CORRECT? CORECT!: Classification of ESG Ratings with Earnings Call Transcript

Haein Lee[1], Hae Sun Jung[2], Heungju Park[4] and Jang Hyun Kim[3]
[1]Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction,
Sungkyunkwan University, Seoul, 03063, Korea
[2]Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, 03063, Korea
[3] Department of Interaction Science/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan
University, Seoul, 03063, Korea
[4]SKK Business School, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: lhi00034@g.skku.edu, jestiriel@g.skku.edu, hj.park@skku.edu, alohakim@skku.edu]
*Corresponding author: Jang Hyun Kim

Abstract

While the incorporating ESG indicator is recognized as crucial for sustainability and increased firm value, inconsistent disclosure of ESG data and vague assessment standards have been key challenges. To address these issues, this study proposes an ambiguous text-based automated ESG rating strategy. Earnings Call Transcript data were classified as E, S, or G using the Refinitiv-Sustainable Leadership Monitor's over 450 metrics. The study employed advanced natural language processing techniques such as BERT, RoBERTa, ALBERT, FinBERT, and ELECTRA models to precisely classify ESG documents. In addition, the authors computed the average predicted probabilities for each label, providing a means to identify the relative significance of different ESG factors. The results of experiments demonstrated the capability of the proposed methodology in enhancing ESG assessment criteria established by various rating agencies and highlighted that companies primarily focus on governance factors. In other words, companies were making efforts to strengthen their governance framework. In conclusion, this framework enables sustainable and responsible business by providing insight into the ESG information contained in Earnings Call Transcript data.

Keywords: BERT, Earnings Call Transcript, ESG, Machine learning, Natural language processing (NLP)

1. Introduction

Over the past few years, the significance of environmental, social, and governance (ESG) factors has experienced a noteworthy surge, especially within the global landscape where collaborative efforts aim to mitigate carbon emissions and confront the intricate challenges posed by climate change [1]. The ongoing debate over the fairness and objectivity of ESG ratings arises due to notable inconsistencies and insufficient transparency exists in the criteria utilized by rating agencies for assessment. [2]. Furthermore, lack of standardized frameworks for disclosing ESG-related metrics significantly intensifies the challenge of ensuring fairness and credibility [3,4]. Considering the escalating volume of data in the ESG domain, the growing importance of advanced natural language processing (NLP) techniques is evident in the effective management and investigation of the comprehensive corpus of unstructured textual data [5,6]. In an endeavor to address these complex challenges, this research engages in the task of proposing a strategy for the objective assessment of corporate open data's ESG scores utilizing automated text-based methodologies. Earnings Call Transcript data were collected from Thomson Reuters and labeled E, S, or by employing the Refinitiv-Sustainable Leadership Monitor, utilizing more than 450 scales. The study employed advanced NLP techniques, including Bidirectional Encoder Representations from Transformers (BERT), A Robustly Optimized BERT Pretraining Approach (RoBERTa), A Lite BERT (ALBERT), Financial BERT (FinBERT), and Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA) models, for precisely classifying ESG documents. Furthermore, the authors averaged the predicted probabilities of each label to determine which ESG factors are more important. This comprehensive approach can help companies understand the factors behind their high ESG scores, providing insights for better decision-making.

2. Related Works

The authors examined previous studies on the importance of ESG. Afterwards, the authors reviewed previous research related to ESG ratings and examined the challenging factors associated with ESG ratings. Lastly, the authors explored studies aiming to apply NLP to ESG-related texts to gain insights.

2.1 ESG Value

As ESG continues to gain importance, associated research is developing. According to [7], enterprises that disregard their social obligations or exhibit inadequate governance encounter significant "hidden" endanger, potentially leading to economic losses or the need to cover costs associated with environmental litigation. According to [8], it was discovered that ESG is successful in lowering operational expenses, including raw material or carbon costs, potentially influencing a company's operating income by as much as 60%. Additionally, research [9,10] has shown that strong ESG practices within a company can draw in skilled employees, boost motivation, and ultimately raise overall productivity. Moreover, companies that prioritize social responsibility have shown a direct relationship to shareholder returns. Overall, research has demonstrated the potential financial benefits of incorporating ESG criteria into business operations. These benefits range from enhanced corporate performance indicators to cost savings and increased productivity. As a result, ESG is becoming increasingly important for companies as they strive for long-term sustainability and success.

