Combat Greenwashing with GoalSpotter: Automatic Sustainability Objective Detection in Heterogeneous Reports

Mohammad Mahdavi
Gisma University of Applied Sciences
Potsdam, Germany
mohammad.mahdavi@gisma.com

Ramin Baghaei Mehr
Gisma University of Applied Sciences
Potsdam, Germany
ramin.baghaeimehr@gisma.com

Tom Debus
Ferris Solutions AG
Cham, Switzerland
tom.debus@ferris.ai

Abstract

Sustainable development is nowadays a prominent factor for the public. As a result, companies publish their sustainability visions and strategies in various reports to show their commitment to saving the environment and promoting social progress. However, not all statements in these sustainability reports are fact-based. When a company tries to mislead the public with its non-fact-based sustainability claims, greenwashing happens. To combat greenwashing, society needs effective automated approaches to identify the sustainability claims of companies in their heterogeneous reports.

In this paper, we present a new sustainability objective detection system, named GoalSpotter, that automatically identifies the environmental and social claims of companies in their heterogeneous reports. Our system extracts text blocks of diverse reports, preprocesses and labels them using domain expert annotations, and then fine-tunes transformer models on the labeled text blocks. This way, our system can detect sustainability objectives in any new heterogeneous report. As our experiments show, our system outperforms existing state-of-the-art sustainability objective detection approaches. Furthermore, our post-deployment results show the significant impacts of our system in real-world business.

CCS Concepts

• Information systems → Data mining; **• Computing methodologies → Natural language processing**.

Keywords

Sustainability, Greenwashing, Machine Learning, Natural Language Processing, Text Mining, Transformers

ACM Reference Format:
Mohammad Mahdavi, Ramin Baghaei Mehr, and Tom Debus. 2024. Combat Greenwashing with GoalSpotter: Automatic Sustainability Objective Detection in Heterogeneous Reports. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3627673.3680110

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3680110

1 Introduction

United Nations adopted the 2030 agenda for sustainable development in 2015, which includes 17 sustainable development goals, such as ending poverty, reducing inequality, and combating climate change [19]. The European Commission presented the European Green Deal in 2019 to set out its commitment to combating environmental issues [7]. Such initiatives have raised the awareness of the public with respect to the sustainability concerns of their lifestyle and consumption behavior [17].

As a result, there is nowadays an increasing demand for products/services that are green, i.e., environmentally friendly [9, 27]. Different stakeholders (e.g., investors, consumers, and policymakers) push companies to disclose their environmental-related information, which is necessary to evaluate their products/services [8]. Companies, therefore, publish their environmental/social visions and strategies in various sustainability documents, such as CSR (Corporate Social Responsibility) and ESG (Environmental, Social, and corporate Governance) reports [8, 17].

However, the environmental/social claims of companies in their sustainability reports and advertisements are not always completely true, i.e., fact-based. Companies often try to look "greener" in the eyes of the public than they really are. A classical example is a hotel that claims to be green because it allows the guests to reuse towels, but it does not have a real strategy in practice to reduce water or energy consumption [13]. This phenomenon is known as greenwashing.

Greenwashing is considered the act of misleading the public with respect to the environmental practices of a company or environmental benefits of a product/service [9]. Greenwashing makes the whole economy less green as it misleads stakeholders and erodes their trust in the whole green market [17]. This concern raises the urge to spot greenwashing in different company reports and product advertisements.

Greenwashing detection consists of two general steps: identifying environmental/social claims of companies and fact-checking them using all the other available data and evidence [20]. The first step of greenwashing detection, i.e., to identify the sustainability claims of companies, is particularly important. The habit of some companies is to slightly alter their sustainability claims over time when they notice that their previously promised goals are not achievable anymore. Therefore, continuous detection and tracking of companies' sustainability claims/objectives are necessary to enable public scrutiny of their sustainability strategies and performance.

Identifying the sustainability objectives of companies is a challenging problem [9] due to the large volume of disclosures and various communication channels that companies employ [6]. Traditionally, domain experts have to manually process hundreds of pages of reports from a company to identify environmental/social claims [20]. This manual process is tedious and time-consuming [29].

Automatic identification of sustainability objectives is particularly necessary for real-world business as it unburdens domain experts a lot. These approaches automatically filter out only those parts of sustainability reports for fact-checking that actually claim something with regard to the environment/society. Existing research has shown that automatic sustainability objective detection can be formulated as a text classification task and effectively addressed when the dataset is homogeneous [29]. However, these approaches are not effective on real-world heterogeneous sustainability reports, which belong to different data domains (e.g., real-estate or pharmaceutical), are usually large (e.g., hundreds of pages), are written in different formats (e.g., HTML and PDF), and contain many noises (in addition to environmental/social information).

In this paper, we propose a novel system, named GoalSpotter, to automatically detect sustainability objectives in heterogeneous reports. In particular, we make the following contributions.

