## FINANSE I PRAWO FINANSOWE

• JOURNAL OF FINANCE AND FINANCIAL LAW •

Marzec/March 2025 • vol. 1(45): 73-91



E

Funding information: Not applicable. Conflicts of interests: None. Ethical considerations: The Author assures of no violations of publication ethics and takes full responsibility for the content of the publication.

# SUSTAINABLE DISCLOSURES OF POLISH BANKS – **TEXT MINING ANALYSIS**

Martyna Białek-Szkudlarek<sup>\*</sup>

https://doi.org/10.18778/2391-6478.1.45.04

#### SUSTAINABLE DISCLOSURES OF POLISH BANKS - TEXT MINING ANALYSIS

#### ABSTRACT

The purpose of the article. Sustainability development issues, particularly Environmental, Social, and Governance (ESG) factors, are becoming increasingly relevant in corporate reporting, driven by rising environmental awareness and regulatory requirements. The aim of the study is the evaluation of ESG disclosures of selected Polish banks. The aim of the paper fills a gap in the Polish academic literature in economics and finance by analyzing the volume and size of non-financial ESG disclosures through computer text mining techniques.

Methodology. This study applies text mining techniques to evaluate the ESG volume and size from 107 financial reports issued by Polish banks between 2006 and 2023. For this purpose, selected tools for computer-based analysis of textual data (text mining) are used. The primary methods include the emotional attitude (sentiment), analysis of the number of words regarding ESG, and analysis of the readability of ESG volume and size contained in company reports.

Results of the research. The study reveals that ESG excerpts are more neutral or less optimistic compared to integrated reports, which tend to have a more positive tone. Additionally, sustainability disclosures are written in a complex language, and the volume of these reports has been increasing over time, likely due to new regulations and growing awareness of sustainability issues. The study focuses on Polish banks but suggests expanding future research to other sectors.

Keywords: sustainability reporting, ESG disclosures, text mining, sentiment analysis, readability index, Polish banking sector.

JEL Class: Q56, Q32, G21, C81, M48.

<sup>&</sup>lt;sup>\*</sup> Mgr, Instytut Ekspertyz Ekonomicznych i Finansowych w Łodzi. Corresponding Author. E-mail: <u>martyna.bialek25@wp.pl</u>

#### Sustainable Disclosures of Polish Banks – Text Mining Analysis

Sustainability development issues have become increasingly important in recent years. This is due, among other things, to an increase in the public's environmental awareness in terms of climate changes, protection of natural resources or biodiversity degradation. There is also a noticeable increase in the public's expectations of environmental actions taken by authorities (both at a local and state level). For years, the emergence and increase of possible risks from the above aspects has been signaled, which could result in serious consequences on a global scale for the environment and human health (EEA Report, 2024; Thuiller et al., 2005; Sheffield & Landrigan, 2010).

In addition to the human impact on the environment, some attention should be paid to corporate behavior, which is also not without impact on the ecosystem, community and corporate governance. In the 1960s, the Club of Rome reports were published, which were the first major work to draw attention to the link between economic development and the state of the natural environment. A large part of the reports focused on aspects of the deteriorating state of ecosystems and emphasized the importance of the finiteness of natural resources, which, if continued to be exploited at similar levels, would be quickly exhausted. In spite of the fact that the reports initially criticized the issue, they stimulated to further studies and helped to raise awareness of the impact on the environment. Other important reports and studies appeared in the following years and included the most significant works and resolutions – the World Conservation Strategy (IUCN, 1980), Agenda 21 (Final Documents of the United Nations Conference on Environment and Development, 1992), the United Nations Framework Convention on Climate Change, with its Kyoto Protocol (1992).

The obligation to disclose non-financial information in company reports, was introduced by European Union Directive 2014/95/EU (NFRD – Non-Financial Reporting Directive). This Directive regards so-called public-interest entities and obliges them to report on the activities that companies undertake within the framework of environmental, social and business ethics activities (ESG). In contrast, 2024 is the first year in which the CSRD (Corporate Sustainability Reporting Directive), one of the elements of the European Green Deal, takes effects. The CSRD Directive, compared to NFRD Directive, expands both the scope of obliged entities and the detail of reported information. The premise of the regulation introduced is to identify risks arising from the above issues in order to prevent them and limit adverse effects in the future (Hristov & Searcy, 2024; Primec & Belak, 2022). Furthermore, with the increasing importance of ESG aspects in the general awareness, reporting of this type of information by companies is increasingly required by stakeholders (e.g., customers, shareholders, employees).

The Sustainable Reports are characterized by their large volume due to the multitude of topics they contain. However, they include mainly text, and possible data numbers only supplementing the textual content, so it can cause difficulties when trying to analyze multiple documents of this type. An important trend, the growing importance of which should not be overlooked, is the use of IT solutions to process text data, which also consists of complex and extensive company reports containing disclosures relating to sustainability issues. Text mining techniques apply modern computational methods to extract valuable information from textual data and detect patterns in the reports' content.

The main objective of the study is to evaluate financial institutions' reports in terms of ESG content. Specifically, there were 107 reports from Polish banks from the years 2006–2023. The following text mining techniques have been applied: the emotional attitude (sentiment) of ESG content (1), volume and size analysis of sustainable reports (2), and analysis of the readability of ESG content contained in company reports (3). This paper fills a gap in the Polish academic literature in the field of economics and finance, by analyzing the content of non-financial information on ESG reported by the banks using computer text mining techniques.

