EVALUATING THE PERCEPTION OF ESG-RELATED PUBLICATIONS AVAILABLE ON PUBLIC DATABASES

Simone PERAZZOLI
Constela Digital
Videira 89564-494, Santa Catarina, Brazil
consteladigital@consteladigital.com

José Pedro DE SANTANA NETO
University of Geneva
Geneva 1227, Switzerland
1jpsneto@gmail.com

Abstract

The contemporary world scenario demands initiatives to address current societal challenges. Among them, the Environmental, Social, and Governance (ESG) criteria have received prominent attention. Traditionally, ESG is adopted to handle investments. Therefore, by assuming interdisciplinary and integrative approaches, ESG has the potential to bring insights, measurements, understandings, and evaluations on how companies and their economic processes affect our society, from a local to regional and global scenery. This study presents a comprehensive bibliometric analysis of 55,000 publications supported by Natural Language Processing (NLP) techniques on the aspects of ESG-related topics available on the Google Scholar database. It follows research tools and methodologies available in the literature, including the steps of data collection, preprocessing, modeling and analysis, and visualization. Results indicate that besides the ESG theme has been studied extensively, the analysis of perception scores for aggregated ESG categories indicates a meaningful decrease in the growth rate for the perception when normalized by the GDP, which may exhibit signs of unsustainable economic growth. Such analysis brings valuable information, serving as a support tool in decision-making processes involving ESG, thus, promoting a more sustainable society and world.

Keywords: ESG, Data analytics, Natural Language Processing, Perception Analysis, Sustainability.

1. Introduction

Within the contemporary demanding world scenario, the academy, industry, and community have demonstrated increasing interest in initiatives aiming to address societal challenges such as resources scarcity, food insecurity, climate change, gender discrimination, misinformation, unemployment, social exclusion, immigration, sustainability transitions among others (Dixson-Declève et al., 2022; Guillen-Royo, 2020; Tíjaro-Rojas et al., 2016).

The environment, social, and governance (ESG) criteria have received distinguished attention in this context. Traditionally, investors use ESG to handle investments, targeting opportunities with favorable risks (Diaye et al., 2022; Freiberg et al., 2021; Kotsantonis & Serafeim, 2019). In addition to it, ESG has the potential to bring insights, measurements, understandings, and evaluations on how companies and their economic processes affect society in several layers, from local to regional and global stances. Thus, it must include interdisciplinary and integrative approaches to reach this potential.

Therefore, innovative methodologies and tools can emerge to support decisions toward a more sustainable world, such as computational strategies applied to extract and process meaningful text information (Perazzoli, de Santana Neto, et al., 2022; Qin et al., 2021; Sun et al., 2017) to support data analysis and discussions (Mina & Barzola, 2019; Salazar-Reyna et al., 2020).

Accordingly, this study forecasts the perception of 55,000 ESG-related publications available on publicly databases aiming to offer an integrative view of the state-of-the-art in this topic as well as bringing valuable information, serving as a support tool in decision-making processes involving ESG, thus, promoting a more sustainable society and world.

2. Methodology

This work demonstrates the preliminary results of a massive analysis of data available in the international literature related to the ESG topic. The research considered the publications available in the Google Scholar database, following methodologies of bibliometric studies already established in the literature (Di Vaio et al., 2022; Perazzoli, de Santana Neto, et al., 2022; Ratinho et al., 2020). The search terms included the following keywords: *'ESG'*, '*climate finance*', '*corporate social responsibility*, '*CSR*', '*environmental governance*', '*socially responsible investments*', '*SRI*', *'corporate sustainability*', '*corporate social performance*', '*social and environmental accounting*', '*sustainable finance*' as well as their grammatical variations.

The following sources of information were considered: peer-reviewed journals, book or book chapters, thesis (specialization, master and doctorate degrees), conference

proceedings, case studies, preprints, and technical reports. The publication dates of returned queries ranged from 2005 [the year ESG term was coined by the United Nations Environment Programme Initiative (UNEP, 2005)] to 2022. The pipeline for data management included the following steps: collection, pre-processing, modeling and analysis, and visualization, as summarized in Fig.1 and, according to the methodology proposed by (Perazzoli, de Santana Neto, et al., 2022).

