



News and ESG investment criteria: What's behind it?

Naiara Pikatza-Gorrotxategi¹ · Jon Borregan-Alvarado¹ · Aitor Ruiz-de-la-Torre-Acha² · Izaskun Alvarez-Meaza¹

Received: 5 September 2023 / Revised: 18 January 2024 / Accepted: 22 January 2024
© The Author(s) 2024

Abstract

News written in the press about different companies generates consumer feelings that can condition the reputation of these companies and, consequently, their financial results. One of the practices that might improve a company's reputation is the Environmental, Social and Governance (ESG) investment criteria. In this research, using Natural Language Processing techniques like Sentiment Analysis and Word2Vec, we detected those ESG-related terms that the written press uses in news articles about companies. Thus, we have been able to discover and analyze those terms that improve sympathy toward companies, and those that worsen it. Our findings show that those terms related to sustainable development, good social practices and ethical governance improve the general public's opinion of a company, while those related to greenwashing and socialwashing worsen it. Therefore, this methodology is valid for enabling companies to detect those terms that improve or worsen their reputation, and thus help them make decisions that improve their image.

Keywords Corporate reputation · ESG · News · Sentiment analysis · NLP

1 Introduction

News published in the written press about different companies originates from the practices and events of these companies themselves. In turn, once these news items are published they project an image of these companies, which influences their reputation. Therefore, business practices (social, productive, economic, environmental and/or corporate) influence social opinion through what is said about

them in the news, which, in turn, causes society to influence such practices through the image that is projected of them. Moreover, society is increasingly demanding social responsibility from companies, and requesting that they account for the social and environmental consequences of their actions. One way of measuring these social and environmental consequences is the Environmental, Social and Governance (ESG) investment criteria. These ESG criteria are a set of standards for a company's behavior and are used as a tool for analysis, with which companies can try to measure their Corporate Social Responsibility (CSR), i.e., the degree of responsibility that the company adopts toward society (Porter and Kramer 2006). The ESG criteria for companies refer to the environmental, social and corporate governance factors that can be taken into account when investing in a company (Initiative 2005), as they influence the company in the form of corporate image. It is therefore a tool for analyzing the company's environmental and social policies, which, in turn, can influence the company's finances, in the form of reputation and image (good or bad).

ESG investment criteria are increasingly relevant when it comes to investing in a company. Indeed, they were priority topics at the *World Economic Forum* and the *Davos Forum 2022 (ESG and Sustainable Finance Data Skills and Capacity Building Directory, 2020)*, (Davos 2022: How Businesses

✉ Naiara Pikatza-Gorrotxategi
naiara.picaza@ehu.eus

Jon Borregan-Alvarado
jon.borregan@ehu.eus

Aitor Ruiz-de-la-Torre-Acha
aitor.ruizdelatorre@ehu.eus

Izaskun Alvarez-Meaza
izaskun.alvarez@ehu.eus

¹ Faculty of Engineering, Industrial Organization and Management Engineering Department, University of the Basque Country, Paseo Rafael Moreno "Pitxitxi" 2, Bilbao, Spain

² Faculty of Engineering, Industrial Organization and Management Engineering Department Vitoria-Gasteiz, University of the Basque Country, Nieves Cano 12, Vitoria-Gasteiz, Spain

Can Deliver on ESG Promises | World Economic Forum, n.d.). In fact, for several years, many authors have studied the relationship between the application of ESG criteria and financial performance of the companies. Thus, Friede et al. (2015) demonstrate, through an exhaustive review, that applying ESG criteria in companies leads to better financial results. According to Amir and Serafeim, the main motivations for companies to use ESG information are, in order of importance: return on investment, customer demand, product strategy and, lastly, ethical considerations (Amir and Serafeim 2018). Brooks and Oikonomou also address the relationship between ESG criteria and financial performance. These authors find a link that is positive and statistically significant—but economically modest—between ESG criteria and financial performance on a company level. According to their article, there is an asymmetry in the financial impacts of ESG, whereby the negative financial effects of corporate social irresponsibility are greater than the positive financial effects of corporate social responsibility (Brooks and Oikonomou 2018). In their research, Fatemi et al. (2018) conclude that the strengths of ESG criteria increase company value and ESG concerns decrease it. Finally, Lee et al. (2016) find a significant positive relationship on a company level between environmental responsibility and financial performance, and between environmental responsibility and operational performance.

In this relationship between ESG investment criteria and the company's financial results, the company's reputation or image is a vitally important variable since it affects consumer satisfaction (Chun 2005). One way of measuring a company's image is by taking into account two indicators: The first one is the sympathy that the company generates in society in general, and the second one is the company's good financial results (Raithel et al. 2010). Society receives this data about companies, from external sources such as word of mouth, news, advertising, etc., and then forms an image of the company's reputation (Kossovsky 2012). That is why, by performing a sentiment analysis (SA) of written news about companies, it is possible to measure the reputation they have in society. A positive Sentiment Analysis of news about companies will generate sympathy toward them, improving their reputation.

In this context, SA—a sub-discipline within data mining and computational semantics—is one way of measuring the image projected by news sentiments. According to Pang and Lee (2008), SA is a dynamic and extensively researched subject in the field of natural language processing (NLP). Its main objective is to computationally process the subjectivity in a given text and analyze the opinions, emotions, evaluations, and feelings of individuals. This powerful technique allows for a deeper understanding of data gathered from sentiment-rich sources such as news articles, social media platforms, reviews, and other similar content (Kim 2015).

As a result, SA serves the purpose of extracting sentiments and emotions from text, finding applications in various domains, ranging from assessing customer satisfaction to understanding political opinions (Mäntylä et al. 2018; Pak and Paroubek 2010).

One limitation of SA is its capacity to score the degree of positivity or negativity within a given sentiment, without explaining the underlying reasons for these sentiments. SA only allows understanding the extent to which a sentiment is better or worse, as it provides degrees of sentiment. Consequently, when extracting sentiments from news articles about companies, the analysis remains incomplete because we really want to understand why those feelings are there. Upon identifying this limitation, a bibliometric review was conducted, revealing no existing research examining the meaning of terms related to ESG within written news articles based on the previously generated sentiment degrees (Liu et al. 2023; Mandas et al. 2023; Park et al. 2022; Salas-Zárate et al. 2017; Zeidan 2022).

The aim of this article, therefore, is to identify from written news those issues related to ESG investment criteria that influence whether a company has a better or worse reputation among consumers.

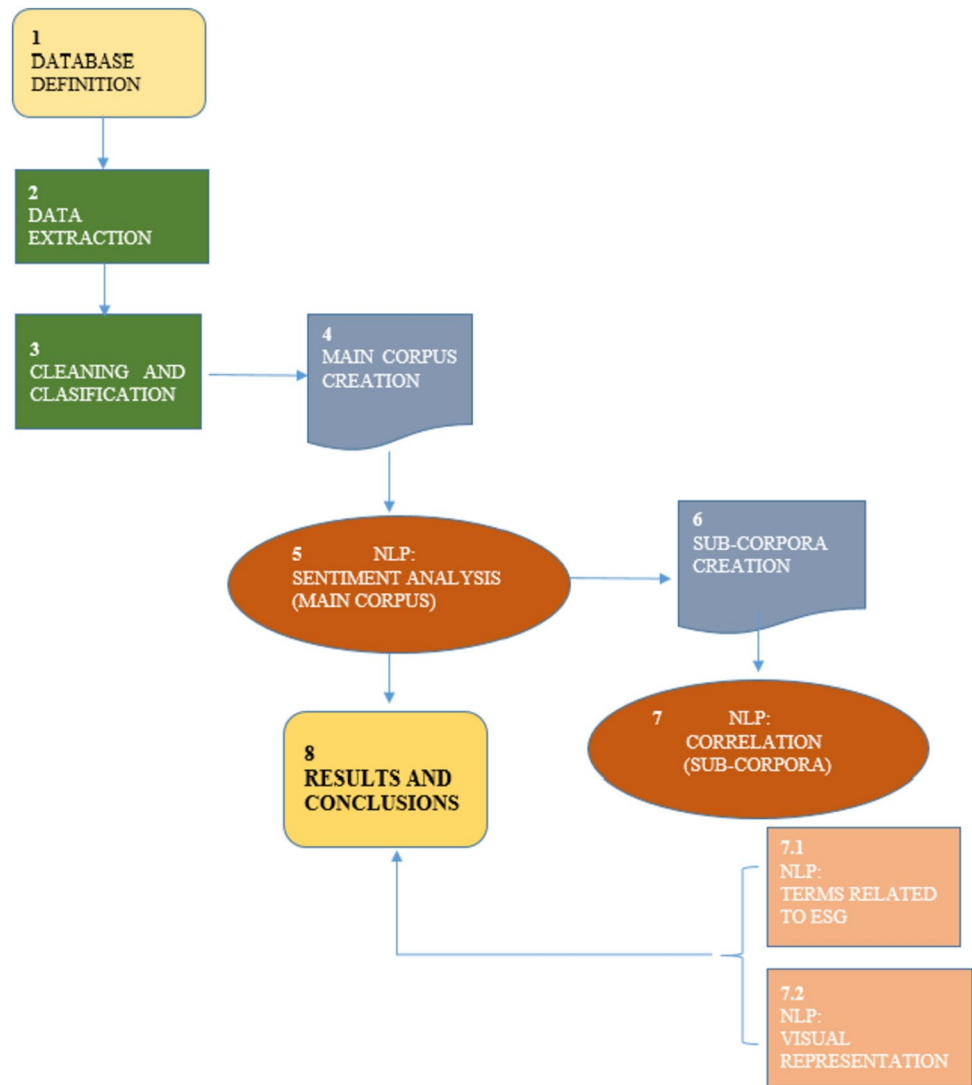
To achieve this, we will firstly identify news written in the press about certain companies. Then, from these news items and using SA techniques, a distinction will be made between those that generate positive and those that generate negative feelings. Finally, we will detect those terms related to ESG investment criteria through Word2Vec techniques executed in Python. It is possible to quantitatively obtain the vector distances between the different terms or words analyzed (word-embeddings), in order to observe those that are closer to—and therefore have greater affinity (Banawan et al. 2023) with—the term or terms of study in this research.

