

Course: Text Mining Domains
Group : Legal Domain (Legals)
Supervisor: Prof. P. Vossen

Detecting News items related to UN Sustainable Development Goals

A Classification Project

1.

Merel de Groot
2607569

Adam Tucker
2671046

Michiel van Nederpelt
2684581

Jan van Casteren
2529356

Peter Caine
2671676

2. Introduction

2.1 Introducing the user problem

One overarching measure of global performance towards a better, more sustainable society is described by the UN's Sustainable Development Goals (SDGs). The SDGs, set in 2015 by the United Nations General Assembly are a collection of 17 global goals (see appendix A). These goals are intended to be achieved by the year 2030, and they are designed to be a "blueprint to achieve a better and more sustainable future for all" (dpicampaigns, 2019).

Companies have acted to incorporate SDG outcomes in their company policy, however, it is insufficient to merely encode these goals in policy documents, and these goals have to be somehow translated into actions and behaviours throughout the company. In large scale corporations, this can be difficult to translate into company culture, and difficult to monitor, especially when subsidiaries and supply chains are taken into account in global enterprises.

From a business consultancy perspective, monitoring performance on how a company is translating policy into actions is crucial to be able to predict future company performance and detect early signals of risk, yet monitoring large complex multinational organisations is no trivial task, especially when it comes to qualitative aims.

One way of monitoring performance is to monitor public perception directly for evidence of actions that relate to these goals. In the bid to provide a constant stream of content, journalists worldwide work to investigate any potential source of information which might be considered newsworthy. As a consequence, a global source of stories are produced on a wide range of topics; many of which relate directly to sustainable development goals.

For a business consultancy focussed on enterprise risk, finding an effective way to monitor the stream of news stories specifically for those stories that relate to the Sustainable Development Goals is of significant value.

2.2 Introducing the proposed application and data

2.2.1 Application

The objective of this project is to design and construct a system which would take the name of a company as input and produce output which reflects the performance of the company with respect to one of the 17 UN SDGs obtained from publicly available global news sources. The intention is for the system to be sensitive to and detect news stories relating to the SDGs and be able to identify whether this is a negative story (posing a risk to any business entities involved) or positive (indicating a positive performance indicator for a company).

As such, a system would have to be able to:

- accept news items as input;
- determine whether a news item is related to particular SDGs, and if so,
- reclassify those items as positively related or negatively oriented towards the SDG.

The end user could use this classifier on a stream of news sources to filter for news stories related to a particular SDG, and if related, to further classify that text according to whether the news was contributing positively to the SDG or otherwise.

In order to evaluate whether a particular company is involved, a database of entities (subsidiaries, suppliers) associated with the company would be compared with the news stories to identify whether there were any SDG-related stories that involved these entities.

While the intention is for the end user to use the system on a source of well structured news sources (such as LexisNexis), licensing constraints meant that this could not be made available to us during system development, and an alternative was sought.

2.2.2 Training Data

System design is often dependent on the quality of the training data. Ideally, custom datasets tailored to the specific task and annotated by human annotators would provide training and test data. For this project, such a dataset would involve an extensive list of news items which had been annotated as unrelated/related for particular SDGs, and if related, whether positively or negatively. Furthermore, training data would ideally consist of news streams from a wide range of sources to resemble the proposed user-input, however, in this instance, such a niche labelled dataset could not be found, and so as a first measure, ideas were explored to aggregate news data.

Given the time constraints and the generally wide-ranging goal of the project, it was decided to make use of existing datasets on which to train our classifiers. Since data acquisition and cleaning could prove a lengthy process, alternatives were suggested by our supervisor, prof. Piek Vossen: the GigaWord dataset for classifying related articles, and the MPQA Opinion Corpus for classifying the subjective relation of the news to the SDG.

The GigaWord dataset (English Gigaword Fifth Edition) is a collection of around ten million documents including over 8 million news articles from seven different news sources (appendix B). It was developed by Johns Hopkins University's Human Language Technology Center of Excellence to provide a standardized corpus for knowledge extraction and distributional semantics for researchers. It is a well-structured data source covering a sufficiently broad range of current affairs topics covering a period of around 15 years from 1994 to 2010¹. While the GigaWord corpus met the requirements for a diverse, well-structured and sufficiently large dataset, it was not labelled for our purposes. This step still required automating.

The MPQA Opinion Corpus contains news articles from a wide variety of news sources manually annotated for sentiment on a sentence level and was chosen for its similarity to the news articles the application would process. Training of the sentiment classifier was done with the MPQA data-set. Only the files containing annotated sentiment (.feat) were used and the explanatory .log files were discarded. .feat files were first transformed to .CSV files to make use of the CSV module and ConuLL module for sentiment feature extraction. After extraction sentences were scored on being negative, neutral or positive with a sum method giving one value (and thus a label) per sentence to be put in the classifier. Further improvement of the classifier could be obtained by first scoring the sentence without attributing a label per sentence but instead use a score, which could "help" the classifier determine sentiment further.

2.3 Structure of the report.

This paper documents the process of preparation of the training data and construction of the two classifiers. The first section will discuss how the GigaWord dataset was filtered and labelled for articles related to particular SDGs, together with a section describing how the data was encoded to be introduced into a classifier. Next a section describing how MPQA Opinion Corpus was similarly prepared and information encoded for the second classifier. The results section will illustrate and describe the output of the classifiers. The final section will discuss the results, evaluating performance including error analysis and make recommendations and suggestions for system refinement.

2.4 Division of responsibilities.

This project consisted of several major components for which work was distributed among group members

Jan: selection of training data for the SDG-relation classifier, report writing, GitHub page editor.

Peter: system design phase 1 (classifier: related/ unrelated; error analysis); report writing.

Merel: Sentiment classifier, BERT-finetuning, report writing

Adam: (sentiment) training data, error analysis, test set collection

Michiel: sentiment training data to classifier, document selection, evaluation of classifier output

¹ Portions © 1994-2010 Agence France Presse, © 1994-2010 The Associated Press, © 1997-2010 Central News Agency (Taiwan), © 1994-1998, 2003-2009 Los Angeles Times-Washington Post News Service, Inc., © 1994-2010 New York Times, © 2010 The Washington Post News Service with Bloomberg News, © 1995-2010 Xinhua News Agency, © 2003, 2005, 2007, 2009, 2011 Trustees of the University of Pennsylvania

3. User requirements

As previously mentioned, the overall goal of our system is to create a tool that could assist companies in finding articles which mention the conduct of these companies, their competitors and/or suppliers with regard to any of the 17 UN Sustainable Development Goals. As a first effort towards this goal, the team developed the SDG-tracker. The tool as presented today has the capability to classify unseen articles on their being related to both one of the companies listed as well as one of the SDGs. It delivers an end-to-end implementation for two of the SDGs, namely SDG 1, to end poverty in all forms, and SDG 12, the realization of sustainable consumption and production.