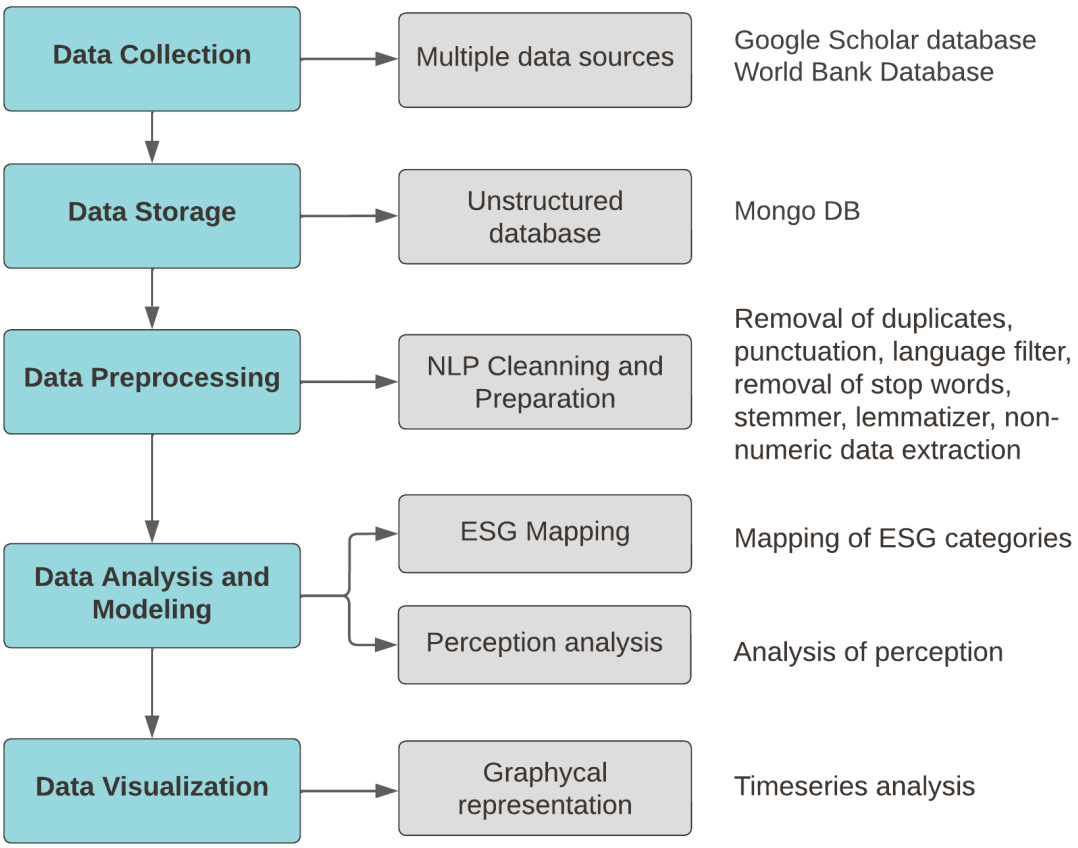

Fig.1. Summary of the data management framework, composed of collection, preprocessing, modeling, analysis, and visualization.

The mapping of the studies evaluated was determined through a deep learning domain-specific BERT Model fine-tuned for text mining in sustainable investing (available in: https://github.com/mukut03/ESG-BERT), providing accuracy higher than and 98% and F-1 score of 0.90. The model provides an overall as well as specific score for individual ESG categories based on the Sustainability Accounting Standards Board's (SASB) materiality map

standards for a particular corpus text. The scores vary from zero to one, in which a higher score indicates stronger contextual relevance or signal of a particular ESG category. For this work, the corpus presenting scores with a maximum value lower than 0.2 were not considered for further analysis. To verify the ability of the model to predict the ESG categories accordingly, a random portion of the dataset was reviewed manually by the authors, given their expertise in this topic, reaching accuracy values similar to those presented above.

