Semantic Clustering of Material ESG Issues: A Comparative Study on Indian Private and Public Sector Banks

Trikal Kumar Paul
Research Scholar,
Department of Commerce,
University of North Bengal
Email: rs_trikal@nbu.ac.in

Dr. Soumitra Sarkar
Associate Professor,
Department of Commerce,
University of North Bengal
Email: soumitrasarkar@nbu.ac.in

ABSTRACT

The role of banks in emerging countries such as India is indubitable, and with sustainable development being recognised as the path forward, this role has now added on the facets of sustainability. Similar to companies in other sectors, banks too report their sustainability practices and performance through sustainability reports, within which the concept of materiality assessments is an important aspect. Materiality assessments provide stakeholders with information on environment, social and governance (ESG) issues that companies deem to be material. While materiality assessments and issues have been studied for a myriad of industries, works comparing the material issues of banks in India remain scarce. This study makes use of multiple Python libraries to semantically cluster material ESG issues of Indian banks and compares the frequencies and proportions of clusters across the categories of private and public sector banks. Since discrepancies were found in material topics reported by banks across various reporting frameworks in a given year, the mandatory Business Responsibility and Sustainability Report (BRSR) was given precedence. Semantic clustering identified eight material ESG issues, with "Managing Climate Change Impacts" being the most frequent issue and "Regulatory Compliance" being the least frequent. Comparing the aggregate frequency of ESG using Wilcoxon-Mann-Whitney Test reveal that private banks report ESG issues more frequently than their public sector counterparts. However, testing differences in the proportions of frequencies across private and public sector exhibit insignificant differences, highlighting that across ownership categories, Indian banks are not qualitatively different in how they view ESG issues they deem material.

Keywords: *ESG Reporting, Semantic Clustering, BRSR, NIFTY Private Bank Index. NIFTY PSU Bank Index*

1. INTRODUCTION

While financial reports convey information to stakeholders with a financial stake in a company, sustainability reporting (or ESG reporting) communicates to stakeholders information surrounding policies and practices of an organisation vis-à-vis sustainability (Sachs & Sachs, 2021). The issues which plague sustainability reporting, such as a lack of standardisation (Sachs & Sachs, 2021; Siew, 2015; Thijssens et al., 2016) and selective reporting (Macellari et al., 2021; Megura & Gunderson, 2022; Stacchezzini et al., 2016) are reduced to some extent by sustainability reporting frameworks. In India, sustainability reporting is mandated by the Securities and Exchange Board of India for the top 1000 publicly listed companies by market capitalisation through the BRSR, or the Business Responsibility and Sustainability Report since the financial year 2022-2023 (Securities and Exchange Board of India, 2021). The BRSR, which is a successor to the Business Responsibility Report, offers stakeholders with certain advantages over other reporting frameworks. These advantages include its compulsory nature and the need for external assurance, although the credibility of the latter is not without limitations (Boiral & Heras-Saizarbitoria, 2020; Pizzi et al., 2024).

While ESG reporting frameworks standardise sustainability reporting, a concrete difference exists between sectors and companies within them, which leads to qualitative differences in what they consider to be worth pursuing as ESG activities. It is here where materiality assessment finds its importance within the ambit of sustainability reporting(Jørgensen et al., 2022; Ngu & Amran, 2021). Through materiality assessments companies enable themselves to report which ESG issues are material, and therefore important to them (Calabrese et al., 2016; Garst et al., 2022; Jørgensen et al., 2022; Taubken & Feld, 2018). Depending on the position of the company and the sector within the larger economy, these issues may vary, thus providing stakeholders with valuable information about the perspective of the company towards ESG topics.

For developing nations, banks are important institutions which enable economies to move towards economic development (Ehlers & Villar, 2015). By financing economic activities, they fill in the scarcity of capital which is endemic in underdeveloped countries. With sustainable development, banks play an even larger role towards development, where there are expectations of them addressing problems of developing economies ranging from social issues (Adefare et al., 2024), to those of environment (Nguyen et al., 2023). In this sense, the role of banks is

unique when compared to other sectors as their role in sustainability goes beyond their direct activities (Jeucken & Bouma, 2001).

