Using Natural Language Processing For Detecting Greenwashing Indicators And Constructing Impact-Focused Index Portfolio

Suyash Saxena

MSc Banking and Digital Finance, University College London (UCL), 2023-24

ABSTRACT

This research aimed to assess the emphasis that FTSE100 companies place on sustainability and ESG in their annual disclosures utilising Natural Language Processing (NLP). These disclosures were analysed both relatively and individually for each company, leading to the development of Greenwashing Indicators that may serve as early warning signs of potential greenwashing behaviour. In addition, Focus Scores based on environment were used as thematic factors to construct impact-focused index portfolios. While these portfolios did not outperform those based solely on ESG ratings in terms of risk and returns, the Modified Impact portfolio achieved the lowest Scope 1+2+3 emissions and had the highest proportion of portfolio revenues aligned with the United Nations Sustainable Development Goals, thereby maximising investment impact.

Keywords: Sustainability, ESG, Natural Language Processing (NLP), FTSE100, Corporate disclosures, Greenwashing, Impact Investing, Climate Change, Biodiversity, Pollution and Waste, GHG Emissions, UN SDGs.

JEL Codes: G11, G28, M14, O16, Q56.

1. INTRODUCTION

Environment, Social and Governance (ESG) is often used as a canvas to conceptualise sustainability. ESG ratings have been the most popular form of communicating a company's ESG performance which industry practitioners have relied on. However, these ratings tend to diverge with increasing level of corporate disclosure which creates short-term volatility in the markets affecting the investors (Senz 2021). ESG ratings can also overshadow one aspect over another, for instance, Tesla was rated highly for its environmental contribution despite its corporate governance issues (Grieve, Halper, and Shriver 2022). Johnson (2023) further demonstrated that funds rated highly on ESG lost up to 90% of their carbon reduction potential when weighted by ESG ratings, underscoring concerns about what these ratings genuinely measure.

The divergence in ESG ratings can also foster greenwashing, where companies exaggerate the environmental-friendly claims related to their products and services. McDonald's promoted its use of paper straws without emphasizing on its environmental footprint (Rose 2024), IKEA sold chairs made of certified timber wood, which was later discovered to be illegally sourced, and Volkswagen installed emission manipulation software in their vehicles (Okunade 2023). These examples suggest how corporates can mislead investors, while improving their ESG ratings and creating a positive market sentiment.

This research uses Natural Language Processing (NLP) to estimate corporate focus on sustainability and ESG in annual disclosures, one of the most common sources of greenwashing. It includes a relative analysis of FTSE100 companies along with an internal analysis of their environmental focus, including claims and actions, and a thematic analysis focusing on climate change, nature and biodiversity, and pollution. Based on these analyses, this research aims to explore the following questions:

- Can NLP techniques detect greenwashing indicators from corporate reports.
- Can NLP-based Focus Scores be used to adjust constituent weights for index funds.

The beneficiaries of this approach include retail investors, asset managers and regulators. For instance, retail investors can ascertain if a company is either skimming or overemphasizing on sustainability and ESG. This approach can also complement the existing methodologies incorporated by asset managers and help them perform a more refined ESG performance analysis of their portfolios. In addition, specific ESG themes and their importance can be identified and used as factors for tilting constituent weights, diversifying the offerings and aligning them with investor preferences.

Regulators are also becoming increasingly aware of greenwashing risks, as evident by the emergence of reporting regulations and frameworks. The European Union (EU) has been the front-runner in introducing regulations regarding sustainability classification and disclosures, for instance the Sustainable Finance Disclosure Regulation, the EU Taxonomy, the Corporate Sustainability Reporting Directive and European Sustainability Reporting Standards. Similar regulations were introduced by the United Kingdom (UK) Financial Conduct Authority in the form of Sustainability Disclosure Requirements. The Greenwashing Indicators framework can complement the existing practices for compliance monitoring of companies. It can be used to detect early warning signs and put those companies under a watchlist. Thus, this research aims to benefit multiple stakeholders within the sustainable finance domain.

This paper is organized as follows: Section 2 reviews the relevant literature background to establish the context for this study. Section 3 outlines the methodology, including the research framework, data collection, and analysis techniques. Section 4 presents the findings and their interpretation. Section 5 discusses the study's limitations and proposes directions for future research. Finally, Section 6 concludes by summarizing the research objectives and key findings.

2. LITERATURE REVIEW

ESG Ratings Conundrum and the Risk of Greenwashing

ESG ratings are often perceived as an indicator of a company's ESG performance. Although this approach seems simplistic, it has its own drawbacks. Berg, Kolbel, and Rigobon (2022) demonstrated that these ratings tend to diverge for different ratings providers due to their respective methodologies. Larcker et al. (2022) further highlighted that investments in companies with higher ESG ratings often does not reduce the market risk for the investors. A study by Raghunandan and Rajgopal (2021) found that ESG funds invested in companies with worse ESG performance compared to that of non-ESG labelled funds and their ratings were correlated more with the volume of ESG disclosure rather than actual emission values.

From an impact measurement viewpoint, Kräussl, Rauh, and Stefanova (2024) showed that higher ESG ratings often result in companies increasing their equity issuance leveraging on the positive market perceptions with no material change in the capex for the companies. Duchin, Gao, and Xu (2022) studied greenwashing within real asset markets and found that usually there is no significant change in the emission levels of plants after their divestiture as the buyers continue to operate them with minimal modifications while the sellers improve their ESG ratings and market perception. Hassani and Bahini (2024) also introduced cross-washing, a form of greenwashing where firms invest in sustainable projects to improve their ESG ratings but do not reduce their prevailing unsustainable core operations.