2.2 ESG Rating

ESG indicators continue to be a topic of debate due to their perceived lack of impartiality. [3] notes that investors face significant challenges when incorporating ESG information into their investment strategies because there are no standardized guidelines for managing the reporting of ESG information. In addition, [4] conducted a comparison of different ESG ratings and information providers, emphasizing the challenges in discerning objectivity or fairness due to the insufficient transparency in the evaluation standard utilized by ESG rating companies. According to [11], there is a relationship between a company's credit evaluation and its corporate social responsibility (CSR). However, different ESG assessment agencies consider dissimilar individual components of CSR to be relevant, resulting in varied ESG ratings. To address these issues, standardized framework is necessary to impartially measuring ESG ratings.

2.3 Natural Language Processing (NLP) and ESG

This section summarizes existing research that has utilized natural language processing (NLP) approaches to extract insights from ESG-related corpus data. Reference [12] employed NLP techniques to translate textual data from social media into ESG scores and utilized BERT for the recognition of ESG risks. In Reference [13], large datasets sourced from newspaper articles and scholarly documents were explored using BERTopic to uncover major keywords and topics related to ESG, proposing a comprehensive ESG management strategy. In Reference [14], the authors investigated past trends in ESG conversations by reviewing corporate earnings call transcripts. Their findings revealed that 15 perecent of statements made during these calls over the past five years were linked to ESG, highlighting the increasing significance of ESG within corporate strategies.

3. Method

In the method part, description on overall process of the experiment is presented. First, details of data collection and preprocessing process are illustrated. The second part describes the models used for the experiment. And the last part outlines the model evaluation methodology and probability analysis.

3.1 Data Collection and Preprocessing

The authors collected Earnings Call Transcript data from Thomson Reuters between January 1, 2014, and December 31, 2018, resulting in a dataset of 3,574,923. Then the authors employed the ESG pillar scores from the Refinitiv-Sustainable Leadership Monitor [15,16,17] to label the data. This metric is formulated for assessing the comparative ESG performance of companies within a particular industry, measuring ESG achievement and efficiency.

3.2 Model

The NLP approach suggested in the study uses a methodology based on transformer architecture language modeling, which is capable of recognizing linguistic features and learning general language characteristics [18]. This approach comprises two stages: pretraining and fine-tuning, where a substantial corpus is preprocessed to grasp the fundamental characteristics of the language. The authors evaluated different models, including BERT, RoBERTa, ALBERT, FinBERT, and ELECTRA to determine their suitability for the text data analysis in ESG domain.

