Prediction of Electric Load Neural Network Prediction Model for Big Data

Chen Siyu, Yuan Ruirui, Lin Xinyuan*

School of Commercial, Wenzhou-Kean University, Wenzhou, China

*Corresponding author: vj162774@163.com

Keywords: ESG, TFIDF, LDA

Abstract: This research aims to investigate the relationship between ESG practices and corporate performance. When exploring relevant academic literature, the specific positive relation between ESG practices and corporate performance have been found. Actually, the moderating role of ESG practices will come into play in specific time and environment. To improve the accuracy of research, this study is analyzed using term frequency-inverse document frequency model and Latent Dirichlet Allocation model. As a result, this study proposes the positive moderating role of ESG practices over the aforementioned relationship, underlining the multiple dimensions effect of ESG practices. Last but not least, this paper gives some operational suggestions for business operators, providing a clear idea of business model innovation.

1. Introduction

Currently, ESG has become a hot topic worldwide. More companies than ever are offering ESG and sustainability reporting using commonly accepted frameworks in their operational strategy, putting ESG theory into operational reality. ESG practices is not only a kind of corporate strategy, it also shows a company's social responsibility and to determine whether a company could gain long-term development. The original intention of company use ESG strategy is to pursue profit in order to improve their corporate performance. Since the acronym "ESG" (environmental, social, and governance) was coined in 2005, until recently, its fortunes and use scale was steadily growing. According to research, for 2020, 80% of publicly traded companies, 79% of venture and private equity-backed companies, and 67% of privately owned companies had ESG initiatives in place (Niemoller, 2023). These data do not only show the increasing passion of companies for ESG practices but also underline that ESG practices have a potentially positive effect on corporate performance so that more and more companies use ESG strategy. For example, a 2022 survey of 1,062 U.S. residents by Green Print, a sustainability tools provider that's now owned by PDI Technologies, found that 66% of the respondents would be willing to spend extra money to buy environmentally friendly products (GreenPrint, 2022). Similarly, 70% of 400 IT professionals surveyed in 2022 by TechTarget's Enterprise Strategy Group division said they think their company would pay more than a 5% price premium for IT products from vendors that have strong ESG practices (TechTarget, 2022). As mentioned above, it is important to integrate ESG in investment decisions when looking for opportunities to create value over time (Admin, 2021), thus more companies adopt ESG practices and formulate ESG reporting. [13]

In collaboration with Rockefeller Asset Management and Casey Clark, CFA (MBA '17), the NYU Stern Center for Sustainable Business examine the relationship between ESG and financial performance in more than 1,000 research papers from 2015 – 2020, these research found that ESG investing provides downside protection, especially during a social or economic crisis and Sustainability initiatives at corporations appear to drive better financial performance due to mediating factors such as improved risk management and more innovation. (Rockefeller Asset Management and Casey Clark, 2020)To summarize, this study plainly indicates that ESG practices could better companies' financial performance through many positive factors. ESG practices have many benefits, measurable ESG practices could increase employee and customer's loyalty and safeguard brand integrity. (Yasar, 2023) Moreover, this practice could attract investors and leaders, making them be

  DOI: 10.25236/icmmct.2024.010

more confident and prospective to build cooperative relationship with the company which adopting ESG practices, in other words, stakeholders are becoming highly interested in organizations that invest in ESG and use ESG disclosures to shed light on their sustainability efforts. The most important benefit is that proper ESG practices could make company gain sustainable operations, for example, because businesses that properly integrate ESG principles into their core operations are better able to identify cost-saving opportunities and enjoy lower energy consumption, reduced resource waste and an overall reduction in operational costs. (Yasar, 2023) All in all, the way how ESG practices positively affect corporate performance are various, but the positive correlation between effective ESG practices and corporate performance is definite.