India is a country with a mixture of both private and majority publicly owned banks. In its history, there have been phases wherein multiple banks have been nationalised and then phases where they have been privatised or at least have seen the government partly divesting from them. New private banks have come up in recent decades and multinational banks have also entered the industry. The differences in ownership between private and public sector banks (PSU banks)lead to how they set up their strategies, which translate into a myriad of differences in how they function (Kundu & Banerjee, 2021). However, whether these differences carry over to how they view ESG issues is a problem that is worthwhile to pursue. This problem can be addressed by analysing the ESG issues that are material to them. Are stakeholders, regardless of these differences, and regardless of their stakes, supposed to see them as two separate sectors only on the basis of their ownership pattern, or are the two categories of banks similar enough to be treated as one, without considering how they are owned. This problem requires a brief survey of the literature to understand what it is that has been found in this regard, and what gaps remain in understanding how one is to view companies from the same sector, differing only in how they are owned.

2. LITERATURE REVIEW

Studies on materiality assessment have been conducted in a plethora of sectors (Gill et al., 2025; Guix et al., 2019), with a preponderance of studies emphasising the process of the assessment itself (Guix et al., 2019; Padilla-Garrido et al., 2024) followed by the concept of materiality assessment (Oll et al., 2025), with lesser emphasis on issues of material topics. As far as sustainability reporting frameworks go, the literature has analysed the contents of the Global Reporting Initiative (GRI) and the Integrated Reporting (<IR>) frameworks (Calabrese et al., 2016; Machado et al., 2021; Padilla-Garrido et al., 2024; Setia et al., 2024; Torelli et al., 2020).For banks, materiality has been the centre of studies in different contexts (Aras & Hacioglu Kazak, 2022; Menicucci, 2025), with studies on Indian banks centring on materiality assessments garnering little attention.

The question of whether there are differences in ESG activities and their implications across multiple categories, including bank ownership, has been addressed and answered in various contexts with varying results (Batae et al., 2020; Dogan Basar et al., 2025; Jaiwani & Gopalkrishnan, 2023; Kouzez et al., 2024).However, differences in ESG reporting on the specific question of material

ESG topics across categories of banks based on the basis of their ownership remain an addressable gap.

3. OBJECTIVES

The objectives of this study are as follows:

- To ascertain material ESG issues for Indian banks.
- To compare the overall recognition of material ESG issues across private and public sector banks in India.
- To compare the priorities of material ESG issues for private and public sector banks in India.

4. METHODOLOGY

Data Sources

The banks analysed in this study make up the NIFTY Private Bank and NIFTY PSU Bank indices as of 1st January 2025.The banks under each index are listed in TABLE 1, which shows that the number of private banks is ten and that of PSU banks is twelve. Therefore, a total of 22 banks were included in this study.

TABLE 1: List of Banks Selected for the Study

NIFTY Private Bank	NIFTY PSU Bank
Axis Bank	Bank of Baroda
Bandhan Bank	Bank of India
Federal Bank	Bank of Maharashtra
HDFC Bank	Canara Bank
ICICI Bank	Central Bank of India
IDFC First Bank	Indian Bank
IndusInd Bank	Indian Overseas Bank
Kotak Mahindra Bank	Punjab & Sind Bank
RBL Bank	Punjab National Bank
Yes Bank	State Bank of India
	UCO Bank
	Union Bank of India

Source: NIFTY Indices

The Business Responsibility and Sustainability Reports (BRSRs), annual reports, and other similar reports from 2022-23 to 2023-24 were looked into for their disclosure of material issues. Where conflicts of information from the same company in the same year arose, primarily due to the difference in how companies

might report on a mandatory framework such as the BRSR and how they might report on voluntary reports, the BRSR was given precedence.

Tools Applied

As per the objectives laid out, this study involves collecting only the ESG topics that companies have stated are material to them in a given year. It should be noted that the material ESG issues were found to be either single words or multi-word phrases, both of which will be collectively referred to as 'phrase(s)' from here on out. Data on the material topics were collected in a comma-separated value (csv) file in the tidy format (Wickham, 2014).This involved placing each variable in its own column with individual observations in rows. The variables included the material topic phrase and its attributes, which were the name of the company, the index to which the company belongs, and the year in which the material topic was stated by the company.