The client preferred the system to adhere to some goals. Firstly, given a data stream, the system should be able to recognise if an article is related to an SDG. If it has found an article it should be able to assess the polarity of this article. The system should be able to classify data from smaller local news sources, because these sources are likely to be firstly informed about events related to suppliers of the clients' customers. Related to this is that the system should be adaptable for different languages. The system should deliver an accurate reporting of any potential risks a company is facing in terms of failure on an SDG commitment. The system aims to be able to define accurately, and at an early stage, perceived company performance across an SDG so that action can be taken before any adverse consequences are realised. The ability for a company to measure its own performance against a competitors, was also stated to be a feature that would be desirable.

4. Data Collection

4.1 Data document classifier

Test set

It was decided to select only a small subset of the SDGs in order to build a proof-of-concept system which could later be expanded. Since we wanted to be able to evaluate real-world data, it was determined that the test set would be leading. The proposed process was to: identify a target company which had ascribed to and had official stated aims regarding achieving SDGs. Explore existing news stories related to those SDGs; determine which SDGs were most prominent in the news stories retrieved; find a competitor in the same sector and conduct a second, narrower search for news stories relating only to those SDGs. Finally, the training set could be tailored to those SDGs. The final evaluation data was manually gathered from the internet by using search engines.

Company selection

Desk research conducted into the SDGs revealed a key factor to consider is the overlap in corporate reporting on individual SDGs. As well as being an issue for corporations reporting on goal performance, it was also a significant factor in the application's design. After reviewing the Communication on Progress (CoP) and Communication on Engagement (CoE) ² it became clear that even large corporations are not focused on achieving all SDGs, or even most of them. It was decided to select a company and competitor that had an official commitment. After some deliberation, the business sector relating to food and beverages was selected as having the wide-ranging potential to impact on a range of SDGs. Furthermore, it was chosen to target a sector that would enable the straightforward collection of multinational supply chain and supplier information.

Starbucks clearly stated that it was committed to all SDGs ("Starbucks Coffee Company – Communication on Progress: UN Global Compact", 2020). A competitor in the same market was selected, which is not registered as a subscribed participant in the SDGs but is a subsidiary of Coca-Cola Company who are committed to 10 SDGs. Starbucks and Costa Coffee also had prominent news stories that relate to SDGs and therefore, from a data collection standpoint, was prudent. In conclusion, it was decided that these two companies would offer a useful starting place to demonstrate a proof of concept.

The most prominent new stories which related to SDGs and involved both companies were:

- SDG 1 - 'End poverty in all its forms everywhere.'
- SDG 12 - 'Ensure sustainable consumption and production patterns.'

² <https://www.unglobalcompact.org/participation/report/coe/create-and-submit/submitted-coe>

4.2 Labelled data for sentiment classifier

Since the data the system has to work on will consist solely of news articles, it was essential to find extensive, reliable training data. However, the options for a free news corpus humanly annotated for sentiment were limited. The most workable choice was the MPQA opinion corpus. The MPQA opinions corpus is a manually annotated news corpus selected from various news sources, which provides, among other features, the sentence level polarities of news items. The described sentiments are negative, positive, both, neutral, uncertain-negative, uncertain-positive, uncertain-both, uncertain-neutral. (Deng and Wiebe, 2015) It was decided to ignore all polarities except for negative and positive. The data is provided in a CoNLL-type format, meaning separate tokens need to be joined to restore the original text because the data gave the sentiment at a sentence level. The data needed to be configured to provide the sentiment at the document level. A count was made of every positive sentence, and every negative sentence subtracted one from the counter. If the counter returned a positive number, a positive label was given, a negative number, a negative label, and in a zero, a neutral label.

4.3 Evaluation data

Articles related to the chosen companies and SDGs were selected by members of the team and annotated. The annotation described the relation between the content of the article and the SDG related or unrelated. If a related article was related, the sentiment, positive or negative was then also added. The choice was made to only annotate the sentiment for positive articles based on the intention of the application. If the articles were not associated with the Sustainable Development Goals, it would not impact the performance on that goal and therefore would not be relevant. There were two evaluation sets created; for SDG 1, No Poverty, and SDG 12, Responsible Consumption and Production. The set for SDG 1 contained: 64 articles, 41 related, split into 17 positive and 24 negative, and 23 unrelated news items. The set for SDG 12 contained: 91 articles, 56 related, split into 38 positive and 18 negative, and 35 unrelated news items. A neutral label was not provided in the evaluation data, because the action of a company related to an SDG always had some positive or negative polarity.

4.3.1 Inter-annotator rules

The test set was compiled by multiple members of the group, using inter-annotator guidelines to ensure consistency of selection and annotation. The following constitute the agreed-upon inter-annotator rules:

- Annotation of related articles was agreed upon to include only articles referencing a certain SDG, at first this consisted only of SDG 1 but was later expanded to also include SDG 12; however the related test sets were kept separate.
- Related articles were decided to be related through either company policy, e.g. tax evasion, worker conditions or through scandals relating to employees or production of either Costa Coffee or Starbucks. These were then sub-divided into a positive contribution (e.g. increase of wages, a direct reference to decrease poverty) or to a negative contribution to achieving either SDG 1 or SDG 12 (e.g. inhumane treatment of workers, tax evasion, returning to plastic cups).
- To distinguish between related and unrelated articles it was agreed that unrelated articles should mention either Costa Coffee or Starbucks. The article should not reference any of the SDGs, as mentions of other SDGs could be used in further developing the system in future work, which means that unrelated articles for one SDG could also be used as unrelated articles for another SDG. Therefore, decreasing the workload of human annotation when incorporating another SDG into the system.

Training set assembly

Once the test set parameters had been specified, the training data could be assembled from the GigaWord corpus. After the selection of the two target SDGs, the next task was to construct a training set of news articles which were related to these particular goals. For that task, a lexical lookup system was developed (see appendix C). The system is provided with a list of keywords related to the SDG and a list of articles. The articles containing

more than 't' unique keywords are returned. Beginning with SDG 1, we created about 40 keywords: for t = 5, we retrieved 880 articles, t = 4 yielded more than 3,000 articles, t = 3 returned over 20,000 articles. At this point, the quality of the articles was unknown, so classification proceeded with all three thresholds, and evaluated for quality.