The rest of this paper is structured as follows. Section 2 provides a comprehensive review of related literature, focusing on the in-depth concept of ESG and benefits of ESG practices on corporate performance. Section 3 introduces some factors push ESG practices be favorable to corporate performance. Section 4 provides details about proper character that effective ESG practices need to possess. Section 5 is the research design, including data description, variable selection, and econometric model establishment. Section 6 presents the analysis of test results. The final section is the conclusion and enlightenment. Researchers in a recent study took advantage of a distinct environmental context and a new dataset to demonstrate a positive correlation between Environmental, Social, and Governance (ESG) performance, associated with the short-term cumulative returns of CSI300 stocks during the COVID-19 crisis. There is an article pointed out that ESG factors matter in terms of corporate financial performance and risk. However, the effects vary across different dimensions such as ESG categories, strength and weakness, and firm size. These findings can potentially help investors understand the underpinning rationale for recent trends in ESG-linked or ESG-integrated investment strategies. Furthermore, a recent study indicated that a discounted cash-flow (DCF) model framework has great benefits on social developmen, such as it can be used to break down the influence of a corporation’s ESG profile on equity valuations, including cash flows, risk, and cost of capital. Whereas, some statistics in another paper showed that the ESG performance still has some challenges, such as different companies report ESG data in a different way when trying to analyse the effects of ESG investment and performance. Recent evidence suggests that to solve the global sustainable development problem, ESG should be integrated into different institutional backgrounds and industry characteristics, instead of only maintaining the mainstream status of the KLD database and Thomson Reuters ASSET 4 database. Since there is still a lack of understanding about the different motivations of their diverse users, further studies on the creation of meaning and language at the ESG data user level will be required in order to ensure that conversations take place with a shared understanding of what we mean when we use different buzz words, such as ESG integration. The increased social consciousness of investors and the favorable regulatory landscape have made investing strategies grounded in corporate social responsibility (CSR) metrics a hot issue for scholars and practitioners alike, however, the bulk of academic and industry studies focus on standards, investment, uncertainty, and corporate management.[22]

The research sector concentrates on the efficacy, precision, and consistency of the ESG assessment standards. According to Dorfleitner et al.[6], there is a clear gap in the convergence of concepts related to ESG measurement. Neither the distribution nor the risk of the various ratings coincide. Thus, it is recommended that all CSP stakeholders critically assess the specific ESG score model's validity. In a study conducted by Auer and Schuhmacher [2], an ESG-based investment strategy's performance is significantly influenced by the industry, regional, and ESG criteria that are used. Choosing socially desirable or less desirable stocks is not a suitable way for ESG-based stock selection to satisfy the demands of profit-seeking investors.

Even at unique times like the COVID-19 outbreak, most studies on ESG investment use stock data analysis to provide recommendations for investors. The study is concentrated on the ESG performance during the global financial crisis. The findings show a positive correlation between the short-term cumulative returns and ESG performance during the COVID-19 crisis. Investors might utilize ESG performance as a risk-reduction or future stock performance prediction during lean times [4]. Research on marketing and accounting in Germany reveals that ESGP has a favorable effect on

ROA but has no effect on Tobin's Q. Governance performance significantly impacts FINP compared to environmental and social performance within ESGP components (Velte, 2017)[25]. Analyzing ESG ratings and stock performance in the COVID-19 setting, Engelhardt et al. [7] find that social score is a significant factor. In low-trust nations with laxer security laws and disclosure requirements, ESG is shown to be valuable.

Uncertainty of ESG can also impact the performance of equities, which could influence investors' choices. ESG uncertainty could affect investors' demand and the risk-return trade-off and reduce economic welfare for ESG-sensitive agents. Our findings echo the growing concerns regarding the need for more consistency in ESG information disclosure and ratings provided by different rating agencies. The impact of company ESG profile uncertainty, a significant obstacle to sustainable investing, on asset pricing and portfolio implications is studied by Avramov et al. [3]. In the face of ESG uncertainty, equilibrium experiences a higher market premium and decreased stock demand.

ESG is crucial for business development and management as well. Corporate profitability is positively impacted by ESG, particularly in larger businesses. Among the ESG categories, corporate governance is essential for poorly managed enterprises. However, the environmental score hurts credit ratings; social impact is the most substantial positive contribution from ESG components (Kim & Li, 2021) [16]. Giese and Lee (2019) [9] confidently establish ESG's positive impact on risk, especially in mitigating tail risks. Limited evidence suggests ESG momentum correlates with portfolio performance, requiring a more extended time series for verifying the ESG risk premium.