Corporate Communication and Natural Language Processing

Dempere, Ebrahim Alamash, and Mattos (2024) found the use of superficial language by companies in their disclosures to be the most common form of greenwashing. It was further highlighted that the emerging regulatory frameworks are still voluntary in many regions and face weak enforcement across all industries. This was also supported by Haggin (2024) by highlighting how managers use exaggerating language in disclosures to portray a positive picture of the company and gain a higher ESG rating, indicating a positive relation between ESG ratings and greenwashing. Based on a study conducted on FTSE350 companies, Purver et al. (2022) demonstrated that the usage of ESG terms has increased since 2015-2016 in the disclosures filed by these companies.

Natural Language Processing (NLP) can be an efficient way to analyse corporate communication, as demonstrated by Schimanski et al. (2023) by developing BERT-based models for the Environment, Social and Governance sentiments of corporate disclosures. Their research also established a relationship between communication made by the companies and their combined ESG rating. Similarly, Lohre et al. (2023) developed a similar model called ControversyBERT which was aimed at detecting derogatory sentiment from news headlines. One of the key findings of this research was the negative relation between controversies and the stock price of the companies. This was also supported by Zhang (2023) by analysing ESG related announcements and their corresponding effect on the market returns. This shows how NLP techniques can be utilised to better understand the textual context of company reports.

Greenwashing Detection and Portfolio Construction

Hu et al. (2024) quantified greenwashing as the difference between Bloomberg ESG scores and SINO ESG scores of Chinese A-share companies. A positive difference between the ESG disclosure of a company to its ESG performance relative to its peers was considered as a sign of greenwashing. Similarly, Ghitti, Gianfrate, and Palma (2020) defined greenwashing for S&P 500 companies as the difference between Newsweek Green Rankings scores and their Thompson Reuters ESG scores. As an alternative to relying on third-party information, Lagasio (2023) came up with the concept of Greenwashing Severity Index (GSI) for a dataset including 702 globally listed companies. The GSI was determined using the TF-IDF scores symbolizing how much emphasis a company puts on ESG through its disclosures relative to other companies. Gorovaia and Makrominas (2024) further highlighted that firms associated with environmental violations were engaged in filing longer and more positive corporate social responsibility (CSR) reports. Overcoming the limitations of statistical NLP methods, Lipenkova, Lu, and Rao (2023) adopted a deep learning approach using large language models (LLMs) to quantify greenwashing for the DAX index companies of Germany. The dataset included both internal and external disclosures of companies and the difference between the sentiment of these two was used to measure greenwashing. Zhao et al. (2023) also used similar approach to measure greenwashing as the difference between the sentiment of internal company disclosures and that of external text sourced from Twitter.

NLP techniques can also be utilized for ESG investing and driving sustainable investments. Goutte et al. (2023) demonstrated that sentiment analysis scores for STOXX600 companies based on ESG-related news can be used to ideate new trading strategies. Antoncic et al. (2020) further highlighted that there can be a positive relation between UN SDGs, ESG performance, and investment alpha by constructing a sector neutral portfolio using MSCI scores and Global AI Corp.'s SDG scores. This suggests that analysing corporate communication using NLP techniques can provide valuable insights which could lead to multiple use cases within sustainable finance.

3. METHODOLOGY

Data

This research was based on FTSE100 companies, including the top 101 companies within the UK based on market capitalisation (m-cap) as of June 2024. The aim was to study companies' overall performance through annual reports (AR) along with analysing sustainability/ESG reports (ESGR) for FY2023. All the reports were first converted into markdown format using the PyMuPDF library in python, and then cleaned using a set of preprocessing steps using the spaCy library in python. For the TF-IDF model, input data was in the form of individual tokens. The tokenization along with lowercasing and stop words removal was implemented using the scikit-learn TF-IDF vectorizer in python. For transformer models, the preprocessing steps included removing page markers, special characters, excessive whitespaces, and unwanted punctuation. Finally, the textual data was split into sentences and stored in two dataframes – one for AR consisting of around 491,294 sentences, and another one for ESGR consisting of 69,618 sentences.

Statistical Natural Language Processing: TF-IDF

The Term Frequency-Inverse Document Frequency (TF-IDF) model is a statistical NLP method to capture the relative importance of keywords within a corpus of documents. The Term Frequency (TF) measures the frequency of keywords in a document, while the Inverse Document Frequency (IDF) highlights the ubiquity of the keywords across all documents. Mathematically, it can be expressed as:

$$TF = \frac{Frequency\ of\ term\ t\ in\ document\ d}{Total\ number\ of\ terms\ in\ that\ document\ d}$$

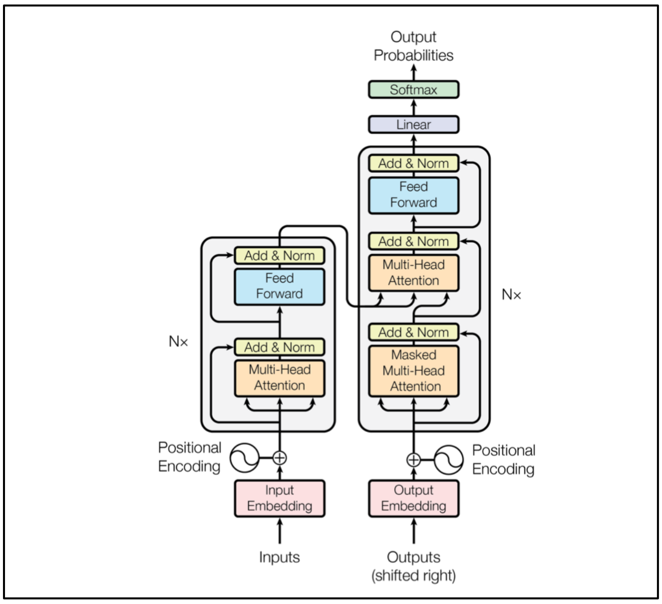
$$IDF = \log\left(\frac{Total\ number\ of\ documents\ in\ the\ corpus}{Number\ of\ document\ containing\ term\ t}\right)$$

$$TF - IDF(t,d) = TF(t,d)\ x\ IDF(t)$$

A low TF-IDF score for a certain keyword in a document indicates that a relatively low focus has been put on that keyword in the given document and vice-versa. Thus, it can identify the relative significance of a topic across all documents (Qaiser & Ali, 2018). In contrast to its simplistic nature, the primary limitation of a TF-IDF model is that it does not account for the position of keywords in the document, and thus is not efficient in capturing the context of the given text. (Ghag & Shah, 2014).