3.2.1 Bidirectional Encoder Representations from Transformers (BERT)

The BERT model is a state-of-the-art transformer-based architecture for natural language processing tasks. BERT introduces a bidirectional context-awareness by leveraging the bidirectional Transformer encoder, enabling the model to consider both preceding and succeeding words when encoding a given word. Pre-trained on extensive text corpora, BERT captures contextualized representations, providing a robust foundation for downstream tasks. During fine-tuning, task-specific layers are added for adaptation to specific applications [19]. This approach has established remarkable success in diverse natural language understanding benchmarks [20] (**Fig. 1**).
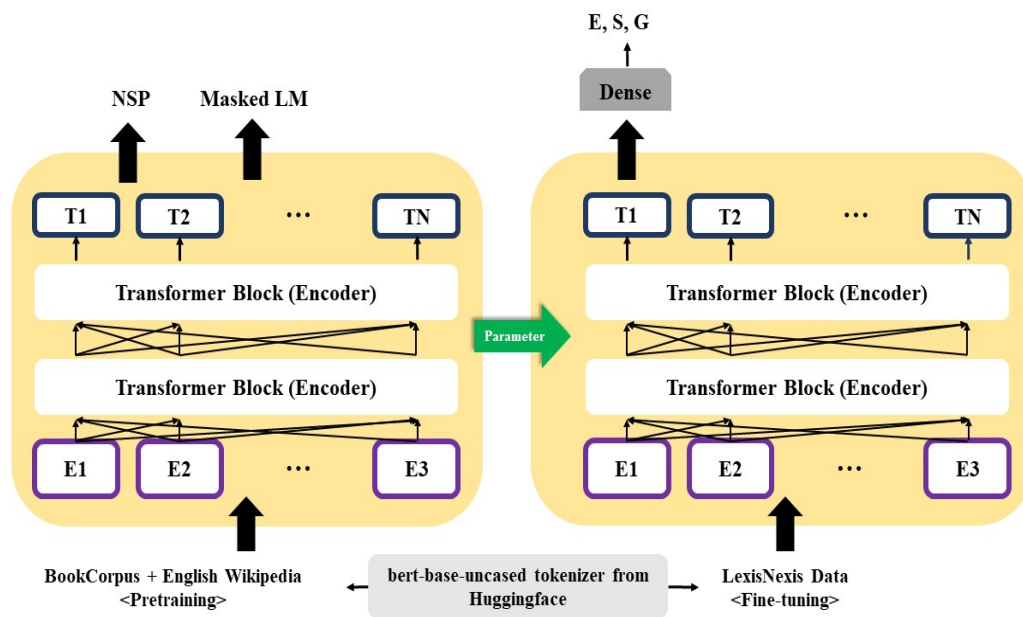

Fig. 1. BERT model architecture (Source: Adapted from [19, 22])

3.2.2 A Robustly Optimized BERT Pretraining Approach (RoBERTa)

RoBERTa is an extension of the BERT architecture designed for improved pre-training and fine-tuning. RoBERTa enhances BERT by excluding the next sentence prediction objective, training with larger mini-batches and learning rates, and increasing the amount of training data. By leveraging dynamic masking during pre-training, RoBERTa captures bidirectional contextual information, enhancing its ability to understand sentence semantics and relationships [21]. It has shown state-of-the-art performance across varied NLP benchmarks [22] (**Fig. 2**).
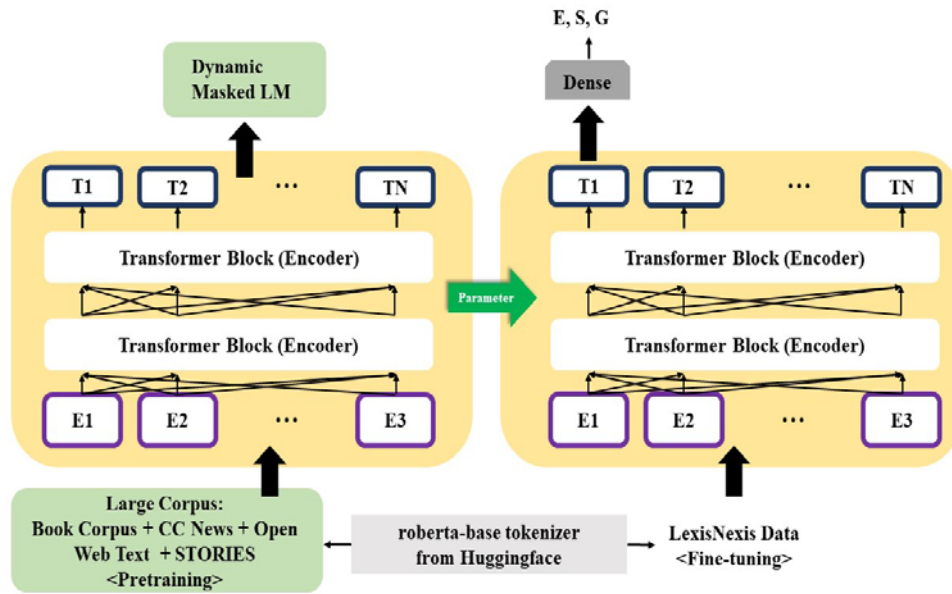
Fig. 2. RoBERTa model architecture (Source: Adapted from [21, 22])