#### **Literature Review**

#### ESG, CSR and Sustainability Development

The sustainable development term was first used in 1987 in the report of the World Commission on Environment and Development and defined as the development that meets the needs of the present without compromising the ability of future generations to meet their own needs. Notably, the existing definitions in the literature from economics and finance focus on corporate profit maximization but also respect environmental and social resources and business ethics in the perspective of both present and future generations (Baborska, 1999; Wilkinson et al., 2001; Dyllick & Hockerts, 2002; Ruggerio, 2021).

From an economic perspective, sustainability is reflected in two concepts – ESG and CSR (Gillan et al., 2021). The ESG acronym was first used in 2004 in the report 'Who Cares Wins' and refers to the areas to which the concept applies (E – environmental, S – social, G – governance). CSR (Corporate Social Responsibility), on the other hand, is most often defined as the concept of managing environmental and social issues in business operations and stakeholder relations (Gray et al., 1987; Hackstone and Milne, 1996; Kiliç et al., 2015; UNIDO, n.d.; for corporate social responsibility standards and norms see Social Accountability International, 2024; *AA1000*, n.d.; ISO, 2024). The main difference between these two concepts is that ESG refers to the overall management and operation of an entity in terms of environmental, social and business ethics, whereas CSR applies to the same processes but only to environmental and community issues (Gillan et al., 2021; Kaźmierczak, 2022).

The reporting of ESG issues may translate into increased reputation among stakeholders (e.g., customers or investors) and consequently improve the companies' value and performance (the impact of ESG disclosures on stakeholder value was studied by Delevingne et al., 2020; Zumente & Bistrova, 2021; in banking, Miralles-Quirós et al., 2019; the value of ESG disclosures on consumer behavior and decisions, Zou et al., 2024; Nugroho et al., 2024). Importantly, these reports mainly consist of textual information and qualitative data while numerical data is significantly smaller. In addition, the lack of

standardization of the data makes it difficult to compare between ESG engagement of companies (Kotsantonis and Serafeim, 2019).

An attempt to address these concerns are ESG ratings – assessing a given company's ESG performance expressed numerically or using a scale (Janicka & Miziołek, 2023). ESG ratings are chosen by investors as an aggregated, quickly and available source of data about an entity when making investment decisions (Berg et al., 2022). ESG ratings are used in empirical research as well (Albuquerque et al., 2019; Abate et al., 2021; Landi and Sciarelli, 2019). Moreover, ESG ratings are developed by independent organizations (the most relevant ESG rating providers – Sustainalytics, KLD, Moody's ESG, Refinitiv, MSCI, S&P Global). However, in practice, again as a result of the lack of standardization, agencies create their own methods and scoring, which still makes it impossible to compare the same companies for different ESG ratings (Billio et al., 2021; Escrig-Olmedo et al., 2019; Charlin et al., 2022; Funk et al., 2024).

The studies on ESG among Polish entities, focus on assessing the level of sustainability data reporting by companies (Balicka, 2014; Kryk, 2017; Majchrzak and Nadolna, 2020) and financial institutions (banks – Wójcik-Jurkiewicz, 2020; Trocka, 2023; insurance companies – Lament, 2017). Less attention is paid on the impact of ESG information on corporate financial performance. It is analyzed based on various aspects of a company's performance, e.g., impact of ESG disclosures on the volatility of a company's share price (Jędrzejka, 2014), analysis of an investment portfolio consisting of socially responsible companies (Wypych, 2018), impact of ESG risks on a discount rate and cash flow value (KPMG, 2022). In the report *ESG in the banking sector* (Gajdka et al., 2024), the authors investigated coefficients whether banks' participation in the RESPECT index on the WSE influences their better stock market performance. A change in corporate lending has been noticeable for some time now, with a tendency among banks to finance companies in industries that are not considered risky. Kosztowniak (2024) presented that commercial banks are more disposed to financing companies which report ESG and belong to "green" industries.

#### Text Mining, Sentiment Analysis and Readability Analysis

The dynamic development of information technology, the Internet and the changing image of society have ushered in the digital age. The huge expansion of information, the volume of which has exceeded human analytical and cognitive capacities, is considered to be the main reason for its emergence of data science field (the term describing techniques from intersection of IT and statistics, such as machine learning, text mining, signal processing, computer vision and knowledge discovery). The ubiquitous access to the Internet, the growing popularity of social networks and the ease of sending and receiving data have resulted in the generation (and continuous growth) of an enormous amount of predominantly textual information (in the form of newspapers, e-mails, books, reports, social media, transcripts of speeches or meeting minutes). Over time, the importance of textual data, which are easily

accessible and relatively inexpensive to obtain, have increased considerably and have been employed in sociology, marketing, finance and economics.

Text mining techniques are used to quickly process and analyze vast amounts of textual data. Text mining, also called computational linguistics or natural language processing, is defined as a process that uses algorithms and methods from the fields of statistics and machine learning to find meaningful patterns in text data (Hotho et al., 2005). Specifically, text mining enables to extract valuable information, look for the relationships and detect patterns using computer techniques (Feldman and Sanger, 2007; Gentzkow et al., 2019; Senave et al., 2023). Gentzkow et al., (2019) draws attention to the relevance of textual data (also in economics) and notes that the popularity of computational linguistics in this field will become increasingly important.