Therefore, thanks to NLP techniques (the combination of SA and Word2Vec methods or models), it is possible to detect, through the terms extracted from the news, the factors that influence whether a company has a better or worse reputation among consumers. As a result, companies will be able to identify, from the published news, those terms close to the ESG investment criteria that have a positive or negative influence on their own image. Among their practices related to ESG criteria, this can be a useful tool for helping companies identify which ones worsen and improve their reputation. In this way, they will be able to make strategic decisions to improve their image and, consequently, their financial results, through consumer behavior.

2 Methodology

The methodological process applied in this research is shown below (Fig. 1):

Fig. 1 Methodological process



2.1 Database definition

The first step in the methodological process was to choose the business sample. A sample of financially consistent companies was sought. For this purpose, we selected the companies from the Eurostoxx 50 that had obtained the best dividend yield at the search date (May 2021). The eight companies with the best financial performance were as follows: Allianz, Basf, BNP Paribas, Daimler, Engie, Eni, ING andIntesa Sanpaolo (*Cotizacion de EURO STOXX 50®—Indice—Resumen—Rentabilidad-Dividendo*, n.d.)

2.2 Data extraction

After choosing the companies, the next objective was to retrieve the news written in the press about those companies. To do this, the original source was used, and these news items located. The query used in each case was the name of each company about which the search was being performed. The 500 most relevant news items per year were chosen for each of the companies from a time period covering 2017 to 2021. Where any company did not reach 500 news items in any year, all of them were chosen. In total, 19,953 news items were downloaded, distributed as follows according to the year (Table 1):

Table 1 Number of news items analyzed

Year	2017	2018	2019	2020	2021	Total
News Items	4104	4047	3838	3952	4012	19,953

Therefore, 2500 news items per company were downloaded (500 news items per year for 5 years) except in three cases: *Intesa San Paolo*, with 1768 news items, and ING with 661.

2.3 Cleaning and classification

Once the news download was done, it was then imported to the data mining software Vantage Point (Liu and Liao 2017). The data were then structured for subsequent export.

2.4 Main corpus creation

Once the data had been cleaned and classified, we then had a corpus with which to proceed to the next step—Sentiment Analysis. The aim here was to detect the topics that influence the reputation of the companies, both positively and negatively. For this purpose, two news corpora were created: the first made up of those news items that obtained a positive Sentiment Analysis, and the second of the news items that had negative results.

2.5 NLP: Sentiment analysis (main corpus)

The news items could then be exported to Orange, a machine learning and data mining suite for data analysis through Python scripting (Demšar et al. 2013). A Sentiment Analysis of the extracted news was performed using the VADER and Hu Liu tools:

- The Phyton tool, Valence Aware Dictionary and Sentiment Reasoner (VADER), is a Sentiment Analysis framework that employs a lexicon-based approach to ascertain the sentiment values of a sentence. VADER has proven to be highly effective in analyzing social media texts, NY Times editorials, movie reviews, and product reviews (Abdul-Rahman et al. 2020). (Thu and Aung 2018; Shapiro et al. 2020; Yu et al, 2021; Medhat et al. 2014). The success of VADER stems from its ability to provide not only Positivity and Negativity scores but also to quantify the degree of positivity or negativity in a given sentiment (Tunca et al. 2023 (Simplifying Sentiment Analysis Using VADER in Python (on Social Media Text) | by Parul Pandey | Analytics Vidhya | Medium, n.d.).)
- The Hu and Liu lexicon is another commonly utilized tool designed specifically for Sentiment Analysis of customer reviews. It classifies words into three resulting categories: Sentiment (a global measure of positivity), Positive, and Negative. The reason for selecting this tool is that it has been predominantly used in studies that do not center around textual production in social media. Its application has shown effectiveness in analyzing cus-

tomers feedback and reviews in various domains (Khoo and Johnkhan 2018).

Given that there are two suitable tools, the first step in measuring the reputation of companies will be through Sentiment Analysis of published news, measured with VADER and Hu Liu.

2.6 Sub-corpora creation

Once the results of the Sentiment Analysis had been obtained, two differentiated corpora were created from the main corpus, with all the news items. The first corpus was comprised of all those news items that had obtained a positive number in the Sentiment Analysis with both tools (VADER and Hu Liu). The second corpus was composed of all those news items that had obtained at least one negative Sentiment Analysis with either of the two tools.

2.7 NLP: correlation

2.7.1 NLP: terms related to ESG

The identification of the terms most related to ESG was carried out in each of the two corpora (positive and negative), via Natural Language Processing (NLP) techniques. Those terms were *environment*, *environmentally*, *social*, *socially* and *government*. This was done through Word2Vec (NLP) models generated and executed in Python, in order to quantitatively obtain the vector distances of several terms, with a value of zero corresponding to the word vectorially closest to the chosen terms, and a value of one to that furthest away

2.7.2 NLP: visual representation

A visual representation of the data was obtained. By means of a conversion to a tabular structure in Python, this new information format, comprising of the vector distances of the words and their metadata, was imported into the TensorBoard Embedding Projector tool; thus obtaining a visual representation of the set of words that make up the word-embedding developed in step 2. The terms obtained were analyzed by comparing both corpora, detecting those terms that may have a positive and negative influence on the company's reputation.

Following the prior generation of two corpora (positive and negative) and their subsequent cleaning, a Word2Vec model—using NLP techniques through Python—was then obtained for each corpus, with information on the set of vectors of the terms that make up the corpus (word-embedding). The set of vectors provides us with the vectorial distance

between the different terms (or terms to be analyzed), so that we can establish those that are most similar to each other (Savytska et al., 2021).

The terms analyzed, in order to know those words that are closer and therefore related (the smaller the vector distance, the greater the affinity), were: “environment,” “environmentally,” “social,” “socially,” and “governance.” These terms were chosen because they are the ones that make up the initial ESG (Environmental, Social and Governance). In addition, when applying NLP techniques using Python, it was observed that the words “environmentally” and “socially” appear with a high frequency in the two generated corpora; so in order to cover the maximum number of terms referring to the ESG concept, these two terms were also analyzed and their corresponding Word2Vec model created.

The most important configuration used in Python during the application of NLP techniques in the generation of Word2Vec models was as follows:

- Vector size: The word vectors used have a dimension of $n = 200$.
- The architecture used to train the algorithm was the so-called skip-gram.
- Negative sampling was used to train the model.
- min_count: All terms with a total frequency of less than five were not taken into consideration.
- Window: The maximum distance between the term to be studied and the word to be predicted within the corpus sentences was five.
- Epochs: The number of iterations performed on each corpus was 10.

Next, Fig. 2 displays the Python code developed, incorporating within it, as an example, the term “social.”

Subsequently, in order to provide another approach, these terms and their related terms were visualized in two dimensions using the Tensorflow Embedding Projector tool (Visualizing Data Using the Embedding Projector in TensorBoard|TensorFlow, 2022). For this purpose, the final Word2Vec models using Python were converted to tabular format, and these were imported into the Tensorflow Embedding Projector for subsequent mapping of the terms to be analyzed. Within this tool, the most important configuration applied was the following:

- Data option: Word2Vec 10 K, as it adjusts to the dimension of $n = 200$ defined above.
- Cosine distance: since the data distribution is unbalanced.
- Number of iterations: 10,000 (stable projection).
- Projection type: t-distributed stochastic neighbor embedding (t-SNE), since it fits correctly to two and three-dimensional displays (Skublov et al. 2022).

- Data points: Since these are corpora with many terms, and in order to eliminate unwanted and non-valuable information, the number of points (terms) was reduced to 1000.