Moving Towards Deep Learning: Transformer Models

Exhibit 1: Transformer Architecture



Source: Vaswani et al. (2017)

The Transformer model was introduced by Google to overcome the limitations of Recurrent Neural Networks (RNNs), which processes input sequences one step at a time and often struggles with long-range dependencies due to vanishing gradient. On the other hand, Transformer models can process the input data in one single timestep and efficiently capture the relationships between tokens irrespective of their position within the given text. A transformer model consists of two key components: encoder and decoder. The encoder processes the input sequence and generates embeddings and positional encoding for input tokens which helps the model to learn the context of the textual data. These input embeddings are converted into Query (Q), Key (K), and Value (V) vectors, which are used to estimate the relevance of each token by taking the dot product of Q and K and using this dot product as weights to sum the V vectors. This distinguishing feature of Transformer models is called 'self-attention' which is performed through multiple attention heads and is then passed through a Feed Forward Neural Network with a ReLU activation function. Each sub-layer is followed by a normalization layer to stabilize the training process. The encoder's output is then used by decoder through cross-attention. The decoder input is processed as the Q vector, while the hidden state from encoder is used as K and V vectors. This allows the decoder to learn from both the previously generated tokens as well as from the current input sequence. The decoder output is passed through a linear layer with a SoftMax activation function, before generating the final output (Vaswani et al. 2017).

The transformer-based models deployed in this research include BERT, RoBERTa, DistilRoBERTa and FinBERT. BERT (Bidirectional Encoder Representations from Transformers) was first introduced by Devlin et al. (2019) and is an encoder-only model. Unlike a traditional transformer model, BERT is bi-directional, i.e., it processes input both right-to-left and left-to-right, which improves its ability to learn underlying context. BERT's pre-training includes Masked Language Modelling (MLM), which masks random words and trains the model to predict them, and Next Sentence Prediction (NSP) which learns the relationships between sentences. Additionally, BERT uses WordPiece tokenization that can handle rare or unseen words by breaking them into sub-words to make better predictions. RoBERTa was developed by Meta as an optimized version of BERT. While both the models are structurally similar, RoBERTa differs in its training approach, using only MLM, which improves the contextual learning between words. In addition, RoBERTa was trained on a much larger dataset than BERT, yielding superior performance but also increased training cost and time (Liu et al. 2019). On the other hand, DistilRoBERTa is a lighter version of RoBERTa, which uses only six layers instead of twelve, making it faster and more efficient but slightly less accurate. It is easier to deploy with limited computational resources. Lastly, FinBERT is variant of BERT, specifically fine-tuned for financial contexts using financial news, reports, and disclosures, enhancing its accuracy in identifying financial sentiment (Kumar and Chaturvedi 2024). These models can be fine-tuned for specific tasks adapting efficiently to domain-specific data.

Focus Scores

One of the key components of this research is estimating Focus Scores for the companies, reflecting the level of importance or 'focus' a company puts on sustainability and ESG in its disclosures. Lagasio (2023) introduced Focus Scores based on the term frequency ratio of ESG-related keywords. However, this research is more nuanced as Focus Scores are further divided into Internal Focus and Relative Focus. In addition, the term frequency approach is replaced by TF-IDF and Transformer models. Since greenwashing closely ties with environment, deep learning models focused on the environmental perspective, whereas the TF-IDF model addressed broader ESG dimensions.

Internal Focus Scores are based on companies' internal reporting practices and reflect the emphasis companies put on environmental issues. Company reports were analysed on a sentence level using ESGBERT/EnvRoBERTa-environmental (EnvRoBERTa) and ESGBERT/EnvironmentalBERT-environmental (EnvBERT), both fine-tuned on 2,000 environment related sentences (Schimanski et al., 2023). Each sentence was classified each as either 'Environmental' or 'none' based on their contextual relevance with environment. Ratios of environmental sentences to total sentences were calculated for each company using both the models. These ratios were averaged across reports for each company, normalised using z-score, and converted into percentile ranks using cumulative distribution function (CDF) to derive the Environment Scores.

The sentences identified as 'Environmental' were then classified separately as environmental 'claims' and environmental 'actions', using ClimateBERT/environmental-claims (EnvClaims) and ESGBERT/EnvironmentalBERT-action (EnvAction), finetuned on a sample of 2,650 sentences and 500 sentences respectively (Stammbach et al. 2022; Schimanski et al. 2023). Ratios of sentences identified as claims and actions were estimated and normalised in a similar way to yield the Claims Scores and Actions Scores.

Thematic scores were derived using FinBERT ESG-9 model, fine-tuned and pretrained on over 14,000 manually annotated sentences (Huang, Wang, and Yang, 2022). Each sentence was classified into nine different categories based on their contextual relevance, including Climate Change, Natural Capital, and Pollution and Waste. Ratios of sentences falling into the aforementioned categories were estimated and normalised, producing Thematic Scores for Climate Change, Natural Capital, and Pollution and Waste.

Exhibit 2: Model Specification Summary

Large Language Model	Architecture	Dataset for Fine-tuning	Scores Derived
EnvRoBERTa	RoBERTa	2,000 sentences	Environment Scores
EnvBERT	BERT	2,000 sentences	Environment Scores
EnvAction	BERT	500 sentences	Action Scores
EnvClaims	DistilRoBERTa	2,650 sentences	Claims Scores
FinBERT ESG-9	BERT	14,000 sentences	Thematic Scores

Source: Stammbach et al. (2022); Schimanski et al. (2023); and Huang, Wang, and Yang, (2022)

Relative Focus Scores were estimated using a TF-IDF model reflecting how much importance a company puts on sustainability and ESG relative to other FTSE100 companies. An extensive vocabulary was prepared using ESG and sustainability relevant keywords. The TF-IDF model then tokenized the content of company reports and transformed these tokens into a vectorized form to estimate the TF-IDF scores for the entire vocabulary. These scores were aggregated by ESG category and then by company. They were normalised twice: first at a company level to adjust for the report length as annual reports are lengthier than sustainability reports, and then using the z-score to derive the Relative Focus Scores. Since the focus on environment might be different within different industries, a broader approach was adopted by addressing all the ESG dimensions.