A notable application in finance includes, among others, central bank communication (Rybiński, 2020; Baranowski et al., 2021a; Baranowski et al., 2021b), financial forecasting (Cao and Tay, 2001; Samonas, 2015; Xing et al., 2017), financial risk detection (Peng et al., 2011; Koyuncugil and Ozgulbas, 2012; Bhatore et al., 2020) and fraud detection (Dornadula and Geetha, 2019). There is also growing literature on the influence of social media content on the performance of companies (impact of information from online forums on stock returns – Antweiler and Frank, 2004; Tetlock, 2007; impact of Twitter sentiment on Dow Jones company value – Bollen et al., 2011).

This paper relies on the content generated and shared by companies, namely companies reports. Reports on non-financial information are a form of communication with stakeholders. It can be presented in various forms. An important element of the financial statements is the Directors' Report. In this report, the company presents, among others, milestones that have affected the company, the company's future, the HR situation. The disclosure of this information is mandatory according to the Accounting Act. With the beginning of the disclosure of ESG data by companies, a new form of reporting of this information has appeared. The integrated report presents both financial and non-financial data in particular on sustainability and the impact of non-financial factors on the organization. The integrated report can vary depending on the information presented (e.g., CSR, Sustainable Development, ESG) (Krzysztofek, 2016).

Klimczak et al. (2023) examine the level of disclosure in sustainability reports. It was found that the level of disclosure, although high, does not focus behind actions taken and completed only mostly speaks about the future and plans in this context. These activities are primarily aimed at satisfying stakeholders. Yoon et al. (2023), used text mining tools to examine sustainability attitudes and trends among the Korean companies.

The growth of the Internet, which allows people to share their opinions quickly, was recognized by corporate executives, who were able to easily verify this information for marketing purposes. However, the plethora of this information challenged human capabilities, resulting in the need for estimating an emotional tone contained within a text. To perform this task, a sentiment analysis, which is one branch of text mining, is used. This technique allows to classify a text document as positive, neutral or negative in terms of emotional tone (Liu, 2010) or to express this tone on a continuous numerical scale (e.g., from -1 do 1). Here, it is worth noting an important study of EUROSTOXX index companies (Funk et. al. 2024), where the tone expressed in companies' reports was compared with their ESG ratings. The study did not confirm a significant relationship between the two variables, which seems to confirm the criticism of ESG ratings.

In the Polish literature, especially in the field of economics, one does not find many studies conducted using sentiment analysis. The most common issues studied using sentiment concern consumer behavior (Kocoń et al., 2019), issues in sociology (Meler, 2023) or marketing and branding in the minds of consumers (Iwińska-Knop & Krystyańczuk, 2016). However, it is worth paying particular attention to the study by Rostek and Młodzianowski (2017) where the impact of information from the Internet on changes in the value of WSE stock market indices was investigated and a correlation was found between data from websites focusing on finance and changes in quotations. Furthermore, Baj-Rogowska's (2019) study found that there is a positive and statistically significant correlation between positive opinions and gross profit and a negative correlation between negative opinions and gross profit. All in all, this suggests that text mining techniques may provide a state-of-the-art solution for controlling changes in consumer sentiment and for responding quickly.

Readability analysis is yet another important research tool used to describe a text on a numerical scale. Over the years, a number of methods have been developed to calculate the readability of a text (among the most popular are the Gunning Fog Index, the Flesch Reading Ease, the Coleman Liau Index, the SMOG index (Łopacińska & Wnuk, 2014)), whose value is most often expressed on a scale of a few or a dozen points, which represents the reader's level of education (DuBay, 2004). Among other things, this method is used to detect fraud in financial statements. Management obfuscation theory assumes that managers of poorly performing companies deliberately use complex vocabulary in reports to camouflage unfavorable information for the company. The technique is also particularly applicable in the audit process, making it easier for auditors to catch deviations and fraudulent practices used in the financial statement production process by management (Othman et al., 2012). A study using readability analysis in the context of sustainability reports of companies from the United States, Europe, Australia and India for a variety of industries, found that sustainability reports were characterized by a difficult language, well beyond the knowledge level of a reader with a secondary education (Smeuninx et al., 2016). In the national body of research, this technique has been used mainly in linguistics and computer science research – far less frequently in economics (study of the readability of advertising messages - Korniichuk and Siminski, 2017; analysis of the readability of evaluation reports of research programs - Łopacińska and Wnuk, 2014; study of financial market analysts' communication with investors - Klimczak and Dynel, 2016).

## **Data and Method**

## Dataset

For the purpose of the study, reports from Polish banks containing ESG disclosures were downloaded. Initially, there were 114 reports; however, seven of these were presented in the form of a dynamic website page from which the data downloaded was not very clear and difficult to standardize correctly, so it was excluded. Finally, 107 reports were used. All were in Polish and covered the period from 2006 to 2023 (however, for most of the companies, the reports were available for a shorter period – see Table 1).

As can be seen in the table below, these reports appear in different forms – depending on the information they contain.