- We design and implement a new clustering-based algorithm to extract text elements of heterogeneous sustainability reports and group them into informative text blocks. This way, we create an extensive labeled dataset for the sustainability objective detection task.
- We design and implement a new system to automatically detect sustainability objectives in heterogeneous reports. We encapsulate the individual components of the system in a way that can leverage new and upcoming foundational large language models. Our system is available online[1].
- We extensively evaluate the effectiveness and efficiency of our system in comparison to state-of-the-art sustainability objective detection approaches. We show that our system outperforms existing solutions on heterogeneous sustainability reports.
- We deploy our system and report its post-deployment impacts for real-world business.

2 Foundations

2.1 Greenwashing

The environmentalist Jay Westervelt coined the term greenwashing in 1986 to report about the hotel industry [8]. Back then, the hotels were asking guests to reuse towels for environmental purposes, but they did not have any environmental strategy in practice [8].

There is not yet a unified agreement on the definition of greenwashing [8, 20]. Greenwashing can be seen as selective disclosure of information, i.e., keeping the companies' negative environmental information undisclosed and disclosing only positive practices and performance metrics [8]. Greenwashing can also be seen as a decoupling behavior, i.e., when a company tries to deflect attention to minor symbolic environmental behaviors with no actual actions [8]. Greenwashing can occur at the firm level (when a company's environmental practices are misleading) or at the product/service level (when the environmental benefits of a product or service are misleading) [8].

[1]https://github.com/Ferris-Solutions/goalspotter_public

Climate change is one of the world's greatest crises, and to address it, the public and private sectors need to act together. **We co-founded The Climate Pledge, a commitment to reach net-zero carbon by 2040.**

Reducing carbon emissions in transportation is a complex challenge for many companies. Businesses also face the challenge of removing carbon emissions from new building construction.

Annotated Sustainability Objectives:
"commitment to reach net-zero carbon by 2040"

Figure 1: A small snippet from a sustainability report. The domain expert annotators extracted and highlighted a text segment that contains an environmental claim.

Our approach is not dependent on any specific definition or type of greenwashing. Due to the learning nature of our approach, it can learn to detect environmental/social claims that potentially commit any type of greenwashing.

2.2 Sustainability Reporting

The core idea behind sustainability reporting is to make companies transparent and accountable for their impact on the environment, society, and economy [21]. In particular, sustainability goals/objectives (e.g., environmental or social claims) are a core part of the sustainability reports that should be effectively communicated to business stakeholders [21].

The heterogeneity of sustainability reports makes them challenging to process. There are various types of sustainability reports in business, such as Corporate Social Responsibility (CSR), Greenhouse Gas Emissions (GHG), Environmental, Social, and corporate Governance (ESG), and UN Sustainable Development Goals (SDG) reports [17]. These sustainability reports often lack international standards or performance metrics for sustainability reporting [21]. Furthermore, these reports vary in several other dimensions, such as their format (e.g., PDF and HTML), template (i.e., whether there is any structure or not), length (e.g., from a few to hundreds of pages), and domain (e.g., pharmaceutical and electronics).

Example 1 (Sustainability report). For example, Figure 1 shows a small snippet from a long sustainability report. Although the entire text is about the environment, only the bold text block (i.e., paragraph) makes a claim, which needs to be extracted and fact-checked. The example also shows the annotation that is provided by domain experts for this document. This annotation can be used to label the bold text block of the document as a sustainability objective, i.e., an environmental claim. □

2.3 Problem Formulation

Our problem is automatically detecting sustainability objectives in any new report. Suppose $D = \{d_1, d_2, \ldots, d_{|D|}\}$ is a set of historical sustainability documents. Let $O_i = \{o_1, o_2, \ldots, o_{|O_i|}\}$ be the set of annotated sustainability objectives for document d_i. Given a new sustainability document d_{new}, the problem is to automatically extract sustainability objectives of the given document, i.e., $O_{\text{new}} = \{o_1, o_2, \ldots, o_{|O_{\text{new}}|}\}$.

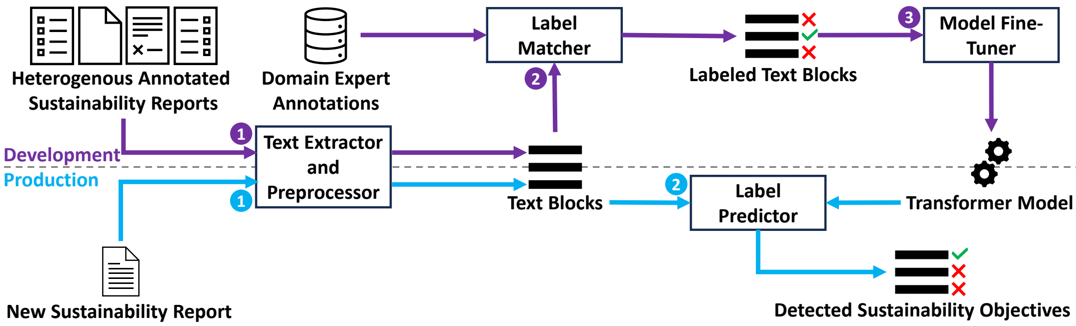


Figure 2: The workflow of GoalSpotter.

