7-12-2021

Are My Stocks Sustainable? Design Principles for Leveraging Information from Analyst Reports

Steffen Bankamp
University of Goettingen, steffen.bankamp@uni-goettingen.de

Jan Muntermann
University of Goettingen, muntermann@wiwi.uni-goettingen.de

Are My Stocks Sustainable? Design Principles for Leveraging Information from Analyst Reports

Completed Research Paper

Steffen Bankamp
University of Goettingen
Goettingen, Germany
steffen.bankamp@uni-goettingen.de

Jan Muntermann
University of Goettingen
Goettingen, Germany
muntermann@wiwi.uni-goettingen.de

Abstract

We address the problem of insufficient information about companies' sustainability, thereby helping investors to incorporate sustainability aspects into investment decisions. Building upon the design science research paradigm, we develop an artifact to extract information on companies' sustainability from text documents (analyst reports). We derive design principles that allow us to extract this information effectively and with a high degree of classification performance. The evaluation of the artifact shows that the proposed approach results in a precise extraction of sustainability-related information. Furthermore, this information is shown to be useful for supporting investors' decision-making.

Keywords: Sustainability, SRI, Design Science Research, Text Mining, Analyst Reports

Introduction

In 2015, the United Nations defined 17 Sustainable Development Goals (SDGs) whose implementation is intended to ensure global sustainable development at the economic, social, and ecological level (UN General Assembly 2015). Companies and their owners have a responsibility to contribute by their actions to achieve these goals. In recent decades, social and ecological indicators have become more important. Also, investors' awareness about sustainability issues has increased considerably (Flammer 2013). Consequently, there is strong growth in financial products that take sustainability issues into account, e.g., by excluding companies that do not fulfill their sustainability criteria (Global Sustainable Investment Alliance 2018). It is also apparent that fund managers rely increasingly on information about sustainability (Amel-Zadeh and Serafeim 2018; van Duuren et al. 2016). This has given rise to a new industry that conducts sustainability assessments and ratings and makes them available to investors. Simultaneously, more than 40% of institutional investors do not fully consider sustainability issues because it is too expensive to obtain and collect this information (Amel-Zadeh and Serafeim 2018). There are also data quality concerns with the Environmental, Social, and Corporate Governance (ESG) ratings produced by professional agencies (Kotsantonis and Serafeim 2019). These scores usually do not provide contextual information about incidents within the company. To make the economy more sustainable, investors can make an important contribution by directing capital to those companies that contribute to the achievement of the SDGs and, in return, withdrawing from unsustainable companies (Pástor et al. 2020). In addition to investors' willingness to invest in sustainable companies, the information base on which capital flows are allocated is crucial.

Against this background, our goal is to develop an artifact to extract sustainability-relevant information from analyst reports. These reports are prepared by financial analysts and distributed to investors. Financial analysts are important information intermediaries in the financial market and discuss a broad range of topics in their reports (Huang et al. 2018). Nilsson et al. (2008) show that analysts also discuss sustainability-related topics in their reports. However, this information is dispersed throughout the analyst reports, making a manual extraction time-consuming for investors. An automated extraction is necessary to make this information easily accessible to investors in an aggregated form. Due to the importance of analyst reports for investors, it can be assumed that these documents are already available to many investors, especially to institutional investors. The extracted information on sustainability should improve the decision-making of investors concerning the assessment of companies' sustainability. Hartzmark and Sussman (2019) show a shift in capital allocation from less sustainable assets to more sustainable assets when sustainability-related information is made easily accessible to investors. The provision of information is thus an elementary building block for achieving a more sustainable economy. To contribute to this goal, we follow the design science research (DSR) paradigm (Hevner et al. 2004) and apply the process model from Kuechler and Vaishnavi (2008) to develop the artifact. The artifact is evaluated with regard to the performance of classification and the informational value of the extracted information.

Research Background on Sustainability in Investing

The idea of taking ethical considerations into account when making investment decisions can be traced back hundreds of years. However, the strong growth in this field has just been observed for several decades and has increased with the general awareness of sustainability issues and with past environmental disasters (Schueth 2003). This type of investment is often called socially responsible investing (SRI) in the financial literature (e.g., Kempf and Osthoff 2007; Nofsinger and Varma 2014). Contrary to what the term suggests, not only social aspects are considered, but also other ethical aspects like environmental issues. According to Schueth (2003), there are three strategies an investor can choose from to implement SRI. First, screening is a strategy by which the investor reduces the investment choice set based on her ethical values. This happens most commonly through negative screening (Amel-Zadeh and Serafeim 2018), where specific companies or industries are excluded that do not fulfill the investor's minimum standards (van Duuren et al. 2016). With positive screening, the investor focuses specifically on sustainable companies or industries (van Duuren et al. 2016). The second strategy proposed by Schueth (2003) is shareholder advocacy. Investors are applying this strategy by influencing the management's decision-making (e.g., through their voting rights at the annual general meeting) to make the company more sustainable. The third strategy from Schueth (2003) is community investing, in which investors provide capital to weaker communities, thereby enabling the financing of low-income housing and small businesses. One major strand of literature within the field of SRI has analyzed the relationship between companies' financial and sustainability performance. Friede et al. (2015) were able to identify over 2,000 studies about this question. Overall, they found a positive correlation between financial and sustainability performance.

Following Hartzmark and Sussman (2019), investors prefer sustainable assets and react to new information regarding assets' sustainability by redirecting their capital. This is in line with the findings of Amel-Zadeh and Serafeim (2018), who found that even 82% of fund managers of conventional (non-sustainable) funds state to consider sustainability aspects in their investment decision. Pástor et al. (2020) show by means of an equilibrium model that these financial investment decisions have a positive impact, as real investments are shifted from non-sustainable to sustainable firms, which makes the economy more sustainable. At the end of 2019, sustainable assets accounted for 15.1% of the total assets held by mutual funds in Europe and are expected to increase to 41-57% by 2025 (PwC 2020). Thus, SRI represents a significant share of the capital market and has large growth prospects.

Methodology and Problem Description

Design Science Research

In order to mitigate the challenge of insufficient data on companies' sustainability (PwC 2020), we build upon the design science research paradigm (Hevner et al. 2004). In DSR, a solution for a problem is developed based on the current state of knowledge (e.g., theories, frameworks, methods). The artifact is developed during the DSR process and "*extend[s] the boundaries of human problem solving and* *organizational capabilities*" (Hevner et al. 2004, p. 76). DSR can contribute to existing knowledge by providing constructs, models, methods, instantiations, and design theories (Gregor and Hevner 2013).