The described configuration is as follows:

Once the NLP analysis in Python has been exported to Word2Vec format, it is uploaded in tabular format to the online tool TensorFlow Embedding Projector, as shown in Fig. 3 below.

With the parameters set according to the defined methodology and after over 10,000 iterations, we obtain, as depicted in Fig. 4, the visual representation of words related to a positive outcome (and vectorially closer) concerning the term “environment.” For the remaining analyzed terms, the steps and configurations used are identical, except that when observing words with vectorially closer negative meanings to a term, the negative Word2Vec model, previously generated, has been loaded in tabular format instead of the positive one. Hence, as the configuration utilized for visual study remains standardized throughout this scientific work, for better reader comprehension and observation, the forthcoming images exclusively capture the visual analysis.

As seen on the right-hand side of Fig. 4, the TensorBoard Embedding Projector also provides us with terms that are vectorially closest to the search word (in this case, “environment”). The limitation present in this case is that, in order for the system to perform adequately within an acceptable computation time, we must significantly reduce the word sample, as indicated by the TensorBoard Embedding Projector itself, as depicted in the following Fig. 5.

The reduction of the sample to a maximum of 10,000 words or points would involve a reduction (or non-utilization) of 68% of the terms from the positive corpus and 59% from the negative corpus. Therefore, by reducing the sample and eliminating such a significant number of terms, the list of terms that are vectorially closest to the search word provided by the TensorBoard Embedding Projector and their vector distances differ from the results of our unfiltered corpora. This is precisely why the NLP methodology was applied using Python. This approach ensures that we consider all terms from our corpora (a wider terminology) and a more accurate calculation of vector distances concerning the term under study.

3 Results and conclusions

The results and conclusions, outlined in their respective sections, were derived from the methodology described earlier. As detailed in the methodology, the primary corpus yielded the initial results. The results for each company were obtained after applying the SA with VADER and Hu Liu to the corpus

```

In [1]: import gensim
        from gensim.models import Word2Vec
        from gensim.test.utils import datapath
        from gensim import utils
        from nltk.tokenize import word_tokenize
        import nltk
        import os
        nltk.download('punkt')

[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!

Out[1]: True

In [2]: class MyCorpus:
        def __init__(self, dir_path):
            self.dir_path = dir_path
        def __iter__(self):
            for file in os.listdir(self.dir_path):
                corpus_path = datapath(os.path.join(self.dir_path, file))
                for line in open(corpus_path, encoding="utf8"):
                    line = line.lower()
                    yield word_tokenize(line, language='english')

In [3]: sentences = MyCorpus('C:\Users\user\AppData\Roaming\nltk_data\corpus\NEG')
        model = Word2Vec(sentences=sentences,
                        vector_size=200,
                        window=5,
                        hs=0,
                        sg=1,
                        negative=10,
                        ns_exponent=-0.5,
                        min_count=5,
                        workers=1,
                        seed=1,
                        epochs=10)

In [4]: model.wv.most_similar("social", topn=20)

Out[4]: [('media', 0.641069769859314),
         ('bundestag', 0.5587275624275208),
         ('bonding', 0.5431100130081177),
         ('pleasure', 0.5398892760276794),
         ('profiles', 0.5318519473075867),
         ('distancing', 0.5282254219055176),
         ('activism', 0.5278491973876953),
         ('well-being', 0.5270755290985107),
         ('welfare', 0.5220524072647095),
         ('socio-economic', 0.5205835103988647),
         ('governance', 0.5162572860717773),
         ('business-friendly', 0.5117393136024475),
         ('cerebral', 0.5109872221946716),
         ('spd', 0.5101480484008789),
         ('butt', 0.4995531737804413),
         ('critic', 0.49738970398902893),
         ('commuting', 0.49736109375953674),
         ('organisational', 0.4957606792449951),
         ('hygiene', 0.49546998739242554),
         ('linkedin', 0.49524563550949097)]

In [5]: word_vectors = model.wv
        word_vectors.save('Negative.w2v')

```

Fig. 2 Python code developed and executed

of news. The relevant conclusions were then drawn based on these results. By utilizing NLP to extract terms from the

sub-corpora and visualizing the data, we interpreted the outcomes to arrive at the final conclusions.

Fig. 3 Data loading in TensorFlow Embedding Projector online tool

Load data from your computer

Step 1: Load a TSV file of vectors.

Example of 3 vectors with dimension 4:

```
0.1\t0.2\t0.5\t0.9
0.2\t0.1\t5.0\t0.2
0.4\t0.1\t7.0\t0.8
```

Choose file

Step 2 (optional): Load a TSV file of metadata.

Example of 3 data points and 2 columns.

Note: If there is more than one column, the first row will be parsed as column labels.

```
Pokémon\tSpecies
Wartortle\tTurtle
Venusaur\tSeed
Charmeleon\tFlame
```

Choose file

The screenshot shows the TensorFlow Embedding Projector interface. On the left, the 'DATA' section shows '5 tensors found' and 'Word2Vec 10K' selected. Below this, there are options for 'Supervise' (set to 'No ignored label'), 'Edit by' (set to 'Tag selection as'), and 'Spherize data' (checked). A 'Checkpoint' is set to 'POSTSV_tensor.tsv' and 'Metadata' to 'POSTSV_metadata.tsv'. At the bottom left, the 'T-SNE' visualization settings are shown, including 'Dimension' (2D), 'Perplexity' (8), 'Learning rate' (10), and 'Supervise' (0). The main visualization area shows a cluster of points labeled 'environment'. At the top right, a status bar indicates 'Points: 1000 | Dimension: 200 | Selected 1000 points'. On the right side, there are controls for 'Show All Data', 'Isolate 1000 points', and 'Clear selection'. Below these, a search bar contains 'environment', and a 'neighbors' slider is set to 99%. The 'distance' metric is set to 'COSINE'. At the bottom right, a table titled 'Nearest points in the original space:' lists names and their cosine distances.

Nearest points in the original space:	
swinburn	0.686
holder	0.695
councillor	0.717
tourism	0.719
stewart	0.723
josh	0.744
visitor	0.748
frydenberg	0.765
wheatley	0.767
youth	0.771
watson	0.774
wang	0.779
patricia	0.780
nelc	0.788
transport	0.792
neutral	0.792
opposition	0.799
escobar	0.801
oxby	0.808
friendly	0.808
cannock	0.812