Example 2 (Sustainability objective detection). For example, given the small document in Figure 1, the goal is to extract the bold text block (i.e., paragraph) as it contains an environmental claim. In real-world scenarios, the documents are much more heterogeneous in many aspects, such as type, format, template, length, and domain, which makes this task much more challenging. □

3 Automatic Sustainability Objective Detection

Figure 2 illustrates the workflow of GoalSpotter. Given a collection of heterogeneous annotated sustainability reports, our system extracts sustainability objectives in any new documents. The workflow consists of a development phase (in purple) and a production phase (in blue).

Development phase. The development phase aims at training a sustainability objective detection model by creating a labeled training dataset using the annotated sustainability reports. In step 1, the system extracts and preprocesses texts of sustainability reports to get the list of text blocks (e.g., paragraphs) in each document. Next, in step 2, the system matches the text blocks of each document to the provided annotations to label each text block, i.e., whether the text block contains a sustainability objective or not. Finally, in step 3, the system leverages the labeled text blocks to fine-tune a transformer model.

Production phase. The production phase aims at extracting the sustainability objectives from any given new report. In step 1, the system again extracts and preprocesses texts of the new sustainability report to get the list of text blocks of the document. Next, in step 2, the system leverages the fine-tuned transformer model to predict the label of each text block, i.e., whether they contain sustainability objectives or not. In the rest of this section, we elaborate on the key components of our system.

3.1 Clustering-Based Text Block Formation

The input sustainability reports to our system are heterogeneous in many aspects, such as type, format, template, and length. In particular, our system processes documents in various formats, including PDF and HTML, which adhere to no predefined templates. This flexibility comes at a cost: We cannot tune our system to extract and preprocess texts for a certain document format and template. Therefore, we have to design the text extraction and preprocessing component of our system in a minimal and generalizable manner to be able to support various document formats and templates.

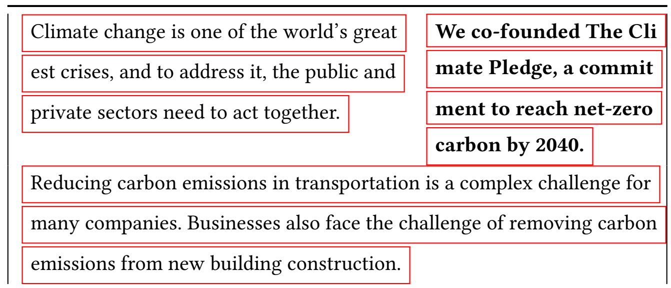


Figure 3: The extracted text elements from the sustainability report are uninformative as the PDF parser tool considers extracted newlines as delimiters.

While some textual file formats, such as HTML, are more straightforward to process, companies release the majority of their sustainability reports in PDF. Clean text extraction from PDFs is challenging because PDF was never designed to be an input data format. Existing PDF parsers extract text elements, such as characters, words, or fragments, which are not directly usable. Especially on difficult documents, such as old PDFs, these extracted text elements are usually uninformative as they are incomplete.

Example 3 (Text element extraction). For example, when we use a PDF parser tool to extract the text of the small document in Figure 3, the tool generates a list of red-framed text elements, which are lines of the original text. Naturally, each of the extracted text elements alone is not complete and informative enough. □

Therefore, we need an algorithm to merge uninformative extracted text elements with their surrounding text elements to form a larger informative text unit. To this end, we define a *text block* as a flexible text unit that can consist of multiple smaller text elements. This flexibility allows us to define text blocks for each document based on its structure. Examples of text blocks could be a heading, a sentence, a paragraph, or a table cell in a PDF/HTML document.

Our clustering-based text block formation algorithm leverages the spatial location of text elements on the documents to form text blocks. First, we use text extractor libraries, such as *Beautiful Soup* for HTML and *pdfminer* for PDF, that keep the spatial information of text elements on documents, such as their line number, their neighboring text elements, the number of newlines around them, and their relative coordinate on the document page. This way, we keep the relative distance of text elements from each other when

Climate change is one of the world's greatest crises, and to address it, the public and private sectors need to act together.

We co-founded The Climate Pledge, a commitment to reach net-zero carbon by 2040.

Reducing carbon emissions in transportation is a complex challenge for many companies. Businesses also face the challenge of removing carbon emissions from new building construction.

Annotated Sustainability Objectives:
"commitment to reach net-zero carbon by 2040"

Figure 4: Our text block formation algorithm clusters extracted text elements into informative text blocks. The domain expert annotations are document level.

extracting them from a document. We then use these features to cluster extracted text elements into larger and more informative text blocks. Finally, we minimally clean the text elements of each cluster (by removing unnecessary characters such as newlines) to make the text block coherent.