In this study, we develop an instantiation that can be classified as improvement research according to (Gregor and Hevner 2013). The artifact extends the problem class (sustainability data provision) by accessing a previously unexploited source of information on companies' sustainability. According to the belief-action-outcome framework of Melville (2010), our artifact should lead to action formation because it is intended to enable investors to consider sustainability issues during their investment-related decision making.

Research Process

We utilize the DSR process model proposed by Kuechler and Vaishnavi (2008). Starting with the *awareness of the problem*, we derive the problem of insufficient sustainability data from the existing literature. Building upon this, we identify three design requirements (DR) related to a potential problem solution. In the second process step (*suggestion*), we propose three design principles (DP). We derive concrete design features (DF) from our design principles and implement them by building the IT artifact in the *development* step. In the fourth step, the *evaluation* of the artifact is conducted in two stages. First, we evaluate the classification performance of the artifact based on different quantitative performance metrics. Finally, we qualitatively evaluate the usefulness of the extracted information for investors. In the final process step (*conclusion*), the acquired design knowledge is summarized and future research opportunities are presented.

Problem Description

Achieving the SDGs is a major task for society. As explained before, investors take sustainability aspects into account. This sustainability-guided allocation function of capital also leads to real investments (e.g., production facilities) being shifted from less sustainable to more sustainable companies (Pástor et al. 2020). These real investments can contribute to the achievement of the SDGs. However, successful SRI relies on available data on companies' sustainability. Only reliable and comprehensive data ensures an effective allocation of capital flows towards more sustainable companies and, in turn, into more sustainable real investments.

However, data about the sustainability of companies poses a significant problem. PwC (2020) found that for 73% of the surveyed asset managers, the lack of data is the largest barrier to implement sustainable products. Also, the high cost of aggregating sustainability information is a major obstacle for taking environmental data into account (Amel-Zadeh and Serafeim 2018). There are two major sources for ESG information. First, self-disclosures made by the companies in the form of Corporate Social Responsibility (CSR) reports, and second, ratings made available by specialized ESG rating agencies. Companies' self-reported data are less objective as negative aspects are inadequately reported (Chauvey et al. 2015). This data is found to be skewed and inaccurate, which questions the reliability and presume greenwashing (PwC 2020). Institutional investors criticize the nonspecificity of this information for using it in a targeted manner (Amel-Zadeh and Serafeim 2018). Designated ESG ratings suffer from data quality issues (Kotsantonis and Serafeim 2019). Also, sustainability principles as "Life-Cycle-Thinking" have not been integrated into ESG Rating agencies' assessments (Escrig-Olmedo et al. 2019). Simultaneously, there is a large dispersion between the assessment results of different rating agencies on the same company (Berg et al. 2020; Dimson et al. 2020; Kotsantonis and Serafeim 2019). This can be in particular attributed to different measurement methods. Berg et al. (2020) call for more transparent ESG ratings regarding their measurement methods. According to PwC (2020, p. 36), "*Traditional ESG data and ESG scoring will no longer suffice*." Dimson et al. (2020) argue that ESG ratings should not be applied blindly but supplemented by the asset manager's own review.

Artifact Design

Design Requirements

In order to support the decision-making of investors in the selection of sustainable companies, we propose analyst reports as an additional source of information that can supplement the existing information sources (CSR reports and ESG ratings). Analyst reports provide important information to investors. They deliver financial analyses about public companies and contain investment recommendations. Financial analysts take the important role of information intermediaries on the capital market (Huang et al. 2018). The reports provided by analysts focus on financial and business aspects of the company. This can be seen from the fact that most of the analysts' reports are published in a narrow timeframe surrounding the quarterly conference calls and earnings announcements (Huang et al. 2018). Nevertheless, Nilsson et al. (2008) show that sustainability aspects are also discussed in analyst reports. It seems appropriate to leverage this information source as well. Compared to the two information sources described above, analyst reports have three advantages. First, unlike the CSR reports, the analyses are not written by the company itself but by a third party, enhancing objectivity. Second, in contrast to ratings from specialized ESG agencies, the analyst reports do not have to be acquired additionally but should already be available to most institutional investors as it is a common source for financial decision-making. Third, the analyst reports discuss sustainability aspects in textual form, which, in comparison to the frequently used quantitative rankings, allows investors to make informed decisions, considering their individual ethical principles and guidelines, and can thus better argue their decision to end-investors or other stakeholders.

As analyst reports cover primarily financial topics, the sustainability topics have to be extracted. Nilsson et al. (2008) found that for companies operating in the oil/gas and chemical industry 35% of the analyst reports contain environmental information. In other industries (e.g., semiconductor or telecommunication), the proportion is lower (Cerin 2010). There is also relatively little environmentally relevant information within a single report (Nilsson et al. 2008). Since sustainability-relevant topics seem to be only sporadically present in analyst reports, and this information has to be gathered across a large universe of companies, especially if negative screening (van Duuren et al. 2016) is conducted, *automated and precise extraction of sustainability-relevant information* is necessary. This builds DR1.

The sporadic content on sustainability (Nilsson et al. 2008) is embedded in a much larger part of financial information (Huang et al. 2018). As a result, there is a significant class imbalance in analyst reports between sustainability-related content and other content. Therefore, the related problem solution must have the *ability to handle extremely imbalanced datasets,* which represents DR2.

Based on the literature, we have identified that in today's ESG ratings, divergence among agencies is a major problem (Berg et al. 2020). At the same time, 43.2% of fund managers say that a lack of standards prevents them from considering ESG-related information in their investment decisions effectively (Amel-Zadeh and Serafeim 2018). Therefore, we consider DR3 elementary, as the extraction of information has to be *based on a common understanding of sustainability*.