Once the set of related text items was established, the set was balanced with an equal number of news items from the same corpus which were not returned in the filtering process described above. This was done by randomly extracting texts from the corpus, dropping duplicates until an equal number of unrelated news items had been established. In order for a classifier to be able to process textual data, the data needs to be translated into a numerical data type. Common ways of doing this involve creating vector representations based on frequencies of words shared by the documents (e.g. Bag of Words vector representations). This approach was adopted as a baseline for comparison for the alternative approach.

More recent innovations in the field of Natural Language Processing involve the use of embeddings, d-length vector representations of words trained on (typically) language modelling tasks which create vector representations based on projections in d-dimensional space. A relatively modern approach involves the use of embeddings. Every word in a text is translated into an embedding, a d-dimensional vector projection of the word in question.

Even more recently, a new class of sequence-to-sequence transformer-based systems have emerged which go beyond representations of individual words and can offer vector-space representations of sentences. Since the SDG monitoring is needed at the document level (i.e. whether a news story is related to a particular SDG or not), it was required to have the same for the training data.

5.Data processing approach

5.1 RoBERTa

RoBERTa³ stand for Robustly optimized BERT approach and is a version of the BERT model (Liu et al., 2019). BERT(Bidirectional Encoder Representations from Transformers) is a language representation model. It uses a masked language model, which means that random tokens from the input get masked and the system has to predict the correct word based on the context. BERT uses a bidirectional approach to this task, taking in account the meaning of words left and right of the masked token. BERT is also trained on text-pair representations (Devlin et al., 2018). So in short, the interesting feature about BERT is that it takes the context into account when generating meanings of words. Furthermore, currently more language models are being added to BERT, which increases its usability.

RoBERTa differs from BERT by using byte pair encoding as a tokenizer. It also underwent more training and varies in the pre-training scheme. It seems it was helpful as it outperforms standard BERT (Liu et al., 2019) (see appendix D for side by side comparison of BERT and RoBERTa on same dataset).

To easily implement the RoBERTa model, HuggingFace sentence transformers⁴ (Wolf et al., 2019) were used. This provides a transformer that can simply call on various pretrained models. For this project the model 'roberta-large-nli-mean-tokens' was selected. This makes use of the large RoBERTa model, which is a 24 layer model, with 1024 hidden layers and 16 heads. It is tuned on Natural Language Inference data and uses mean-pooling.

It was decided not to train a RoBERTa model on our own data, because of the high capacity hardware or cloudspace it would require. Some time was spent trying to finetune an existing model, as this would be less intensive, however this did not yield any usable results as the current models are still changing.

5.2 Classifiers

The classifiers were all provided by the Sci-kit learn library. A selection was made; Naive Bayes, Random Forest, Support Vector Machine and Multilayer Perceptron. Naive Bayes was chosen, because it has been a popular algorithm in text classification, regardless of its simplicity (Rennie et al., 2003). Random Forest because it is used for classification tasks and should combat overfitting on the training data. Support Vector Machines as they are performing well on text classification tasks (Joachims,1998). Lastly, the multilayer perceptron, because it can focus more on the underlying patterns of the text, which might improve the system (Corrêa & Ludermir,

³ <https://github.com/pytorch/fairseq/tree/master/examples/roberta>

⁴ <https://github.com/huggingface/transformers>

2002). They were tested internally. It was found that the Support Vector Machine and the Multi-Layer Perceptron using stochastic gradient descent performed the best for both the document classifier and the sentiment classifier. However for the latter, the Naive Bayes was close behind. Based on these findings the SVM and MLP were selected to be trained on the full dataset and evaluated on the test data.

5.3 Baseline systems

The baseline systems differ from the proposed system by not using the BERT-embeddings. Instead the text was processed into tokens and fed to a Countvectorizer as a Bag of Words approach. Afterwards a Support Vector Machine was used to serve as the classifier.

Figure 1 below shows an overview of the major components of the first phase of the classification system.

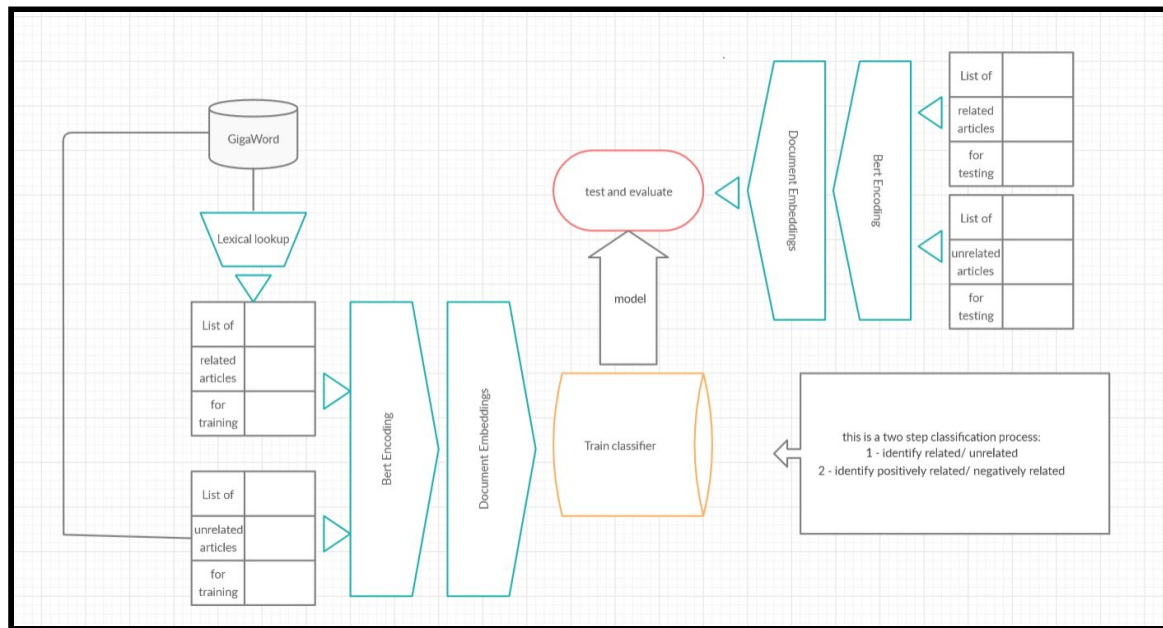


Fig.1 The GigaWord corpus is filtered using a lexical lookup approach to extract news articles related to a particular SDG. A training set is compiled which consists of a list of related articles and an equal number of unrelated articles. This is BERT encoded at the document level for introduction to a classifier which is subsequently evaluated on a test set, which has undergone a similar vectorization process.