Perception analysis was inferred utilizing both data mining and NLP. In this study, the negative perception is understood as a quantitative measure of how much a factor degenerates in its cohesion and coherence; meanwhile, positive perception is considered as a quantitative measure of how much a factor is consolidated in terms of cohesion and coherence. To perform such analysis, the corpus text was analyzed through the FinBERT model (Araci, 2019). It provides a score for three sentiment categories (positive, negative, and neutral), varying from zero to one, where a higher score indicates a stronger probability of the corpus belonging to a given sentiment category. By taking it into account, the Perception Score (PS) was computed by the difference between positive and negative categories. Thus, values lower than zero are related to negative perceptions, and, those higher than zero are related to positive perceptions. The obtained values were further normalized, giving us the Normalized Perception Score (NPS). NPS varies between zero (negative) and one (positive) values.

Python Programming Language (Lee, 2015) was selected to execute the data process workflow, including visualization tasks, due to a wide variety of well-established scientific libraries (Harris et al., 2020; Morris et al., 2020).

3. Results

To leverage the benefits of increasing data availability, it is essential to capture its importance and describe it concisely (Perazzoli, de Santana Neto, et al., 2022). The information retrieved from collected data is presented and discussed in this section. After the preprocessing step, 55,000 scholarly publications were selected, serving as input for this study. A sample of the dataset containing the titles of the scholarly records is available under request at the Zenodo database (Perazzoli, Joshi, et al., 2022).

Fig. 2 shows the perception of publications for ESG. Overall, there is a slight improvement in perception along with time for environmental criteria, which has gained more attention, especially concerning energy management aspects, due to the various initiatives focusing on renewable energy transition and reducing dependence on fossil fuels. In the case of governance, as it is generally aligned with economic growth, it follows a similar trend.

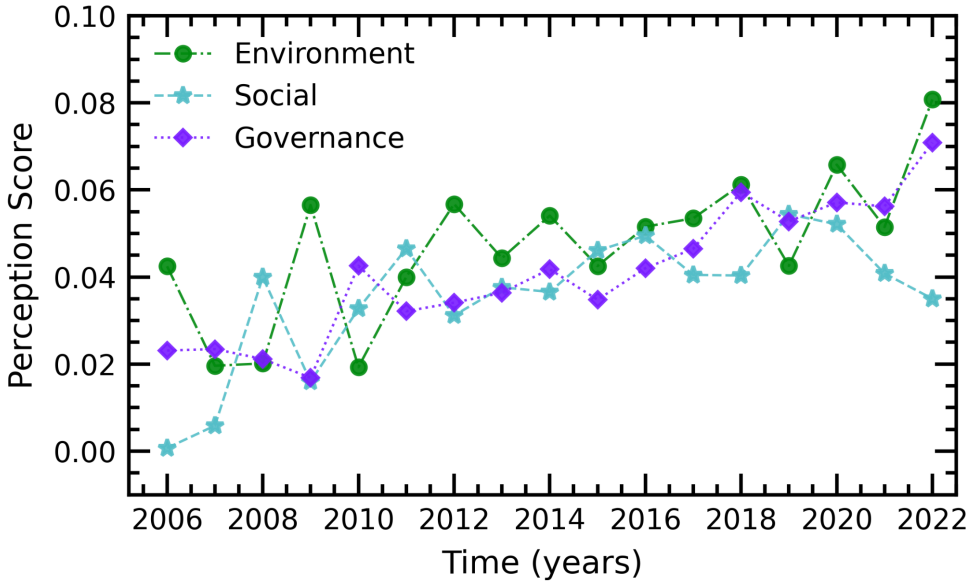

Fig. 2. Forecast of perception analysis through time for Environment (green, circle shape with a dash-dotted line), Social (blue, star shape with dashed line), and Governance (purple, diamond shape with dotted line).