For the first objective, this study employed multiple Python libraries, namely, pandas, sentence-transformers and scikit-learn, to generate clusters of phrases with similar meanings (Pedregosa et al., 2011; Reimers & Gurevych, 2019; The pandas development team, 2025).The all-MiniLM-L6-v2 embedding model was used to assign vectors (list of numbers) to each phrase such that phrases with closer meanings were closer in value (Vergou et al., 2023; Yin & Zhang, 2024).The phrases that were now numerically close were then clustered using K-means clustering and assigned a representative phrase with NumPy(Harris et al., 2020).For the process of clustering, the default value of eight clusters was used, and the seed was set arbitrarily at 42 to ensure reproducibility.

For the second objective, this study compared the aggregate number of issues disclosed by private and PSU banks, viewed as two sets. The aim here was to test whether differences in the average frequencies of material topics exist between the two sets of banks, without regard to the frequencies within the issue clusters. The results of this test would enable us to infer whether one category of banks as a whole reports more material issues than the other category. In this regard, the two categories were viewed as two independent samples, and the normality of the data was not assumed. Therefore, to test the differences between the two categories of banks, this study applies the asymptotic Wilcoxon-Mann-Whitney test using the wilcox_test command in the coin package in R (Hollander et al., 2015; Hothorn et al., 2005; R Core Team, 2025) for the following hypothesis:

H01: There is no difference between the average number of ESG issues deemed material across the categories of banks.

The attainment of the third objective consisted of comparing private and PSU banks by testing whether the proportion of clusters of material issues was independent of the bank category. Since the data being analysed is categorical, testing for the existence of a difference would require applying Chi-square test (Bajpai, 2020; R Core Team, 2025) for the following hypothesis:

H02: There are no differences in the proportions of clusters of material ESG issues across the categories of banks.

Both the asymptotic Wilcoxon-Mann-Whitney test and the Chi-square test were applied at a 5% level of significance.

5. FINDINGS

Semantic Clusters

The list of eight clusters along with their representative phrases is presented in TABLE 2. This number of clusters is the default maximum value of clusters made by the tools specified in the METHODOLOGY section. The output from the tools used assigns whole numbers to clusters; that is, the first cluster would be "0: Employee Development and Engagement." For better comprehension, the number 1 has been added to the whole numbers and a prefix of C has been assigned, therefore converting the first cluster to "C1: Employee Development and Engagement" and so on.

TABLE 2: Semantic Clusters of Material ESG Issues and their Representative Phrases

Cluster	Representative Phrase	Frequency	Relative Frequency
C1	Employee Development and Engagement	61	14.63%
C2	Managing Climate Change Impacts	83	19.90%
C3	Privacy and Data Security	54	9.59%
C4	Business Ethics and Governance	55	11.75%
C5	Sustainable Finance	67	16.31%
C6	Regulatory Compliance	27	6.47%
C7	Financial Inclusion	26	6.24%
C8	Customer Satisfaction	44	8.87%
	Total	**417**	**100.00%**

Source: Authors' calculations on data obtained from company reports

As evident from TABLE 2, the banks altogether raised 417 material topics over the period of the study. The highest frequency was observed for cluster C2, represented by the phrase "Managing Climate Change Impacts" with 83 instances where issues with similar meanings were highlighted by banks to be material to them. Following this, the cluster represented by "Sustainable Finance" is the most mentioned cluster, with semantically similar terms being stated 68 times across the reports being analysed for this study. These are followed by clusters C1, C4, and C3, represented by "Employee Development and Engagement", "Business Ethics and Governance", and "Privacy and Data Security", respectively. Thereafter, the phrases "Customer Satisfaction" and "Financial Inclusion", followed by the last ranked phrase "Regulatory Compliance", serve as representative terms for the clusters that they are in.