6.Result analysis

6.1 The document classifier (related/ unrelated)

In fig 1 below, the classifiers were trained on a dataset containing 6741 unique news articles from the GigaWord corpus (50% related and 50% unrelated to SDG 1). The results show that both the Support Vector Machine (LinearSVM) and Multi-layer Perceptron (MLP) classifiers significantly outperform the Bag of Words baseline system, with the MLP classifier performing the best across the board.

Baseline	P	R	f1
related	1.00	0.61	0.76
unrelated	0.59	1.00	0.74
accuracy			0.75

SVM	P	R	f1
related	0.88	0.90	0.89
unrelated	0.82	0.78	0.80
accuracy			0.86

MLP	P	R	f1
related	0.90	0.93	0.92
unrelated	0.86	0.83	0.84
accuracy			0.89

fig. 1 System output for 3 classifiers: SVM, MLP and Baseline (BoW).
Classifier trained on SDG 1 data with keyword threshold filter set at 4 (training set ~6000 documents)

Baseline	P	R	f1
related	0.89	0.88	0.88
unrelated	0.81	0.83	0.82
accuracy			0.86

SVM	P	R	f1
related	0.88	0.89	0.88
unrelated	0.82	0.80	0.81
accuracy			0.86

MLP	P	R	f1
related	0.89	0.89	0.89
unrelated	0.83	0.83	0.83
accuracy			0.87

fig 2. System output for 3 classifiers: SVM, MLP and Baseline (BoW).
Classifier trained on SDG 12 data with keyword threshold filter set at 4 (training set ~6000 documents)

In fig 2 above, the classifiers were trained on a smaller dataset containing 2084 articles from the same dataset (again split 50:50 related:unrelated to SDG 12). In this case it is not immediately clear which system performs best, although once again the MLP-based classifier seems to have a slight edge on the competing systems.

In order to perform a qualitative analysis, both the input data and the output predictions were inspected for quality. What follows is a short summary of this analysis:

Training data - manual inspection

75 random articles from the training corpus labelled 'related' for SDG 1 and SDG 12 were inspected to confirm the accuracy of the automated labelling system based on a keyword filter with the threshold set at 4.

For SDG 1, It was determined that approximately 11% (8/75) were not actually related to the topic of poverty, in that the article happened to mention poverty, but the articles themselves were not particularly concerned with the topic.

For SDG 12 there were far few inconsistencies, with only 4% (3/75) of the news stories were not directly related to the SDG in question.

System output - manual inspection

SDG 1 - SVM

incorrectly classified 9/ 64

- 5 with low confidence (distance to hyperplane less than 0.3)
- In 2 cases the gold was mislabelled
- In 2 cases I do not see why the classifier misclassified

SDG 1 - MLP

incorrectly classified 7/ 64

- 1 with low confidence (55%)
- In 3 cases the gold was mislabeled
- In 3 cases I do not see why the classifier misclassified

(for detailed analysis see appendix E)

SDG 12 - SVM

13/91 misclassified

- 3 with confidence under 0.3
- 5 cases where i think an expanded definition of sustainable production might be required, since they might be included in this SDG despite being more strongly related to SDG 1.
- 4 cases are understandable (3 related to shark fin soup - which was not possible to classify given our narrow lexicon; 1 article not sufficiently elaborated in 5 sentences to be able to detect the topic)
- 1 case which is clearly misclassified

SDG 12 - MLP

13/91 misclassified

- 1 with low confidence (<60%)
- 5 cases where i think an expanded definition of sustainable production might be required, since they might be included in this SDG despite being more strongly related to SDG 1.
- 4 understandable cases (3 related to shark fin soup - which was not possible to classify given our narrow lexicon; 1 article not sufficiently elaborated in 5 sentences to be able to detect the topic)
- 2 cases which are clearly misclassified

(for detailed analysis see appendix F)

6.2 The sentiment classifier

The sentiment classifier does not perform as well as the document classifier. The second phase classifier is trained on a dataset of 506 elements. When focussed on the SDG 1, the baseline system barely outperforms random. As presented in table 1, all systems are an improvement on the baseline. The Support Vector machine has the best results, with the exception of the negative precision. However, when comparing the results of the classifiers on SDG 12, the Naive Bayes outperforms the others. The other classifiers all perform worse than their counterparts on SDG 1. The Multilayer Perceptron even has a lower recall for a positive polarity than the baseline system.

		SDG 1			SDG 12		
		Precision	Recall	F1	Precision	Recall	F1
NB	Positive	0.68	0.76	0.72	<u>0.94</u>	<u>0.76</u>	<u>0.84</u>
	Negative	<u>1.00</u>	0.58	0.74	<u>1.00</u>	<u>0.61</u>	<u>0.76</u>
SVM	Positive	<u>0.82</u>	<u>0.82</u>	<u>0.82</u>	0.85	0.61	0.71
	Negative	0.86	<u>0.75</u>	<u>0.80</u>	0.50	0.50	0.50
MLP	Positive	0.73	0.47	0.57	0.89	0.42	0.57
	Negative	0.82	0.58	0.68	0.53	0.56	0.54
Baseline	Positive	0.50	0.47	0.48	0.83	0.53	0.65
	Negative	0.47	0.29	0.36	0.47	0.50	0.49

Table 1: The results for the different systems per SDG. Neutral is ignored, as it is not present in the test set. Full classification reports can be found in Appendix G.

6.2.1 Error analysis

An error analysis was performed on the results of the different classifiers of SDG 1. The SVM had 9 misclassifications and the Multilayer Perceptron 19. Most mistakes were based on the classifier incorrectly tagging articles as neutral. Other errors were based on negative descriptions of problems, with the positive effect of the event not outweighing this; on the term 'at first sight'; describing positive events in past tense the classifier not picking up it has ended or were understandable and beyond the scope of a basic classifier. A more extensive report can be found in Appendix H and I.

7. Final application result: SDG-tracker

The end product of this research project is a system to find articles related to a certain company and SDG 1 or 12. The application is the first step towards a fully-fledged SDG-tracker. The obtained results mostly point towards the potential of further development of this research project. For the related/unrelated-classifier, the application was optimized for recall, so it misses as little relevant news articles as possible. For the sentiment classifier, the system was developed to focus on precision as once the article has been labeled related or unrelated recall is largely immaterial.

The system makes use of an embedding system, BERT, that has many different versions. There are some that are specially trained for different languages and there is also a multilingual version available. This should satisfy the clients need for a system that can process several different languages.

