 2 - K-means clustering results of social score

Cluster	Average S score	Stock change	Nr. Of companies
0	5,89	20,69	27
1	33,74	6,12	1
2	0,58	16,54	9
3	15,46	-98,93	10
Total	4,79	5,53	47

Cluster 2 is the largest cluster, comprising 24 companies. These companies have the lowest average S Score of approximately 0.59, indicating a minimal emphasis on social issues. Despite this, these companies have seen a significant average stock price increase of approximately 16.54% over the period.

Cluster 3 includes 5 companies with a high average S Score of approximately 15.47, suggesting that social issues were addressed in the report. However, these companies have experienced a dramatic average stock price decrease of approximately 98.93% from 2018 to 2023.

These results suggest a complex relationship between a company's emphasis on social issues and its stock price performance. While clusters 0 and 2 show a positive relationship, Cluster 3 presents a contrasting scenario. This discrepancy suggests of other stock price determinants that might not be represented.

As expected, the world cloud results for the Social (S) scores are a lot less broad. The result show a strong emphasis on the terms 'diversity' and 'inclusion'. These terms are key indicators of a company's commitment to social responsibility, particularly in relation to its workforce.

For instance, the company International Flavors & Fragrances Inc. had the second highest Social score, an example sentence that contributed to their high Social score is "Positive Principles - We seek to embed the principles of eliminating the concept of waste, using clean renewable energy, and celebrating diversity into our company and culture".



Figure 7: Word cloud results of social score

4.3 Governance scores

The average Governance (G) score of approximately 5.01 is higher than the average Social (S) score of approximately

4.79, but significantly lower than the average Environmental (E) score of approximately 24.67. The highest governance score is 38.89.

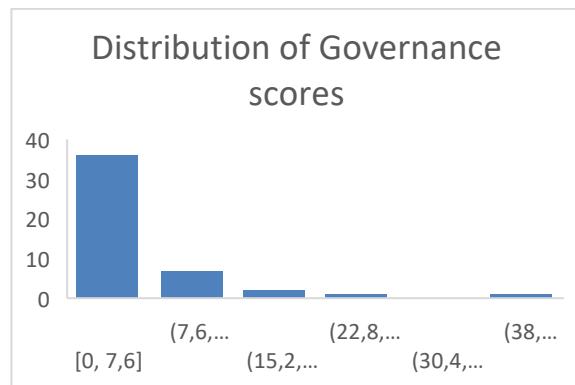


Figure 8: Distribution of Governance scores

The distribution of Governance (G) scores are comparable to the distribution of Environment (E) scores, suggesting that only few companies choose to address social and corporate governance issues in their annual reports.

The K-means clustering algorithm has been applied to the governance scores of the companies, resulting in the formation of four distinct clusters. The clusters are differentiated based on the governance score ('G Score').

Table 3- K-means clustering results of governance score

Cluster	Average G score	Stock change	Nr. Of companies
0	7,51	8,20	16
1	0,42	0,81	27
2	38,88	58,38	1
3	21,61	16,21	3
Total	5,01	5,53	47

Cluster 0 comprises 16 companies with an average G score of approximately 7.51, indicating a moderate emphasis on governance issues in their annual reports. The average stock price change for this cluster is approximately 8.20%, suggesting a positive market performance over the period from 2018 to 2023.

Cluster 1 is the largest cluster, encompassing 27 companies. These companies have a relatively low average G score of approximately 0.43, indicating a minimal emphasis on governance issues in their annual reports. The average stock price change for this cluster is approximately 0.82%, indicating relatively stable market performance.

Cluster 2 consists of a single company with a high G score of approximately 38.89, suggesting a strong emphasis on governance issues. This company also experienced a substantial stock price change of approximately 58.39%, indicating a strong market performance. An example sentence from the highest scoring company's annual report that contributed to their high score: "The Company must comply with other general business regulations covering areas such as accounting and income taxes, anti-corruption,

anti-bribery, global trade, trade sanctions, environmental, product safety, and handling and production of regulated substances".

Cluster 3 includes three companies with a high average G score of approximately 21.61, indicating a strong emphasis on governance issues. The average stock price change for this cluster is approximately 16.21%, suggesting a positive market performance.