Fig. 4 TensorBoard Embedding Projector menu

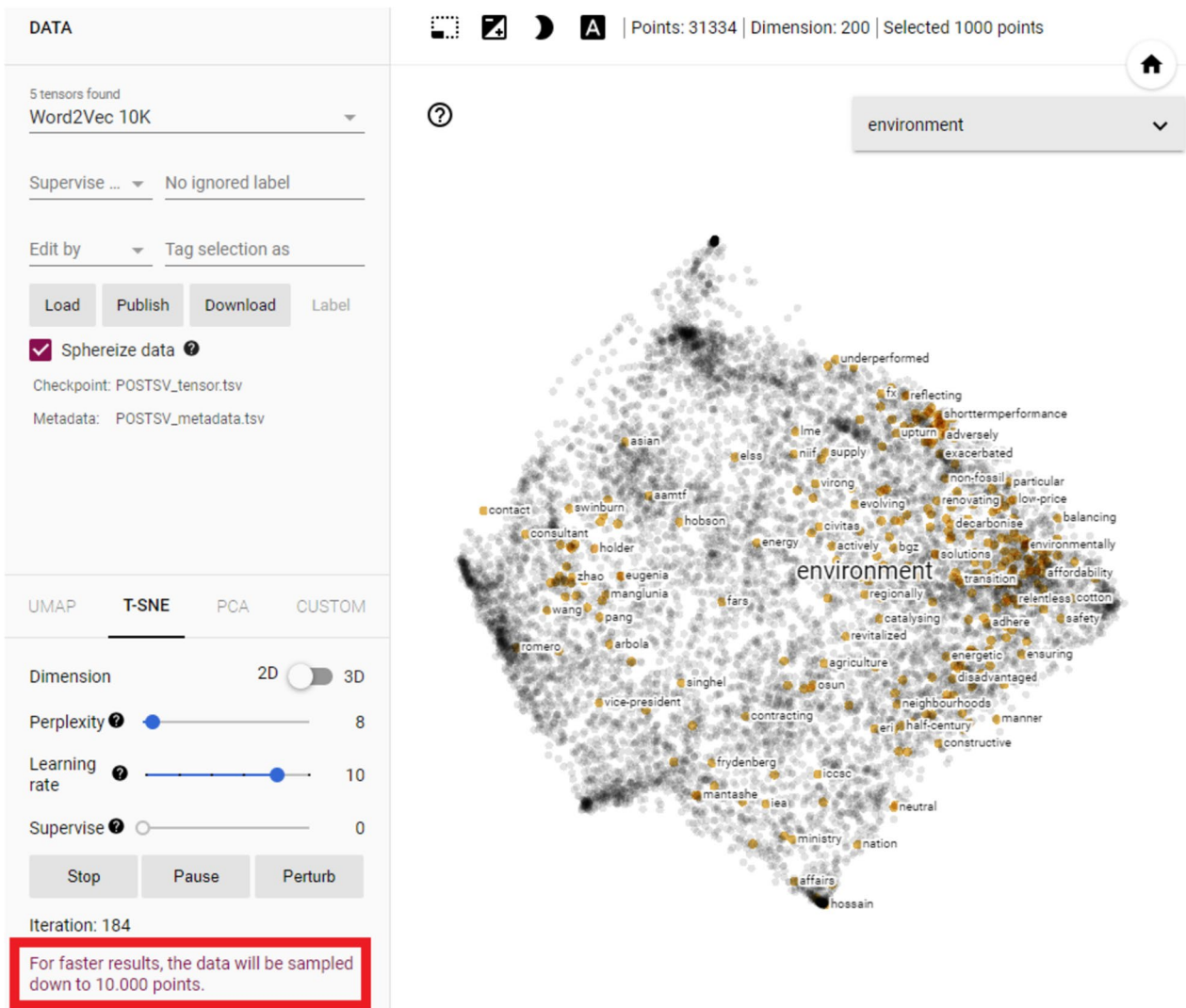


Fig. 5 Filtered in TensorBoard Embedding Projector

4 Results

4.1 NLP: sentiment analysis (main corpus)

Table 2 shows the results obtained from applying Sentiment Analysis to the different news corpora. In this case they have been divided by company and year, from 2017 to 2021. The numbers indicate the degree of “sentiment” obtained by each company each year, when applying the two SA techniques—Vader and Hu Liu. Figures below zero (shown in red) indicate a negative result, i.e., the sentiments extracted from those news items were negative. On the contrary, if the figure is greater than zero or positive, those news items generated positive sentiments or connotations.

In order to study the reliability of the two Sentiment Analysis tools, the Pearson correlation coefficient was calculated, with the results giving a coefficient between VADER and

Hu Liu of 0.5624. Pearson’s correlation coefficient ranges from minus one to one. A value close to one indicates a strong positive correlation, while a value close to minus one indicates a strong negative correlation. A value close to zero indicates a weak or no correlation. In this case, a correlation coefficient of 0.5624 suggests that there is a moderately positive relationship between the two columns of VADER and Hu Liu numbers.

4.2 NLP: correlation

Word2Vec (NLP) techniques were used in each of the two corpora obtained by applying SA (the one formed from news that obtained a positive result and the one formed from news with a negative result). This was done by introducing terms related to ESG in the Python code. The terms were: *environment*, *environmentally*, *social*, *socially* and *governance*.

Table 2 Sentiment analysis applied to the news corpus

Company	SA tool	2017	2018	2019	2020	2021
Axa	Vader	0.5854	0.5020	0.6067	0.4868	0.5246
	Hu Liu	0.4247	0.6175	0.9739	0.7558	0.7642
Eni	Vader	0.4324	0.5069	0.292	0.2458	0.2379
	Hu Liu	− 0.159	0.1346	0.0826	0.0451	0.1123
Intesa Sanpaolo	Vader	0.3902	0.2359	0.2491	0.4422	0.3852
	Hu Liu	− 0.2161	− 0.1158	− 0.029	− 0.0998	− 0.0258
ING	Vader	0.4959	0.4649	0.4582	0.4191	0.4726
	Hu Liu	− 0.0508	− 0.1559	− 0.0911	− 0.0967	− 0.0238
Engie	Vader	0.6451	0.7615	0.66807	0.7289	0.8412
	Hu Liu	0.5361	0.7889	1.06375	− 0.1089	0.0233
BNP Paribas	Vader	0.4526	0.3498	0.22695	0.1446	0.1654
	Hu Liu	0.1447	− 0.2711	− 0.5982	− 0.6652	− 0.6522
BASF	Vader	0.78113	0.7784	0.4003	0.4244	0.3365
	Hu Liu	0.51298	0.5016	0.1311	0.2914	0.3819
Allianz	Vader	0.9974	0.6413	0.5682	0.5014	0.6379
	Hu Liu	1.2698	0.3388	0.1541	0.2663	0.1554
Daimler	Vader	0.7238	0.6545	0.5607	0.5485	0.5982
	Hu Liu	1.2554	1.0816	1.0316	0.9613	1.0253

4.3 Environment and environmentally

The first study terms corresponding to this scientific work—environment and environmentally—were introduced into the execution of code in Python. In this way, we quantitatively obtained the terms “positive” and “negative” with lower vectorial distance (see Table 3), synonymous with related words, due to the continuous and constant appearance by proximity to the terms *environment* and *environmentally*, within the different sentences that make up the different corpora generated by news in the written press about companies.

It should be noted that the greater the existing affinity, the closer the vectorial distance is to the value of zero; and, consequently, the lower the affinity, the closer the value will be to one.

4.4 Social and socially

The same process was then carried out, but this time introducing the terms *social* and *socially* into the model. The terms that were retrieved according to the vectorial distance in each corpus (positive and negative) are shown in Table 4.

4.5 Governance

Finally, the process was repeated, but this time with the third component of the initials ESG, *Governance*. Once again, the terms that were retrieved according to the vector distance in each corpus (positive and negative) were those shown in Table 5.

In order to draw conclusions about these terms, it was decided to classify them. Terms obtained in each corpus (positive and negative) were classified by topics: on the one hand, those terms related to ESG investment criteria were grouped together; on the other hand, those related to the ECONOMY, and finally, those with POSITIVE and NEGATIVE connotations were also grouped together. Those terms that did not belong to any of these sections were grouped in the “NON CLASSIFIED TERMS” section. Any term belonging to more than one section, appears in all of the sections to which it belongs. This process was carried out three times: first with the data obtained from the terms “environmental” and “environmentally” (from Tables 3, 4, 5 and 6); secondly, with the results obtained by introducing the terms “social” and “socially” into the model (from Tables 4, 5, 6 and 7); and finally, the same process was carried out with the data obtained by introducing the term “government” into the model (from Tables 5, 6, 7 and 8). The results obtained in each of the three cases are as follows:

In order to provide a visual appreciation of the vectorial distances, thanks to the conversion to tabular format using Python and the subsequent import into the TensorFlow Embedding Projector tool, different analyses were carried out on the basis of the new perspectives and/or visual models (Figs. 6, 7 and 8).

Table 3 Terms classified by ESG term (environment/environmentally) and corpus

Positive corpus		Negative corpus	
Term vector distance		Term vector distance	
<i>Environment</i>			
Tourism	0.433445572	Prevailing	0.421923995
Wellbeing	0.445513129	Hypothetical	0.434905469
Fluctuations	0.459128320	Disadvantage	0.437582671
Prosperity	0.460092365	Influenced	0.440388023
Digitalization	0.463633954	Affordability	0.444092392
Socially	0.464314401	Tense	0.447711765
Visitor	0.464512288	Low-rate	0.449414432
Low-price	0.465399384	Low-interest-rate	0.449502766
Lifestyles	0.470048427	Manageable	0.450833976
Relentless	0.470940172	Extremely	0.456005692
Realizing	0.471468210	Macroeconomic	0.456753492
Shaping	0.471738159	Affects	0.457300484
Fundamentals	0.474626302	Unfavorable	0.459510564
Promotes	0.474775016	Philosophy	0.459606707
Simplest	0.159067630	Commercially	0.170735061
<i>Environmentally</i>			
Harnessing	0.177143156	Facilitating	0.183501064
Socially	0.185342192	Owning	0.222322463
Cost-effective	0.189878940	Fool	0.223687827
Lower-carbon	0.190832138	Calculate	0.225822567
Prosperity	0.193319737	Inefficient	0.226818144
Sdgs	0.208558917	Rewarding	0.229097247
Minded	0.210308313	Straightforward	0.229637563
Cost-efficient	0.213971376	Qualitative	0.231361687
Zero-carbon	0.223791599	Conscious	0.232301533
Culturally	0.227308213	Readily	0.234510004
Industry leading	0.227360427	Geared	0.235493063
Dignified	0.231191873	Unviable	0.237469732
Trustworthy	0.231316983	Systemically	0.240056097
Cleanest	0.232036113	Socially	0.240715801
Value-add	0.239728987	Define	0.242525458

5 Discussion and conclusions

5.1 NLP: sentiment analysis. main corpus

To check the reliability of the data obtained from the Sentiment Analysis of the news, we first analyzed the tools used, in this case VADER and Hu Liu. For this purpose, the Pearson correlation was calculated between the data obtained with VADER and Hu Liu. In this case, the correlation coefficient of 0.5624 suggests that there is a moderately positive relationship between the two columns of numbers in the two tools. As the analysis coincides, it can be concluded that both techniques are valid for calculating news Sentiment Analysis, and therefore the data obtained are reliable.