Little of the currently available literature integrates ESG with NLP. In the research, Pasch and Ehnes (2022) [19] propose to use text-based AI techniques for ESG. Transformer-based model accuracy is considerably improved by pre-screening texts for ESG content. In order to supplement the work of domain experts, Pasch and Ehnes (2022) suggest using the most recent NLP techniques to automatically extract knowledge from text. The study fills in the transparency gaps by presenting new NLP models for accurate ESG measurement. Through the evaluation of corporate disclosures, these models promote sustainable development. Stakeholders can assess organizations' ESG communication more thoroughly when datasets and models are made publicly available (Schimanski et al., 2023) [23]. The capacity of natural language processing (NLP) to automatically interpret and analyze human discourse is causing it to gain popularity in management research (Kang et al., 2020)[14]. Natural language understanding is merely the initial step towards word- and concept-level methods to NLP. Computational paradigms driven by language and biology that facilitate story comprehension and, consequently, 'sensemaking' hold the key to the future of natural language processing (Cambria & White, 2014)[5]. Neural network-based models are the norm in research and application systems as natural language processing enters a new era (Zhou et al., 2020)[26]. On a range of NLP tasks, neural models perform on par with or better than other state-of-the-art systems (Li et al., 2015)[17]. New research and application scenarios will also arise from combining NLP with multi-modal tasks including speech recognition, image/video captioning, and quality assurance (Zhou et al., 2020)[26].

2. Models

2.1 TFIDF

TF-IDF stands for term frequency-inverse document frequency and it is a measure, used in the fields of information retrieval (IR) and machine learning, it can quantify the importance or relevance of string representations (words, phrases, lemmas, etc.) in a document amongst a collection of documents (also known as a corpus). (Simha, 2021) The simplified formula for TF-IDF is:$TF - IDF(term, document) = TF(term, document) \times IDF(term)$ (Robertson, 1970), where TF(term, document) is the frequency of the term in the document, and IDF(term) is the inverse document frequency of the term, which is calculated as follows: $IDF(term) = log\left(\frac{N}{DF(term)}\right)$ $TF(term) = \frac{n}{M}$ (Robertson, 1970), where N is the total number of documents in the corpus, DF(term) is the number of documents that contain the term, n is the number of times the term

appears in the document, and M is the total number of terms in the document. According to the formula, the TF-IDF score for a term in a document is high if the term appears frequently in the document and is rare across all other documents in the corpus. This formula contains two parts, the TF part means term frequency, showing how often a term occurs in a document, the other part: IDF means inverse document frequency of a term, reflecting the proportion of documents in the corpus that contain the term. (Karabiber, 2022)

2.2 LDA

LDA is the primary way to model documents in the field of machine learning, and a topic model used to discover hidden topic information from a large number of documents, which is essentially a text clustering model. The goal of LDA is to find a short description of the members of a collection so that large collections can be handled efficiently while preserving basic statistical relationships. It is an unsupervised machine learning technique that uses statistical methods to infer underlying topic structures from textual data. LDA needs to collect and preprocess the data text, including word segmentation, stop removal, stem extraction, etc., and then determines the number of topics and other model hyperparameters. Then LDA algorithm is used to train the text data, and then analyzes the topics and their distribution in the document to get a deeper understanding. Finally, scores such as confusion and topic consistency were used to evaluate the model. The advantage of this algorithm is that compared with the traditional PLSA algorithm, it applies the Bayesian method and adds conjugate prior distribution to the topic distribution and word distribution, so it has more statistical advantages in the case of a small amount of data or small samples.

$$\text{Perplexity (D)} = \exp\frac{\sum_{d=1}^{M} \log P\,(W_d)}{\sum_{d=1}^{M} N_d}$$

3. Results

The terms "Environment," "Social," and "Corporate governance" (ESG) describe the three main components used to assess an investment's long-term viability and social impact. These factors aid in more accurately predicting a company's future financial success (risk and return). In this analysis, text from an online ESG report in PDF format is extracted, NLP is applied to the data, significant ESG initiatives are summarized using WordClouds and TDIDFs, and topics are found by constructing a Latent Dirichlet Allocation (LDA) model. Just one ESG report is being used in order to keep this exercise as straightforward as feasible. In particular, the 2019 ESG report from Citibank.