These results suggest a potential relationship between the emphasis on governance issues in annual reports and market performance. Compared to the relationship seen for the Environmental (E) and Social (S) scores, this relationship seems to be more pronounced. Though it's crucial to remember that these conclusions are based on a small subset of companies.

The word cloud results for the Governance (G) scores highlight a significant focus on 'corruption' and 'bribery', with some mentions of 'lobbying'. These terms are key indicators of a company's commitment to ethical business practices and good governance.



Figure 9: Word cloud results of governance score

5. DISCUSSION

The analysis of the Environmental, Social, and Governance (ESG) scores in the food and beverage industry reveals some interesting patterns. The focus on environmental issues surpassed the frequency of references to 'Social' and 'Governance' categories. This finding suggests that, while social and governance issues are undoubtedly important, environmental considerations are currently at the forefront of corporate disclosure in the food and beverage industry.

This trend could be indicative of the industry's response to the global emphasis on environmental sustainability and improvement of brand image. However, it also raises questions about the relative lack of emphasis on social and governance issues, which are equally critical components of corporate responsibility. The observed emphasis on environmental terminology within the annual reports of food and beverage companies could also be viewed through the lens of 'greenwashing'. Companies could be inherently incentivized to promote their green initiatives to be seen more attractive to investors and consumers, while downplaying other significant aspects that have environmental impact. The high average Environmental (E) score observed in our analysis could potentially be indicative of such practices. Companies may be emphasizing certain environmental aspects in their annual reports to project an image of environmental responsibility. However, this could result in an inaccurate depiction of the company's actual environmental performance in the absence of a thorough and balanced representation of all environmental impacts.

It's important to note that a high frequency of environmental terms does not necessarily equate to greenwashing. It could also reflect genuine efforts by these companies to reduce their environmental impact and operate more sustainably. There are however risks that some aspects of the seven sins of greenwashing as defined by TerraChoice are present. These companies may emphasize their use of environmental terms in an attempt to project a green image without providing tangible proof of their efforts.

As Xie et al. (2019) suggest, the green image of a brand can mediate the relationship between green technology innovation and financial performance. Therefore, if the environmental claims made within these reports are substantive and verifiable, it could enhance the company's green image, leading to improved financial performance and stakeholder trust.

The emergent themes should also be looked through the VUCA framework and the Technical Uncertainty Matrix. By understanding the degree of technical and market uncertainty surrounding these ESG initiatives, we can gauge the potential risks and rewards associated with them.

Therefore, while the word frequency analysis provides valuable insights into the discourse within these reports, further investigation would be needed to determine the extent to which these environmental claims are substantiated by the companies' actual practices.

This emphasizes the significance of reliable and consistent ESG reporting methodologies. With the use of such frameworks, businesses can make sure that the data they report on is reliable and verifiable and covers a wide range of environmental indicators. This can assist stakeholders in identifying businesses that are sincerely devoted to environmental sustainability from those who might be using greenwashing.

The relationship between the ESG scores (E score, S score, and G score) and stock price change was not clear in our study. Although some businesses with strong ESG scores saw increases in stock price, others did not follow this pattern. This implies that there may be more nuance than first thought in the relationship between ESG practices and financial performance.

In the context of the research question - 'To what extent can the growth and financial prosperity made by organisations be explained by their efforts on sustainable performance?', it becomes apparent that the connection between ESG scores and financial prosperity, as indicated by stock price, isn't as direct or transparent as initially hypothesized.

5.1 Theoretical implications

There are various theoretical implications of this study, text mining combined with data analysis techniques could be used as valuable tools in ESG research. ESG research could greatly benefit from incorporating more data points that could be gathered through text mining.

While the research by Xie et al. (2019) suggest that green brand image can have a positive impact on firms' financial performance, the findings indicate that the relationship between ESG practices and stock performance isn't consistently clear.

The relationship appears to be complex, suggesting that investors may consider other factors when deciding to invest. While ESG scores provide a way to measure a company's commitment to environmental, social, and governance issues, they are just one piece of the puzzle.

Investors are also considering other factors such as the company's financial health, market position, growth potential, and the broader economic environment when making investment decisions.