Greenwashing Indicators

The concept of Greenwashing Indicators is inspired from the Greenwashing Severity Index introduced by Lagasio (2023) by quantifying the greenwashing potential of companies using their TF-IDF scores. The underlying logic was that higher the TF-IDF score, higher will be the likelihood of greenwashing as these companies could be overemphasising their ESG communication in their reports. However, this research introduced a novel three-level framework to raise potential greenwashing flags which starts with a broader analysis and then narrows down the scope in subsequent steps.

First, the Relative Focus Scores were used to reflect the level of focus a company puts on ESG relative to other companies. However, based on the differences in the nature of industry and the size of a company, it was deemed plausible that its TF-IDF score could be exceptionally high or low. In addition, a company with a high TF-IDF score might be a leader in adopting sustainable practices whereas other companies are laggards. Hence, two additional steps were added to tackle this issue.

After comparing these companies on a relative level, they were then compared specifically on their focus on environment using the Environment Scores. However, it was still deemed plausible that a company might have a high Environment Score due to its size as large-cap companies usually have more resources to spare compared to their smaller competitors. Therefore, a third step was introduced by defining the Net Action as the difference between Actions Score and Claims Score. The Net Action would highlight the gap between a company's claims and its actions towards the environment.

Finally, the Greenwashing Indicators were estimated by combining the three-level analysis mentioned above. This research did not aim specifically at quantifying the greenwashing potential of a company to one score as that can be potentially misleading for the stakeholders. Hence, a more granular approach was adopted by observing these greenwashing indicators from three different dimensions. To the best of our knowledge, no prior work has examined greenwashing with such granularity.

Portfolio Construction: Adjusting Index Fund for Impact

ESG ratings are often inconsistent in terms of the underlying methodologies adopted by the rating providers. Their methodologies are usually a black box, posing a challenge for asset managers to make informed investment decisions. This research introduced another novel approach for constructing index portfolios and measuring their impact. This approach could also help tackle the divergence in ESG ratings and can complement the existing practices adopted by different asset managers.

Constituent Weights: This research focused on two portfolio construction strategies using constituents from the FTSE100 index. The first, called the Base Impact portfolio, estimated the base weights corresponding to the m-cap which were then modified to align more closely with the United Nation's Sustainable Development Goals (SDGs) 12, 13, 14, and 15. Constituent weights were adjusted by using the Thematic Focus Scores as per the tilt-based methodology introduced by FTSE Russell (2017). The second portfolio, called Modified Impact, prioritised Environment Scores over m-cap for the base weight to avoid size bias towards large-cap companies. In both the portfolios, the final tilt was derived using the Actions Scores to counter greenwashing by rewarding companies with stronger environmental efforts. To the best of our knowledge, this represents a novel approach to modifying constituent weights in an index portfolio to enhance environmental and social impact.

Backtesting: The base m-cap weights were separately tilted as per the ESG scores from Refinitiv and Sustainalytics to construct two additional portfolios for comparison. All four portfolios were backtested January 2024 to September 2024 using the Portfolios and Lists (PAL) and Portfolio Analytics (PORTF) applications within LSEG Workspace. The timeframe for backtesting was deemed appropriate to correspond with the FY2023 reports. In practice, the timeframe for backtesting and the frequency of rebalancing would depend on the regulations around company reporting and the discretion of asset managers.

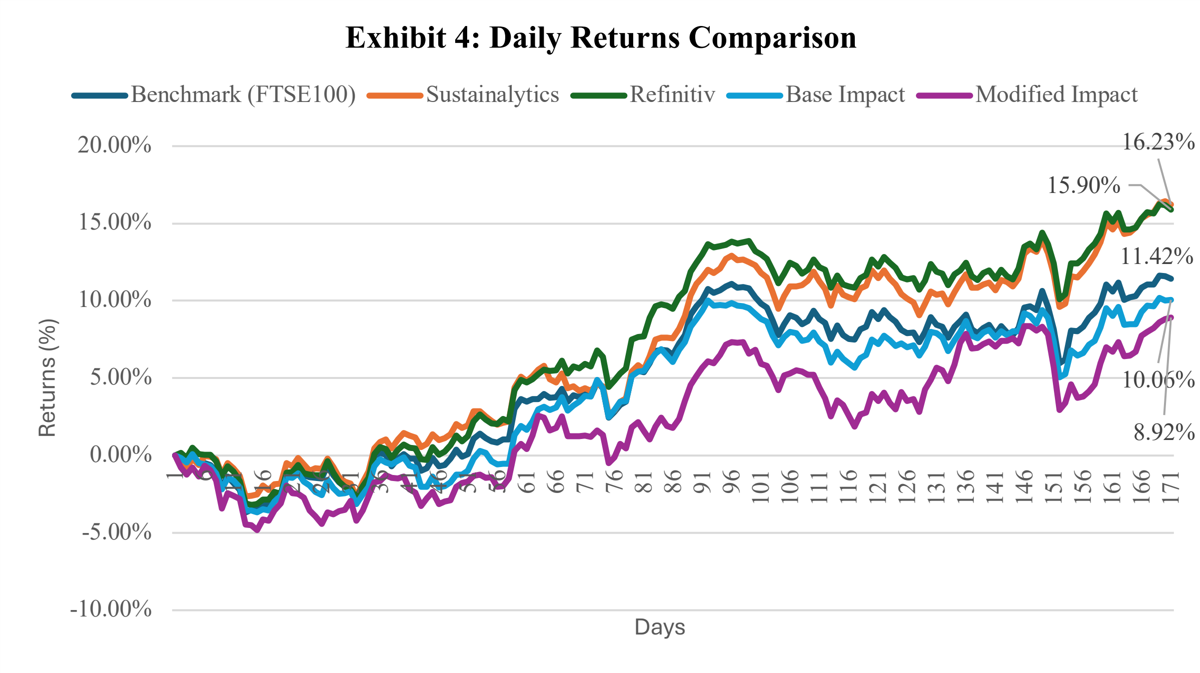
Impact Measurement: To measure the overall impact of portfolios, Matter's SDG Revenue Fundamentals dataset was utilised. This dataset consisted of a range of metrics associated with UN SDGs, including aligned/potentially aligned revenues and misaligned/ potentially misaligned revenues, denoted as percentages of overall company revenue. Weighted averages were estimated for all four portfolios using constituent data. Precisely, the following metrics were estimated: percentage of portfolio's revenue – a) aligned/misaligned with one or more targets across all SDGs (denoted as SDG_ALL_A / SDG_ALL_M); b) aligned/misaligned with one or more environmental targets across all SDGs (denoted as SDG_ALL_E_A / SDG_ALL_E_M); c) aligned/misaligned with one or more social targets across all SDGs (denoted as SDG_ALL_S_A / SDG_ALL_S_M).