3.2.3 A Lite BERT (ALBERT)

The ALBERT model is an optimized variant of the BERT architecture designed to achieve similar or even superior performance with significantly fewer parameters. ALBERT employs a parameter-sharing strategy that reduces model size while retaining the expressive power of the original BERT model. This is achieved through cross-layer parameter sharing and factorized embedding parameterization. By efficiently utilizing parameters, ALBERT scales well to larger datasets and demonstrates competitive performance on various natural language processing tasks [23] (**Fig. 3**).
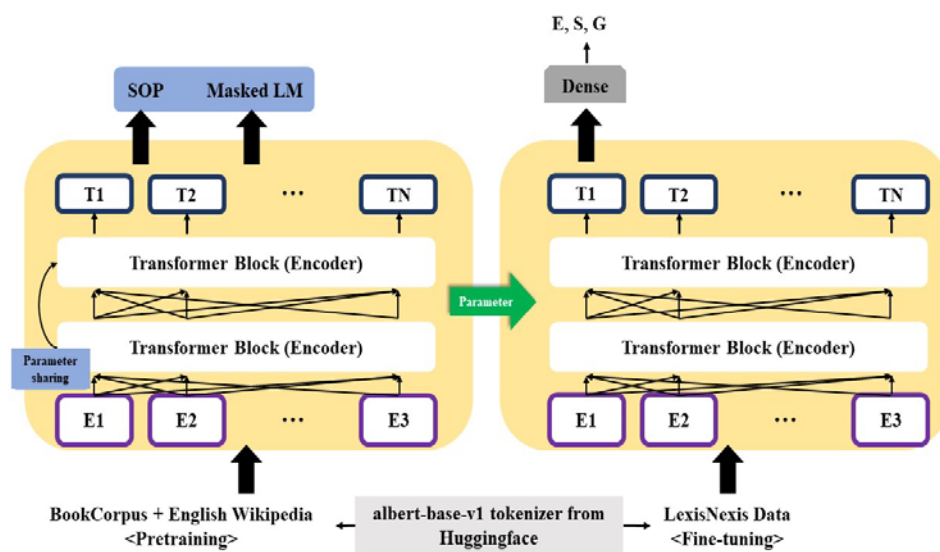

Fig. 3. ALBERT model architecture (Source: Adapted from [22, 23])

3.2.4 Financial BERT (FinBERT)

FinBERT is trained on financial text data to capture subtle language patterns specific to the finance domain. Leveraging transformer architecture, FinBERT excels in contextual language understanding, sentiment identification, and financial insights extraction from textual data. Pre-training on financial corpora enhances domain-specific knowledge and improves performance in sentiment analysis tasks within financial contexts [24] (**Fig. 4**).
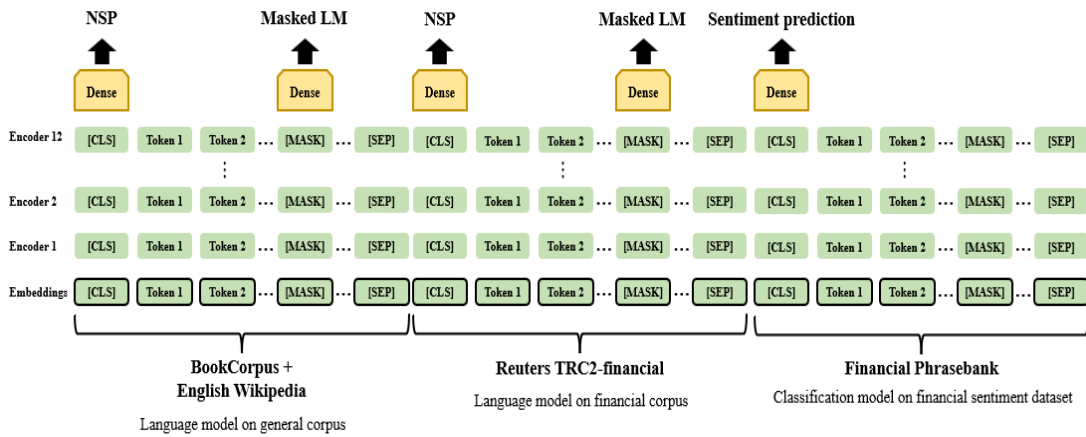