The results raise important questions about the nature of ESG disclosures. The observed emphasis on environmental terminology and the potential for 'greenwashing' highlight the need for more critical way of analysing company ESG initiatives.

While the research by study (Bauer & Hann, 2010), suggests that companies who don't address ESG issues adequately might be in risk of the legal, reputational, and regulatory risks associated with environmental incidents. The findings were quite complex and companies with lower scores for ESG variables did not always experience a noticeable stock price decrease. This could be partially explained by investors not fully pricing in the risks associated with poor ESG performance until an incident occurs. The ESG activities undertaken by company might take some time to materialize and there could be time lag between companies ESG performance and financial performance that is larger than the time gap set in this study. These factors could also help explain the lack of clear correlation.

5.2 Practical implications

To analyse the practical implications of our findings, a decision was made to compare the calculated Environment scores with the audited ESG scores available from Refinitiv Eikon. Environment variable was chosen as it was by far most well represented in the report when compared to Social and Governance aspects. This method of analysis will allow us to assess the accuracy of our text analysis approach. If there is a high correlation between our calculated Environment scores and the audited ESG scores, it would suggest that our method of extracting and quantifying environmental considerations from annual reports is reliable.

Table 4 - Environment scores compared to audited ESG scores

Cluster	Audited ESG score	Calculated E score	Nr. Of companies
0	58,25	21,16	27
1	35	128,64	1
2	64,55	44,64	9
3	50,37	5,8	10
Total	57,54	24,67	47

Upon examining the data, we can observe a general trend that higher calculated E scores are associated with higher audited ESG scores. This suggests that companies that emphasize environmental issues in their annual reports tend to have better overall ESG practices, as reflected in their audited ESG scores.

Excluding the outlier cluster 1, there seems to be a positive relationship between the calculated E scores and the audited ESG scores. This implies that a higher emphasis on

environmental issues in company reports generally aligns with better overall ESG practices.

These results show that text mining and data analysis tools could be used to create a better understanding of a company's ESG performance. By analysing the frequency and context of ESG-related keywords in company reports, we can gain insights into the areas that companies prioritize and emphasize. This, in turn, can provide a more nuanced understanding of a company's ESG practices beyond what is reflected in their audited scores.

The practical implications of these findings are that text mining and data analysis methods can aid investors in portfolio construction by identifying companies with similar ESG profiles, as well as help analyse companies that might not have an audited ESG report. Regulators and other industry bodies can also explore this as a viable way to monitor their competitors and other market players.

5.3 Limitations

While this study provides valuable insights into the use of text mining and data analysis in ESG research, it is not without its limitations.

Firstly, the study is limited to companies within the food and beverage industry. While this allows for a more focused analysis, it also means that the findings may not be generalizable to other industries. Different industries may have different ESG priorities and practices, and further research would be needed to determine whether the patterns observed in this study hold true in other contexts.

Secondly, the study is based on a relatively small sample size of 47 companies. While this was sufficient for the purposes of this study, a larger sample size would allow for more robust statistical analysis and could potentially reveal additional patterns and trends.

The study is based on data from 2018 annual reports. While this provides a snapshot of ESG practices at a particular point in time, it does not capture changes in these practices over time. Future research could benefit from a longitudinal

approach, examining changes in ESG disclosures and practices over multiple years.

Finally, the study relies on the company annual reports to calculate ESG scores. While these scores may provide useful insights into a company's ESG performance, they are not without their limitations. They are based on the information that companies choose to address in their annual reports, which may not fully encompass all of the ESG activities undertaken by a company.

5.4 Future research

The results of this study provide the opportunity for various directions of future research. One potential direction would be to explore the use of other machine learning methods in analysing ESG disclosures. While k-means clustering provided useful insights in this study, other more complex techniques could be potentially more reliable and reveal additional insights present in company annual reports.

Future studies could also consider integrating other data sources into their analyses. For instance, integrating data from companies' social media accounts, news articles etc.

As the study was focused on analysing companies within a specific industry and region, it would be interesting to extend this research to other industries and regions as well. This would allow for a more detailed understanding of ESG practices across different contexts and could potentially reveal industry-specific or region-specific trends.