Table 4 Terms classified by ESG term (social/socially) and corpus

Positive corpus		Negative corpus	
Term vector distance		Term vector distance	
<i>Social</i>			
Security	0.368864775	Media	0.358930230
Environmental	0.446575463	Bundestag	0.441272438
Media	0.453666687	Bonding	0.456889987
Distancing	0.464979828	Pleasure	0.460110724
Governance	0.496470928	Profiles	0.468148053
Poppy	0.507857412	Distancing	0.471774578
Influencers	0.51081413	Activism	0.472150803
Wellbeing	0.537917942	Wellbeing	0.472924471
Gustafsson	0.540181846	Welfare	0.477947593
Interactions	0.54255316	Socio-economic	0.47941649
Inequality	0.547308713	Governance	0.483742714
Nurture	0.552197189	Business-friendly	0.488260686
Cohesion	0.552558929	Cerebral	0.489012778
Co-investment	0.55849416	Spd	0.489851952
g4	0.561163307	Butt	0.500446826
		Critic	0.502610296
		Commuting	0.502638906
		Organizational	0.504239321
		Hygiene	0.504530013
		Linkedin	0.504754364
<i>Socially</i>			
Culturally	0.187335849	Banana	0.117113769
Minded	0.219793856	Useful	0.156410575
Trustworthy,	0.228682399	Tilbury	0.184528291
Value-add	0.229019701	Conscious	0.194831014
Distanced	0.231041253	Automated	0.198270082
Lower-carbon	0.235871136	Misunderstood	0.199140251
Technologically	0.237680018	Athletes	0.199868977
Structurally	0.239706397	Genuinely	0.203133225
Greener	0.240239739	Treating	0.203298748
Professionally	0.241836727	Philosophy	0.205927908
Prosperity	0.243514419	Remind	0.214348078
Emotionally	0.243970513	Readily	0.21535629
Custodians	0.245280921	Qualitative	0.216073036
Proactive	0.247604966	Systematic	0.217741847
Bottom-up	0.247947633	Behavior	0.219204187
Commoditized	0.25001663	Irresponsible	0.219710112
Conscious	0.25328964	Discriminate	0.220253944
Fiscally	0.254751265	Advertise	0.221538901
Healthier	0.25714457	Utilise	0.221774638

Another result which allows us to conclude that the data obtained in the Sentiment Analysis are reliable is that negative results were only obtained in 14 out of 45 total cases, i.e., in 31.1%. The companies that the news reports refer to are financially consistent, and those news reports produce sentiments with positive connotations. In other words, financially consistent companies “produce” positive sentiments,

Table 5 Terms classified by ESG term (governance) and corpus

Positive corpus		Negative corpus	
Term	vector distance	Term	vector distance
<i>Governance</i>			
Oversight	0.3287686110	stewardship	0.325320423
Societal	0.3314985633	Landell-mills	0.325662017
Accountability	0.3594197035	Sarasin	0.357682705
Chairmanship	0.3608509302	Considerations	0.360547602
Diversity	0.3731201887	ESG	0.371784985
Environmental	0.3738816977	Emphasis	0.379451513
Private-sector	0.3806357980	Incorporate	0.384524584
Sturgeon	0.3824636936	Transparency	0.404077232
Specificity	0.3825304508	Diversity	0.40459466
Agenda	0.3849596977	Environmental	0.409209251
Responsibility	0.3859971762	Engagement	0.410869479
Supervisory	0.3866764903	Ethical	0.415782571
Socio-economic	0.3874014020	Incorporating	0.420585036
		Zeb	0.424612105
Boardrooms	0.3886923790	Deka	0.428588629
Inclusivity	0.3893466592	Black-rock	0.429401100

and one of the variables for measuring the good reputation or image of a company is its financial consistency (Raithel et al. 2010). From this, it can be concluded once again that the data obtained are reliable.

5.2 NLP: Correlation. sub-corpora: terms related to ESG

5.2.1 Environmental and environmentally

If we look at the data visualization of the term *environment*, with regard to the positive terms (green box), three

main clusters can be observed. One of these clusters is composed of the term “environment” together with its related words. In addition, this cluster includes a considerable number of related terms, thus generating a significant and noteworthy area, synonymous with the importance and influence it generates and its high frequency of appearance in the different news items in the written press. As for the negative terms (red box), two main clusters can be seen, which indicates a lower segmentation, but maintaining the same explanations as above; i.e., generated by the term the cluster “environment” and its related terms is relevant and, therefore, remarkable within the “negative” corpus of news in the written press.

Regarding the terms related to “environmental” and “environmentally,” the following was highlighted: The positive corpus contains many terms associated with ESG investment criteria, and several of them have a positive connotation (*wellbeing, cleanest, lower-carbon, zero-carbon*); in turn, the negative corpus has only one term associated with ESG criteria, and it has a neutral connotation (*socially*). As for the terms associated with ECONOMY, there are several characteristics: Among the terms extracted from the positive corpus, some of them have a positive connotation (*cost-effective, cost-efficient, industry leading, value-add*), and several of them are related to productivity. In the negative corpus, on the other hand, some economic terms refer to capital or property (*owning, rewarding*). Moreover, terms associated with intentionality, i.e., actions that can help to achieve a desired result, were also detected: *influenced, manageable, calculate, geared, and facilitating*. Finally, the positive corpus contains many terms with positive connotations (8), and none with negative connotations. The negative corpus, on the other hand, despite containing several positive terms (4), has many more negative ones (13).

Table 6 Terms from Table 3 classified by section and corpus

	Environmental and environmentally	
	Positive corpus	Negative corpus
ESG	Socially, wellbeing, lifestyles, culturally, lower-carbon, sdgs, zero-carbon, cleanest	Socially
Economy	Fluctuations, prosperity, low-price, cost-effective, cost-efficient, industry-leading, value-add //tourism, digitalization, visitor	Affordability, low-rate, low interest-rate, macroeconomic, commercially, owning, rewarding
Positive	Wellbeing, prosperity, promotes, harnessing, cost-effective, dignified, trustworthy, sdgs, lower-carbon, zero-carbon, cleanest	Affordability, facilitating, rewarding, readily
Negative	∅	Disadvantage, tense, unfavorable, fool, affects, inefficient, unviable//terms indicating intentionality: influenced, manageable, calculate, facilitating, geared
Non-classified terms	Relentless, realizing, shaping, fundamentals, simplest	Prevailing, hypothetical, extremely, philosophy, straightforward, qualitative, conscious, systemically, define

Table 7 Terms from Table 4 classified by section and corpus

	Social and socially	
	Positive corpus	Negative corpus
ESG	Environmental, governance, wellbeing, inequality, culturally, lower-carbon, greener, healthier,	Wellbeing, governance, organizational, hygiene,
Economy	Co-investment, value-add, technologically, prosperity, commoditized, fiscally // terms connected to social media: media, influencers,	socio-economic, business-friendly, organizational//terms related to social media: media, linkedin,
Positive	Security, wellbeing, nurture, cohesion, trustworthy, lower-carbon, greener, professionally, prosperity, proactive, healthier,	Pleasure, wellbeing, welfare, business-friendly, hygiene, useful,
Negative	Inequality,	Butt, critic, misunderstood, irresponsible, discriminate, // Terms indicating intentionality: influenced, manageable, calculate, facilitating, geared
Non-classified terms	distancing, poppy, Gustafsson, interactions, G4, minded, distanced, structurally, emotionally, conscious, custodians, bottom-up,	Bundestag, bonding, profiles, distancing, activism, cerebral, SPD, commuting, banana, Tilbury, athletes, genuinely, treating, philosophy, remind, readily, qualitative, conscious, systematic, behavior, advertise, utilize

Table 8 Terms from Table 5 classified by section and corpus

	Governance	
	Positive corpus	Negative corpus
ESG	Societal, environmental, socio-economic, inclusivity	ESG, transparency, environmental
Economy	Societal, accountability, Chairmanship, private sector, socio-economic, boardrooms	Stewardship, Company or entrepreneur names: Landed-mills, Sarasin, Zeb, Deka, Black-Rock
Positive	Diversity, responsibility, inclusivity	Transparency, diversity, engagement, ethical,
Negative		
Non-classified terms	Oversight, sturgeon, specificity, agenda, supervisory	Considerations, emphasis, incorporate, incorporating,

Several conclusions can be drawn from the results obtained. On the one hand, the fact that there are terms with a positive connotation in the positive corpus and terms with a negative connotation in the negative corpus confirms the reliability of the data and of the methodological process. On the other hand, terms related to the ESG criteria appear in the positive corpus, meaning that ESG criteria are associated with good practices. Moreover, the fact that there are so many terms associated with the economy indicates the close relationship between the environment (keyword) and the economy, supporting the initial thesis that ESG investment criteria are closely linked to the company's reputation and, therefore, to its financial results. It can also be seen that several of the economic terms extracted from the positive corpus indicate good results in terms of productivity; i.e., they focus on the process, on how to do, which, linked to ESG terms, can be related to *sustainable development*. The concept of sustainable development implies imposing limits on technology and the social organization of environmental resources to absorb the effects of human activity (Kates et al. 2005; Geissdoerfer et al. 2016). In contrast, the economic terms in the negative corpus refer to raising capital. If we

relate this to the fact that there are also many terms that indicate intentionality, it can be associated with the "use" of the environment as a reputation-enhancing tool, i.e., with *greenwashing*, or how companies deceive consumers about their environmental performance. Such practices can have negative effects on consumer and investor confidence (Delmas and Burbano, 2011; Strauß, 2022; Mendonça et al. 2023).