TABLE 3: Frequency of Material Issues raised by Banks (by Bank, separated by Index)

Private Bank Name	Material Issues Mentioned	PSU Bank Name	Material Issues Mentioned
Axis Bank	14	Bank of Baroda	21
Bandhan Bank	20	Bank of India	20
Federal Bank	26	Bank of Maharashtra	11
HDFC Bank	19	Canara Bank	12
ICICI Bank	7	Central Bank of India	6
IDFC First Bank	54	Indian Bank	13
IndusInd Bank	8	Indian Overseas Bank	14
Kotak Mahindra Bank	35	Punjab & Sind Bank	4
RBL Bank	46	Punjab National Bank	12
Yes Bank	28	State Bank of India	27
		UCO Bank	12
		Union Bank of India	8
Total	**257**	**Total**	**160**

Source: Authors' calculations on data obtained from company reports

As TABLE 3 indicates, the number of material issues raised by banks in the private sector is higher than that of the PSUs, with the frequency of the former being 378 across all clusters and that of the latter being 160 irrespective of clusters. Across all banks, IDFC First Bank raised more issues and is therefore the bank mentioning the highest number of material ESG issues in the private bank category as well. Across both categories of banks, it is followed by RBL Bank, Kotak Mahindra Bank and then the State Bank of India, which leads in the number of ESG issues mentioned as

material within the PSU category. On the other end, the Punjab & Sind Bank is at the bottom, highlighting only four issues that are material to them.

TABLE 4: Asymptotic Wilcoxon-Mann-Whitney Test

Data: Material Issue by Index (PVT, PSU)	
Z	-2.0151
p-value	0.0439

Source: Authors' calculations on TABLE 3

Not only is there an apparent difference in the number of material issues being relayed by the categories of banks in TABLE 3, the difference is also statistically significant. TABLE 4, displaying the results of the asymptotic Wilcoxon-Mann-Whitney test, reveals the difference between private banks and PSU banks to be significant at a 5% level of significance. Therefore, the hypothesis H01 stands rejected.

The second aspect of comparing private and PSU banks involved testing hypothesis H02, which tested whether any difference between the proportion of material topic clusters across bank categories existed. TABLE 5 lists the clusters and classifies their mentions across bank categories in a contingency table, along with their expected frequencies in parentheses. Notably, both categories of banks mentioned issues that populate all the clusters, leaving not a single cluster to be thought of as material by either category exclusively. Moreover, private banks have consistently mentioned more material ESG issues in every cluster when compared to their PSU counterparts.

TABLE 5: Contingency Table of Observed and Expected Mentions of Material ESG Topic Clusters

Cluster	Private Bank Mentions	PSU Bank Mentions	Total
C1	42 (37.5947)	19 (23.4053)	61
C2	46 (51.1535)	37 (31.8465)	83
C3	33 (33.2806)	21 (20.7194)	54
C4	38 (33.8969)	17 (21.1031)	55
C5	39 (41.2926)	28 (25.7074)	67
C6	16 (16.6403)	11 (10.3597)	27
C7	13 (16.0240)	13 (9.97602)	26
C8	30 (27.1175)	14 (16.8825)	44
Total	**257**	**160**	**417**

Source: Authors' calculations on data obtained from company reports

On the question of proportion of clusters however, which would inform as to whether private and PSU banks have different priorities in ESG issues, it is found that the visible difference is not statistically significant at the 5% level, as evidenced by a p-value of 0.219 stated in TABLE 6.Therefore, hypothesis H02 cannot be rejected, and the differences in proportions between private and PSU banks are insignificant.

TABLE 6: Chi-square Test

Data: Contingency Table	
χ^2	6.681
df	7
p-value	0.463

Source: Authors' calculations on TABLE 5

A summary of the results of testing the two hypotheses set for this study and their respective p-values is outlined in TABLE 7. What is evident is that while there is some difference across the two categories of banks in how many ESG issues they view as material, the proportions of the clusters of material ESG issues are not.

TABLE 7: Summary of Results for Testing Hypotheses

Hypothesis	**p-value**	**Rejection Status of null**
H_{01}: There is no difference between the average number of ESG issues deemed material across the categories of banks.	0.0439	Rejected
H_{02}: There are no differences in the proportions of clusters of material ESG issues across the categories of banks.	0.463	Failed to reject

Source: TABLE 3 and TABLE 6

6. CONCLUSION

This study derived clusters of material ESG issues of Indian banks from 2022-23 to 2023-24 using semantic clustering and compared the subsets of private and PSU banks in this regard. Based on the findings and results of this study, the conclusion to be drawn is multifaceted. The banks analysed in this study stated that 417 ESG issues were material to them. Among these topics, issues of climate change being the most talked about exhibit an acknowledgement of environmental issues by Indian banks. This is followed by a myriad of ESG issues which both categories of banks have stated, showing the span of ESG issues that banks in India, regardless of whether they are private banks or PSUs, deem to be material to them.