What draws attention is the behavior of the social criterion, which can be divided into 3 phases. There is a certain growth between 2006 and 2010 mainly due to the technological advent, in which the technology itself democratizes access to information and dissemination of knowledge. However, it is observed a period of stagnation from 2010 until 2019. This is because, even with the adoption of technology in favor of process optimization, there is no longer gain concerning the social issue after a certain point. Ultimately, a decrease started in 2019. It is explained due to the exposition of the social problems accentuated by the Covid-19 crisis, especially concerning the neglect of social and human aspects of this period (He & Harris, 2020; Kateb et al., 2022; Mendez-Lopez et al., 2022; Patuelli et al., 2021).

By analyzing the sole perception score (in terms of NPS) for aggregated ESG categories and comparing it to the respective NPS/GDP ratio, it is noticed a meaningful decrease in the growth rate for the perception when it is normalized by the GDP (Table 1). As can be observed in Table 1, this reduction varies from 60.71% for environment, 18.42% for social, and 30.77% for governance categories, respectively. These findings may exhibit signs of unsustainable economic growth, in which several challenges still need to be addressed (Diaye et al., 2022).

Table 1. Statistical coefficients related to the comparison among the NPS and NPS/GDP ratio between 2006 to 2022.

		Perception growth ($ps.t^{-1}$)*	Squared error (r^2)	Variance (σ^2)
NPS	E	0.028	0.4849	0.042
	S	0.038	0.4743	0.077
	G	0.039	0.8515	0.045
NPS/ GDP ratio	E	0.011	0.065	0.051
	S	0.031	0.210	0.113
	G	0.027	0.614	0.031

*ps refers to the perception (NPS or NPS/GDP ratio) for E, S or G category, in a given time t, in years.

4. Conclusions

Due to the contemporary challenging world scenario, there is an increasing demand for the development of initiatives to address societal challenges. Among them, the ESG criteria have received prominent attention. Traditionally, it is adopted to handle investments, targeting opportunities with favorable risks. In addition to the context of investments, it has the potential to bring insights, measurements, understandings, and evaluations on how companies and their economic processes affect society in several layers, from a local to regional and global view. To achieve this perspective and potential, it is necessary to possess interdisciplinary and integrative approaches.

By considering it, a comprehensive analysis of 55,000 publications supported by NLP on the aspects of ESG-related topics was presented. Particularly, the analysis of perception scores for aggregated ESG categories shows a meaningful decrease in the growth rate for the perception when normalized by the GDP, which may indicate signs of unsustainable economic growth. Lastly, it is possible to conclude that based on robust literature analysis, the adopted methodology can bring quantitatively valuable information, serving as a support tool

in decision-making processes involving ESG. Thus, promoting a more sustainable society and world.

5. Acknowledgment

This study is an initiative that does not depend on third-party funding, and the authors are thankful to all those who motivated us to carry out this research.

6. References

Araci, D. (2019). *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models* (P. Ren & Z. Genc (eds.)) [Master of Science in Information Studies: Data Science, University of Amsterdam]. https://doi.org/10.48550/arXiv.1908.10063

Diaye, M.-A., Ho, S.-H., & Oueghlissi, R. (2022). ESG performance and economic growth: a panel co-integration analysis. *Empirica*, *49*(1), 99–122.

Di Vaio, A., Hassan, R., & Alavoine, C. (2022). Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change*, *174*, 121201.