To make use of the SDG tracker, the following resources are required:

- A training corpus for the SDG-relation classifier. The classifier was trained on the English Gigaword corpus.⁵ It

⁵ <https://catalog.ldc.upenn.edu/LDC2003T05>

should be possible to provide the classifier with a different training set; the application assumes all training texts are concatenated together in one .json-file, with the article's headline and first 5 sentences in the 'text' column. If users are fond of their computer system, they could try to have the whole text parsed, but the runtime consequences will be vast.

- A training corpus for the sentiment classifier (we used the MPQA opinion corpus)⁶.
- Python 3.7; the required libraries and more information can be found in the requirements file on the GitHub page of our project.
- An embedding model. Our system uses RoBERTa-embeddings; the model specifications can be found in section 5.1.

8. Discussion and Conclusion

The results for the first phase classifier (categorising news stories according to their relevance to a particular SDG) were interesting. A simple question that can be answered readily pertains to whether the use of concatenated sentence level embeddings for the headline and first five sentences of a news story is a reasonable proxy for document-level representation. The results of using off-the-shelf BERT and RoBERTa vector representations without further fine-tuning or re-training are quite remarkable - with results consistently around 90% for SVM or MLP classifiers, despite the fact that these representations were trained on very different data. There is some indication that occasional errors in classification occur due to the fact that while the story itself was related, this was not captured in the first paragraph or so of the news article. However, the norms of the genre dictate that the most pertinent information would be at the beginning of the article, and the results were consistent in this regard. The same holds true for the sentiment classification.

Also with regards to the first phase classifier, although the Neural Network (MLP) classifier approach performed best for both SDGs tested, a qualitative inspection of results showed the strengths of the SVM classifier to perhaps be of great utility, especially if precision is of paramount importance. First of all, using the distance to the hyperplane parameter as a proxy for confidence, it was possible to be able to filter out a great deal of erroneous results by setting the threshold for this at around 0.3.

In the case of SDG 1, this would mean that five of the nine classification errors could be filtered out. Since two of the nine classification errors were actually mislabelled gold data, this would make fine tuning the SVM classifier simple. This is in contrast to the MLP classifier, which often gave high confidence scores when misclassifying, with only one of the seven misclassifications being able to be filtered out in this way (once again, this seven actually contained three mislabelling errors according to the inspector who was not directly involved in the test-set creation).

In the case of SDG 12, 3 of the 13 misclassifications for the SVM could be filtered out using a distance function of > 0.3 as a threshold for quality, post classification, whereas once again, only one of the MLP misclassification instances could be resolved in this way.

This approach however would be a double edged sword, with positive results also being culled by the SVM, however if precision was the primary concern, this would be a simple way to refine the results for the SVM.

The second phase classifier does perform slightly worse than the first phase one. The SVM receives the best results for SDG 1. However, for SDG 12 the Naive Bayes is the most optimal. Based on the results for both classifiers it seems ideal to use the Naive Bayes classifier as a general system to classify the polarity of the articles. It also seems that the inclusion of the neutral label leads to unnecessary errors. It can be argued that there is never a neutral mention of a company in relation to the SDGs, therefore it might be possible to drop the neutral labels from the training data to get a better result from the second phase classifier.

It may seem surprising that the gold test-set data was mislabelled in several instances, but this speaks more to the interdependent nature of the SDGs. The SDGs were not designed to be distinct and it is difficult when tasked with collecting texts related to one specific SDG to not select articles that at first glance do not appear to be related to that particular SDG. Take for example an article that reports how a farmer who was growing coffee was faced with difficulties when the price of coffee crashed. That article is clearly related to SDG 1, poverty, but is it

⁶ http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/

not also related to SDG 12 ,sustainable production, since this farmer is mono-cropping, which could be easily interpreted as unsustainable production. However, according to that logic almost any report about coffee farmers could be seen through this lens, and ultimately is due to the subjective knowledge and beliefs of the individual annotator.

Some of the classification errors were labelled as 'understandable' in part due to the fact that the original lexicon upon which the classifier was constructed, was not designed with a particular issue in mind. For example, several articles which related to Starbucks selling shark's fin soup, in hindsight is clearly related to sustainability. Yet looking at the lexicon used for SDG 12 (see appendix C) , the creator(s) of the lexicon had a much narrower definition of sustainability in mind (waste, recycling etc.). This speaks somewhat to the rather arbitrary approach to selection of the lexicon. The lexicon selection is clearly an area which requires a more refined approach, whether an iterative approach or otherwise.

The contrast in results of the inspection of the automated labelling of training data, is also noteworthy. That an automated search based on keywords consisting of bigrams and unigrams occasionally returns unrelated texts is not surprising, but that results for SDG 12 were much more consistent than for SDG 1 gives pause for thought, especially since: 1) the lexicon for SDG 12 was significantly smaller, with a narrow focus (as mentioned above); 2) the threshold for number of unique keywords needing to be present was lower in the SDG 12 training dataset (t=3) compared to the training set for SDG 1 (t=4). There may be some effect by having a lexicon with a narrow focus, and further experimentation might reveal that a limited, highly focussed lexicon requires a smaller threshold than a more broad ranging one.

Future work

One area for improvement is on the training data. The way the data is collected for the training set for the document classifier is not optimal. As touched upon above there were some unrelated articles in the set that were tagged as related, but were in fact unrelated. A way to combat this might be to create a more objective way to collect a lexicon for the keyword look-up, using either an iterative process of examining resulting articles to extract more related words; or as suggested above, using similarity scores from word embeddings to generate keywords. Further experimentation would have to take place to find appropriate thresholds.

The training data for the sentiment classifier can be improved as well. The data now is an aggregation of sentiments on a sentence level. It is possible this resulted in loss of information. Additionally, the inclusion of the neutral label seems to lead to some errors in classifying the evaluation data. The system might improve when this label is dropped, as it is currently not used.

One further improvement would relate to the RoBERTa representations used. These can be fine-tuned on relevant datasets; in this case news articles. If the hardware is available, training one of these models from scratch would be recommended, since this will ensure the embeddings are most optimised for news articles.

Once these modifications are in place, adjustments can be made to further expand the system. Adding more SDG's, preferably completing the set of seventeen would make a complete system. However, that said, the system might benefit from redefining the Sustainable Development Goals themselves, since they overlap in their missions. It would increase usability if one classifier could classify to which if any of the goals an article is related to, and it would be a neater system if the SDGs were made distinct.