While both categories of banks have stated issues across all clusters, private banks have been found to state ESG issues more frequently than PSU banks. This, in conjunction with the results which exhibit no differences in the proportions across clusters of material ESG topics, displays that the difference in materiality reporting between Indian banks across categories based on ownership is limited to aggregate quantitative differences only. Therefore, banks in India, regardless of whether they are privately owned or are PSUs, view similar material ESG issues in an order that is not too dissimilar.

The results are pertinent to stakeholders, which encompass actors in all economic activities in the banking sector. First, the list of material issues and their order of mention provide stakeholders with the ability to determine whether their ESG concerns are congruent with those of the banking sector. The utility of such clusters is sector indifferent, and clusters for any sector would provide its stakeholders with similar value. Given the banking sector's position in the economy, the findings have unique implications. With the priorities of material ESG issues of banks, businesses can gauge the ESG issues they must address in their activities to better align with the ESG issues that are material to an actual, or at the very least potential, major source of their funds. Furthermore, because priorities do not vary substantially across the two types of banks, stakeholders need not consider the two categories of banks differently regarding which ESG issues are material to them.

Within the literature on ESG issues, this study adds to the literature on material issue identification (Papafloratos et al., 2023) and the application of text mining to ESG reports (Yoon et al., 2023).Furthermore, these results, which indicate a similarity in the priorities of ESG issues and, as a result, make them comparable, provide context for the comparison of banks across other aspects of ESG activities.

REFERENCES

Adefare, T., Adeola, O., Mogaji, E., Nguyen, N. P., & Mogaji, S. A. (2024). Empowering women agriculture entrepreneurs: Banks' role in achieving sustainable development goals. *International Journal of Bank Marketing*, *42*(4), 692–724. https://doi.org/10.1108/IJBM-03-2023-0128

Aras, G., & Hacioglu Kazak, E. (2022). Enhancing Firm Value through the Lens of ESG Materiality: Evidence from the Banking Sector in OECD Countries. *Sustainability*, *14*(22), 15302. https://doi.org/10.3390/su142215302

Bajpai, N. (2020). *Business Statistics* (3rd edn). Pearson India.

Batae, O.-M., Dragomir, V. D., & Feleaga, L. (2020). Environmental, social, governance (ESG), and financial performance of European banks. *Journal of Accounting and Management Information Systems*, *19*(3), 480–501. https://www.ceeol.com/search/article-detail?id=899304

Boiral, O., & Heras-Saizarbitoria, I. (2020). Sustainability reporting assurance: Creating stakeholder accountability through hyperreality? *Journal of Cleaner Production*, *243*, 118596. https://doi.org/10.1016/j.jclepro.2019.118596

Calabrese, A., Costa, R., Levialdi, N., & Menichini, T. (2016). A fuzzy analytic hierarchy process method to support materiality assessment in sustainability reporting. *Journal of Cleaner Production*, *121*, 248–264. https://doi.org/10.1016/j.jclepro.2015.12.005

Dogan Basar, B., Ekşi, İ. H., & Yudaruddin, R. (2025). Is ownership structure effective in the relationship between ESG and bank performance? *Journal of Financial Regulation and Compliance*, *33*(3), 347–358. https://doi.org/10.1108/JFRC-10-2024-0194

Ehlers, T., & Villar, A. (2015). *The Role of Banks* (SSRN Scholarly Paper No. 2692224). Social Science Research Network. https://papers.ssrn.com/abstract=2692224

Garst, J., Maas, K., & Suijs, J. (2022). Materiality Assessment Is an Art, Not a Science: Selecting ESG Topics for Sustainability Reports. *California Management Review*, *65*(1), 64–90. https://doi.org/10.1177/00081256221120692

Gill, S., Kohli, M., & Satija, P. (2025). Materiality in sustainability disclosures and investment portfolio performance in the Indian context. *Journal of Sustainable Finance & Investment*, *15*(3), 757–779. https://doi.org/10.1080/20430795.2022.2086844