4. RESULTS AND DISCUSSION

Exhibit 3: Greenwashing Indicators - Industrials

Company	ICB Industry	Relative Focus Score	Environment Score	Claims Score	Actions Score	Net Action Direction
Ashtead Group Plc	Industrials	85.23	59.14	34.93	35.19	Positive
BAE Systems Plc	Industrials	44.01	12.23	57.33	26.72	Negative
Bunzl Plc	Industrials	8.67	88.40	48.53	86.93	Positive
DCC Plc	Industrials	24.41	70.16	69.16	5.83	Negative
Diploma Plc	Industrials	29.00	27.97	79.80	81.66	Positive
Experian Plc	Industrials	26.67	5.12	30.81	36.58	Positive
Halma Plc	Industrials	34.52	14.39	11.27	11.47	Positive
IMI Plc	Industrials	39.05	25.44	68.51	88.94	Positive
Intertek Group Plc	Industrials	35.24	43.08	27.37	21.44	Negative
Mondi Plc	Industrials	85.06	92.83	97.71	88.93	Negative
Melrose Industries Plc	Industrials	73.67	89.60	73.74	57.77	Negative
Rolls-Royce Holdings Plc	Industrials	35.37	14.35	26.85	8.87	Negative
Rentokil Initial Plc	Industrials	13.71	7.00	53.60	25.42	Negative
Smurfit Kappa Group Plc	Industrials	91.61	95.37	96.01	55.87	Negative
DS Smith Plc	Industrials	94.50	95.98	96.66	83.99	Negative
Smith & Nephew plc	Industrials	75.58	53.75	58.55	72.43	Positive
Severn Trent Plc	Industrials	36.41	97.45	83.51	97.25	Positive
The Weir Group Plc	Industrials	26.35	17.74	39.22	38.54	Negative

The Greenwashing Indicators were based on a three-level framework – identifying companies with disproportionately high relative focus scores, supported by high environment scores, along with a negative Net Action direction, i.e., having Claims score higher than Action score. These indicators seemed to be in alignment for Mondi plc, Smurfit Kappa Group Plc and DS Smith plc. These companies fall in the top 20% for their Relative Focus scores and Environment scores. Additionally, the Net Action direction for these companies is also negative. This might pose a critical question for the external stakeholders, that if the net action direction is negative, why these companies are emphasizing disproportionately on environment and ESG compared to other companies. This can be flagged as an early warning sign which can be investigated further.



Source: LSEG Workspace

In terms of portfolio construction, the daily pricing movements of all four portfolios were plotted against that of FTSE100. Refinitiv and Sustainalytics portfolios generated excess returns than the benchmark resulting in a positive alpha. On the other hand, the impact portfolios could not generate excess returns than the market. Although impact investing strategies are not usually focused on maximising alpha, asset managers should abide by their fiduciary duty to act in the best interest of investors.