In essence, a word cloud is a visual representation of text. A word cloud is an electronic picture displaying the words found in a specific electronic text or text series. Here, words are sized differently according to how frequently they occur (KABIR et al., 2020) [15]. It is clear from Figure 1 that the terms "employee," "client," and "include" have the highest volume in the report. By briefly scanning over the word cloud at the start of the study, readers can immediately gain a general notion of the term space of a document. The most common terms and subjects are communicated right away, providing a solid foundation for the analysis (Heimerl et al., 2014) [10]. The company's ESG reporting and practices place a high priority on these three elements.

In today's information retrieval systems, one of the most widely used term weighting techniques is "term frequency–inverse document frequency" (tf–idf) [1]. TF counts the number of times a term appears in a document (Gautam & Kumar, 2013)[8]. The inverse document frequency gives words that occur frequently less weight, while words that occur infrequently are given more weight (Qaiser & Ali, 2018)[20]. The 30 most common uni- and bi-grams from the 2019 Citibank ESG report were sorted using TF-IDF, and it is clear that from Figure 2, the three terms with the highest TF-IDF number are "risk," "employee," and "human right." High TF-IDF words indicate a close association between the term and the document in which it appears, indicating that the user may find the document interesting if the word appears in a query [21]. The reader can roughly deduce the report's topic from these three words. The word "risk" could refer to an emphasis on locating, evaluating, and reducing hazards related to social, political, or environmental variables. An emphasis on

employee-related concerns, such as working conditions, employee well-being, diversity and inclusion, or human capital management, may be indicated by the term "employee." The "human right" concept designates a commitment to addressing and disclosing human rights issues during the business's supply chain and operations. TF-IDF is a numerical statistic that illustrates the significance of keywords to particular documents or offers those keywords, allowing for the identification or classification of certain documents (Qaiser & Ali, 2018). To summarize, stakeholders can gain insight into the company's priorities, concerns, and areas of emphasis related to risk management, employee relations, and human rights by looking at the terms with the highest TF-IDF scores in the ESG report.

Figure 1 Word cloud for ESG report to compute corpus wide term frequencies.

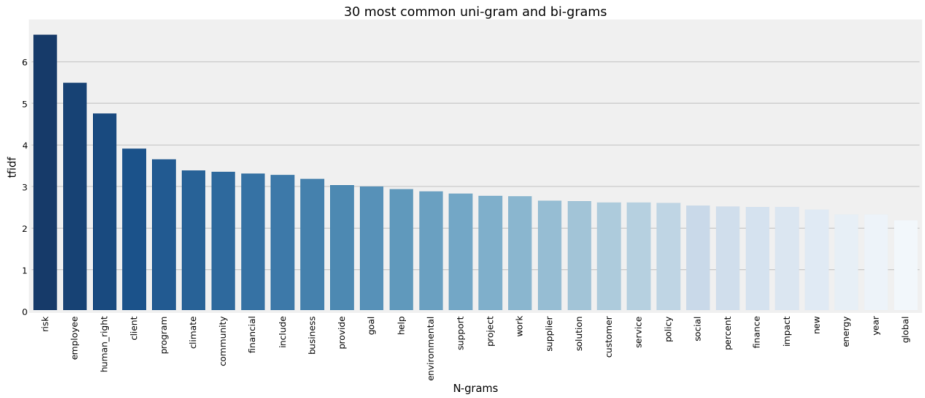

Figure 2 30 most common uni-gram and bi-gram created by TF-IDF.