In addition to the suggestions mentioned above, future research could also consider analysing financial variables other than stock price change. While stock price is a commonly used indicator of a company's financial performance, it may not always be representative of the company's overall success. Other financial metrics could be incorporated, such as revenue or profit. Future studies could also analyse the impact on non-financial indicators, these could be customer satisfaction, brand loyalty, perceived brand image etc.

6. REFERENCES

- Aaker, D. A. (2009). *Managing Brand Equity*. In *Google Books*. Simon and Schuster. https://books.google.nl/books?hl=lv&lr=&id=r_TSY5sxnO8C&oi=fnd&pg=PT10&dq=brand+equity+aaker&ots=Ax5HmiEWS6&sig=CgrR8qe2pFAN7M6vPwtU4fJdpkl#v=onepage&q=brand%20equity%20aaker&f=false
- Bauer, R., & Hann, D. (2010). Corporate Environmental Management and Credit Risk. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1660470>
- Bennett, N., & Lemoine, G. J. (2014, January). *What VUCA Really Means for You*. Harvard Business Review. <https://hbr.org/2014/01/what-vuca-really-means-for-you>
- Davila, T. (2014, June 5). *The Innovation Strategy Big Companies Should Pursue*. Harvard Business Review. <https://hbr.org/2014/06/the-innovation-strategy-big-companies-should-pursue>
- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–

- <https://doi.org/10.1080/20430795.2015.1118917>
- Gatti, L., Seele, P., & Rademacher, L. (2019). Grey zone in – greenwash out. A review of greenwashing research and implications for the voluntary-mandatory transition of CSR. *International Journal of Corporate Social Responsibility*, 4(1).
- Green Business Bureau. (2021, December 16). *The Seven Sins of Greenwashing*. Green Business Bureau. <https://greenbusinessbureau.com/green-practices/the-seven-sins-of-greenwashing/>
- Hou, X., & Li, M. (2010). The Relationship between R&D and Performance of Listed Company in China. *2010 International Conference on E-Product E-Service and E-Entertainment*. <https://doi.org/10.1109/iceee.2010.5660774>
- IBM. (2023). *What is the k-nearest neighbors algorithm?* | IBM. [Www.ibm.com. https://www.ibm.com/topics/knn](https://www.ibm.com/topics/knn)
- Kearney. (2022, June 23). *Closing the consumer aspiration gap*. Kearney. <https://www.kenaninstitute.unc.edu/kenan-insight/esg-measurement-a-surprisingly-complex-issue/>
- Kenan Institute of Private Enterprise. (2022, February 17). *ESG Measurement: A Surprisingly Complex Issue*. Kenaninstitute.unc.edu. <https://kenaninstitute.unc.edu/kenan-insight/esg-measurement-a-surprisingly-complex-issue/>
- Our World In Data. (2019). *Productivity: output per hour worked*. Our World in Data. <https://ourworldindata.org/grapher/labor-productivity-per-hour-pennworldtable?tab=chart>
- PricewaterhouseCoopers. (2021). *Companies failing to act on ESG issues risk losing investors, finds new PwC survey*. PwC. <https://www.pwc.com/lt/en/about/press-room/pwc-global-investor-esg-survey.html>
- Rathore, D. S., Jain, R. C., & Ujjainiya, B. (2013). A text mining method for research project selection using KNN. *2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*. <https://doi.org/10.1109/icgce.2013.6823562>
- Schwartz, M. (2021, August 31). *Council Post: How Technology Can Provide A More Sustainable Future For The Industrial Sector*. Forbes. <https://www.forbes.com/sites/forbestechcouncil/2021/08/31/how-technology-can-provide-a-more-sustainable-future-for-the-industrial-sector/>
- Statista.com. (2022, August 17). *ESG ETF assets 2022*. Statista. <https://www.statista.com/statistics/1297487/assets-of-esg-etfs-worldwide/>
- The World Bank. (2020). *Life expectancy at birth, total (years) | Data*. Worldbank.org. <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>
- United Nations. (2015). *The 17 Sustainable Development Goals*. United Nations. <https://sdgs.un.org/goals>
- Xie, X., Huo, J., & Zou, H. (2019). Green process innovation, green product innovation, and corporate financial performance: A content analysis method. *Journal of Business Research*,

101,

697–706.

<https://doi.org/10.1016/j.jbusres.2019.01.010>