Example 4 (**Text block formation**). For example, our text block formation algorithm clusters the extracted text elements of the previous example into 3 text blocks, which correspond exactly to the three paragraphs in Figure 4. □

3.2 Labeling Text Blocks Using Annotations

Each input sustainability report comes with a set of annotated sustainability objectives as shown in Figure 1. Since our system should be able to detect sustainability objectives at the text block level, we need to assign a label to each text block using document-level annotations. This is not trivial because documents are not annotated in a structured way. The quality of the text block labels is essential as it directly affects the performance of our system in detecting sustainability objectives in the end.

We label text blocks by matching them with domain expert annotations. Let $B_i = [b_1, b_2, \ldots, b_{|B_i|}]$ and $O_i = [o_1, o_2, \ldots, o_{|O_i|}]$ be the list of text blocks and annotated sustainability objectives for document d_i, respectively. For each pair of a text block $b \in B_i$ and an annotated sustainability objective $o \in O_i$, we first preprocess b and o to remove spaces, punctuation, and special characters, and lower the letters. We then match b and o. If string b contains string o, we label the original text block b as a sustainability objective. Otherwise, we label it as noise.

Example 5 (**Text block labeling**). After preprocessing text blocks and the domain expert annotation for the document in Figure 4, the annotation matches the bold text block. Therefore, this bold text block will be labeled as a sustainability objective and the rest of the text blocks will be labeled as noise. □

3.3 Fine-Tuning the Transformer Model

We can now train a text classification model using the labeled text blocks dataset. While traditional models, such as naive Bayes and random forest, are fast to train, recent transformers have been shown to be more effective in various scenarios [25]. Therefore, we focus on fine-tuning pretrained transformers on our dataset.

As shown in the experiments, we have tried out several transformers with various hyperparameters. In our current prototype, we use a *DistilRoBERTa* model but, due to the modular design of our system, it is straightforward to replace this model with any newer models in the future. By default, we fine-tune the model on our dataset for up to 10 epochs and set the learning rate to 5e-5, the batch size to 16, and the optimizer to Adam.

4 Evaluation

Our experiments aim to answer the following questions.

- How does our system compare to the existing sustainability objective detection approaches?
- How do our design decisions, including data preprocessing, model selection, and hyperparameter tuning, affect the system performance?

We first introduce our experimental setting and then detail our experiments.

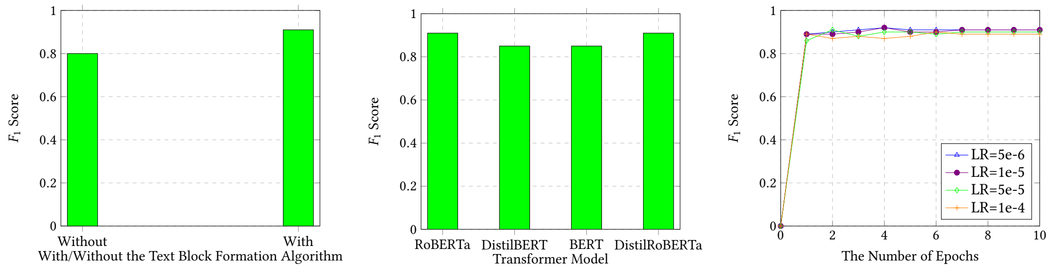
4.1 Setup

Datasets. We evaluate our system on 2 datasets.

- *Green Claims* [29] is a dataset containing 773 tweets for the period 2017 to 2020 from 48 companies in the cosmetics and electronics domains. Domain experts annotated each tweet with one of the following labels: (1) explicit green claim, (2) implicit green claim, or (3) not green claim [29]. Therefore, the binary classification task is to distinguish green claims (either explicit or implicit) from the rest of the tweets. We use this homogeneous dataset to evaluate our system in straightforward scenarios.
- *Sustainability Goals* is our own proprietary dataset containing 218876 extracted text blocks, out of which only 5071 are labeled as sustainability objectives. To create this dataset, we collected 718 sustainability reports from 422 companies. Domain experts annotated each sustainability report with a few textual goals that are mentioned somewhere in the document. Our data collocation and labeling approach extracts text blocks of these documents and matches them to the domain expert annotations to create our own proprietary labeled dataset. We use this highly heterogeneous and imbalanced dataset to evaluate our system in more challenging real-world scenarios.

Baselines. We compare our system to 3 baseline approaches.

- *BERTClaimBuster* is a fine-tuned BERT model on the ClaimBuster dataset, which is labeled for the task of detecting check-worthy factual claims in the US general election debate transcripts [10]. We use this pretrained model on our datasets without any further fine-tuning steps to check the performance of general pretrained claim detection models for the sustainability objective detection task.
- *TF-IDF+Random Forest* is a traditional text classification pipeline using *TF-IDF* featurization and a *Random Forest* model. We use this approach to check the performance of the traditional approaches for the challenging sustainability objective detection task.