Dixson-Declève, S., Balland, P.-A., Bria, F., Charveriat, C., Dunlop, K., Giovannini, E., Tataj, D., Hidalgo, C., Huang, A., Isaksson, D., Martins, F., Mir Roca, M., Morlet, A., Renda, A., & Sylvia, S. S. (2022). *Industry 5.0: A transformative vision for Europe* (ESIR Policy Brief No. 3). European Commission. https://ec.europa.eu/info/news/industry-50-transformative-vision-europe-2022-jan-13_en

Freiberg, D., Park, D. G., Serafeim, G., & Zochowski, T. R. (2021). Corporate Environmental Impact: Measurement, Data and Information. *Harvard Business School Working Paper*, *20-098*. https://doi.org/10.2139/ssrn.3565533

Guillen-Royo, M. (2020). Applying the fundamental human needs approach to sustainable consumption corridors: participatory workshops involving information and communication technologies. *Sustainability: Science Practice and Policy*, *16*(1), 114–127.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362.

He, H., & Harris, L. (2020). The impact of Covid-19 pandemic on corporate social responsibility and marketing philosophy. *Journal of Business Research*, *116*, 176–182.

Kateb, S., Ruehle, R. C., Kroon, D. P., van Burg, E., & Huber, M. (2022). Innovating under pressure: Adopting digital technologies in social care organizations during the

COVID-19 crisis. *Technovation*, *115*, 102536.

Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, *31*(2), 50–58.

Lee, K. D. (2015). *Python Programming Fundamentals*. Springer.

Mendez-Lopez, A., Stuckler, D., McKee, M., Semenza, J. C., & Lazarus, J. V. (2022). The mental health crisis during the COVID-19 pandemic in older adults and the role of physical distancing interventions and social protection measures in 26 European countries. *SSM - Population Health*, *17*, 101017.

Mina, M. A. E., & Barzola, D. D. P. G. (2019). Data Scientist: A Systematic Review of the Literature. *Technology Trends*, 476–487.

Morris, J., Yoo, J. Y., & Qi, Y. (2020). TextAttack: Lessons learned in designing Python frameworks for NLP. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*. https://doi.org/10.18653/v1/2020.nlposs-1.18

Patuelli, A., Caldarelli, G., Lattanzi, N., & Saracco, F. (2021). Firms' challenges and social responsibilities during Covid-19: A Twitter analysis. *PloS One*, *16*(7), e0254748.

Perazzoli, S., de Santana Neto, J. P., & de Menezes, M. (2022). Systematic analysis of constellation-based techniques by using Natural Language Processing. *Technological Forecasting and Social Change*, *179*, 121674.

Perazzoli, S., Joshi, A., Ajayan, S., & de Santana Neto, J. P. (2022). *Evaluating Environmental, Social, and Governance (ESG) from a Systemic Perspective - Dataset* [Data set]. https://doi.org/10.5281/zenodo.6981641

Qin, X., Liu, J., Wang, Y., Liu, Y., Deng, K., Ma, Y., Zou, K., Li, L., & Sun, X. (2021). Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. *Journal of Clinical Epidemiology*, *133*, 121–129.

Ratinho, T., Amezcua, A., Honig, B., & Zeng, Z. (2020). Supporting entrepreneurs: A systematic review of literature and an agenda for research. *Technological Forecasting and Social Change*, *154*, 119956.

Salazar-Reyna, R., Gonzalez-Aleu, F., Granda-Gutierrez, E. M. A., Diaz-Ramirez, J., Garza-Reyes, J. A., & Kumar, A. (2020). A systematic literature review of data science, data analytics and machine learning applied to healthcare engineering systems. In *Management Decision: Vol. ahead-of-print* (Issue ahead-of-print). https://doi.org/10.1108/md-01-2020-0035

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *An International Journal on Information Fusion*, *36*, 10–25.

Tíjaro-Rojas, R., Arce-Trigatti, A., Cupp, J., Pascal, J., & Arce, P. E. (2016). A Systematic and Integrative Sequence Approach (SISA) for mastery learning: Anchoring Bloom's Revised Taxonomy to student learning. *Education for Chemical Engineers*, *17*, 31–43.

UNEP. (2005). *A legal framework for the integration of environmental, social and governance*

issues into institutional investment. UNEP. https://www.unepfi.org/publications/investment-publications/a-legal-framework-for-the-integration-of-environmental-social-and-governance-issues-into-institutional-investment/