Therefore, we have detected the practices related to the environment within the ESG investment criteria that improve and worsen the reputation of companies in the news: those related to sustainable development improve it while those related to greenwashing worsen it.

5.2.2 Social and socially

Regarding the visualization of the data with the term *social*, among the positive terms (green box), the term *social* belongs to the main cluster, but does not stand out as an independent cluster. Therefore, it is an important but not crucial term in the various news items analyzed. This visual information is consistent with the analysis of vectorial distances (see Table 4), which also shows that most of the

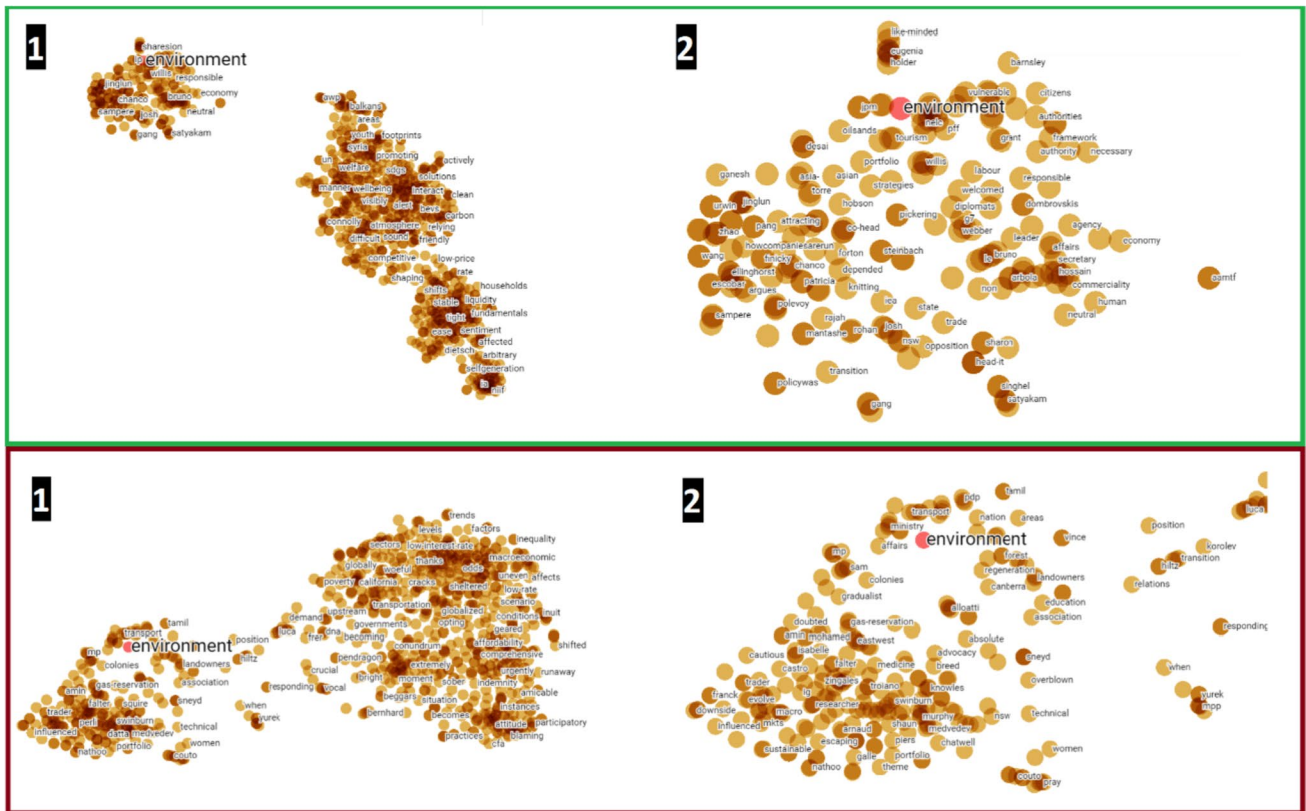


Fig. 6 Visualization of the words and clusters associated to the term *environment* (positive above and negative below)

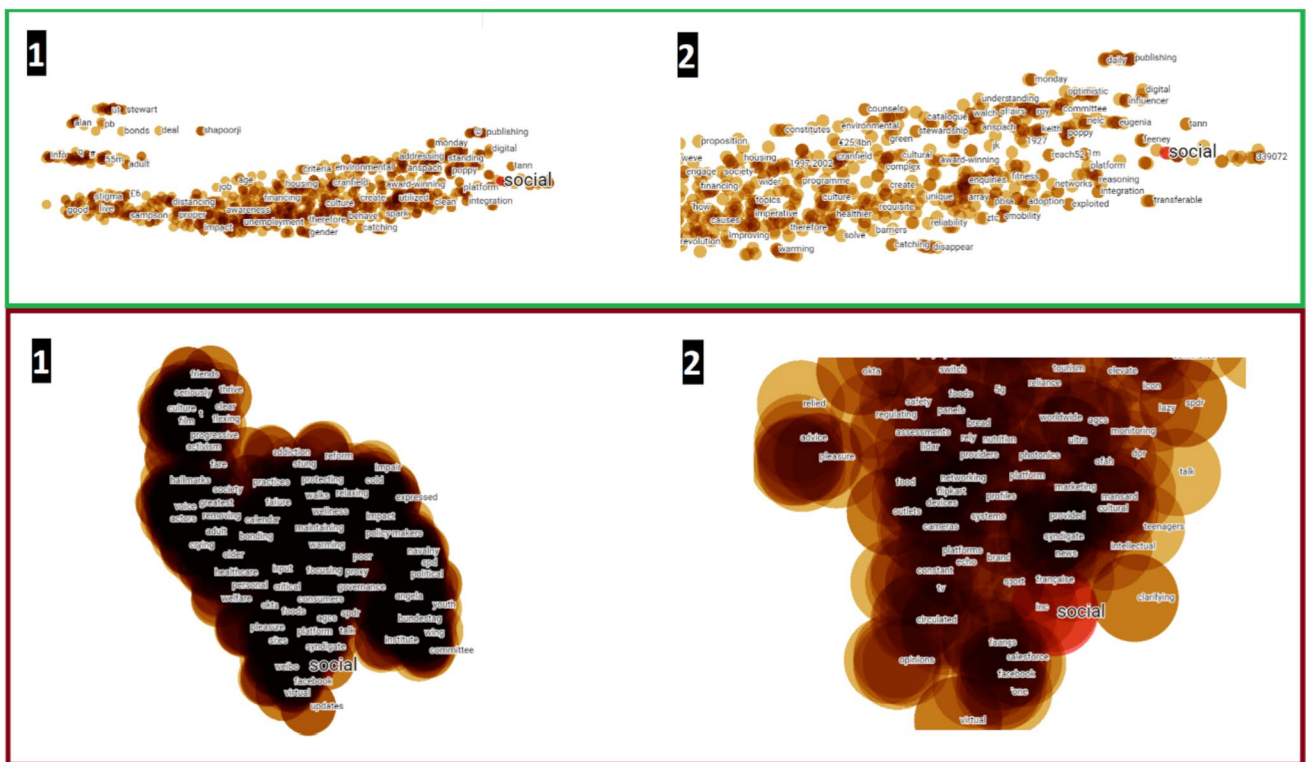


Fig. 7 Visualization of the words and clusters associated to the term *social* (positive above and and negative below)