(a) The Effect of the Clustering-Based Text Block Formation

(b) The Effect of the Transformer Model

(c) The Effect of the Epochs Count and Learning Rate

Figure 5: System effectiveness with different internal design decisions on the *Sustainability Goals* dataset.

Table 1: System effectiveness and efficiency in comparison to the baselines. Time is measured in minutes.

Approach	Green Claims				Sustainability Goals			
	P	R	F	T	P	R	F	T
BERTClaimBuster	0.32	0.59	0.42	< 1	0.10	0.69	0.17	< 1
TF-IDF + Random Forest	0.97	0.69	0.80	< 1	0.92	0.37	0.53	5
Bin_RoBERTa	0.82	0.91	0.86	1	0.72	0.68	0.70	72
GoalSpotter	0.86	0.94	**0.90**	1	0.93	0.89	**0.91**	48

- *Bin_RoBERTa* [29] is a pretrained *RoBERTa* model that was recently proposed for the environmental claim detection task. Same as the original approach [29], we first preprocess the datasets using the *texthero* library. We then fine-tune this model for the binary classification task of detecting sustainability objectives using our datasets. We use this approach to compare our system to the state-of-the-art sustainability objective detection approaches.

Evaluation measures. The sustainability objective detection task is naturally imbalanced, i.e., the number of sustainability objectives in the reports is far lower than the rest of the text blocks. Therefore, we report precision, recall, and the F_1 score of the main target class, i.e., being a sustainability objective, to evaluate the effectiveness. Precision is the number of correctly detected sustainability objectives divided by the number of all text blocks that are flagged as sustainability objectives. Recall is the number of correctly detected sustainability objectives divided by the number of all annotated sustainability objectives. The F_1 score is the harmonic mean of precision and recall. All evaluation metrics are measured on an unseen test set, which forms 20% of the original dataset. For each evaluation measure, we report the mean of 5 independent runs. For the sake of readability, we omit the standard errors as they are always small numbers close to zero (< 1%).

We also report the training/fine-tuning time in minutes to evaluate the efficiency. Note that We run the experiments on an Ubuntu 18.04 LTS machine with 8 2.3 GHz CPU cores, 30 GB memory, and 1 NVIDIA Tesla T4 GPU.

4.2 Comparison with the Baselines

We compare the effectiveness and efficiency of our system with the baselines in Table 1. Our system outperforms all the baselines in terms of the F_1 score on all the datasets due to our effective text block formation algorithm and careful model fine-tuning process.

Even though some of the baselines, such as *TF-IDF + Random Forest*, achieve high precision, they cannot match the F_1 score of GoalSpotter. This is because these baselines can only identify a few sustainability objectives, resulting in low recall. Obviously, our system is not the fastest approach as we have to fine-tune a transformer model on our datasets. However, the fine-tuning process can be done in a fraction of an hour for all the datasets. Even though simpler approaches, such as applying a pretrained *BERTClaimBuster* model without any fine-tuning, are faster, they cannot achieve a high F_1 score.

4.3 Experimenting with the Design Decisions

We analyze the effect of the most important design decisions on the effectiveness of our sustainability objective detection system. Figure 5 shows the results of the corresponding experiments with data preprocessing, model selection, and hyperparameters on the *Sustainability Goals* dataset.

The effect of the clustering-based text block formation. As shown in Figure 5, our data preprocessing component, including our clustering-based text block formation algorithm, has a significant positive effect on the effectiveness. Our system has to process heterogeneous sustainability reports of different companies, which are published in various types, formats, templates, and sizes. Without our text block formation algorithm, the extracted text units are not informative enough to be detected as sustainability objectives. However, once we equip our system with the text block formation algorithm, it forms more informative text units to be processed by the transformer model.

The effect of the transformer model. As shown in Figure 5, the choice of the model, as long as it is a transformer, does not affect the effectiveness significantly. We observe that *RoBERTa* models achieve slightly higher F_1 scores than *BERT* models. By default, we use a *DistilRoBERTa* model as its fine-tuning speed is faster than other alternatives.

The effect of the number of epochs and learning rate. As shown in Figure 5, the number of epochs and learning rate do not affect the model convergence speed significantly if they are chosen from their typical ranges. By setting the learning rate to $5e-5$, our model achieves its highest F_1 score in a few, i.e., 2, epochs.

Table 2: Summary of post-deployment data.

Company	#Documents	#Pages	#Extracted Objectives
C1	20	2131	150
C2	18	3172	642
C3	41	3560	447
C4	19	2488	102
C5	17	1298	113
C6	29	3278	343
C7	23	2208	247
C8	22	5012	764
C9	64	4791	379
C10	16	1202	79
C11	17	1229	95
C12	64	1721	71
C13	18	3250	105
C14	12	2531	43
Total	380	37871	3580

5 Post-Deployment Impacts

In accordance with the guidelines of the applied research track, we would like to also report on the post-deployment impacts of our sustainability objective detection system in practice.