Latent Dirichlet Allocation assumes that papers are produced using a specific probabilistic model to learn the correlations between words, themes, and documents (Stevens et al., 2012)[24]. Figure 3 shows that LDA is divided into two components: perplexity and coherence score. The perplexity is ranged from -7.25 to -7.5. The complexity of voice recognition tasks is measured by perplexity (Jelinek et al., 1977)[12]. Lower perplexity values indicate better performance. A topic model may better predict words in unseen texts based on the topics it has learned if its perplexity is lower (Huang et al., 2017).[11] Since the value of perplexity in this report is less, the model performs better when making predictions on data that hasn't been seen. The coherence score is ranged from 0.34 to 0.43. The degree of semantic similarity between high-scoring terms within a topic is measured to determine the topic coherence score (Stevens et al., 2012). Higher coherence scores indicate greater defined and easier-to-understand concepts. High coherence scores indicate that a topic's assigned words are semantically related and constitute coherent, meaningful themes (Mimno et al., 2011). [18]The high coherence score in this report suggests that the subjects produced by the model are more coherent and interpretable.

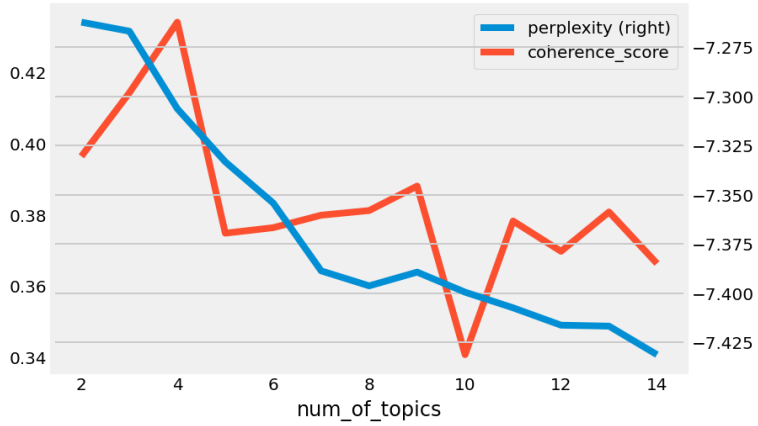

Figure 3 Perplexity and coherence score in Evaluaton Metrics produced by LDA.

From Figure 4, LDA visualizes 4 topics and selects the top 10 keywords for each topic. The four topics are support community, value employees, code of conduct and ethical investments. These terms are probably used in an ESG report to emphasize the company's efforts and dedication to moral, environmentally friendly, and socially conscious business practices. To support the community and help create a more sustainable and inclusive future, the report may go into detail about particular programs, laws, and results associated with each keyword.

Figure 4 Distinguishing the different ESG topics.

4. Conclusions

This study examines the connection between NLP and ESG by utilizing word clouds, TF-IDF, and LDA, three different types of technology, to analyze Citibank's 2019 ESG report. Quickly identify essential terms with a quick scan using word cloud research. TF-IDF is the statistical frequency. LDA is more inclined to summarize keywords. These three techniques are all part of natural language processing (NLP), and it is clear that NLP retrieves text content accurately and quickly. Natural language processing (NLP) is critical in advancing environmental, social, and governance (ESG) initiatives by enhancing data analysis and communication. News articles, social media, reports, and other unstructured data sources can all be used to extract and analyze unstructured data. This process allows NLP to provide significant new insights into the attitudes and trends underlying ESG. Businesses may assess their sustainability performance more efficiently by enabling the automation

of ESG data processing. Additionally, NLP supports transparency and standardizes ESG reporting by assisting stakeholders in understanding complex ESG information through more straightforward language, facilitating informed decision-making and responsibility in pursuing sustainable practices. Developing sophisticated predictive analytics models, improving sentiment analysis, and raising the quality of ESG data are anticipated to be the main areas of study for NLP and ESG research. Incorporating AI technologies will enhance corporate responsibility and sustainable decision-making even more.