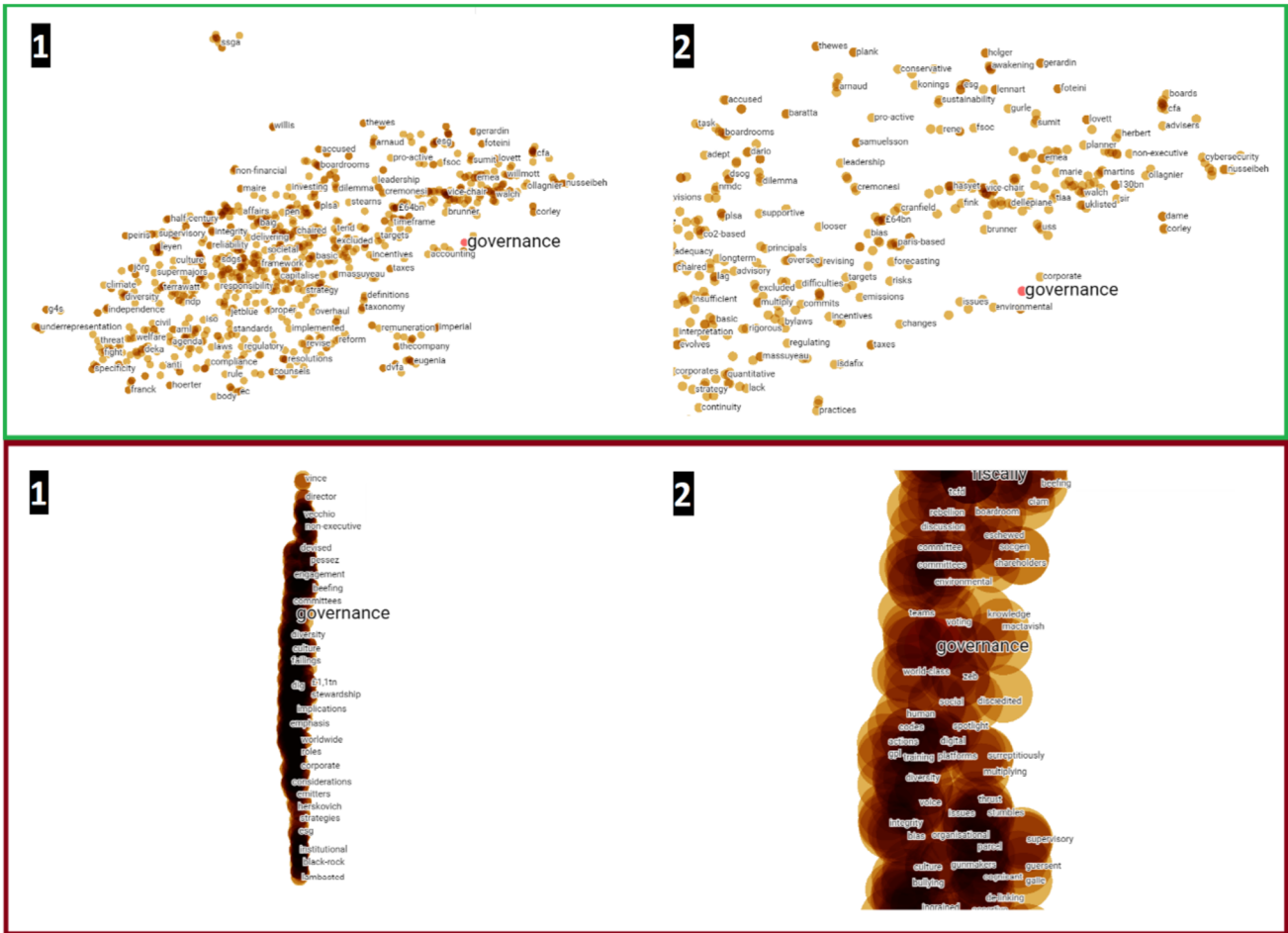


Fig. 8 Visualization of the words and clusters associated to the term *governance* (positive above and negative below)

words related to the term *social* have vector distances greater than 0.5. As for the negative terms (red box), there is no segmentation since there is only one cluster, which includes the term “social.” In this case, as with the positive terms, it is a notable but not crucial term, which coincides with the quantitative analysis corresponding to the vectorial distances.

Once again, the reliability of the data and of the methodological process is confirmed. On the one hand, in the positive corpus there are more terms with positive connotations (15) than in the negative corpus (9). On the other hand, in the negative corpus there are more terms with negative connotations (6) than in the positive corpus (1). As for the terms associated with ESG investment criteria and the environment, almost all of them appear in the positive corpora (6) (*environmentally, culturally, lower-carbon, greener, environmental, governance*), while in the negative corpora only one associated term appears—*governance*. In other words, ESG investment criteria have a positive connotation in the press, and this can have an impact on the good image of the company.

If we focus on the positive terms in the positive corpus, they can be classified into three large blocks: those related to ESG (*greener, healthier and nurture*); those related to the economy (*value-add, prosperity, security, wellbeing and cohesion*); and finally, those terms related to ways of doing or of acting (*minded, professionally, trustworthy, proactive, conscious, emotionally*). These terms can be related mainly to a strong work ethic, and to positive environmental, social and economic results. Therefore, news items that positively evaluate ESG investment criteria relate work ethics to good environmental and social performance and financial prosperity. As for the terms with a negative connotation (almost all of which appear in the negative corpus), once again we can see that they are terms that indicate intentionality (influenced, manageable, calculate, facilitating, geared) or bad practices (butt, critic, misunderstood, irresponsible, discriminate). Considering that all these terms come from the keywords *social*, and *socially*, we can relate a “use” of the social aspect of the company to achieving a good image, i.e., “socialwashing.” In fact, Nardi suggests

that CSR communication can be decisive in discouraging “socialwashing” (Nardi 2022).

It is therefore clear that good social practices in companies get “good press” and, consequently, improve their image. On the other hand, social practices whose sole objective is to improve their image have the opposite effect.

5.2.3 Governance

Finally, with regard to the term *governance*, among the positive terms (green box), four clusters can be observed. Two of these clusters are practically insignificant (“ssga” and “not-for-profit”), and the third one (“thresholds,” “glow,” “values,” etc.) has little influence on the main one. In the main cluster (and with the largest area), we find the term *governance* together with its related words, this being considered a cluster and term that is notable and influential in the different news items analyzed. As for the negative terms (red box), there is only one single cluster, which includes the term *governance*, so there is no segmentation whatsoever. In this case, and in contrast to the negative terms associated with *social*, the quantitative analysis corresponding to the vectorial distances supports the importance of the term *governance* and its related terms, and, therefore, its high frequency of appearance and notoriety within the “negative” corpus generated by the analyzed news.

When we introduce the term *governance*, unlike in the two previous cases (environmental/environmentally and social/socially), the differences between terms with positive and negative connotations are not apparent. In fact, in neither of the two corpora are there any terms with negative connotations. Once again, most of the extracted terms can be classified into ESG and ECONOMY.

The terms in the positive corpus are related to corporate management on the one hand (*accountability, chairmanship, boardrooms*), and to social responsibility on the other (*diversity, responsibility, inclusivity*). In other words, they deal with the responsible management of companies. The terms of the negative corpus also deal with corporate social responsibility (*transparency, diversity, engagement, ethical*). Among these, the terms *transparency* and *ethical* stand out, in clear reference to a “clean” management of the company. However, they do not do the same in a general context as in the positive corpus, but focus on specific companies and entrepreneurs (Landed-mills, Sarasin, Zeb, Deka, Black-Rock). They deal with the ethical and transparent management of specific companies. In other words, the focus is on the responsible and ethical management of certain companies. As there are no adjectives or names with a negative connotation in the corpus, it is not possible to know the term of the criticism, whether it is in a positive or negative sense. It can therefore be concluded that when

news items refer to corporate governance, the focus is on the ethical and responsible management of certain companies.

5.3 NLP: Correlation—Sub-corpora—visual representation

In terms of data visualization—via the NLP technique and Word2Vec models—as expected the results obtained are in accordance with the graphical representations observed in the TensorFlow Embedding Projector tool. The concepts or terms “environment” and “governance” can be seen both in the positive and negative variants, where they have general vectorial distances between 0.325 and 0.474, and are always part of the main clusters or large clusters. Therefore, their frequency of use, and consequently, importance and influence in the news about different companies in the written press, is quite remarkable. The term “social,” on the other hand, has an overall vectorial distance between 0.441 and 0.561 (except for the words “security” and “media”), and is not always part of the main clusters or large clusters. Therefore, although it may appear in the news, its frequency of appearance, and consequently its influence compared to the terms “environment” and “governance,” is not as high. This is synonymous with the fact that companies today are giving greater importance, within ESG, to the environmental and governance aspect than to the social aspect.

In any case, it can be concluded that the news items about companies that appear in the written press deal with the issue of ESG business investment criteria. On the one hand, it has been shown that when talking about the environment in news related to companies, those business practices related to sustainable development improve the company’s image; but on the other hand, those related to greenwashing worsen it. On the other hand, with regard to corporate social practices, we can conclude that good corporate social practices improve the company’s image, while social practices whose sole objective is to improve their image—known as socialwashing—have the opposite effect. Finally, when news items refer to corporate governance, the focus is on the ethical and responsible management of certain companies.