Design Principles

Based on the three design requirements discussed above, we derive design principles and link them with related requirements (DRn→DPn). In order to ensure the automated extraction of sustainability-relevant information required by DR1, various methods are conceivable. First, it can be done based on dictionaries (wordlists). A text section is classified as sustainability-relevant if it contains one or more words from the dictionary. For example, the studies of Nilsson et al. (2008) and Cerin (2010) used predefined keywords to search for environmentally relevant content in analyst reports. Dictionary-based methods are also frequently used for sentiment analysis in the finance and accounting literature, where sentiment is extracted from texts (Loughran and McDonald 2016). The advantages of such a dictionary are that it can be applied straightforward once it has been created, it can be easily transferred to other documents, and it is replicable (Albaugh et al. 2014). A disadvantage of dictionaries is that it only recognizes the exact terms it contains. Dictionaries are not able to recognize synonyms unless they are already included in the dictionary. It is therefore particularly promising if there is a rigid terminology for the corresponding topic (Albaugh et al. 2014). Second, machine learning methods such as a support vector machine or a neural network can be applied for a classifier. The training of the classifier is conducted on a labeled dataset of text sections containing sustainability-relevant information and text sections not containing this information. A trained classifier can then be applied for inference, classifying unlabeled text sections automatically. Huang et al. (2014) showed for the analyst domain that a sentiment classifier based on machine learning (naïve Bayes) is more accurate than a dictionary-based classification. However, a sufficiently large training dataset is required for training the classifier. Due to the expected class imbalance (DR2), it is questionable whether it is possible to manually label enough text sections containing sustainability-relevant information within a reasonable amount of time. In addition, DR3 requires that the sequence extraction is built upon a common understanding of sustainability. This can be achieved more adequately by using a domain-specific dictionary. In order to provide high extraction performance (DR1) and to address DR2 and DR3, we propose a combination of these two approaches. Related to all design requirements (DRn→DP1), we derive DP1, according to which a *hybrid approach that combines a dictionary with machine learning* should be followed. Eickhoff (2015) shows that dictionary-based and machine learning-based classifiers can be successfully combined in a hybrid approach.

To achieve a precise extraction (DR1), not only the sole word count as in bag-of-words models should be considered, but also the textual context. The finance and sustainability domain share many words that have different meanings. For example, the word “disposal” is commonly used in finance contexts to refer to the sale of parts of a company. In the context of sustainability, it instead refers to the discharge of waste. To address this problem (DR1→DP2), we are *considering the context in textual data* (DP2).

To further address DR3, we derive another design principle (DR3→DP3). To increase the generalizability, we rely on existing knowledge from SRI and from frameworks for sustainability reporting that we consider as kernel theory informing our artifact design. This ensures that the extraction of text sections is grounded on a broadly accepted understanding of sustainability and that the results are thus accepted by a majority of users. DP3 is the *theory-based design of the artifact that utilizes existing frameworks and knowledge from sustainability reporting*.

Design Features

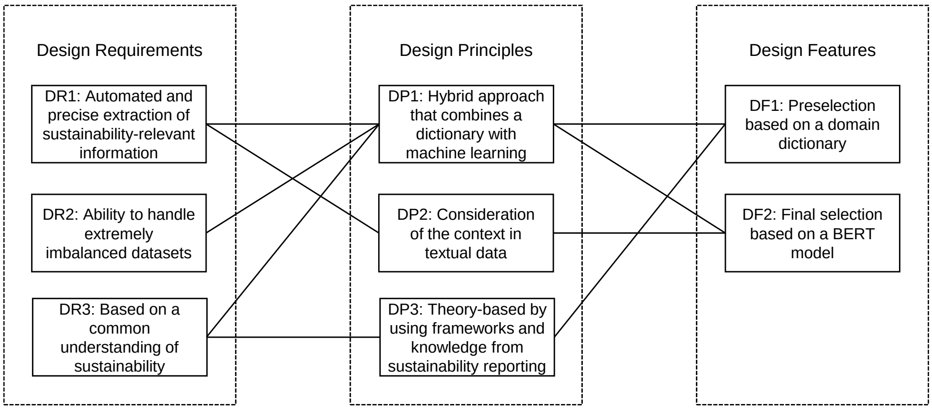

Figure 1. Design Requirements, Principles, and Features

Based on the three design principles, we derive and link artifact-related design features (DPn→DFn). The mapping between design requirements, principles, and features is illustrated in Figure 1. Design features describe the specific technical design of the artifact. This distinguishes design features from design principles, which describe the artifact properties at a higher level of abstraction (Meth et al. 2015). Incorporating the design features into the artifact should enable the artifact to fulfill the design requirements. As the first design feature (DP1/3→DF1), we define the *preselection of text passages using a domain dictionary*. We select the sentence level as the granularity of the text passages. A sentence is included in the preselection if this sentence contains at least one word or n-gram of the dictionary. The selection of the dictionary is an important parameterization of the artifact. A review of the knowledge base shows that two sustainability dictionaries have been developed. Deng et al. (2017)

present a dictionary for environmental sustainability within the IT industry. However, since our approach is to develop an artifact that is not restricted to a specific industry, this dictionary does not fit. Pencle and Mălăescu (2016) have developed a dictionary on four CSR dimensions (employee, environment, human rights, and social community). This dictionary is based on a deductive (derived from literature) and an inductive (derived from IPO prospectuses) approach. The dictionary is general in terms of the industry but is also tied to the finance literature. Therefore, it seems to be adequate in the context of our problem class. This design feature (DF1) is the first module of the hybrid approach (DP1) and draws on existing sustainability-related knowledge (DP3).

To increase the precision of the sequence extraction, a *final selection of sequences is conducted by applying a Bidirectional Encoder Representations from Transformers (BERT) language representation model* (DP1/2→DF2). This approach replaces the manual evaluation done by Nilsson et al. (2008) following the keyword search. For this purpose, we build a binary classifier based on the BERT model (Devlin et al. 2019). This model consists of a deep neural network that is unsupervised pre-trained on the English Wikipedia and a large dataset of books (Devlin et al. 2019). The pre-training is based on a cloze task and a next sentence prediction. For the cloze task, the model is trained to predict masked tokens within a sentence. For this prediction, the entire sentence (except for the masked token) is available to the model, which is why it is called a bidirectional model. This is an important difference from so-called unidirectional architectures, where a word is predicted only from the preceding or following tokens. This allows the model to learn the entire context of a word. In addition, the next-sentence prediction is used in pre-training, where the model is trained to predict whether a chain of two sentences consists of consecutive sentences or not. The pre-training reduces the computational effort for task-specific training (fine-tuning) substantially (Devlin et al. 2019). In fine-tuning, a classification layer is added to the model. With a binary classification problem (sustainable/non-sustainable), this layer will have two outputs. In fine-tuning, the model is trained for the specific task based on the labeled dataset. However, with the BERT-architecture, not only the weights of the classification layer are adjusted, but also those of the entire model.