Ferris Solutions[2] is a Switzerland-based company focusing on building artificial intelligence solutions for sustainability problems. In particular, they build data science applications on ESG reports to support domain experts in various scenarios, such as evaluating target companies in terms of their sustainability performance.

One of the big challenges of Ferris Solutions is to detect and track the sustainability objectives of its target companies over time. This requires domain experts to continuously collect and manually inspect sustainability reports of the target companies. Table 2 shows the size of the current repository of sustainability reports for target companies at Ferris Solutions. Overall, we have collected 380 new sustainability reports from 14 new target companies over the period of the last decade. Note that we could not reveal the names of target companies due to the confidentiality of information. Manual inspection of all these sustainability reports, which in total goes above 37871 pages, is extremely costly for Ferris Solutions.

We have deployed GoalSpotter at Ferris Solutions to automate this process. Overall, GoalSpotter extracts 3580 sustainability objectives from the above sustainability reports in only 12 hours. Table 3 shows the top 2 extracted sustainability objectives per company and the confidence of our system in their detection.

We observe that all extracted text blocks are sustainability goals and objectives. These results again showcase the effectiveness of our system on new real-world heterogeneous data. Domain experts leverage these insights to compare different target companies based on the specificity of their sustainability objectives. Furthermore, domain experts store these sustainability objectives in databases to track them over time and to ensure that companies are fulfilling their historic sustainability claims.

6 Related Work

Our system automatically extracts sustainability objectives (i.e., environmental/social claims) from heterogeneous reports to detect greenwashing. Therefore, we first review existing (manual and automated) approaches for detecting greenwashing. Then, we discuss general claim detection approaches. Finally, we review text extraction approaches for heterogeneous documents.

[2]https://www.ferris.ai

6.1 Greenwashing Detection

Manual approaches. Due to the lack of a unique definition, there is no standard approach to detect greenwashing [20]. Previous research has proposed various frameworks, guidelines, and criteria to avoid/detect greenwashing [20]. TerraChoice lists 7 sins of greenwashing in formulating environmental claims. This list includes (1) hidden trade-offs, (2) lack of proofs, (3) vagueness, (4) irrelevance, (5) lesser of two evils, (6) fibbing, and (7) worshiping false labels [27]. Other researchers also expanded this list to more sins [8]. BSR and Futerra propose a framework to check three criteria when communicating environmental initiatives [12]: (1) Is the initiative impact real and significant? (2) Is the initiative aligned with other functions within the company? (3) Is the initiative communicated clearly and transparently?

Although these frameworks allow domain experts to systematically assess environmental claims, they need manual effort. A domain expert has to manually find an environmental claim and check it against a list of indicator questions in the framework [20]. In contrast, we propose an automated approach to detect environmental/social claims without involving domain experts.

Automated approaches. A recent idea is to use artificial intelligence to automate greenwashing detection [5]. However, to the best of our knowledge, there are only a few research papers that proposed automatic environmental/green claim detection approaches [26, 29]. Some authors formulated the green claim detection problem as a classification task [29]. They collected and annotated green claim data from Twitter on two domains (i.e., cosmetics and electronics). They fine-tuned pretrained models (e.g., RoBERTa) on this dataset to classify tweets into explicit, implicit, and non-green claim classes. Another study introduced an expert-annotated sentence-level dataset for real-world environmental claims made by listed companies [26]. The authors also fine-tuned pretrained transformers (e.g., RoBERTa and ClimateBERT) to detect whether a sentence is an environmental claim or not.

Existing approaches assume the input data is homogeneous, i.e., small tweets [29] or clean sentences [26]. In contrast, we do not make any assumptions about the heterogeneous input data. Our sustainability objective detection system learns from real-world heterogeneous sustainability reports. Therefore, our approach is more generalizable for real-world data as shown in the experiments.

6.2 Automatic Claim Detection

The investigative reporting genre that fact-checks political claims has become more common in recent years, especially during elections [11]. In journalism, fact-checking is the task of verification of a claim, which requires fact-checkers to evaluate the claim against existing facts and evidence in order to reach a final verdict [4]. The goal of computational journalism is to automate the fact-checking process [11]. Automated fact-checking aims at unburdening the human in assessing the veracity of a claim as the fact-checking process may take even a few days for complex claims [28]. Automatic claim detection is one of the important steps of fact-checking.

Table 3: The top 2 extracted sustainability objectives per company from the post-deployment data.