5.4 Implications and limitations of the study and future research

The implications of the study for scientists, business and society have been identified. For academics, as it is a new methodology, it opens up a new perspective on SA research. In terms of interdisciplinary research, it facilitates collaboration between areas such as linguistics, computer science and social sciences by merging text analysis and sentiment processing, thus fostering the exchange of knowledge and approaches. Moreover, by being applicable to a wide range of subjective texts, from news to

social media posts, it broadens the scope of research in areas such as psychology, sociology, and communication. For companies, it becomes a strategic tool to understand and improve their brand image. By identifying terms that generate negative sentiment, companies can adjust their communication and marketing strategies to address issues and improve their brand perception. It also provides an agile tool to monitor brand reputation in real time, enabling a rapid response to changing trends and perceptions. Finally, the implications for society have been analyzed. By enabling SA in various types of texts, society can better understand perceptions, opinions and reactions to issues, products or companies. This promotes greater transparency in the information that is disseminated and helps society to make more conscious consumer decisions and engage in informed discussions in social networks and other media. In addition, society can influence companies to act more responsibly, as public perception can affect their image and reputation. Finally, this analysis can provide information on emerging social trends, changes in cultural perceptions and evolving attitudes toward different issues. This can be useful for governments, non-profit organizations and other actors in decision-making and strategic planning.

In terms of the limitations of the study, the feelings generated by certain topics, and their associated words, can evolve over time. This requires constant updating of the models and studies carried out, as they may become obsolete. On the other hand, when analyzing texts, there may be difficulties in accessing them. In addition, privacy and ethical concerns must be considered, as misuse of personal data or misidentification of emotions could lead to unintended consequences. It is important to consider these limitations when applying this methodology, as they could affect the accuracy, applicability and ethics of the results obtained.

Future research in this field could focus on several aspects to improve and broaden its application. On the one hand, exploring how this model can be automatically adapted and updated to reflect changing trends. On the other hand, research could be extended to address linguistic and cultural diversity by developing SA models that are applicable to different languages and cultures. Finally, the integration of research with other areas, such as artificial intelligence, psychology or sociology, could be explored to gain a deeper understanding of how emotions relate to other human aspects.

Author contributions NP and JB wrote the main manuscript of the text. NP and AR made the analysis and interpretation of the Sentiment Analysis. NP and JB made the analysis and interpretation of the Sentiment NLP. IA and AR prepared figures. AR and IA revised it critically. All authors reviewed the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest There are no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amir AZ, Serafeim G (2018) Why and how investors use ESG information: evidence from a global survey. *Financ Anal J* 74(3):87–103. <https://doi.org/10.2469/faj.v74.n3.2>
- Banawan MP, Shin J, Arner T, Balyan R, Leite WL, McNamara DS (2023) Shared language: linguistic similarity in an algebra discussion forum. *Computers* 12(3):53. <https://doi.org/10.3390/computers12030053>
- Brooks C, Oikonomou I (2018) The effects of environmental, social and governance disclosures and performance on firm value: a review of the literature in accounting and finance. *Br Account Rev* 50(1):1–15. <https://doi.org/10.1016/J.BAR.2017.11.005>
- Chun R (2005) Corporate reputation: meaning and measurement. *Int J Manag Rev* 7(2):91–109. <https://doi.org/10.1111/j.1468-2370.2005.00109.x>
- Cotizacion de EURO STOXX 50® - Indice - resumen - Rentabilidad-Dividendo. (n.d.). https://www.economista.es/indice/EUROS_TOXX-50/resumen/Rentabilidad-Dividendo. Accessed on 14 July 2023
- Delmas MA, Burbano VC (2011) The drivers of greenwashing
- Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, Možina M, Polajnar M, Toplak M, Starič A, Štajdohar M, Umek L, Žagar L, Žbontar J, Žitnik M, Zupan B (2013) Orange: data mining toolbox in python. *J Mach Learn Res* 14(August):2349–2353
- ESG & Sustainable Finance Data Skills and Capacity Building Directory. (2020). <https://initiatives.weforum.org/sustainable-finance-data-skills-and-capacity-building/home>. Accessed on 13 June 2023
- Fatemi A, Glaum M, Kaiser S (2018) ESG performance and firm value: the moderating role of disclosure. *Glob Financ J* 38:45–64. <https://doi.org/10.1016/J.GFJ.2017.03.001>
- Friede G, Busch T, Bassen A (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *J Sustain Financ Invest* 5(4):210–233. <https://doi.org/10.1080/20430795.2015.1118917>
- Geissdoerfer M, Savaget P, Bocken NM, Hultink EJ (2016) The circular economy—a new sustainability paradigm? *J Clean Prod* 143:757–768. <https://doi.org/10.1016/j.jclepro.2016.12.048>
- Initiative F (2005) UNEP FI 2005 overview.

- Kates RW, Parris TM, Leiserowitz AA (2005) What is sustainable development? Goals, indicators, values, and practice. *Environment* 47(3):8–21. <https://doi.org/10.1080/00139157.2005.10524444>
- Khoo CSG, Johnkhan SB (2018) Lexicon-based sentiment analysis: comparative evaluation of six sentiment lexicons. *J Inf Sci* 44(4):491–511. <https://doi.org/10.1177/0165551517703514>
- Kim Y (2015) Convolutional neural networks for sentence classification. Master's thesis, University of Waterloo
- Kossovosky N (2012) Reputation, stock price, and you: Why the market rewards some companies and punishes others. In: Reputation, stock price, and you: why the market rewards some companies and punishes others, vol 9781430248. <https://doi.org/10.1007/978-1-4302-4891-0>
- Lee KH, Cin BC, Lee EY (2016) Environmental responsibility and firm performance: the application of an environmental, social and governance model. *Bus Strateg Environ* 25(1):40–53. <https://doi.org/10.1002/BSE.1855>
- Liu W, Liao H (2017) A Bibliometric analysis of fuzzy decision research during 1970–2015. *Int J Fuzzy Syst.* <https://doi.org/10.1007/s40815-016-0272-z>
- Liu M, Luo X, Lu W-Z (2023) Public perceptions of environmental, social, and governance (ESG) based on social media data: evidence from China. *J Cleaner Prod.* <https://doi.org/10.1016/j.jclepro.2022.135840>
- Mandas M, Lahmar O, Piras L, De Lisa R (2023) ESG in the financial industry: what matters for rating analysts? *Res Int Bus Financ.* <https://doi.org/10.1016/j.ribaf.2023.102045>
- Mäntylä MV, Graziotin D, Kuutila M (2018) The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. *Comput Sci Rev* 27:16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>
- Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J* 5(4):1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Mendonça AMB, Leal Filho W, Alves F (2023) Written Press's approach to climate change in the autonomous region of madeira and the autonomous community of the canary islands. *Clim Change Manag Part F5*:459–474. https://doi.org/10.1007/978-3-031-28728-2_22
- Nardi L (2022) The corporate social responsibility price premium as an enabler of substantive CSR. *Acad Manag Rev* 47(2):282–308. <https://doi.org/10.5465/AMR.2019.0425>
- Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. <http://tumblr.com>. Accessed on 29 July 2023
- Pang B, Lee L (2008) Opinion mining and sentiment analysis. *Found Trends Inf Retr* 2(2):1–135
- Park J, Choi W, Jung S-U (2022) Exploring trends in environmental, social, and governance themes and their sentimental value over time. *Front Psychol.* <https://doi.org/10.3389/fpsyg.2022.890435>
- Porter ME, Kramer MR (2006) Strategy & society: the link between competitive advantage and corporate social responsibility. *Harv Bus Rev* 84(12):78–92. <https://doi.org/10.1108/sd.2007.05623ead.006>
- Raithel S, Wilczynski P, Schloderer MP, Schwaiger M (2010) The value/relevance of corporate reputation during the financial crisis. *J Prod Brand Manag* 19(6):389–400. <https://doi.org/10.1108/10610421011085703>
- Salas-Zárate MDP, Valencia-García R, Ruiz-Martínez A, Colomo-Palacios R (2017) Feature-based opinion mining in financial news: an ontology-driven approach. *J Inf Sci* 43(4):458–479. <https://doi.org/10.1177/0165551516645528>
- Savyt'ska L, Vnukova N, Bezugla I, Pyvovarov V, Turgut Sübay M (2021) Using Word2vec technique to determine semantic and morphologic similarity in embedded words of the Ukrainian language.
- Skublov SG, Gavrilchik AK, Berezin AV (2022) Geochemistry of beryl varieties: comparative analysis and visualization of analytical data by principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). *J Min Inst* 255(3):455–469. <https://doi.org/10.31897/PMI.2022.40>
- Strauß N (2022) Covering sustainable finance: Role perceptions, journalistic practices and moral dilemmas. *Journalism* 23(6):1194–1212. <https://doi.org/10.1177/14648849211001784>
- Tunca S, Sezen B, Balcioglu YS (2023) Content and sentiment analysis of the New York times coronavirus (2019-nCoV) articles with natural language processing (NLP) and leximancer. *Electronics* 12(9):1964. <https://doi.org/10.3390/ELECTRONICS12091964>
- Visualizing Data using the Embedding Projector in TensorBoard | TensorFlow. (2022) https://www.tensorflow.org/tensorboard/tensorboard_projector_plugin?hl=en. Accessed on 10 July 2023
- Zeidan R (2022) Why don't asset managers accelerate ESG investing? A sentiment analysis based on 13,000 messages from finance professionals. *Bus Strategy Environ* 31(7):3028–3039. <https://doi.org/10.1002/bse.3062>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.