The BERT-model is suitable for many text mining tasks, including sentence classification making it suitable for the problem at hand. In contrast to bag-of-words models, it also considers the sequence within a sentence and contextual information (DP2). Furthermore, based on BERT, significantly better results on text mining tasks could be achieved than by prior methods (Devlin et al. 2019). We consider it useful to use BERT for the final selection of the preselected sentences and thus to be the second module of the hybrid approach (DP1).

Artifact Evaluation

Dataset

To evaluate the artifact, we use a comprehensive dataset of analyst reports. As a company sample, we select all companies of the Dow Jones Industrial Average (major US index) and the EuroStoxx 50 (major European index) that have been a constituent at any time during our investigation period ranging from 01-01-2015 to 12-31-2019. This results in a sample of 90 companies. This sample includes companies across a wide range of industries, including chemical and gas companies, for which a relatively large amount of information on environmental issues has been found by prior research in analyst reports, as well as companies in the telecommunications industry, for which significantly less information has been found (Cerin 2010). Both the selection of two large capital markets and the large variety of industries increase our study's generalizability. We collect all analyst reports available from Refinitiv Thomson ONE about the 90 companies during the investigation period, resulting in a sample of 95,665 reports. To clean the dataset, we remove duplicates, automatically generated analyst reports, analyst reports with more than 50 pages (typically industry analysis) and short updates with less than 300 words. This leaves 61,592 analyst reports as final sample. Table 1 shows the ten companies with the most reports of the two stock indices as well as the ten brokers who published the most reports. From the analyst reports (PDF files), we extract the text and remove charts, tables, diagrams, and boilerplate such as disclaimers. Since information extraction is done at the sentence level, the text is split into sentences. This results in 3,410,598 sentences that are used for further analysis. To develop and evaluate the instantiated artifact, we limit the scope to the extraction of information regarding the environmental dimension of sustainability. According to Hartzmark and Sussman (2019), the environmental dimension is most strongly associated with sustainability (79%). Despite the multifaceted nature of sustainability, we therefore consider it appropriate to first focus on the environmental dimension. For this reason, we use only the environmental word list of Pencle and Mălăescu (2016). This consists of 451 entries, including 323 uni-grams, 114 bi-grams, 10 tri-grams, and four entries consisting of more than three words.

Table 1. Top Companies and Brokers within the Sample

Top 10 Europe		Top 10 US		Top 10 Broker	
Company	N Reports	Company	N Reports	Broker	N Reports
SAP	1,156	Apple	2,085	JP Morgan	5,323
Volkswagen	1,041	Intel	1,380	Morgan Stanley	4,323
AB InBev	914	Boeing	1,188	Morningstar	3,960
ASML	820	Walmart	1,177	UBS	3,755
Nokia	805	Microsoft	1,177	Deutsche Bank	3,682
Bayer	801	Caterpillar	1,171	RBC	3,293
Sanofi	767	Cisco Systems	1,156	Barclays	3,018
Airbus	760	Johns. & Johns.	1,105	Société Générale	2,819
Daimler	753	General Electric	1,060	Credit Suisse	2,689
Telefonica	741	Merck & Co	954	Jefferies	2,480

Evaluation Results

In the first step, we evaluate whether the hybrid approach is even necessary. Alternatively, a training dataset could be labeled directly and the classifier trained based on this training dataset. For this purpose, 1,000 randomly selected sentences of the whole corpus were manually labeled. The labeling is based on the Global Reporting Initiative (GRI) framework (GRI 2020), which is the globally dominant standard in sustainability reporting (KPMG 2020). Many companies apply the GRI framework when preparing their sustainability report. The GRI Standards can be divided into series. The GRI 100 series define basic universal standards. The GRI 200, GRI 300, and GRI 400 series contain topic-specific standards. The GRI 300 series deals with environmental aspects, while GRI 200 focuses on economic and GRI 400 on social issues. Each series contains a set of standards. GRI 302-3, for example, defines the reporting of the organizations' energy intensity (GRI 2020). We classify a sentence as environmentally relevant if it contains an aspect of the GRI 300 series. Labeling this way ensures that the allocation is based on a common understanding of sustainability. Only three of the 1,000 randomly labeled sentences contained sustainability information. If this proportion of 0.3% ($CI_{0.95} = [0.11\%; 0.87\%]$) corresponds to the true proportion in the overall sample, over 33,000 sentences would have to be labeled to obtain a sample containing 100 environmentally relevant sentences. Even this training dataset with 100 sentences of the minority class would still be relatively small. To solve this problem and generate a training and validation dataset more efficiently, we consider the hybrid approach to be appropriate. The hybrid approach is illustrated in Figure 2, which shows the process of prototype development (i.e., artifact instantiation) and evaluation.