Once these wrinkles are ironed out, combining the document classifier and the sentiment classifier to some visualiser that automatically adjusts an SDG score for a company based on an news input stream would see the system complete, although at the time of writing, this seems a distant goal.

9. Wordcount

6146

10. References

Corrêa, R. F., & Ludermit, T. B. (2002, November). Automatic text categorization: case study. In *VII Brazilian Symposium on Neural Networks, 2002. SBRN 2002. Proceedings.* (p. 150). IEEE.

Deng, L., & Wiebe, J. (2015). Mpqa 3.0: An entity/event-level sentiment corpus. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1323-1328).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

dpicampaigns (2019). "About the Sustainable Development Goals". *United Nations Sustainable Development*.

English Gigaword Fifth Edition : <https://catalog ldc.upenn.edu/LDC2011T07>

Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O. Lewis, M. Zettlemoyer, L. & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 616-623).

Starbucks Coffee Company – Communication on Progress: UN Global Compact. (n.d.). Retrieved March 27, 2020, from <https://www.unglobalcompact.org/participation/report/cop/create-and-submit/active/350961>

The Coca-Cola Company. (n.d.). Retrieved March 27, 2020, from <https://www.unglobalcompact.org/what-is-gc/participants/9195>

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R. Funtowicz, M. & Brew, J. (2019). Huggingface's transformers: State-of-the-art natural language processing. *ArXiv, abs/1910.03771*.

Appendix A - list of Sustainable Development Goals

1. No Poverty
2. Zero Hunger
3. Good Health and Well-being
4. Quality Education
5. Gender Equality
6. Clean Water and Sanitation
7. Affordable and Clean Energy
8. Decent Work and Economic Growth
9. Industry, Innovation, and Infrastructure
10. Reducing Inequality
11. Sustainable Cities and Communities
12. Responsible Consumption and Production
13. Climate Action
14. Life Below Water
15. Life On Land
16. Peace, Justice, and Strong Institutions
17. Partnerships for the Goals

Appendix B - news sources for GigaWord

- Agence France-Presse, English Service
- Associated Press Worldstream, English Service
- Central News Agency of Taiwan, English Service
- Los Angeles Times/Washington Post Newswire Service
- Washington Post/Bloomberg Newswire Service
- New York Times Newswire Service
- Xinhua News Agency, English Service

Appendix C - keywords used for lexical lookup

SDG 1 - Keywords used:

less than 1.25 a day social protection protection systems protection measures vulnerable equal rights economic resources poor ownership control property resilience social security aid	exposure shocks disasters mobilization development cooperation adequate means policy frameworks pro-poor poverty eradication human rights working conditions inhuman development poverty	workers fair wage extortion slavery slave famine minimum wage child labor hunger self-governance entrepreneur free education low wage natural disaster
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

SDG12 - Keywords used:

procurement awareness harmony with nature sustainable tourism fossil-fuel subsidies local culture local products	local production waste footprint material extraction hazardous waste landfill dumping
------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------

Appendix D - Comparison of BERT and RoBERTa on the same training set (SDG1, t=4)

BERT (RoBERTa)-based System - SVM					BERT - NLI-based System				
	precision	recall	f1-score	support		precision	recall	f1-score	support
related	0.88	0.90	0.89	41	related	0.83	0.83	0.83	41
unrelated	0.82	0.78	0.80	23	unrelated	0.70	0.70	0.70	23
accuracy			0.86	64	accuracy			0.78	64
macro avg	0.85	0.84	0.85	64	macro avg	0.76	0.76	0.76	64
weighted avg	0.86	0.86	0.86	64	weighted avg	0.78	0.78	0.78	64

Appendix E - Error analysis - system predictions: SDG 1

SVM-classifier

incorrectly classified 9/ 64

- 5 with low confidence (distance to hyperplane less than 0.3)
- In 2 cases the inspector agreed with the classifier
- In 2 cases the reason for misclassification was not clear to the human annotator

Text:	System prediction:
Coronavirus Closed This Starbucks, but Seattle Needs Its Coffee SEATTLE â€” Machele Miely, on her way to meet a friend for coffee in downtown Seattle on Saturday, stopped to snap a photo on her phone of a sign posted on the door of a Starbucks saying that the store was temporarily closed and that it was sorry for the inconvenience. The sign said nothing about why it was closed, but Ms. Miely knew the reason: coronavirus. â€œPeople are trying not to panic, while preparing to panic,â€ said Ms. Miely, 41, who works as a project manager for a construction company and had a small bottle of hand sanitizer clipped to her belt â€” a must-have accessory in recent days. The coronavirus has hit the Seattle area harder than anywhere else in the United States, with 16 deaths in Washington State, most from an outbreak at a nursing home northeast of the city. The virus and its public health implications and consequences have rippled out, affecting how and where people work and where and when they gather in groups.	Related Confidence 1.4 (high)
Starbucks Barista Fired After Officerâ€™s Cup Had â€”Pigâ€™ on the Label An Oklahoma police officer working on Thanksgiving made a Starbucks run for emergency dispatchers to thank them for working during the holiday. When he got his order, he said, he saw the word â€œPIGâ€ printed on a cupâ€™s label. The officer, a member of the Kiefer, Okla., police department, took the order to the dispatchers and contacted the police chief, Johnny Oâ€™Mara, who was on vacation at the time, according to Jory Mendes, a Starbucks spokesman. The word â€œPIGâ€ was printed on all five cups, KTUL-TV reported. When Chief Oâ€™Mara called the Starbucks, in Glenpool, Okla., the manager offered to reprint the cups, which the chief said was an insufficient apology, the station reported.	Related Confidence 1.5 (high)

MLP

incorrectly classified 7/ 64

- 1 with low confidence (55%)
- In 3 cases the inspector agreed with the classifier
- In 3 cases it was unclear to the human annotator why the classifier misclassified:

Text:	System prediction:
Is Starbucks hurting Ethiopian Coffee Farmers? ADDIS ABABA, Ethiopia Ethiopia and British charity Oxfam on Thursday accused Starbucks of stopping the Horn of Africa country trademarking its coffee, denying farmers potential income of about \$94 million.Oxfam said the U.S. coffee shop giant, which had turnover of \$7.8 billion in the year to Oct. 1, prevented Ethiopia from securing trademark protection for two of its best-known beans, Sidamo and Harar.Had Ethiopia, one of the worldâ€™s poorest countries, been successful, it would have allowed the country to control the use of the beans in the market, giving its farmers more of the retail price and securing an estimated extra 48 million pounds, the charity said.â€œSecuring the trademark for its Sidamo, Harar and Yirgacheffe coffee beans could have allowed the country to increase its negotiation leverage through control of the names and ultimately drive a greater share of the retail price in the global market,â€ Ethiopiaâ€™s Foreign Ministry said in a statement.Starbucks denied being behind the blocking bid by the U.S. National Coffee Association (NCA) at the U.S. Patent and Trademark Office (USPTO).	Unrelated Confidence 95.6%