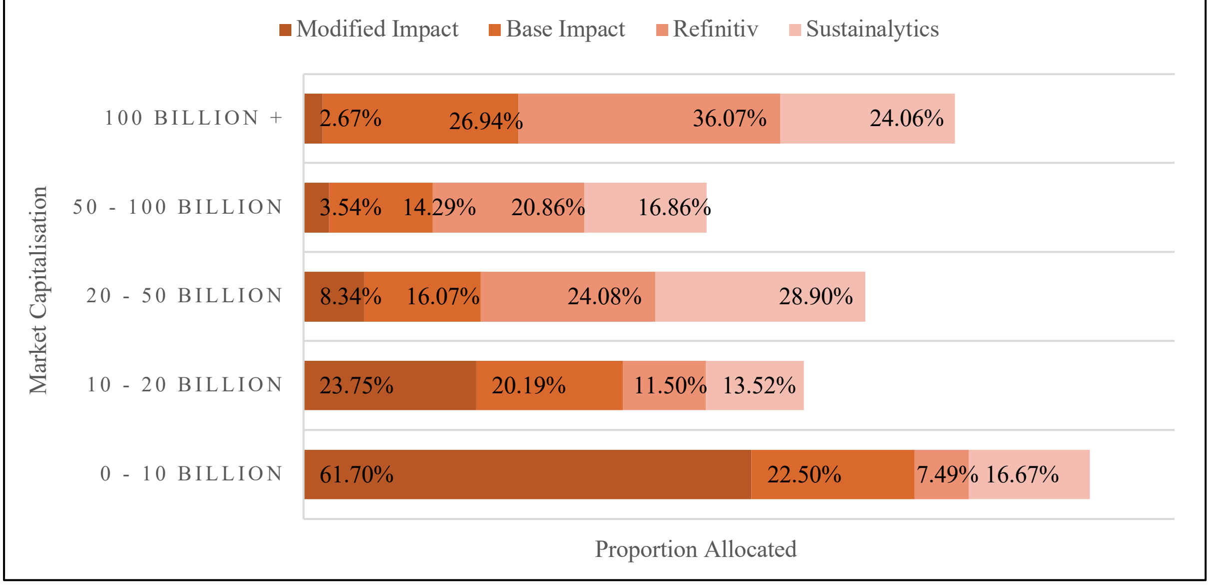
Exhibit 5: Portfolio Risk

Portfolio	Sharpe Ratio	Standard Deviation (%)	Tracking Error (%)
FTSE100	1.12	9.28	0.00
Sustainalytics	1.79	9.70	2.96
Refinitiv	1.77	9.49	1.56
Base Impact	0.88	9.59	2.90
Modified Impact	0.61	11.20	6.37

Source: LSEG Workspace

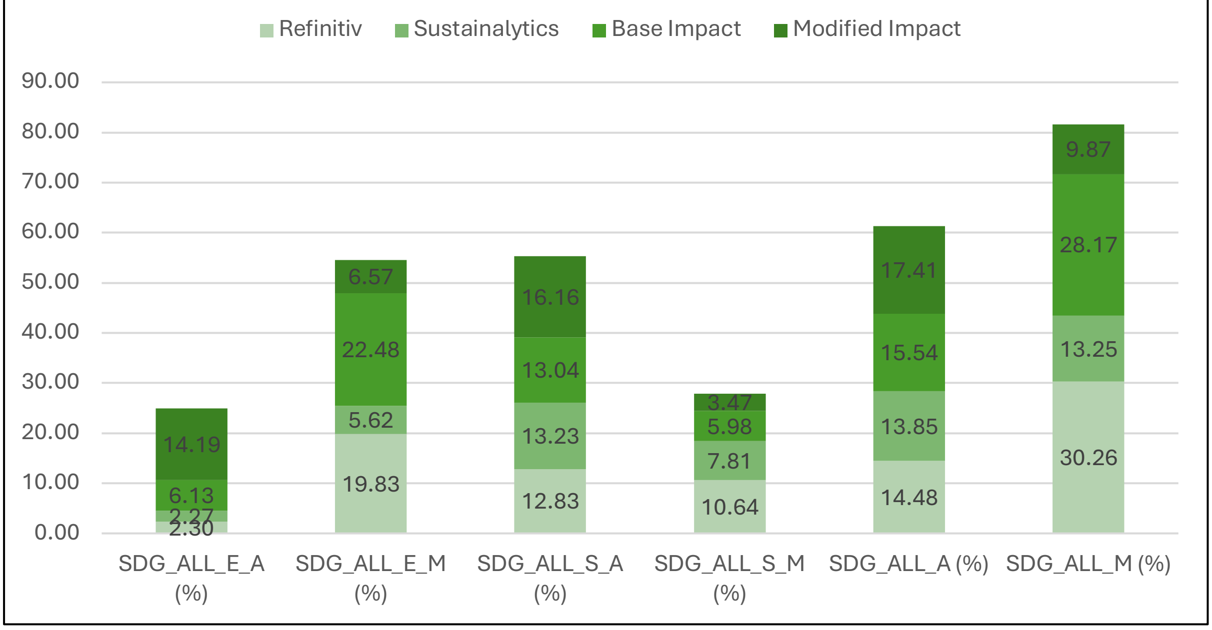
From a risk perspective, Refinitiv and Sustainalytics portfolios had a lower standard deviation, higher sharpe ratio, and lower tracking error than that of the impact portfolios. However, upon analysing the portfolio allocation, it was observed that the impact portfolios were more heavily weighted in small-cap and mid-cap companies, whereas Refinitiv and Sustainalytics portfolios were weighted more towards large-cap companies. This possibly explains the lower standard deviation of these portfolios as large-cap companies are deemed as safer and less volatile investment options.