Guix, M., Font, X., & Bonilla-Priego, M. J. (2019). Materiality: Stakeholder accountability choices in hotels' sustainability reports. *International Journal of Contemporary Hospitality Management*, *31*(6), 2321–2338. https://doi.org/10.1108/IJCHM-05-2018-0366

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. https://doi.org/10.1038/s41586-020-2649-2

Hollander, M., A. Wolfe, D., & Chicken, E. (2015). *Nonparametric Statistical Methods* (1st edn). Wiley. https://doi.org/10.1002/9781119196037

Hothorn, T., Winell, H., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2005). *coin: Conditional Inference Procedures in a Permutation Test Framework* (p. 1.4-3) [Data set]. https://doi.org/10.32614/CRAN.package.coin

Jaiwani, M., & Gopalkrishnan, S. (2023). Do private and public sector banks respond to ESG in the same way? Some evidences from India. *Benchmarking: An International Journal*, *32*(1), 194–221. https://doi.org/10.1108/BIJ-05-2023-0340

Jeucken, M., & Bouma, J. J. (2001). The Changing Environment of Banks*. In *Sustainable Banking*. Routledge.

Jørgensen, S., Mjøs, A., & Pedersen, L. J. T. (2022). Sustainability reporting and approaches to materiality: Tensions and potential resolutions. *Sustainability Accounting, Management and Policy Journal*, *13*(2), 341–361. https://doi.org/10.1108/SAMPJ-01-2021-0009

Kouzez, M., Lee, J. Y., & Leroux, J. M. (2024). Investment in ESG activities and bank performance: Does bank ownership matter. *International Journal of Business Governance and Ethics*, *18*(4/5), 357–374. https://doi.org/10.1504/IJBGE.2024.139628

Kundu, S., & Banerjee, A. (2021). Operational and policy efficiency: A comparison between public and private Indian banks. *International Journal of Productivity and Performance Management*, *71*(4), 1537–1558. https://doi.org/10.1108/IJPPM-06-2020-0322

Macellari, M., Yuriev, A., Testa, F., & Boiral, O. (2021). Exploring bluewashing practices of alleged sustainability leaders through a counter-accounting analysis. *Environmental Impact Assessment Review*, *86*, 106489. https://doi.org/10.1016/j.eiar.2020.106489

Machado, B. A. A., Dias, L. C. P., & Fonseca, A. (2021). Transparency of materiality analysis in GRI-based sustainability reports. *Corporate Social Responsibility and Environmental Management*, *28*(2), 570–580. https://doi.org/10.1002/csr.2066

Megura, M., & Gunderson, R. (2022). Better poison is the cure? Critically examining fossil fuel companies, climate change framing, and corporate sustainability reports. *Energy Research & Social Science*, *85*, 102388. https://doi.org/10.1016/j.erss.2021.102388

Menicucci, E. (2025). Double Materiality and ESG for Banks. In E. Menicucci (Ed.), *ESG Integration in the Banking Sector: Navigating Regulatory Frameworks and Strategic Challenges for Financial Institutions* (pp. 85–106). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-81677-2_5

Ngu, S. B., & Amran, A. (2021). Materiality Disclosure in Sustainability Reporting: Evidence from Malaysia. *Asian Journal of Business and Accounting*, *14*(1), 225–252. https://doi.org/10.22452/ajba.vol14no1.9

Nguyen, A. H., Do, M. H. T., Hoang, T. G., & Nguyen, L. Q. T. (2023). Green financing for sustainable development: Insights from multiple cases of Vietnamese commercial banks. *Business Strategy and the Environment*, *32*(1), 321–335. https://doi.org/10.1002/bse.3132

Oll, J., Spandel, T., Schiemann, F., & Akkermann, J. (2025). The concept of materiality in sustainability reporting: From essential contestation to research opportunities. *Sustainability Accounting, Management and Policy Journal*, *16*(2), 321–350. https://doi.org/10.1108/SAMPJ-03-2024-0296