Fig. 4. FinBERT model architecture (Source: Adapted from [22, 24])

3.2.5 Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA)

ELECTRA redefines pre-training by implementing a new method for identifying replacement tokens. Unlike conventional token masking and prediction, ELECTRA substitutes a portion of input tokens with incorrect tokens, prompting the model to distinguish between actual tokens and replaced tokens. This approach promotes a deeper understanding of contextual clues and improves the model's understanding of intricate linguistic details. Operating mechanism of ELECTRA has proven effective across diverse NLP tasks, such as sentiment analysis and text classification [25] (**Fig. 5**).
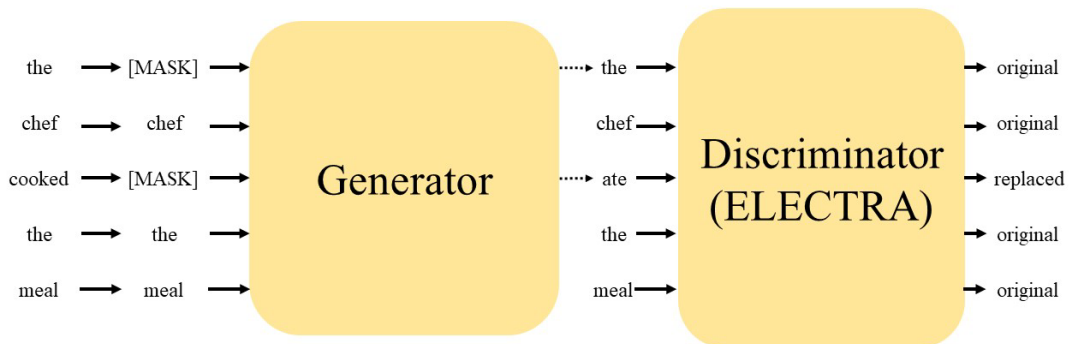

Fig. 5. ELECTRA model architecture (Source: Adapted from [22, 25])

3.3 Model Evaluation and Scaled Probability Analysis

After the label classification, the optimized model was employed to estimate the probability of each ESG label. This probability estimation serves as a crucial metric, offering a quantitative measure of the model's certainty in attributing specific ESG classifications.

By assigning probabilities to each label, our methodology not only provides categorical predictions but also quantifies the level of confidence associated with each classification. This approach improves the transparency of the decision-making process, enabling stakeholders to gauge the reliability of the ESG assessments and make decisions based on the generated probabilities. The entire process is illustrated in **Fig. 6**.
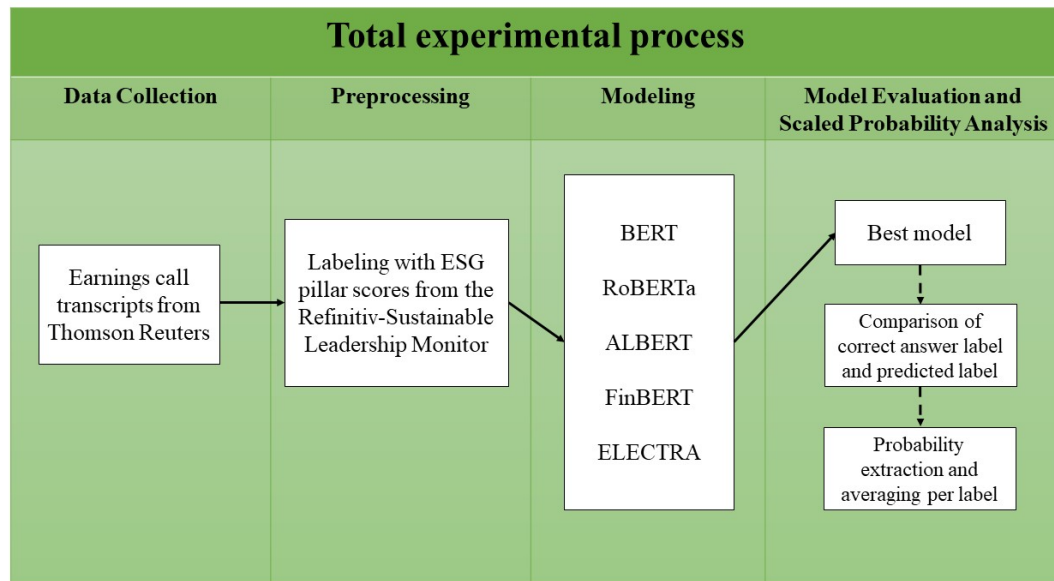