Company	Extracted Sustainability Objectives	System Confidence
C1	Integrate sustainability information into their reporting cycle	0.99
C1	Voluntary turnover rate in 2021: 8.1%	0.99
C2	Substitute F-gases for low GWP alternatives	0.99
C2	Commitments to double environmental efficiency with new energy, water and waste targets	0.99
C3	We are committed to empowering 100 million smallholder farmers in low to middle	0.99
C3	Transition all Consumer Health products to 100% recyclable or reusable packaging	0.99
C4	Explore innovative value-based approaches	0.99
C4	Reduce employees' risk of a serious incident or fatality	0.99
C5	Expand principles of sustainability and performance indicators at key suppliers	0.99
C5	250 students in STEM awareness activities	0.99
C6	Define sustainability strategies, goals and policies in consultation with key stakeholders	0.99
C6	Join industry peers, UN entities and/or other stakeholders in initiatives contributing to solving common challenges and dilemmas at the global and/or local levels	0.99
C7	Uses 25 percent PCR content in bottle	0.99
C7	100 percent of major brands share product sustainability information on their websites	0.99
C8	Mote the proportion of women in leadership positions at the company	0.99
C8	Perform waste audits to identify ways to reduce waste or increase recycling efficiency	0.99
C9	Align strategies, goals and incentive structures of all business units and subsidiaries with corporate sustainability strategy	0.99
C9	Implement water saving programs at the top-10 sites with highest water footprint and water scarcity	0.99
C10	Demonstrate added value of new products	0.99
C10	Pursue leadership in HAEMOPHILIA	0.99
C11	Incorporate environmental sustainability across all aspects of our organization	0.99
C11	Integrate immunization delivery and family planning services in an effort	0.99
C12	30% increase in the representation of women in key leadership roles	0.99
C12	Reached goal of 20% of women in key positions a year ahead of schedule	0.99
C13	By 2025: priority sites located near sensitive natural areas shall implement a biodiversity protection	0.99
C13	By 2025, pilot projects will be implemented for promoting further the sustainable use and responsible	0.99
C14	Share high-quality medical resources, provide material and technical support to poor areas	0.99
C14	Make monthly contributions to the schemes at approximately 7% to 10% of the relevant income	0.99

Automatic claim detection systems have been developed for various problems, such as fake news detection [4], argument mining [16], and scientific discourse tracking [2]. The main goal of these systems is to estimate the check-worthiness of a piece of content to decide whether it contains an important factual claim that is worth fact-checking or not [23]. Detecting check-worthy claims is the first step of fact-checking as fact-checkers are flooded with statements [18]. Claim detection aims to unburden fact-checkers in processing a large volume of online content to find claims [30].

Existing automatic claim detection systems leverage both traditional machine learning models, such as logistic regression [15] and support vector machines [16], and recent transformers, such as BERT and RoBERTa [1, 4, 23, 25]. Not surprisingly, top-performing systems use transformers and transfer learning [25]. The recent public benchmark datasets for claim/evidence detection, such as CheckThat! [25] and NEWSCLAIMS [23], significantly advanced research in this area [2, 3, 15].

Although our environmental/social claim detection task is conceptually similar to general claim detection, the main difference is in available resources. As mentioned, there are various labeled datasets and trained models for general claim detection tasks as this is a well-researched topic. However, the recently emerged problem of environmental/social claim detection still requires more research and resources as it cannot directly benefit from datasets and models that are developed for general claim detection tasks. This is why we collect our own proprietary dataset and build a specific system for the sustainability objective detection problem.

6.3 Text Extraction from Heterogeneous Documents

Existing research seeks to extract text from heterogeneous documents by first extracting their templates or layouts. Template extraction has been performed for web data using HTML tags [14], for PDFs using structural rules [22], and for visually rich documents using visual cues [24].

Since we work with a large-scale collection of both heterogeneous HTML and PDF files, we cannot assume HTML tags, structural rules, or visual cues are always available to extract templates of documents first. That is why we developed our clustering-based text block formation algorithm.

7 Conclusion

We proposed a new sustainability objective detection system, called GoalSpotter, that identifies environmental/social claims in heterogeneous reports. Our system extracts text blocks from heterogeneous sustainability reports. It then preprocesses and labels text blocks using domain expert annotations. Our system next fine-tunes a pretrained transformer model on the created dataset to predict sustainability objectives in any new report. As our experiments show, our system outperforms existing sustainability objective detection approaches. As our post-deployment results show, our system also makes a significant impact in real-world business.

There are still future directions for improvement. We plan to extend our system to extract fine-granular information, such as the baseline and deadline, from each detected sustainability objective.

References

[1] Amani S Abumansour and Arkaitz Zubiaga. 2022. Check-worthy claim detection across topics for automated fact-checking. *arXiv preprint arXiv:2212.08514* (2022).

[2] Titipat Achakulvisut, Chandra Bhagavatula, Daniel Acuna, and Konrad Kording. 2019. Claim extraction in biomedical publications using deep discourse model and transfer learning. *arXiv preprint arXiv:1907.00962* (2019).

[3] Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. 2014. A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In *Proceedings of the first workshop on argumentation mining*. 64–68.

[4] Javier Beltrán, Rubén Míguez, and Irene Larraz. 2021. ClaimHunter: An unattended tool for automated claim detection on Twitter.. In *KnOD@ WWW*.

[5] Cristina Chueca Vergara and Luis Ferruz Agudo. 2021. Fintech and sustainability: Do they affect each other? *Sustainability* 13, 13 (2021), 7012.