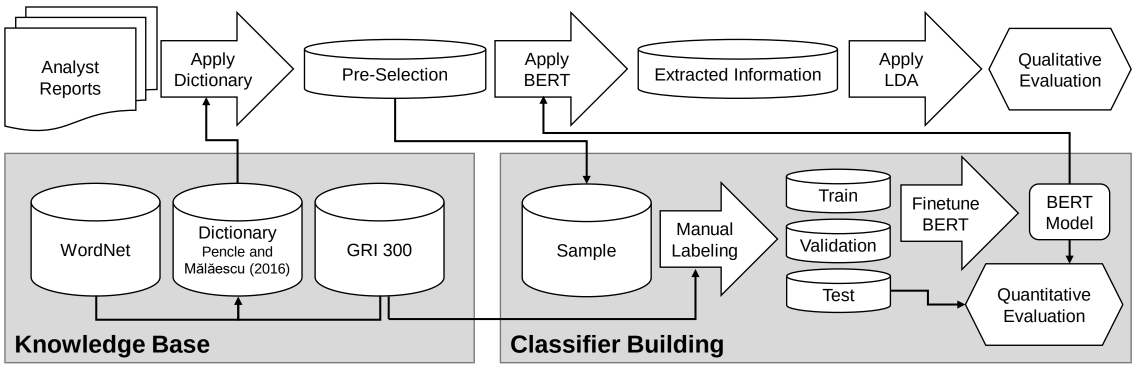

Figure 2. Process Map for Prototype Development and Evaluation

Pencle and Mălăescu (2016) developed their dictionary to evaluate the level of CSR language in financial documents. It is therefore not explicitly designed to use for the task of information extraction. For this reason, we adapt the dictionary to increase both sensitivity and precision in recognizing sentences with environmental content. To increase sensitivity, we first add 70 expressions (e.g., "fuel consumption") that are central within the GRI 300 standards but are not part of the Pencle and Mălăescu (2016) dictionary. By adding relevant words to the dictionary, the likelihood of sustainability-related sentences being recognized by the dictionary is increased, which improves sensitivity. We further extend the dictionary by adding synonyms and lemmas of all elements from the dictionary by using the lexical database WordNet (Fellbaum 1998). This results in a dictionary containing 1,941 entries. To increase precision, we then exclude words or n-grams that cannot be directly assigned to the standards of the GRI 300 series. The excluded words are very generic (e.g., "accept", "design" or "grow") and are therefore not suitable for an information extraction system with a high degree of precision. By removing these generic words, the probability of false-positive results is reduced. This in turn improves precision. The remaining dictionary consists of 402 words or n-grams, where 341 (84.83%) entries can be mapped to a single standard within the GRI 300 series (e.g., "energy efficiency" → GRI 302). The remaining 61 entries have a meaning that covers several of the sub-standards.

We apply the dictionary to the whole dataset. 65,848 of the total 3,410,598 sentences (1.93%) are preselected by the dictionary. These preselected sentences contain at least one word from the dictionary. We then manually labeled 4,000 randomly selected sentences of the preselection. 752 (18.8%) sentences were identified as environmentally relevant. This shows the suitability of the preselection and that, despite less labeling, a relatively large training dataset in terms of the positive class could be obtained. The preselection rate (1.93%) and the positive rate in the preselection (18.8%) result in an expected value for the total extraction of 0.364%. Since this value is above the positive rate of 0.3% found in the initially labeled sample of 1,000 sentences from the population, the preselection does not lead to a substantial decrease of sensitivity.

The 4,000 labeled sentences are randomly split into a training (N=2,400), validation (N=600), and test dataset (N=1,000). The training dataset is used to fine-tune the BERT model. As this is done in epochs, the validation dataset is used to stop the learning process when the highest predictive performance based on the validation dataset is achieved. This procedure prevents the model from overfitting.

Based on the test dataset, the BERT model achieves a precision of 0.7573 (75.73% of the sentence classified as environmentally relevant are actually environmentally relevant) and a recall of 0.8342 (83.42% of sentences within the preselection that are environmentally relevant are classified accordingly). Different precision-recall combinations of the classifier (solid black line) and receiving operator characteristics (ROC) can be derived from Figure 3.

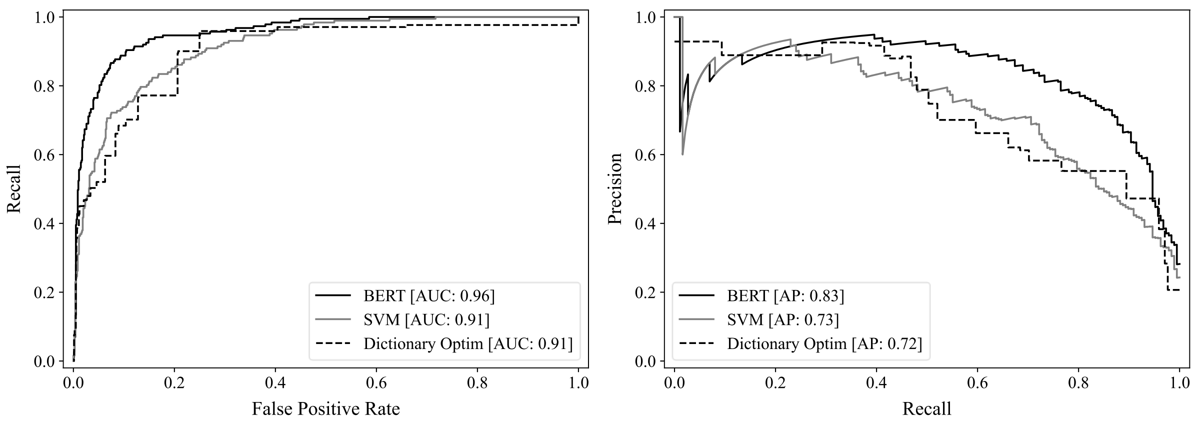

Figure 3. ROC and Precision-Recall Curves

To contextualize these performance measures, we compare the developed artifact with two alternative approaches. The first approach is the optimization of the dictionary. Instead of using a relatively unspecific but sensitive dictionary for preselection and increasing the precision through the machine learning module, only a dictionary is used, which is optimized based on the training and validation dataset (N=3,000) to increase its precision. For this purpose, we calculate for each word or n-gram of the dictionary how often it occurs in non-environmentally relevant sentences and how often it occurs in environmentally relevant sentences. The ratio between these two counts is calculated. A high ratio means, that the word or n-gram frequently occurs in non-environmentally-relevant sentences. These entries with a particularly high ratio could lead to many false-positives. Subsequently, the entries with the highest ratio (high occurrence in non-relevant sentences) are removed from the dictionary step by step. For evaluation, the sentences from the test dataset are classified based on this optimized dictionary. A sentence is considered environmentally relevant if it contains at least one of the entries. The performance of this benchmark is shown in Figure 3 by the dashed line (Dictionary Optim). To achieve the same recall as with the hybrid approach (0.8342), the precision must be reduced to 0.4722. This benchmark shows that applying a hybrid approach is indeed necessary and that it is not sufficient to reduce the entries in the dictionary until the precision has reached the desired value. As a second benchmark, we use a support vector machine trained on sentence-level averaged GloVe word embeddings (Pennington et al. 2014), represented by the grey line in Figure 3. It has been shown that the use of these word embeddings can improve supervised NLP systems (Pennington et al. 2014). The computational effort to train an SVM based on GloVe embeddings is significantly lower than fine-tune the BERT model on sentence classification. We chose the SVM because it is particularly suitable for high-dimensional and imbalanced classification problems (Goudjil et al. 2018), which is in line with our requirement (DR2). This benchmark is used to evaluate whether it is necessary to use such a complex model as BERT. The achievable precision is 0.5324 if a minimum value for the recall is set to 0.8342. The area under the curve (AUC) and the average precision (AP) show that the BERT model is superior to both benchmarks for this application.