<p>Coronavirus Closed This Starbucks, but Seattle Needs Its Coffee SEATTLE â€” Machele Miely, on her way to meet a friend for coffee in downtown Seattle on Saturday, stopped to snap a photo on her phone of a sign posted on the door of a Starbucks saying that the store was temporarily closed and that it was sorry for the inconvenience. The sign said nothing about why it was closed, but Ms. Miely knew the reason: coronavirus. â€œPeople are trying not to panic, while preparing to panic,â€ said Ms. Miely, 41, who works as a project manager for a construction company and had a small bottle of hand sanitizer clipped to her belt â€” a must-have accessory in recent days. The coronavirus has hit the Seattle area harder than anywhere else in the United States, with 16 deaths in Washington State, most from an outbreak at a nursing home northeast of the city. The virus and its public health implications and consequences have rippled out, affecting how and where people work and where and when they gather in groups.</p>	<p>Related</p> <p>Confidence 99.3%</p>
<p>Starbucks Barista Fired After Officerâ€™s Cup Had â€”Pigâ€” on the Label An Oklahoma police officer working on Thanksgiving made a Starbucks run for emergency dispatchers to thank them for working during the holiday. When he got his order, he said, he saw the word â€œPIGâ€ printed on a cupâ€™s label. The officer, a member of the Kiefer, Okla., police department, took the order to the dispatchers and contacted the police chief, Johnny Oâ€™Mara, who was on vacation at the time, according to Jory Mendes, a Starbucks spokesman. The word â€œPIGâ€ was printed on all five cups, KTUL-TV reported. When Chief Oâ€™Mara called the Starbucks, in Glenpool, Okla., the manager offered to reprint the cups, which the chief said was an insufficient apology, the station reported.</p>	<p>Related</p> <p>Confidence 100%</p>

Cases where inspector agreed with the classifier

<p>Starbucks C.A.F.E. Practices vs. Fair Trade Certified Coffee Hello supporters of all things fair trade! My name is Jack Ramaika and I am a student here at PSU Brandywine. I am also a lover of coffee, which would lead me to find a job at Starbucks Coffee. I have been a partner at Starbucks since 2006. In my time at Starbucks, I have been very proud of the company and the principles on which we stand. When I joined the Fair Trade movement at PSU Brandywine, I began to take a serious look at how Starbucks purchases and deals with the producers of their coffee beans.</p>	<p>related</p> <p>perhaps later in the article the connection to poverty becomes clear</p> <p>also not news</p>
<p>Starbucks Falls Short After Pledging Better Labor Practices Last year, Starbucks vowed to provide store employees with more consistent schedules from week to week, and to post their schedules at least 10 days in advance. The company said it would stop asking workers to endure the sleep-depriving ritual known as a â€œclopensing,â€ which requires them to shut down a store at night only to return early the next morning to help open it. But Starbucks has fallen short on these promises, according to interviews with five current or recent workers at several locations across the country. Most complained that they often receive their schedules one week or less in advance, and that the schedules vary substantially every few weeks. Two said their stores still practiced clopenings.</p>	<p>related</p> <p>not immediately obviously related to poverty</p>
<p>From Starbucks to FedEx, Coronavirus Upends Businesses That Depend on China The coronavirus outbreak in China has rattled the global economy, disrupting virtually every major industry, from food, fashion and entertainment to automobiles and technology. And more than two weeks into the crisis, there is no sign that the economic impact is about to ease up. Much of China was supposed to have reopened by now. But the countryâ€™s empty streets, quiet factories and legions of idle workers suggest that weeks or months may pass before this motor of global growth is running at full speed. Over the years, companies as varied as Disney, Nike, McDonaldâ€™s and Hyundai have come to rely on Chinaâ€™s efficient factories and increasingly affluent consumers.</p>	<p>unrelated</p> <p>i would say this is at least tangentially related to poverty although not Starbucks as cause of poverty</p>

Appendix F - Error analysis - system predictions: SDG 12

SVM

13/90 misclassified

- 3 with confidence under 0.3
- 5 cases where i think an expanded definition of sustainable production might be required, since they might be included in this SDG despite being more strongly related to SDG 1.
- 4 cases are understandable (3 related to shark fin soup - which was not possible to classify given our narrow lexicon; 1 article not sufficiently elaborated in 5 sentences to be able to detect the topic)
- 1 case which is clearly misclassified

Example of clearly misclassified text

Text:	System prediction:
Starbucks and Our Plastic Pollution Problem. The globe is covered in plastic waste. The "to-go" coffee culture is a big part of this problem. Take Starbucks, for example. Starbucks uses more than 8,000 paper cups a minute, which adds up to more than four billion a year. 1.6 million trees are harvested every year for all of those single-use cups. Since these cups are lined with plastic, they are not really recyclable â€” only four U.S. cities even accept Starbucks paper cups for recycling.	unrelated confidence : 0.51 (low- mid)

Examples of texts which were found to be (indirectly) related to SDG12 upon reflection

Text:	System prediction:
Starbucks, home of the \$4 latte, is moving into poor areas. Starbucks has a point to prove: Thereâ€™s more to the company than selling \$4 lattes to rich people. The Seattle-based coffee giant that has cultivated a reputation for being socially responsible said Thursday it is expanding its effort to put more coffee shops â€” and create more jobs â€” in poor neighborhoods. Starbucks plans to open or remodel 85 stores by 2025 in rural and urban communities across the U.S. Each store will hire local staff, including construction crews and artists, and have community event spaces. The company will also work with local United Way chapters to develop programs at each shop, such as youth job training classes and mentoring. The effort will bring to 100 the number of â€œcommunity storesâ€ Starbucks has opened since it announced the program in 2015.	related confidence: 0.58 (mid)
If You Love Coffee... You Can Help Eradicate Poverty. Every day Americans enjoy more than 580 million cups of coffee, while worldwide, more than 3 billion cups are consumed. Coffee is produced by more than 26 million growers, with over 250 million people in coffee producing countries depending directly or indirectly on coffee production for their daily survival. Sadly, most of them live in poverty or in extreme poverty. Without the daily struggle of the growers and their families there would be no beans for the coffee enjoyed daily by one billion coffee lovers around the world. In the last three years drought and crop pests like coffee rust have caused losses worth billions of dollars to coffee farmers in Latin America alone.	related confidence: 0.72 (mid)
â€”Thereâ€™s no money in coffee any moreâ€”: Guatemalans driven to US migration. From his wooden hut in the foothills of the Sierra Madre, Rodrigo Carrillo can see the product of his life savings: a vast green sea of coffee plants, sprouting red berries like tiny Christmas ornaments. Those plants once seemed like a life-changing investment. Carrillo joined a cooperative that sells beans to Starbucks and several certified fair-trade organisations. In Guatemalaâ€™s fertile highlands, there was no faster way out of poverty than to supply American coffee drinkers. But in recent years, the price of coffee has crashed, leaving Carrillo, 48, with a choice to make.	related confidence : 0.84 (mid) is this an example of unsustainable production
Slave Labor Discovered At Starbucks, Verified Coffee Plantation In Brazil. The farm appears to be a model property. Itâ€™s Brazil, and the CÃ³rego das Almas farm in Piumhi, in rural Minas Gerais state has signs posted that slave labor is not allowed. These signs also display international certifications as testament to the fact that the labor practices and production values on the farm are above reproach. Mongabay reports on the	related confidence: 0.35 (low)