Exhibit 6: Portfolio Allocation



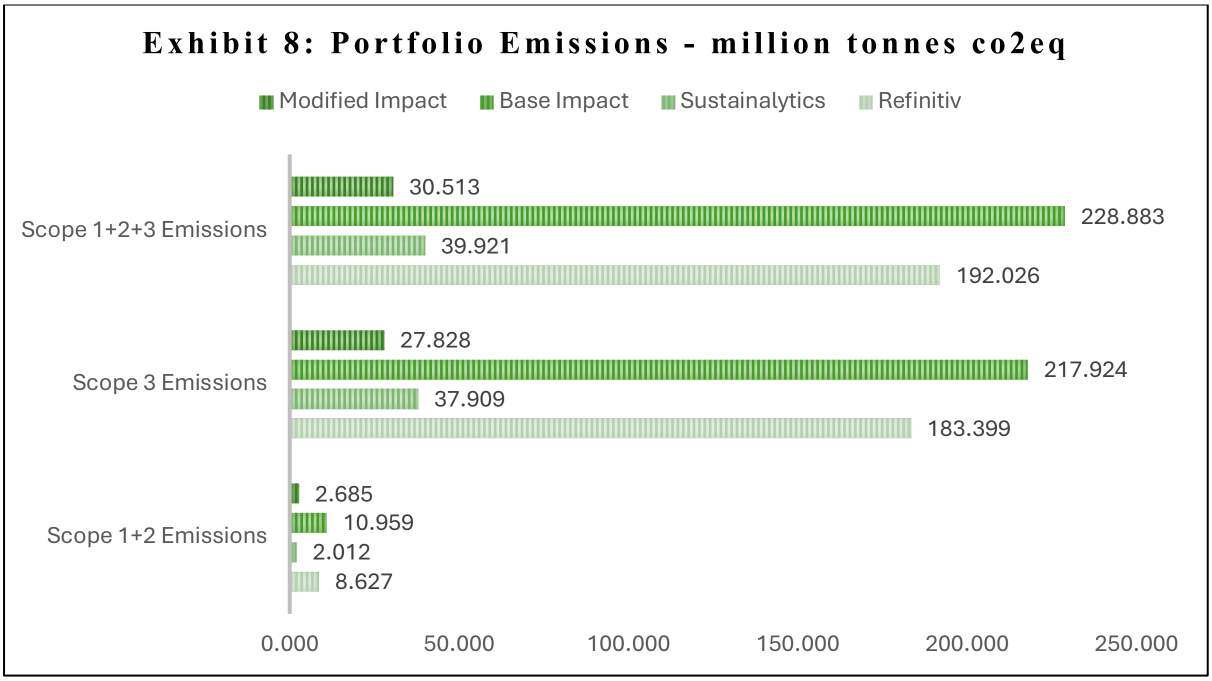
Source: LSEG Workspace

Exhibit 7: Impact Measurement - % Revenues Aligned with UN SDGs



Source: Matter SDG Revenue Fundamentals

From an impact viewpoint, all four portfolios were evaluated using Matter's SDG Revenue Fundamentals dataset. The modified impact portfolio had higher SDG-aligned portfolio revenues across all the categories, thus achieving a higher impact than other three portfolios. Measuring the impact using SDG-aligned revenues can be an effective way to conduct sustainability analysis of investments as it compensates for the size-bias. A large organisation with higher revenues would not be considered sustainable unless its revenues are SDG-aligned.



Source: LSEG Workspace

Thus, the novel portfolio construction approach introduced in this research proved effective in increasing the alignment of investments with SDGs. Additionally, the modified impact portfolio was found to have the lowest amount of Scope 1+2+3 emissions. This indicates that using Environment Scores to determine base weights proved to be more efficient in rewarding lower emission generating companies.

5. LIMITATIONS AND FURTHER RESEARCH

As NLP models work with textual data and can be highly resource extensive, this research was limited to FY2023 and utilised pretrained deep learning models due to limited computational resources. Additionally, only publicly available, university subscribed, and freely accessible proprietary databases were used for this research. Nevertheless, this research can be regarded as a prototype for conducting extended research at a larger scale. Precisely, the frameworks introduced in this research can be tested on bigger sample of companies. Further research can also focus on a timeseries by analysing historical reports and identifying the underlying trends. The Focus Scores can also be combined with sentiment analysis of news headlines to test new trading strategies. Lastly, the deep learning models used in this research can be fine-tuned further using additional training data for different purposes.

6. CONCLUSION

With growing awareness about responsible and impact investing, it is important to address the limitations of how ESG is integrated with investment decision-making. Industry practitioners have long relied on ESG ratings provided by a few dominant rating agencies. However, the divergence between ESG ratings from different providers along with their unclear methodologies not only creates a state of confusion for investors and regulators but also gives rise to the possibility of greenwashing. As corporate communication is a popular source of greenwashing, this research was aimed at analysing the annual and sustainability reports of 101 companies from the FTSE100 using NLP models and deriving potential greenwashing indicators. Additionally, the internal focus scores estimated using NLP models were further utilised to construct impact-focused index portfolios, which were then compared against portfolios tilted as per the ESG scores from Refinitiv and Sustainalytics. The novel framework introduced in this research resulted in a negative alpha but achieved a higher overall impact. The modified impact portfolio of this research achieved a higher proportion of SDG-aligned portfolio revenues and lesser total emissions compared to other portfolios.

REFERENCES

Antoncic, Madelyn, Geert Bekaert, Richard V Rothenberg, and Miquel Noguer. 2020. "Sustainable Investment - Exploring the Linkage between Alpha, ESG, and SDG's." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3623459.

Berg, Florian, Julian F Kolbel, and Roberto Rigobon. 2022. "Aggregate Confusion: The Divergence of ESG Rating." *Review of Finance* 26 (6): 1315–44. https://doi.org/10.1093/rof/rfac033.

Dempere, Juan, Ebrahim Alamash, and Paulo Mattos. 2024. "Unveiling the Truth: Greenwashing in Sustainable Finance." *Frontiers in Sustainability* 5 (May). https://doi.org/10.3389/frsus.2024.1362051.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." ArXiv.org. May 24, 2019. https://doi.org/10.48550/arXiv.1810.04805.

Dominik Stammbach, Nicolas Webersinke, Julia Anna Bingler, Mathias Kraus, and Markus Leippold. 2022. "Environmental Claim Detection," September. https://doi.org/10.48550/arxiv.2209.00507.

Duchin, Ran, Janet Gao, and Qiping Xu. 2022. "Sustainability or Greenwashing: Evidence from the Asset Market for Industrial Pollution." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4095885.

FTSE Russell. 2017. "Multi-Factor Indexes: The Power of Tilting." FTSE Russell. https://www.lseg.com/content/dam/ftse-russell/en_us/documents/research/multi-factor-indexes-power-of-tilting.pdf.

Ghag, Kranti, and Ketan Shah. 2014. "SentiTFIDF – Sentiment Classification Using Relative Term Frequency Inverse Document Frequency." *International Journal of Advanced Computer Science and Applications* 5 (2). https://doi.org/10.14569/ijacsa.2014.050206.

Ghitti, Marco, Gianfranco Gianfrate, and Lorenza Palma. 2020. "The Agency of Greenwashing." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3629608.

Gorovaia, Nina, and Michalis Makrominas. 2024. "Identifying Greenwashing in Corporate-Social Responsibility Reports Using Natural-Language Processing." *European Financial Management*, July. https://doi.org/10.1111/eufm.12509.

Goutte, Stephane, Fei Liu, Hoang Viet Le, and Hans-Jorg von Mettenheim. 2023. "ESG Investing: A Sentiment Analysis Approach." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4316107.