## Table 1

The structure of the reports by the quantity and the period they represent

Bank	Type of report	Number of reports	Period
Alior Bank S.A.	Non-financial information	3	2021-2023
Bank Millennium S.A.	CSR	11	2006–2016
	Non-financial information	4	2017-2020
	ESG	3	2021-2023
Bank Ochrony Środowiska S.A.	ESG	3	2020-2022
	Management Board's on the Company's activities	1	2023
Bank Pocztowy S.A.	Management Board's on the Company's activities	4	2019–2022
Bank Gospodarstwa Krajowego	Management Board's on the Company's activities	5	2017–2021
	Integrated	2	2022-2023
BNP Paribas Bank Polska S.A.	CSR	4	2011–2013, 2018
	CSR and ESG	2	2019-2020
	ESG	1	2021
	Integrated	2	2022-2023
Bank Handlowy w Warszawie S.A.	Non-financial information	6	2017–2022
Credit Agricole Bank Polska S.A.	CSR	7	2017–2023
ING Bank Ślaski S.A.	Management Board's on the Company's activities	6	2011–2016
	Integrated	7	2017-2023
mBank S.A.	Management Board's on the Company's activities	4	2015–2018

Martyna Białek-Szkudlarek

	Integrated	3	2019–2021
	ESG	2	2022–2023
Powszechna Kasa Oszczędności Bank Polski S.A.	Application of corporate governance principles	1	2007
	Management Board's on the Company's activities	16	2008–2023
Santander Bank Polska S.A.	CSR	5	2013–2014, 2019
	Management Board's on the Company's activities	1	2018
	ESG	4	2020–2023
	Summary	107	

Source: own study based on data collection from the companies' websites.

The companies' reports from their websites were used in PDF form. These were then generated in a text form, cleaned of characters that cause machine-reading errors. The integrated reports were also cleaned of strictly financial information (mostly figures). In additional, in order to analyze sentiment, the data was categorized into two groups:

- 1. ESG excerpt which were extracted from the companies' general reports on non-financial information, mostly from Management Board's report on the Company's activities).
- 2. Integrated reports which are a combination of financial report and sustainable data. They differ from the standard Management Board's report in the way that the two types of information are integrated with each other, the overall tone of the report focusing on ESG issues rather than business issues.

#### Method

For each text the authors calculated text sentiment and readability index. The study used a readymade sentiment model that had been pre-tested on a corpus of several thousand Polish texts. The sentiment of these texts had been evaluated by experts (Kocoń et al., 2019). This tool uses a neural network, the main element of which is the so-called large language model of the BERT class (Bidirectional Encoder Representations from Transformers). Such models process the text as a whole, considering grammar and context. Although this model shows a high accuracy in sentiment prediction, its complex architecture (more than a dozen layers, non-linear activation functions, more than 100 million parameters; see Mroczkowski et al., 2021) does not allow for a written sentiment formula. The resulting sentiment indicator is normalized from -1 to 1.

As for readability, the authors used Polish texts' readability index inspired by FOG-index (Broda et al., 2014). More specifically, they used an extended formula that takes into account, e.g., average sentence or word length, the ratios between different parts of speech or their share in the text, the proportions of predefined complex words from Imiołczyk's list. The resulting readability score ranges between 1 (classes 1–3 from elementary school) to 7 (postgraduate studies or expert in the field).

Additionally, to illustrate the trend in ESG reporting the volume and size analysis all of all obtained reports is presented.

#### **Results**

#### Sentiment analysis for sustainable disclosures

As it was mentioned above, there were not enough studies in literature using a sentiment analysis combined with sustainable disclosures. However, a sentiment analysis was previously used as a tool to test whether the tone of qualitative information in reports corresponds to the sentiment of qualitative information (Huang et al., 2014). Due to the subjectivity of sustainability reports, it is difficult to assess whether the information presented in them matches reality. For this reason, a sentiment analysis is considered to be a tool to help assess the reliability of reports. There are cases where managers have manipulated readers through an appropriate tone and selection of words – the disproportionate tone of qualitative disclosures in relation to quantitative disclosures may indicate attempts to manipulate the reader (Huang et al., 2014).

#### Figure 1



Results for sustainable disclosures sentiment analysis in Polish banks

Note: INTGR = mean sentiment index from integrated reports, ESG = mean sentiment index in each year for ESG excerpts from non-financial reports.

Source: own study based on data collection from the companies' websites.

Figure 1 shows the evolution of sentiment for sustainable disclosure of Polish banks from 2006 to 2023. The INTGR and ESG represents integrated reports and the ESG excerpts, respectively. Importantly it is found that the sentiment indicator for the integrated reports is higher and much more optimistic than for the ESG excerpts. INTGR sentiment is mostly above 0 (the only drop below 0 in 2009) and ESG sentiment takes values below 0 or close to 0 in studied period. This indicates that the

extracts taken from the general reports are more factual, and probably the intention to demonstrate the company's successes is expressed through quantitative information and financial results rather than sustainability issues. More specifically, average sentiment score of integrated report and ESG excerpts is: 0,00578 and 0,00069, respectively. We used two-sided *t* test for the difference between the sentiment and ESG excerpts. The t-statistics was 3.77 (p = 0.0002), which shows that the difference is significant, even though it is small in magnitude. Consequently, it has been observed that integrated reports are more optimistic than ESG excerpts.