After it has been shown that the artifact can extract environmentally relevant sentences from analyst reports with adequate recall and precision, we evaluate whether the extracted sentences are relevant for investors. For this purpose, we extract all environmentally relevant sentences from the corpus utilizing the artifact and identified 12,884 sentences as environmentally relevant. The preselection contains 65,848 sentences. Figure 4 shows a word cloud based on the whole corpus on the left-hand side and based on environmentally relevant sentences on the right-hand side. This gives a first impression of the content discussed in the extracted sentences.

Figure 4. Word Cloud from the Full Corpus and the Extracted Sentences

To analyze the topic structure of these sentences in more detail, we apply Latent Dirichlet Allocation (LDA) and develop a topic model (Blei et al. 2003). This methodology allows us to find topics that are discussed across different documents. Based on the coherence score and a visual inspection of varying topic models, we set the number of topics to 10, which is the central hyperparameter in LDA. Each sentence is considered to be a separate document. Table 2 shows the 20 most important words per topic. Based on these top words, we have assigned a label to each topic that describes it.

Table 2. Top Words of Topics from Extracted Sentences

Topic: Label	Top 20 Words per Topic
1: CO_2 target	target, company, reduction, carbon dioxide emission, management, give, additional, current, achieve, follow, performance, meet, require, recent, line, reach, policy, regard, deliver, highlight
2: Expansion of renewables	increase, year, capacity, total, offshore wind, order, estimate, asset, addition, add, mw, unit, offshore, revenue, share, expect, turbine, strong, portfolio, farm
3: Efficiency of products	emission, cost, reduce, sale, level, change, carbon dioxide, term, start, fuel, assume, test, challenge, cut, show, standard, ahead, close, fleet, peer
4: Fuel and emission market	market, high, price, continue, demand, due, carbon, low, rise, remain, drive, decline, coal, offset, mix, margin, positive, trend, volume, expect
5: Regulatory	risk, result, potential, lead, view, regulation, future, key, time, benefit, opportunity, sector, industry, average, return, government, move, point, exist, company
6: Legal issues	impact, car, model, relate, include, issue, sell, fine, environmental, number, major, hybrid, provision, range, face, state, pay, launch, concern, brand
7: Clean products and production	technology, product, production, improve, customer, solution, waste, build, efficiency, offer, process, source, great, good, water, work, consumer, building, produce, aim
8: Diesel scandal	month, vehicle, diesel, make, report, system, european, engine, software, limit, announce, accord, case, german, control, full, investigation, measure, fix, today
9: Power generation	wind, power, energy, solar, project, gas, large, global, electricity, service, plant, supply, base, operation, generation, country, engie, construction, world, power generation
10: Power grid investments	renewable, growth, business, investment, network, focus, plan, grow, strategy, capex, grid, generation, area, segment, activity, period, exposure, group, main, represent

In addition, we provide for each of the ten topics an example sentence in Table 3. The sentences have been chosen based on the probability for the respective topic assigned by LDA. Overall, it can be seen that CO_2 emissions and the resulting climate change play an important role in the extracted text sections. Topic 1 deals with it directly. Jonson et al. (2019) discuss the CO_2 reduction targets of the analyzed company (see topic 1 in Table 3). Topics 2, 3 and 4 are also related indirectly to the problem of greenhouse gas emissions. Topic 2 focuses on specific projects to reduce emissions. These are not only investments in wind or solar farms of large energy providers. Ferry and Letzeler (2018) for example reported that a brewery shifted towards an own renewable energy supply (see topic 2 in Table 3). The third topic thematized the energy efficiency and emissions of products. This is particularly noticeable in analyst reports on automotive manufacturers. This reveals another important characteristic of the environmental relevance of information for investors. Ellinghorst et al. (2019) link the environmental aspects with financial indicators (see topic 3). This link is perceived as very important by investors (IIRC & Kirchhoff 2020). Topics 5 and 6 deal with environmental regulation and legal disputes. Aguilar (2018) discusses a court settlement over a penalty for pollution. Again, this shows the linkage of environmental and financial issues. Topic 7 looks at transformation steps towards more sustainable production and product design. The focus is not only on CO_2 emissions but also on waste reduction and water consumption. Vasilescu et al. (2018) describe in their analyst report a production concept that reduces waste. Topic 8 summarizes statements related to the VW emissions scandal. As this had significant consequences for VW and was perceived worldwide, it is discussed extensively in many analyst reports. Topic 9 contains information about energy generation and its sources while topic 10 deals with grid investment projects. Here, for example, Mackie (2015) points out in his analyst report that growth opportunities for smart grids arise for the analyzed company due to the growing share of renewable energy.

Overall, it can be seen that the developed artifact can be useful for extracting targeted information on sustainability. These text fragments cover a broad range of environmental issues and link sustainability with financial indicators and are therefore highly relevant for investors. The information extracted by the artifact can provide an additional source of information to the quantitative ESG ratings and the self-reported information from CSR reports.

Table 3. Per-Topic Examples of Environmental Sentences

Topic	Example Sentences per Topic
1	*"On track to achieve its current target of a 25% reduction in carbon by 2020, in the coming months CRH will update the market on its ESG objectives putting out targets for both 2025 and 2030."* (Jonson et al. 2019, p. 1)
2	*"Renewable energy: ABInBev has finalised a purchase agreement for on-site solar equipment at five South African breweries, which equates to 10% of South African electricity requirements*." (Ferry and Letzeler 2018, p. 34)
3	*"Based on the industry's 2018 CO2 footprint, we estimate it will cost an aggregate €15.1bn to comply, assuming a €60 cost per gram to reduce CO2 emissions for premium carmakers and €40 for volume OEMs (our discussions with companies suggest that this is a reasonable rule of thumb)."* (Ellinghorst et al. 2019, p. 14)
4	*"Mining is undergoing significant distress due to coal's declining cost-competitiveness relative to unconventional natural gas production and more carbon emissions legislation."* (Schoonmaker 2018, p. 9)
5	*"There are four ways in which the new President could benefit the industry: […] 4) new measures to improve energy efficiency could benefit Saint-Gobain."* (Gardiner 2017, p. 2)
6	*"There was also the February legal settlement with the state of Minnesota that amounted to a pretax charge of $897 million, inclusive of legal fees and other obligations, related to its natural resource damages lawsuit concerning certain perfluorocarbons present in the environment."* (Aguilar 2018, p. 15)
7	*"To reduce waste, much of Nike's focus over the past few years has been on additive versus deductive manufacturing."* (Vasilescu et al. 2018, p. 6)
8	*"The agencies accused the company of deliberately manipulating through software algorithms in roughly 428,000 diesel-equipped vehicles, the activation of anti-pollution controls during emissions tests only."* (Hilgert 2015, p. 17)
9	*"Last year, in Texas >70m MWh of power was generated from renewable sources, enough to power 2mln homes for an entire year, and today AL announced it has signed an agreement for the supply of 50MW of renewable power, which will come on line at the end of 2020."* (Walsh et al. 2018, p. 2)
10	*"In contrast, rising penetration of renewable energy generation in North America, Europe and Asia has supported growth and returns for Medium-Voltage products, Energy Automation and Smart Grid Solutions and Services, segments where we see 4% compound growth potential."* (Mackie 2015, p. 15)