Fig. 6. Overall experimental diagram

4. Results

The authors compared the performance of BERT-based methodologies that were utilized to classify ESG in the result section. The authors performed experiments varying the batch sizes. The results highlighted that the BERT achieved the highest accuracy at 78.62% when employing a batch size of 16 (**Table 1**).

Table 1. Performance comparison of BERT-based models for ESG label classification based on different batch size

Model	Accuracy (%)
BERT (Batch size: 8)	74.58
RoBERTa (Batch size: 8)	75.86
ALBERT (Batch size: 8)	76.45
FinBERT (Batch size: 8)	53.59

ELECTRA (Batch size: 8)	53.21
BERT (Batch size: 16)	78.62
RoBERTa (Batch size: 16)	75.37
ALBERT (Batch size: 16)	76.75
FinBERT (Batch size: 16)	69.23
ELECTRA (Batch size: 16)	68.91
BERT (Batch size: 32)	76.85
RoBERTa (Batch size: 32)	77.54
ALBERT (Batch size: 32)	75.37
FinBERT (Batch size: 32)	68.76
ELECTRA (Batch size: 32)	68.39

The subsequent additional experiment involved utilizing the BERT model with a batch size of 16, which had been proven as the most efficient model. The authors extracted instances where actual values matched predicted values and computed the average predicted probability for each label. This additional experiment was conducted to gain insights into the robustness of the identified best model in this study. As a result, accuracy of 81.07% for label E, 87.63% for label S, and 88.40% for label G were observed (**Fig. 7**).
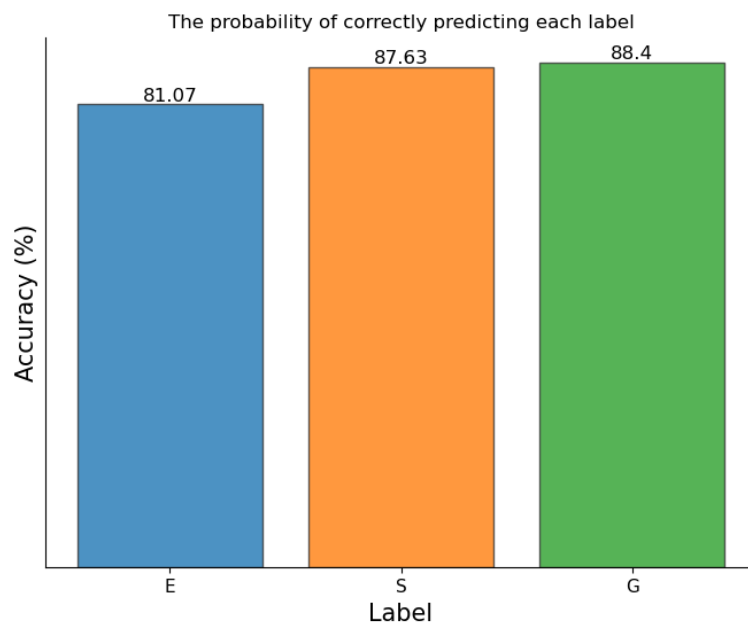

Fig. 7. The probability of correctly predicting each label

5. Conclusions

This research discusses the importance of ESG ratings in assessing a company's long-term investment potential. However, the inconsistency among companies and standards in managing ESG reporting have been identified as primary obstacles in incorporating ESG ratings into investment procedures. To address these limitations, the authors collected Earnings Call Transcript from Thomson Reuters and used advanced NLP approaches for retrieving relevant information from substantial text data to complement ESG assessment standards. We employed a language model based on transformers for accurate categorization of ESG, demonstrating its effectiveness.