References

[1] Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. Information Processing & Management, 39(1), 45–65. https://doi.org/10.1016/s0306-4573(02)00021-3

[2] Auer, B. R., & Schuhmacher, F. (2016). Do socially (ir) responsible investments pay? New evidence from international ESG data. The Quarterly Review of Economics and Finance, 59, 51–62. https://doi.org/10.1016/j.qref.2015.07.002

[3] Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2021). Sustainable Investing with ESG Rating Uncertainty. Journal of Financial Economics, 145(2), 642–664. https://doi.org/10.1016/j.jfineco.2021.09.009

[4] Broadstock, D. C., Chan, K., Cheng, L. T. W., & Wang, X. (2020). The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. Finance Research Letters, 38(38), 101716.

[5] Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. IEEE Computational Intelligence Magazine, 9(2), 48–57. https://doi.org/10.1109/mci.2014.2307227

[6] Dorfleitner, G., Halbritter, G., & Nguyen, M. (2014). Measuring the Level and Risk of Corporate Responsibility - An Empirical Comparison of Different ESG Rating Approaches. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2536265

[7] Engelhardt, N., Ekkenga, J., & Posch, P. (2021). ESG Ratings and Stock Performance during the COVID-19 Crisis. Sustainability, 13(13), 7133. https://doi.org/10.3390/su13137133

[8] Gautam, J., & Kumar, E. (2013). An integrated and improved approach to terms weighting in text classification. International Journal of Computer Science Issues (IJCSI), 10(1), 310.

[9] Giese, G., & Lee, L. E. (2019). Weighing the evidence: ESG and equity returns. MSCI Research Insight.

[10] Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January). Word cloud explorer: Text analytics based on word clouds. In 2014 47th Hawaii international conference on system sciences (pp. 1833-1842). IEEE.

[11] Huang, L., Ma, J., & Chen, C. (2017). Topic Detection from Microblogs Using T-LDA and Perplexity. 2017 24th Asia-Pacific Software Engineering Conference Workshops (APSECW). https://doi.org/10.1109/apsecw.2017.11

[12] Jelinek, F., Mercer, R. L., Bahl, L. R., & Baker, J. K. (1977). Perplexity—a measure of the difficulty of speech recognition tasks. The Journal of the Acoustical Society of America, 62(S1), S63–S63. https://doi.org/10.1121/1.2016299

[13] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2018). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

[14] KABIR, A. I., AHMED, K., & KARIM, R. (2020). Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. Informatica Economica, 24(4/2020),

55–71. https://doi.org/10.24818/issn14531305/24.4.2020.05

[15] Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics, 7(2), 139–172. https://doi.org/10.1080/23270012.2020.1756939

[16] Kim, S., & Li, Z. (Frank). (2021). Understanding the Impact of ESG Practices in Corporate Finance. Sustainability, 13(7), 3746.

[17] Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2015). Visualizing and understanding neural models in NLP. arXiv preprint arXiv:1506.01066.

[18] Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. In Proceedings of the 2011 conference on empirical methods in natural language processing (pp. 262-272).

[19] Pasch, S., & Ehnes, D. (2022). NLP for Responsible Finance: Fine-Tuning Transformer-Based Models for ESG. 2022 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata55660.2022.10020755

[20] Qaiser, S., & Ali, R. (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

[21] Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning (Vol. 242, No. 1, pp. 29-48).

[22] Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. Journal of Banking & Finance, 32(9), 1723–1742. https://doi.org/10.1016/j.jbankfin.2007.12.039

[23] Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2023). Bridging the Gap in ESG Measurement: Using NLP to Quantify Environmental, Social, and Governance Communication. Social Science Research Network. https://doi.org/ 10.2139/ssrn. 4622514

[24] Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012, July). Exploring topic coherence over many models and many topics. In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning (pp. 952-961).

[25] Velte, P. (2017). Does ESG Performance Have an Impact on Financial performance? Evidence from Germany. Journal of Global Responsibility, 8(2), 169–178. https://doi.org/10. 1108/jgr-11-2016-0029

[26] Zhou, M., Duan, N., Liu, S., & Shum, H.-Y. (2020). Progress in Neural NLP: Modeling, Learning, and Reasoning. Engineering, 6(3), 275–290. https://doi.org/10.1016/j.eng.2019.12.014