Discussion

Our study shows that the proposed design principles and design features are suitable for extracting sustainability-relevant information from analyst reports. This contributes to solving the lack of meaningful and reliable information on corporate sustainability. The hybrid approach that combines a dictionary with a state-of-the-art machine learning model is a central solution component of the developed artifact. The dictionary used for preselection ensures a high level of sensitivity and at the same time a grounding on existing knowledge and generally accepted definitions of sustainability. Further, due to the extreme class imbalance, the preselection allows a comprehensive training and test dataset to be manually labeled in a reasonable amount of time. Through the machine learning component, a highly specific extraction is achieved. By taking the context into account, the model can deal with the frequent ambiguity of terms used in the sustainability and finance context.

Through a comprehensive evaluation of the classification performance and the content analysis of extracted sentences, it becomes apparent that the artifact can help investors to include sustainability aspects in their investment decisions. Because it provides investors with concrete and qualitative information, it supplements the often used and less transparent ESG scores. Unlike CSR reports, the extracted information is not self-reported and should therefore provide a closer look at critical aspects.

However, our proposed solution is also subject to some limitations. Only environmentally relevant topics were extracted during prototype development and evaluation. The other two dimensions of sustainability following GRI *economic* (GRI 200) and *social* (GRI 400) are not considered. The design principles and features developed should also be transferable to these, but a related replication has to be carried out first, as little can be said about the scope of this information in analyst reports. Thereby, future related research can contribute to more complete and mature knowledge and nascent design theory (Gregor and Hevner 2013).

Another limitation is the evaluation of sensitivity. Based on the sample, we estimate that 0.3% ($CI_{0.95} = [0.11\%; 0.87\%]$) of the records contain sustainability information. However, the confidence interval is relatively large. If the proportion of sustainability sentences is actually 0.87%, many sustainability-relevant sentences have been overlooked during the preselection and the sensitivity of the artifact would be very low. Due to a high degree of uncertainty about the underlying proportion of environmental sentences in the total population, a reliable statement about the sensitivity can only be made starting from the preselection.

The evaluation is also not yet exhaustive. Despite the comprehensive content analysis of the extracted sentences, an evaluation involving investment practitioners has to be carried out in order to assess the decision usefulness of the information obtained in a real-world scenario.

Conclusion

In this paper, we present a problem solution and related design principles for extracting relevant information on sustainability from analyst reports, which are available to many investors. In doing so, we address the obstacles that prevent investors from fully integrating sustainability considerations into their investment strategy. These obstacles are, among others, the lack of data, the costs of information gathering and the nonspecificity of self-disclosed information on sustainability (Amel-Zadeh and Serafeim 2018; PwC 2020). Building on a hybrid architecture combining a knowledge grounded preselection and a state-of-the-art machine learning model, the artifact can improve investors' information base. This should improve the decisions of investors concerning the sustainability assessment of companies. Better informed decisions will improve the allocation of capital, shifting real investments from non-sustainable to sustainable companies. As a result, it can contribute to a more sustainable economy and thus to the achievement of the SDGs.

We contribute to the literature by proposing design principles and features that address the specific problems of class imbalance and the need for a unified understanding of sustainability. The generalizability of our approach, however, allows theorists and practitioners alike to apply this approach to related problems. Especially against the background of ever-increasing data volume, information extraction will become more and more important. In addition to its use in investment practice, our artifact might be helpful for researchers conducting research on corporate sustainability or SRI. They can apply the artifact to extract sustainability-relevant information from financial documents.

Our work offers starting points for subsequent design science research. First, the artifact should be extended with respect to the dimensions *economy* and *social*. Second, a field study with investors should evaluate the extent to which the extracted information is useful for decision-making. Third, the functionality of the artifact should be further developed and corresponding design principles and features derived. There is particular potential for improvement in the presentation of the extracted sentences. One extension that should be discussed with potential users would be to display a snippet from the PDF document for each extracted sentence so that the investor can read the sentence in context and supplementing graphics and tables can thus also be considered.

References

Aguilar, J. 2018. "We think 3M is a wide-moat business but also think the market has recognized its earning power," Morningstar Equity Research.

Albaugh, Q., Soroka, S., Joly, J., Loewen, P., Sevenans, J., and Walgrave, S. 2014. "Comparing and Combining Machine Learning and Dictionary-Based Approaches to Topic Coding," 7th Annual Comparative Agendas Project (CAP) Conference, Konstanz, Germany.

Amel-Zadeh, A., and Serafeim, G. 2018. "Why and How Investors Use ESG Information: Evidence from a Global Survey," Financial Analysts Journal (74:3), pp. 87-103.

Berg, F., Kölbel, J. F., and Rigobon, R. 2020. "Aggregate Confusion: The Divergence of ESG Ratings," Working Paper. Retrieved 03-01-2021, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3438533

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," Journal of Machine Learning Research (3), pp. 993-1022.

Cerin, P. 2010. "Analysing the Environmental Content of Financial Analyst Reports by developing an ESG Framework that incorporates Business Opportunities and the Product Perspectives," Working Paper. Retrieved 03-01-2021, from https://econpapers.repec.org/paper/hhbsicgwp/2010_5f004.htm

Chauvey, J.-N., Giordano-Spring, S., Cho, C. H., and Patten, D. M. 2015. "The Normativity and Legitimacy of CSR Disclosure: Evidence from France," Journal of Business Ethics (130:4), pp. 789-803.

Deng, Q., Hine, M., Ji, S., and Sur, S. 2017. "Building an environmental sustainability dictionary for the IT industry," Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 950-959.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT 2019, pp. 4171-4186.

Dimson, E., Marsh, P., and Staunton, M. 2020. "Divergent ESG Ratings," The Journal of Portfolio Management (47:1), pp. 75-87.