Furthermore, by extracting the probability predictions for labels from the best-performing model and averaging them based on Earnings Call Transcript data, it was confirmed that companies primarily focus on the governance (i.e., G) element. In other words, it can be inferred that companies are emphasizing or making efforts to strengthen their governance frameworks. These results derived from Earnings Call Transcript data underscore the significance of corporate financial performance and disclosures as crucial sources of ESG-related information.

The findings of this study provide valuable insights into where companies should concentrate their efforts when considering strategies to enhance their ESG. Additionally, the Earnings Call Transcript is updated regularly. Therefore, evaluating a company's ESG rating using this data emphasizes the importance for companies to continually enhance their governance factors. In summary, the implications of this study suggest a strategic priority in improving governance practices as enterprises prepare to address the challenges and considerations linked to ESG. Moreover, by utilizing Earnings Call Transcript data, companies can reinforce their governance efforts and further aligning strategies with ESG principles.

However, there are some limitations that need to be addressed in the future. This paper solely utilized BERT-based models. Future research could explore other state-of-the-art language models beyond BERT and consider employing ensemble techniques to enhance performance. Additionally, while this study solely utilized Earnings Call Transcript data, future research could collect data from various sources to establish a more robust framework.

Acknowledgement

References

[1] A. G. Hoepner, A. A. Majoch and X. Y. Zhou, "Does an asset owner's institutional setting influence its decision to sign the principles for responsible investment?," *Journal of Business Ethics*, vol. 168, no. 2, pp. 389–414, 2021. Article (CrossRef Link)

[2] S. Kotsantonis and G. Serafeim, "Four things no one will tell you about ESG data," *Journal of Applied Corporate Finance*, vol. 31, no. 2, pp. 50-58, 2019. Article (CrossRef Link)

[3] A. Amel-Zadeh and G. Serafeim, "Why and how investors use ESG information: Evidence from a global survey," *Financial Analysts Journal*, vol. 74, no. 3, pp. 87-103, 2018.
Article (CrossRef Link)

[4]   E. Escrig-Olmedo, M. Á. Fernández-Izquierdo, I. Ferrero-Ferrero, J. M. Rivera-Lirio and M. J. Muñoz-Torres, "Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles," *Sustainability*, vol. 11, no. 3, pp. 915, 2019. Article (CrossRef Link)

[5]   H. Lee, S. H. Lee, D. Nan and J. H. Kim, "Predicting user satisfaction of mobile healthcare services using machine learning: confronting the COVID-19 pandemic," *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 6, pp. 1-17, 2022. Article (CrossRef Link)

[6]   H. S. Jung, S. H. Lee, H. Lee and J. H. Kim, "Predicting Bitcoin Trends Through Machine Learning Using Sentiment Analysis with Technical Indicators," *Computer Systems Science & Engineering*, vol. 46, no. 2. pp. 2231-2246, 2023. Article (CrossRef Link)

[7]   V. Díaz, D. Ibrushi and J. Zhao, "Reconsidering systematic factors during the COVID-19 pandemic–The rising importance of ESG," *Finance Research Letters*, vol. 38, pp. 101870, 2021. Article (CrossRef Link)

[8]   W. Henisz, T. Koller and R. Nuttall, "Five ways that ESG creates value," *McKinsey*, 2019. Article (CrossRef Link)

[9]   A. Edmans, "Does the stock market fully value intangibles? Employee satisfaction and equity prices," *Journal of Financial economics*, vol. 101, no. 3, pp. 621-640, 2011. Article (CrossRef Link)

[10] A. Edmans, "The link between job satisfaction and firm value, with implications for corporate social responsibility," *Academy of Management Perspectives*, vol. 26, no. 4, pp. 1-19, 2012. Article (CrossRef Link)