[6] Theodor Cojoianu, Andreas GF Hoepner, Georgiana Ifrim, and Yanan Lin. 2020. Greenwatch-shing: Using AI to detect greenwashing. *AccountancyPlus-CPA Ireland* (2020).

[7] European Commission. 2023. A European Green Deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en. Accessed: 20.02.2023.

[8] Sebastião Vieira de Freitas Netto, Marcos Felipe Falcão Sobral, Ana Regina Bezerra Ribeiro, and Gleibson Robert da Luz Soares. 2020. Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe* 32, 1 (2020), 1–12.

[9] Matthias Gräuler and Frank Teuteberg. 2014. Greenwashing in online marketing–Investigating trust-building factors influencing greenwashing detection. *Multikonferenz Wirtschaftsinformatik* (2014).

[10] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1803–1812.

[11] Naeemul Hassan, Chengkai Li, and Mark Tremayne. 2015. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th acm international on conference on information and knowledge management*. 1835–1838.

[12] Rina Horiuchi, Ryan Schuchard, Lucy Shea, and Solitaire Townsend. 2009. Understanding and preventing greenwash: A business guide. *London: Futerra Sustainability Communications* (2009), 1–39.

[13] Greenwashing Index. 2023. About Greenwashing. https://web.archive.org/web/20130426232159/http://www.greenwashingindex.com/about-greenwashing/. Accessed: 20.02.2023.

[14] Chulyun Kim and Kyuseok Shim. 2010. Text: Automatic template extraction from heterogeneous web pages. *IEEE Transactions on knowledge and data Engineering* 23, 4 (2010), 612–626.

[15] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 1489–1500.

[16] Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

[17] Wayne Moodaley and Arnesh Telukdarie. 2023. Greenwashing, sustainability reporting, and artificial intelligence: A systematic literature review. *Sustainability* 15, 2 (2023), 1481.

[18] Preslav Nakov, David Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, and Giovanni Da San Martino. 2021. Automated fact-checking for assisting human fact-checkers. *arXiv preprint arXiv:2103.07769* (2021).

[19] United Nations. 2023. Sustainable Development Goals. https://sdgs.un.org/goals. Accessed: 20.02.2023.

[20] Noémi Nemes, Stephen J Scanlan, Pete Smith, Tone Smith, Melissa Aronczyk, Stephanie Hill, Simon L Lewis, A Wren Montgomery, Francesco N Tubiello, and Doreen Stabinsky. 2022. An integrated framework to assess greenwashing. *Sustainability* 14, 8 (2022), 4431.

[21] Oluyomi A Osobajo, Adekunle Oke, Ama Lawani, Temitope S Omotayo, Nkeiruka Ndubuka-McCallum, and Lovelin Obi. 2022. Providing a roadmap for future research agenda: A bibliometric literature review of sustainability performance reporting (SPR). *Sustainability* 14, 14 (2022), 8523.

[22] Cartic Ramakrishnan, Abhishek Patnia, Eduard Hovy, and Gully APC Burns. 2012. Layout-aware text extraction from full-text PDF of scientific articles. *Source code for biology and medicine* 7 (2012), 1–10.

[23] Revanth Gangi Reddy, Sai Chetan Chinthakindi, Zhenhailong Wang, Yi Fung, Kathryn Conger, Ahmed Elsayed, Martha Palmer, Preslav Nakov, Eduard Hovy, Kevin Small, et al. 2022. Newsclaims: A new benchmark for claim detection from news with attribute knowledge. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 6002–6018.

[24] Ritesh Sarkhel and Arnab Nandi. 2019. Visual segmentation for information extraction from heterogeneous visually rich documents. In *Proceedings of the 2019 international conference on management of data*. 247–262.

[25] Shaden Shaar, Alex Nikolov, Nikolay Babulkov, Firoj Alam, Alberto Barrón-Cedeno, Tamer Elsayed, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Giovanni Da San Martino, et al. 2020. Overview of CheckThat! 2020 English: Automatic identification and verification of claims in social media. *CLEF (Working Notes)* 2696 (2020).

[26] Dominik Stammbach, Nicolas Webersinke, Julia Anna Bingler, Mathias Kraus, and Markus Leippold. 2022. A Dataset for detecting real-world environmental claims. *arXiv preprint arXiv:2209.00507* (2022).

[27] TerraChoice. 2010. The sins of greenwashing: Home and family edition. *Underwriters Laboratories* (2010).

[28] James Thorne and Andreas Vlachos. 2018. Automated fact checking: Task formulations, methods and future directions. *arXiv preprint arXiv:1806.07687* (2018).

[29] Vinicius Woloszyn, Joseph Kobti, and Vera Schmitt. 2021. Towards automatic green claim detection. In *Forum for Information Retrieval Evaluation*. 28–34.

[30] Xia Zeng, Amani S Abumansour, and Arkaitz Zubiaga. 2021. Automated fact-checking: A survey. *Language and Linguistics Compass* 15, 10 (2021), e12438.