#### Readability analysis for sustainable disclosures

Sustainable disclosures are, by definition, aimed at external users. They are a source of knowledge for them e.g., when making investment decisions (Berg et al., 2022). It is important that their readability is accessible to the regular reader. The previous studies conclude that entity reports, especially those containing sustainable disclosures, are rated as difficult – more difficult than standard financial statements (Smeuninx et al., 2016). Similar results were obtained in the presented study (see Figure 2).

#### Figure 2



Readability of sustainable reports

Source: own study based on data collection from the companies' websites.

As can be seen from the graph, reports containing sustainable disclosure have a high level of readability. The average value for the period under study is around the maximum value of the indicator (which may be due to prediction failure). This means that the text consists of information at the expert level. This may be due to several reasons. Firstly, the complexity of the report – it is a combination of financial and non-financial information, contains a lot of information from a variety of areas and often uses advanced vocabulary (e.g., in the context of carbon emissions or shareholder structure information, it requires more advanced knowledge to be interpreted correctly). Furthermore, as can be seen from the

graph, there is no significant improvement over the studied period. Reports characterized by a difficult language can negatively influence investment decisions and reduce the usefulness of the reports, which is contrary to its communication purpose (Adhariani and du Tiot, 2020).

#### Volume and size analysis for sustainable disclosures

The organizational legitimacy concept assumes that the company uses its own sources of communication (e.g., various reports or press releases) to communicate about activities which are decent and expected by its environment. It also allows to respond to pressure by adjusting the volume or tone of disclosures. This approach is particularly relevant for environmental, community and governance matters (Aerts and Cormier, 2008). Communication strategies are also used by managers when faced with the need to gain, rebuild or strengthen the company's legitimacy in society in the mentioned areas (Suchman, 1995; Cho and Patten, 2006). Text mining is used in content analysis in relation to sustainable disclosures (Aureli et al., 2016; Yoon et al., 2023). This technique allows the extraction of relevant information from the text, e.g., the number of words in the reports allowing for the evaluation of its volume and quality.

Selected text mining tools were used in this study. The authors extracted the number of words for each sustainable report in a particular year (based on 107 reports on integrated, sustainable disclosures and ESG excerpts from annual reports). The numbers of words and reports in each year was then summarized. The resulting data is presented in Figure 3.



#### Figure 3

(W) Number of words in sustainable reports and (M) mean number of words per 1 sustainable report

Source: own study based on data collection from the companies' websites.

Figure 3 presents a significant increase in the number of reports and the number of words in each year. From 2006, the low number of strict ESG reports may have been caused by a lack of awareness and low popularity of sustainability aspects among the Polish banks. Over the years, there

has been an increase in disclosures by banks, which may have been due to peer pressure and a desire to build a positive brand image. The rise was also influenced by the obligations to present these disclosures (2017 was the first year when non-financial data was obliged to disclosure and a significant increase compared to the previous year can be seen then). As mentioned earlier, from 2024 the CSRD Directive covering more companies than the NFRD Directive, which will probably result in an increase in the indicators studied in the following years. However, there is a noticeable increase in the number of words per report. The process of increasing volume is due to several issues - the implementation of new regulations, and the detailing of existing ones (CSRD and NFRD Directives), investors' pressure (Delevingne et al., 2020), consumers' expectations (Zou et al., 2024; Nugroho et al., 2024) or risk management (Landi et al., 2022).

### Conclusion

Reporting on ESG issues is important because it helps companies demonstrate their commitment to sustainability, risk management and transparency, which increases investor and stakeholder trust. As the reported non-financial information is mainly textual data, the use of *text mining* for its analysis seems entirely justified. The study aimed to investigate the sentiment, readability, volume and size of sustainability disclosures using text mining techniques.

The study shows that the difference between the sentiment of the two types of texts studied – (1) ESG excerpts from annual company reports, and (2) annual company reports where financial and non-financial information is integrated – is strong and statistically significant. Sentiment for ESG excerpts is more neutral or less optimistic than sentiment for annual company reports. This may be due to a tendency to present strict ESG information in a less optimistic or more neutral way (as a result of more information about possible risks, climate disasters or corruption) than annual company reports, for which the tone of the narrative tends to be more optimistic (which may be due to a willingness to improve the company's image) also It was also found that integrated reports are more optimistic than ESG excerpts and the difference is statistically significant. Sentiment analysis of sustainability disclosures has rarely been studied which makes this study one of the first in this field (especially in the Polish literature). As mentioned above, a sentiment analysis could also be a tool for management and other stakeholders to assess a company's commitment to sustainability and overall performance. However, it should be noted that an overly optimistic tone may indicate attempts to manipulate reported data, especially in the case of corporate reports. On the other hand, an overly negative tone may be perceived as poor company performance. It is therefore important to balance a sentiment appropriately to reflect the actual performance and condition of banks. This is undoubtedly a tool that deserves further study, as it provides the public with a lot of relevant information in an aggregated form.