[11] N. Attig, S. El Ghoul, O. Guedhami and J. Suh, "Corporate social responsibility and credit ratings," *Journal of business ethics*, vol. 117, pp. 679-694, 2013. Article (CrossRef Link)

[12] A. Sokolov, J. Mostovoy, J. Ding and L. Seco, "Building machine learning systems for automated ESG scoring," *The Journal of Impact and ESG Investing*, vol. 1, no. 3, pp. 39-50, 2021. Article (CrossRef Link)

[13] H. Lee, S. H. Lee, K. R. Lee and J. H. Kim, "Esg discourse analysis through bertopic: comparing news articles and academic papers," *Computers, Materials & Continua*, vol. 75, no. 3, pp. 6023-6037, 2023. Article (CrossRef Link)

[14] N. Raman, G. Bang and A. Nourbakhsh, "Mapping ESG trends by distant supervision of neural language models," *Machine Learning and Knowledge Extraction*, vol. 2, no. 4, pp. 453-468, 2020. Article (CrossRef Link)

[15] H. Lee, S. H. Lee, H. Park, J. H. Kim and H. S. Jung, "ESG2PreEM: Automated ESG grade assessment framework using pre-trained ensemble models," *Heliyon*, vol. 10, no. 4, pp. e26404, 2024. Article (CrossRef Link)

[16] F. Berg, K. Fabisik and Z. Sautner, "Is history repeating itself? The (un) predictable past of ESG ratings," *The (Un) Predictable Past of ESG Ratings (August 24, 2021), European Corporate Governance Institute–Finance Working Paper*, vol. 708, 2020. Article (CrossRef Link)

[17] Refinitiv, [Online]. Available: https://www.refinitiv.com/en/products/sustainability-reporting-on-leadership

[18] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," 2018. Article (CrossRef Link)

[19] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018. Article (CrossRef Link)

[20] H. S. Jung, H. Lee and J. H. Kim, "Unveiling Cryptocurrency Conversations: Insights From Data Mining and Unsupervised Learning Across Multiple Platforms," *IEEE Access*, vol. 11, pp. 130573-130583, 2023. Article (CrossRef Link)

[21] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, ... and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019. Article (CrossRef Link)

[22] H. Lee, H. S. Jung, S. H. Lee and J. H. Kim, "Robust Sentiment Classification of Metaverse Services Using a Pre-trained Language Model with Soft Voting," *KSII Transactions on Internet & Information Systems*, vol. 17, no. 9, pp. 2334-2347, 2023. Article (CrossRef Link)

[23] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," 2019. Article (CrossRef Link)

[24] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," 2019.
       Article (CrossRef Link)
[25] K. Clark, M. T. Luong, Q. V. Le and C. D. Manning, "Electra: Pre-training text encoders as
       discriminators rather than generators," 2020. Article (CrossRef Link)

Haein Lee is Ph.D Candidate at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. Her research focuses on Natural Language Processing, Deep Learning, and Machine Learning.

Hae Sun Jung is Ph.D researcher (i.e., Postdoctoral Researcher) and obtained his Ph.D from the Department of Applied Artificial Intelligence, Sungkyunkwan University. His research focuses on Natural Language Processing, Computer Vision, Deep Learning, and Machine Learning.

Heungju Park is an assistant professor at SKK Business School, Sungkyunkwan University. His research focuses on Empirical Asset Pricing and Corporate Finance, Banking, Chinese Financial Market.

Jang Hyun Kim is a professor at Department of Human-Artificial Intelligence Interaction/Department of Interaction Science/Department of Applied Artificial Intelligence, Sungkyunkwan University. His research focuses on social/semantic data analysis, social media, and future media. He has authored over 50 papers in major journals such as Information Processing & Management, Telematics & Informatics, Cities, Government Information Quarterly, Technological Forecasting and Social Change, and Journal of Computer-Mediated Communication. Jang Hyun Kim is the corresponding author and can be contacted via alohakim@skku.edu.