Eickhoff, M. 2015. "Enabling Reproducible Sentiment Analysis: A Hybrid Domain-Portable Framework for Sentiment Classification," International Conference on Design Science Research in Information Systems, pp. 215-229.

Ellinghorst, A., Yoward, M., and McNally, C. 2019. "Time for the Shareholders," Evercore ISI.

Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., and Muñoz-Torres, M. J. 2019. "Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles," Sustainability (11:3), p. 915.

Fellbaum, C. 1998. WordNet: An electronic lexical database. Cambridge: MIT press.

Ferry, E., and Letzeler, J. 2018. "ABInBev CMD: more brains than brawn," Exane BNP Paribas Research.

Flammer, C. 2013. "Corporate Social Responsibility and Shareholder Reaction: The Environmental Awareness of Investors," Academy of Management Journal (56:3), pp. 758-781.

Friede, G., Busch, T., and Bassen, A. 2015. "ESG and financial performance: aggregated evidence from more than 2000 empirical studies," Journal of Sustainable Finance & Investment (5:4), pp. 210-233.

Gardiner, R. 2017. "Wendel places part of its stake, creates another overhang," Davy Research.

Global Sustainable Investment Alliance. 2018. "Global Sustainable Investment Review." Retrieved 03-01-2021, from http://www.gsi-alliance.org/wp-content/uploads/2019/06/GSIR_Review2018F.pdf

Goudjil, M., Koudil, M., Bedda, M., and Ghoggali, N. 2018. "A Novel Active Learning Method Using SVM for Text Classification," International Journal of Automation and Computing (15:3), pp. 290-298.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quarterly (37:2), pp. 337-355.

GRI. 2020. "Consolidated Set of GRI Sustainability Reporting Standards 2020." Retrieved 03-01-2021, from https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language/.

Hartzmark, S. M., and Sussman, A. B. 2019. "Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows," The Journal of Finance (74:6), pp. 2789-2837.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," MIS Quarterly (28:1), pp. 75-105.

Hilgert, R. 2015. "Volkswagen Update on Diesel Scandal Provides Little New Information; EUR 190 Fair Value Unchanged," Morningstar Equity Research.

Huang, A. H., Lehavy, R., Zang, A. Y., and Zheng, R. 2018. "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach," Management Science (64:6), pp. 2833-2855.

Huang, A. H., Zang, A. Y., and Zheng, R. 2014. "Evidence on the Information Content of Text in Analyst Reports," The Accounting Review (89:6), pp. 2151-2180.

IIRC & Kirchhoff. 2020. "Closing the gap: The role of integrated reporting in communicating a company's value creation to investors." Retrieved 03-01-2021, from https://integratedreporting.org/news/iirc-and-kirchhoff-consult-study-investors-rely-on-non-financial-information-corporate-reports-not-there-yet/

Jonson, G., Woolf, P., and Bala, S. 2019. "CRH Feedback From Sell-Side Dinner with Mgmt," Jeffries.

Kempf, A., and Osthoff, P. 2007. “The Effect of Socially Responsible Investing on Portfolio Performance,” European Financial Management (13:5), pp. 908-922.

Kotsantonis, S., and Serafeim, G. 2019. “Four Things No One Will Tell You About ESG Data,” Journal of Applied Corporate Finance (31:2), pp. 50-58.

KPMG. 2020. “The Time has Come: The KPMG Survey of Sustainability.” Retrieved 03-01-2021, from https://home.kpmg/xx/en/home/insights/2020/11/the-time-has-come-survey-of-sustainability-reporting.html

Kuechler, B., and Vaishnavi, V. 2008. “On theory development in design science research: anatomy of a research project,” European Journal of Information Systems (17:5), pp. 489-504.

Loughran, T., and McDonald, B. 2016. “Textual Analysis in Accounting and Finance: A Survey,” Journal of Accounting Research (54:4), pp. 1187-1230.

Mackie, W. 2015. “Power down,” Kepler Cheuvreux.

Melville, N. P. 2010. “Information Systems Innovation for Environmental Sustainability,” MIS Quarterly (34:1), pp. 1-21.

Meth, H., Mueller, B., and Maedche, A. 2015. “Designing a requirement mining system,” Journal of the Association for Information Systems (16:9), pp. 799-837.

Nilsson, H., Cunningham, G. M., and Hassel, L. G. 2008. “A study of the provision of environmental information in financial analysts’ research reports,” Sustainable Development (16:3), pp. 180-194.

Nofsinger, J., and Varma, A. 2014. “Socially responsible funds and market crises,” Journal of Banking & Finance (48), pp. 180-193.

Pástor, Ľ., Stambaugh, R. F., and Taylor, L. A. 2020. “Sustainable investing in equilibrium,” Journal of Financial Economics (in press).

Pencle, N., and Mălăescu, I. 2016. “What’s in the Words? Development and Validation of a Multidimensional Dictionary for CSR and Application Using Prospectuses,” Journal of Emerging Technologies in Accounting (13:2), pp. 109-127.

Pennington, J., Socher, R., and Manning, C. D. 2014. “GloVe: Global Vectors for Word Representation,” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532-1543.

PwC. 2020. “The growth opportunity of the century.” Sustainable Finance Series Retrieved 03-01-2021, from https://www.pwc.lu/en/sustainable-finance/docs/pwc-esg-report-the-growth-opportunity-of-the-century.pdf

Schoonmaker, K. 2018. “Caterpillar Grows 2Q Top Line an Impressive 24%, Attains Record 2Q EPS,” Morningstar Equity Research.

Schueth, S. 2003. “Socially Responsible Investing in the United States,” Journal of Business Ethics (43:3), pp. 189-194.

UN General Assembly. 2015. “Transforming our World: the 2030 Agenda for Sustainable Development.” A/RES/70/1 Retrieved 03-01-2021, from https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

van Duuren, E., Plantinga, A., and Scholtens, B. 2016. “ESG Integration and the Investment Management Process: Fundamental Investing Reinvented,” Journal of Business Ethics (138:3), pp. 525-533.

Vasilescu, L., Hwang, W., Wakeham, S., Huang, L., Chang, T., and Jian, C. 2018. “The Moonshot: How Nike Changes its Footprint,” Macquarie Research.

Walsh, P. R., Webb, C. L., Thrum, A., and Rajan, N. 2018. “Solving for the CO2 problem,” Morgan Stanley.