In case of readability, it was found that sustainable disclosures are characterized by a difficult language (rated as a college level). Furthermore, there is no clear improvement in the simplicity of the

language over the period studied. As mentioned, this may be due to the need to hide unfavorable circumstances for the company. In addition, difficult company disclosures, which are by definition addressed to the stakeholders, can be misleading. Similarly to sentiment analysis, there is a lack of studies in the Polish literature evaluating the readability of sustainable reports. The results may provide guidance for managers when creating such reports in the future.

The volume and size analysis showed that not only the words of the sustainability reports are increasing year on year, but also the average volume per report. This is due to new regulations, increased awareness of sustainability, shareholder pressure or legitimacy.

The present study focused only on the Polish banks and their sustainable disclosures between 2006 and 2023. Future studies should also focus on companies from other sectors to be able to fully present the state and structure of sustainable disclosures in Poland.

#### Acknowledgments

My sincere thanks go to Paweł Baranowski for his mentorship, which was crucial in shaping the direction of this research. His guidance was invaluable in the formulation of conclusions and in writing this paper.

#### **Disclosure Statement**

The authors report no conflicts of interest.

#### References

- Abate, G., Basile, I., & Ferrari, P. (2021). The level of sustainability and mutual fund performance in Europe: An empirical analysis using ESG ratings. *Corporate Social Responsibility and Environmental Management*, 28(5).
- AA1000 AccountAbility Principles. (n.d.). <u>https://www.accountability.org/standards/aa1000-accountability-principles</u>
- Adhariani, D., & Du Toit, E. (2020). Readability of sustainability reports: Evidence from Indonesia. Journal of Accounting in Emerging Economies, 10(4), 621–636. <u>https://doi.org/10.1108/jaee-10-2019-0194</u>
- Aerts, W., & Cormier, D. (2008). Media legitimacy and corporate environmental communication. Accounting, Organizations and Society, 34(1), 1–27. https://doi.org/10.1016/j.aos.2008.02.005
- Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, *65*(10).
- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3).

- Aureli, S., Medei, R., Supino, E., & Travaglini, C. (2016). Sustainability disclosure after a crisis. *International Journal of Social Ecology and Sustainable Development*, 7(1), 35–49. https://doi.org/10.4018/ijsesd.2016010102
- Baborska, B. (1999). The concept of sustainable agriculture: Some theoretical and practical issues. *Argumenta Oeconomica*, 2(8).
- Baj-Rogowska, A. (2019). Analiza sentymentu jako narzędzie monitorowania wyników finansowych przedsiębiorstwa. *Przedsiębiorczość i Zarządzanie*, 20(12.1).
- Balicka, A. (2014). Informacje środowiskowe w raportach rocznych przedsiębiorstw branży motoryzacyjnej. Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse, Rynki Finansowe, Ubezpieczenia, 69.
- Baranowski, P., Doryń, W., Łyziak, T., & Stanisławska, E. (2021a). Words and deeds in managing expectations: Empirical evidence from an inflation targeting economy. *Economic Modelling*, 95, 49–67. https://doi.org/10.1016/j.econmod.2020.12.003
- Baranowski, P., Bennani, H., & Doryń, W. (2021b). Do the ECB's introductory statements help predict monetary policy? Evidence from a tone analysis. *European Journal of Political Economy*, 66.
- Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, *26*(6).
- Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: A systematic literature review. *Journal of Banking and Financial Technology*, 4(1).
- Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5), 1426–1445. <u>https://doi.org/10.1002/csr.2177</u>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1).
- Broda, B., Niton, B., Gruszczynski, W., & Ogrodniczuk, M. (2014). Measuring readability of Polish texts: Baseline experiments. In *LREC* (Vol. 24, pp. 573–580).
- Cao, L., & Tay, F. E. H. (2001). Financial forecasting using support vector machines. *Neural Computing & Applications*, *10*(2).
- Charlin, V., Cifuentes, A., & Alfaro, J. (2022). ESG ratings: An industry in need of a major overhaul. *Journal of Sustainable Finance & Investment* [Preprint].
- Cho, C. H., & Patten, D. M. (2006). The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society, 32*(7–8), 639–647. https://doi.org/10.1016/j.aos.2006.09.009
- Delevingne, L., Gründler, A., Kane, S., & Koller, T. (2020, February 12). The ESG premium: New perspectives on value and performance. *Insights on Sustainability*.

- Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups.
- Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *Procedia Computer Science*, *165*.
- DuBay, W. H. (2004). The principles of readability. Impact Information.
- Dyllick, T., & Hockerts, K. (2002). Beyond the business case for corporate sustainability. *Business Strategy and the Environment*, *11*(2).
- Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles. *Sustainability*, 11(3), 915. <u>https://doi.org/10.3390/su11030915</u>
- European Environment Agency. (2024). *Europe's air quality status 2024*. <u>https://www.eea.europa.eu/publications/europes-air-quality-status-2024/europes-air-quality-status-2024</u>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Funk, C., Tönjes, E., & Haas, C. (2024). Exploring the predictive capacity of ESG sentiment on official ratings: A few-shot learning perspective. MAGKS Joint Discussion Paper Series in Economics, 12-2024.
- Gajdka, J., Brzeszczyński, J., Marcinkowska, M., & Kulińska-Sadłocha, E. (2024). Aspekty ESG w działalności banków w Polsce rozwój, stan obecny, perspektywy. PAB/WIB 1/2024.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3).

- Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, 66.
- Gray, R., Owen, D., & Maunders, K. T. (1987). *Corporate social reporting: Accounting and accountability*. Prentice-Hall International.
- Hackston, D., & Milne, M. J. (1996). Some determinants of social and environmental disclosures in New Zealand companies. *Accounting, Auditing &; Accountability Journal, 9*(1).
- Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of Text Mining. *Journal for Language Technology and Computational Linguistics*, 20(1).
- Hristov, I., & Searcy, C. (2024). Integrating sustainability with corporate governance: A framework to implement the Corporate Sustainability Reporting Directive through a balanced scorecard. *Management Decision*.

Huang, X., Hong Theo, S., & Zhang, Y. (2014). Tone Management. The Accounting Review, 89(3).

International Organization for Standardization. (2024). *ISO* 26000:2010 – *Guidance on social responsibility*. <u>https://www.iso.org/standard/42546.html</u>

- IUCN. (1980). World conservation strategy: Living resource conservation for sustainable development.
- Iwińska-Knop, K., & Krystyańczuk, H. (2016). Wykorzystanie Big Data w badaniu wizerunku marki w świadomości konsumentów. *Ekonomika i Organizacja Przedsiębiorstwa*, *9*, 28–42.
- Janicka, M., & Miziołek, T. (2022). Finanse zrównoważone: ESG–Przedsiębiorstwa–Sektor finansowy. Polskie Wydawnictwo Ekonomiczne.
- Jędrzejka, D. (2014). Ujawnienia ESG polskich spółek giełdowych w świetle wyników badań. *Studia Ekonomiczne, 186*, 13–26.
- Kiliç, M., Kuzey, C., & Uyar, A. (2015). The impact of ownership and board structure on corporate social responsibility (CSR) reporting in the Turkish banking industry. *Corporate Governance*, 15(3), 357–374.
- Klimczak, K. M., & Dynel, M. (2016). Linguistic attributes of equity analysts' reports. *Financial Sciences*, 1(26), 9–19.
- Klimczak, K. M., Hadro, D., & Meyer, M. (2023). Executive communication with stakeholders on sustainability: The case of Poland. *Accounting in Europe*, *20*(3), 305–331.
- Kocoń, J., Miłkowski, P., & Zaśko-Zielińska, M. (2019, November). Multi-level sentiment analysis of PolEmo 2.0: Extended corpus of multi-domain consumer reviews. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)* (pp. 135–146).
- Korniichuk, R., & Simiński, R. (2017). Wspomaganie prognozowania wpływu tekstów reklamowych z wykorzystaniem systemu Promovot. *Studia Informatica*, *38*(2), 57–69.
- Kosztowniak, A. (2024). The impact of ESG regulations and taxonomy on the credit process in commercial banks. *Finanse i Prawo Finansowe*, 2(42), 78–95.
- Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, *31*(2), 50–58.
- Koyuncugil, A. S., & Ozgulbas, N. (2012). Financial early warning system model and data mining application for risk detection. *Expert Systems with Applications*, 39(6), 6238–6253. https://doi.org/[DOI]
- KPMG Poland. (2022). *ESG w wycenie przedsiębiorstw*. https://kpmg.com/pl/pl/home/insights/2022/08/esg-w-wycenie-przedsiebiorstw.html
- Kryk, B. (2017). Informacje środowiskowe w sprawozdaniach z działalności. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, (479),* 108–117.
- Krzysztofek, A. (2016). Dyrektywa 2014/95/UE oraz wynikające z niej zmiany / Directive 2014/95/EU, and changes resulting from it. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 450, 89–101.
- Lament, M. (2017). Raportowanie informacji niefinansowych w zakładach ubezpieczeń w Polsce. Zeszyty Teoretyczne Rachunkowości, 91, 45–63.