Grieve, Duncan, Jason Halper, and Timbre Shriver. 2022. "ESG Ratings: A Call for Greater Transparency and Precision." The Harvard Law School Forum on Corporate Governance. November 10, 2022. https://corpgov.law.harvard.edu/2022/11/10/esg-ratings-a-call-for-greater-transparency-and-precision/.

Haggin, Imri. 2024. “Significance of Transparency and Accountability in ESG Reporting.” https://cs.brown.edu/media/filer_public/ce/25/ce258e58-578e-4f71-bc5d-f49945648bfe/hagginimri.pdf.

Hassani, Bertrand Kian, and Yacoub Bahini. 2024. “Crosswashing in Sustainable Investing: Unveiling Strategic Practices Impacting ESG Scores.” *ArXiv (Cornell University)*, March. https://doi.org/10.48550/arxiv.2407.00751.

Hu, Peng, Xuming Li, Nian Li, Yiying Wang, and Derek D Wang. 2024. “Peeking into Corporate Greenwashing through the Readability of ESG Disclosures.” *Sustainability* 16 (6): 2571–71. https://doi.org/10.3390/su16062571.

Huang, Allen H., Hui Wang, and Yi Yang. 2022. “FinBERT: A Large Language Model for Extracting Information from Financial Text†.” *Contemporary Accounting Research*, September. https://doi.org/10.1111/1911-3846.12832.

Johnson, Steve. 2023. “Companies with Good ESG Scores Pollute as Much as Low-Rated Rivals.” *Financial Times*, July 31, 2023, sec. Exchange traded funds. https://www.ft.com/content/b9582d62-cc6f-4b76-b0f9-5b37cf15dce4.

Kraeussl, Roman, Joshua D Rauh, and Denitsa Stefanova. 2024. “Is ESG a Sideshow? ESG Perceptions, Investment, and Firms’ Financing Decisions,” January. https://doi.org/10.2139/ssrn.4867910.

Kumar, Saurabh, and Rahul Chaturvedi. 2024. “Evaluating the Efficacy of Distilled Transformer Models for Sentiment Analysis in Financial Texts: A Comparative Study.” *International Journal of Current Science* 14 (2): 2250–1770. https://www.rjpn.org/ijcspub/papers/IJCSP24B1335.pdf.

Lagasio, Valentina. 2023. “Measuring Greenwashing: The Greenwashing Severity Index.” Social Science Research Network. Rochester, NY. September 25, 2023. https://doi.org/10.2139/ssrn.4582917.

Larcker, David F., Lukasz Pomorski, Brian Tayan, and Edward M. Watts. 2022. “ESG Ratings: A Compass without Direction.” Papers.ssrn.com. Rochester, NY. August 2, 2022. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4179647.

Lipenkova, Janna, Guang Lu, and Susie Rao. 2023. “Detecting Greenwashing Signals through a Comparison of ESG Reports and Public Media.” https://www.swisstext.org/wp-content/uploads/2023/09/Greenwashing.pdf.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar S Joshi, Danqi Chen, Omer Levy, Michael Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” *ArXiv (Cornell University)* 1 (July). https://doi.org/10.48550/arxiv.1907.11692.

Lohre, Harald, Sandra Nolte, Lakshmi Ranganathan, Carsten Rother, and Margit Steiner. 2023. “ControversyBERT: Detecting Social Controversies and Their Impact on Stock Returns.” *Journal of Impact & ESG Investing* 4 (1): 87. https://doi.org/10.3905/jesg.2023.1.076.

Okunade, Ibrahim. 2023. “Green Lies: The Most Infamous Greenwashing Examples Revealed.” Www.greenhive.io. November 24, 2023. https://www.greenhive.io/blog/greenwashing-examples.

Purver, Matthew, Matej Martinc, Riste Ichev, Igor Lončarski, Katarina Sitar Šuštar, Aljoša Valentinčič, and Senja Pollak. 2022. “Tracking Changes in ESG Representation: Initial Investigations in UK Annual Reports.” *ACL Anthology*, June, 9–14. https://aclanthology.org/2022.csrnlp-1.2.

Qaiser, Shahzad, and Ramsha Ali. 2018. “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents.” *International Journal of Computer Applications* 181 (1): 25–29. https://doi.org/10.5120/ijca2018917395.

Raghunandan, Aneesh, and Shivaram Rajgopal. 2021. “Do ESG Funds Make Stakeholder-Friendly Investments?” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3826357.

Rose, Amelia. 2024. “Greenwashing Examples 2024: Top 10 Greenwashing Companies Misleading Consumers - LytHouse | Empowering Sustainability.” Lythouse. June 27, 2024. https://www.lythouse.com/blog/top-10-greenwashing-companies-examples#:~:text=The%20examination%20of%20McDonald%27s%20paper.

Schimanski, Tobias, Andrin Reding, Nico Reding, Julia Bingler, Mathias Kraus, and Markus Leippold. 2023. “Bridging the Gap in ESG Measurement: Using NLP to Quantify Environmental, Social, and Governance Communication.” *Social Science Research Network*, January. https://doi.org/10.2139/ssrn.4622514.

Senz, Kristen. 2021. “What Does an ESG Score Really Say about a Company?” HBS Working Knowledge. July 21, 2021. https://hbswk.hbs.edu/item/what-does-an-esg-score-really-say-about-a-company.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” ArXiv.org. December 5, 2017. https://doi.org/10.48550/arXiv.1706.03762.

Zhang, Yimeng. 2023. “ESG-Based Market Risk Prediction and Management Using Machine Learning and Natural Language Processing.” https://cdn.shanghai.nyu.edu/sites/default/files/yimeng_zhang_thesis_nyush_honors_2023.pdf.

Zhao, Yue, Leon Kroher, Maximilian Engler, and Klemens Schnattinger. 2023. “Detecting Greenwashing in the Environmental, Social, and Governance Domains Using Natural Language Processing.” *SciTePress*, January, 175–81. https://doi.org/10.5220/0012155400003598.