Padilla-Garrido, N., Aguado-Correa, F., Rabadán-Martín, I., López-Jiménez, J. M., De La Vega-Jiménez, J. J., & Peletier-Ribera, I. (2024). Materiality analysis in sustainability reporting: Insights from large Spanish companies. *Corporate Social Responsibility and Environmental Management*, *31*(6), 5391–5412. https://doi.org/10.1002/csr.2866

Papafloratos, T., Markidis, I., Kotzaivazoglou, I., & Fragidis, G. (2023). Sustainability Material Topics and Materiality Analysis in the Chemical Industry. *Sustainability*, *15*(18), 14014. https://doi.org/10.3390/su151814014

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pizzi, S., Venturelli, A., & Caputo, F. (2024). Restoring trust in sustainability reporting: The enabling role of the external assurance. *Current Opinion in Environmental Sustainability*, *68*, 101437. https://doi.org/10.1016/j.cosust.2024.101437

R Core Team. (2025). *R: a language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing. https://www.R-project.org/

Reimers, N., & Gurevych, I. (2019, November). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. http://arxiv.org/abs/1908.10084

Sachs, J. D., & Sachs, L. E. (2021). Business alignment for the "Decade of Action". *Journal of International Business Policy*, *4*(1), 22–27. https://doi.org/10.1057/s42214-020-00090-6

Securities and Exchange Board of India. (2021). *Business Responsibility & Sustainability Reporting Format*. https://www.sebi.gov.in/sebi_data/commondocs/may-2021/Business%20responsibility%20and%20sustainability%20reporting%20by%20listed%20entitiesAnnexure1_p.PDF

Setia, N., Abhayawansa, S., Joshi, M., & Wasantha Pathiranage, N. (2024). Shifting perspectives: Unveiling the dual nature of sustainability materiality in integrated reports. *Meditari Accountancy Research*, *32*(4), 1291–1323. https://doi.org/10.1108/MEDAR-08-2023-2128

Siew, R. Y. J. (2015). A review of corporate sustainability reporting tools (SRTs). *Journal of Environmental Management*, *164*, 180–195. https://doi.org/10.1016/j.jenvman.2015.09.010

Stacchezzini, R., Melloni, G., & Lai, A. (2016). Sustainability management and reporting: The role of integrated reporting for communicating corporate sustainability management. *Journal of Cleaner Production*, *136*, 102–110. https://doi.org/10.1016/j.jclepro.2016.01.109

Taubken, N., & Feld, T. Y. (2018). Impact measurement and the concept of materiality—New requirements and approaches for materiality assessments. *NachhaltigkeitsManagementForum | Sustainability Management Forum*, *26*(1–4), 87–100. https://doi.org/10.1007/s00550-018-0483-x

The pandas development team. (2025). *pandas-dev/pandas: Pandas* (Version v2.3.2) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.3509134

Thijssens, T., Bollen, L., & Hassink, H. (2016). Managing sustainability reporting: Many ways to publish exemplary reports. *Journal of Cleaner Production*, *136*, 86–101. https://doi.org/10.1016/j.jclepro.2016.01.098

Torelli, R., Balluchi, F., & Furlotti, K. (2020). The materiality assessment and stakeholder engagement: A content analysis of sustainability reports. *Corporate Social Responsibility and Environmental Management*, *27*(2), 470–484. https://doi.org/10.1002/csr.1813

Vergou, E., Pagouni, I., Nanos, M., & Kermanidis, K. L. (2023). Readability Classification with Wikipedia Data and All-MiniLM Embeddings. In I. Maglogiannis, L. Iliadis, A. Papaleonidas, & I. Chochliouros (Eds), *Artificial Intelligence Applications and Innovations. AIAI 2023 IFIP WG 12.5 International Workshops* (pp. 369–380). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34171-7_30

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10). https://doi.org/10.18637/jss.v059.i10

Yin, C., & Zhang, Z. (2024). A Study of Sentence Similarity Based on the All-minilm-l6-v2 Model With "Same Semantics, Different Structure" After Fine Tuning. In Y. Wang (Ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)* (Vol. 115, pp. 677–684). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-540-9_69

Yoon, J., Han, S., Lee, Y., & Hwang, H. (2023). Text Mining Analysis of ESG Management Reports in South Korea: Comparison With Sustainable Development Goals. *Sage Open*, *13*(4), 21582440231202896. https://doi.org/10.1177/21582440231202896