- Landi, G. C., Iandolo, F., Renzi, A., & Rey, A. (2022). Embedding sustainability in risk management: The impact of environmental, social, and governance ratings on corporate financial risk. *Corporate Social Responsibility and Environmental Management*, 29(4).
- Landi, G., & Sciarelli, M. (2019). Towards a more ethical market: The impact of ESG rating on corporate financial performance. *Social Responsibility Journal*, *15*(1).
- Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkhya & F. J. Damerau (Eds.), *Handbook of Natural Language Processing* (2nd ed., pp. 1–38). Chapman and Hall/CRC.
- Łopacińska, L., & Wnuk, U. (2014). Analiza indeksu mglistości tekstu w raportach ewaluacyjnych strategicznych programów badawczych w obszarze innowacji technicznych. *e-mentor*, 5(57), 45–53.
- Kaźmierczak, M. (2022). A literature review on the difference between CSR and ESG. Scientific Papers of Silesian University of Technology, Organization and Management Series, 2022(162), 275–289. <u>https://doi.org/10.29119/1641-3466.2022.162.16</u>
- Majchrzak, I., & Nadolna, B. (2020). Environmental information in the consolidated financial statements of selected energy companies in Poland a comparative analysis. *Zeszyty Teoretyczne Rachunkowości*, *110*(166), 89–107.
- Meler, A. (2023). Parlamentarne ćwierkanie o pandemii. Analiza sentymentu tweetów parlamentarzystów publikowanych podczas pierwszych 12 miesięcy pandemii koronawirusa w Polsce. *Studia Socjologiczne*, 2(245), 57–82.
- Miralles-Quirós, M. M., Miralles-Quirós, J. L., & Redondo Hernández, J. (2019). ESG performance and shareholder value creation in the banking industry: International differences. *Sustainability*, *11*(5), 1405.
- Mroczkowski, R., Rybak, P., Wróblewska, A., & Gawlik, I. (2021). HerBERT: Efficiently pretrained transformer-based language model for Polish. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing (BSNLP)* (pp. 120–134).
- Nugroho, D. P. D., Hsu, Y., Hartauer, C., & Hartauer, A. (2024). Investigating the interconnection between environmental, social, and governance (ESG), and corporate social responsibility (CSR) strategies: An examination of the influence on consumer behavior. *Sustainability*, 16(2), 3421. https://doi.org/10.3390/su16020614
- Othman, I. W., Hasan, H., Tapsir, R., Rahman, N. A., Tarmuji, I., Majdi, S., Masuri, S. A., & Omar, N. (2012). Text readability and fraud detection. 2012 IEEE Symposium on Business, Engineering and Industrial Applications, 72–77.
- Peng, Y., Wang, G., Kou, G., & Shi, Y. (2011). An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing*, 11(2), 2906–2915. <u>https://doi.org/10.1016/j.asoc.2010.11.028</u>
- Primec, A., & Belak, J. (2022). Sustainable CSR: Legal and managerial demands of the new EU legislation (CSRD) for the future corporate governance practices. *Sustainability*, *14*(24), 16530.

- Rostek, K., & Młodzianowski, D. (2017). Współzależność informacji sieciowych oraz zmian indeksów zachodzących na Giełdzie Papierów Wartościowych w Warszawie. Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach, 42(115), 89–104.
- Ruggerio, C. A. (2021). Sustainability and sustainable development: A review of principles and definitions. *Science of The Total Environment*, 786. https://doi.org/10.1016/j.scitotenv.2021.147481

Rybiński, K. (2020). The forecasting power of the multi-language narrative of sell-side research: A machine learning evaluation. *Finance Research Letters*, *34*. https://doi.org/10.1016/j.frl.2019.08.009

- Samonas, M. (2015). *Financial forecasting, analysis, and modelling: A framework for long-term forecasting.* John Wiley & Sons.
- Senave, E., Jans, M. J., & Srivastava, R. P. (2023). The application of text mining in accounting. *International Journal of Accounting Information Systems*, 50. https://doi.org/10.1016/j.accinf.2023.100624
- Sheffield, P. E., & Landrigan, P. J. (2010). Global climate change and children's health: Threats and strategies for prevention. *Environmental Health Perspectives*, 119(3), 291–298. <u>https://doi.org/10.1289/ehp.1002233</u>
- Smeuninx, N., De Clerck, B., & Aerts, W. (2016). Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and NLP. *International Journal of Business Communication*, 57(1), 52–81. <u>https://doi.org/10.1177/2329488416675456</u>

Social Accountability International. (2024). SA8000 Standard. https://sa-intl.org/programs/sa8000/

- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. Academy of Management Review, 20(3), 571. <u>https://doi.org/10.2307/258788</u>
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3).
- Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, 102(23), 8245– 8250. https://doi.org/10.1073/pnas.0409902102
- Trocka, M. (2023). Raportowanie informacji związanych z klimatem na przykładzie banków. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 85(2). https://doi.org/10.14746/rpeis.2023.85.2.15
- UNIDO. (n.d.). What is CSR? United Nations Industrial Development Organization. <u>https://www.unido.org/our-focus-advancing-economic-competitiveness-competitive-trade-</u> <u>capacities-and-corporate-responsibility-corporate-social-responsibility-market-integration/what-</u> <u>csr</u>
- Wilkinson, A., Hill, M., & Gollan, P. (2001). The sustainability debate. *International Journal of Operations & Production Management*, 21(12).

- Wójcik-Jurkiewicz, M. (2020). Non-financial reporting of banks from a socially responsible perspective – examples from wig-ESG. Zeszyty Teoretyczne Rachunkowości, 108(164).
- World Commission on Environment and Development. (1987). *Our common future*. Oxford University Press.
- Wypych, M. (2018). Efektywność inwestycji w spółki społecznie odpowiedzialne. *Przedsiębiorczość i Zarządzanie*, *19*(10, cz. 1), 553–569.
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2017). Natural language-based financial forecasting: A survey. *Artificial Intelligence Review*, 50(1), 49–73. <u>https://doi.org/10.1007/s10462-017-9588-9</u>
- 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). <u>https://doi.org/10.1177/21582440231202896</u>
- Zou, X., Jiang, J., Zhang, H., & He, H. (2024). ESG performance, media coverage, and brand value. *Asia Pacific Journal of Marketing and Logistics*, *37*(4). <u>https://doi.org/10.1108/APJML-04-2024-0441</u>
- Zumente, I., & Bistrova, J. (2021). ESG importance for long-term shareholder value creation: Literature vs. practice. *Journal of Open Innovation: Technology, Market, and Complexity, 7*(2). <u>https://doi.org/10.3390/joitmc7020127</u>

Received: 05.11.2024 Accepted: 23.03.2025 Available online: 31.03.2025