possible illegal conditions uncovered at the farm. Brazil Labor Ministry officials raided the farm and rescued 18 workers who were living and working there under conditions akin to slavery. The farm holds the C.A.F.E. Practices certification, which is owned by Starbucks in a partnership with SCS Global Services.	is slave labour sustainable?
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------

MLP

13/91 misclassified

- 1 with low confidence (<60%)
- 5 cases where i think an expanded definition of sustainable production might be required, since they might be included in this SDG despite being more strongly related to SDG 1.
- 4 understandable cases (3 related to shark fin soup - which was not possible to classify given our narrow lexicon; 1 article not sufficiently elaborated in 5 sentences to be able to detect the topic)
- 2 cases which are clearly misclassified

Misclassified

Text:	System prediction:
Starbucks and Our Plastic Pollution Problem. The globe is covered in plastic waste. The "to-go" coffee culture is a big part of this problem. Take Starbucks, for example. Starbucks uses more than 8,000 paper cups a minute, which adds up to more than four billion a year. 1.6 million trees are harvested every year for all of those single-use cups. Since these cups are lined with plastic, they are not really recyclable – only four U.S. cities even accept Starbucks paper cups for recycling.	related confidence : 90%
Starbucks launches vegan pumpkin spice latte sauce in Europe. Pumpkin Spice Season is about to get more vegan, at least in Europe. Coffee giant Starbucks launched a new vegan pumpkin sauce with spice flavouring in all its stores located across the continent. The Starbucks mix is not a vegan product since it comes from condensed milk, which left many customers clamouring for a vegan option. This included a formal request in 2013 that received the support of over 11,000 people. The launch comes right after Starbucks presented two new cold brew coffee protein smoothies for the U.S market.	related confidence: 67%

Appendix G - Classification reports sentiment classifier

Sentiment Classifier SDG 1:

Naive Bayes - RoBERTa				
	precision	recall	f1-score	support
negative	1.00	0.58	0.74	24
neutral	0.00	0.00	0.00	0
positive	0.68	0.76	0.72	17
micro avg	0.66	0.66	0.66	41
macro avg	0.56	0.45	0.49	41
weighted avg	0.87	0.66	0.73	41

SVM - RoBERTa				
	precision	recall	f1-score	support
negative	0.86	0.75	0.80	24
neutral	0.00	0.00	0.00	0
positive	0.82	0.82	0.82	17
micro avg	0.78	0.78	0.78	41
macro avg	0.56	0.52	0.54	41
weighted avg	0.84	0.78	0.81	41

MLP (stochastic gradient descent) - RoBERTa				
	precision	recall	f1-score	support
negative	0.82	0.58	0.68	24
neutral	0.00	0.00	0.00	0
positive	0.73	0.47	0.57	17
micro avg	0.54	0.54	0.54	41
macro avg	0.52	0.35	0.42	41
weighted avg	0.78	0.54	0.64	41

Baseline				
	precision	recall	f1-score	support
negative	0.47	0.29	0.36	24
neutral	0.00	0.00	0.00	0
positive	0.50	0.47	0.48	17
micro avg	0.37	0.37	0.37	41
macro avg	0.32	0.25	0.28	41
weighted avg	0.48	0.37	0.41	41

Sentiment Classifier SDG 12:

Naive Bayes - RoBERTa				
	precision	recall	f1-score	support
negative	1.00	0.61	0.76	18
neutral	0.00	0.00	0.00	0
positive	0.94	0.76	0.84	38
micro avg	0.71	0.71	0.71	56
macro avg	0.65	0.46	0.53	56
weighted avg	0.96	0.71	0.81	56

SVM - RoBERTa				
	precision	recall	f1-score	support
negative	0.50	0.50	0.50	18
neutral	0.00	0.00	0.00	0
positive	0.85	0.61	0.71	38
micro avg	0.57	0.57	0.57	56
macro avg	0.45	0.37	0.40	56
weighted avg	0.74	0.57	0.64	56

MLP (stochastic gradient descent) - RoBERTa				
	precision	recall	f1-score	support
negative	0.53	0.56	0.54	18
neutral	0.00	0.00	0.00	0
positive	0.89	0.42	0.57	38
micro avg	0.46	0.46	0.46	56
macro avg	0.47	0.33	0.37	56
weighted avg	0.77	0.46	0.56	56

Baseline				
----------	--	--	--	--

	precision	recall	f1-score	support
negative	0.47	0.50	0.49	18
neutral	0.00	0.00	0.00	0
positive	0.83	0.53	0.65	38
micro avg	0.52	0.52	0.52	56
macro avg	0.44	0.34	0.38	56
weighted avg	0.72	0.52	0.59	56

Appendix H - Error analysis sentiment classifier

Sentiment Classifier Error analysis :

SDG 1 (Costa Coffee & Starbucks)

SVM

- 9 in total misclassified
- 3 misclassified as negative which were positive
- 3 misclassified as positive which were negative
- 3 misclassified as Neutral which were negative

Text:	misclassified as:
<p>1. If You Love Coffee... You Can Help Eradicate Poverty Every day Americans enjoy more than 580 million cups of coffee, while worldwide, more than 3 billion cups are consumed. Coffee is produced by more than 26 million growers, with over 250 million people in coffee producing countries depending directly or indirectly on coffee production for their daily survival. Sadly, most of them live in poverty or in extreme poverty. Without the daily struggle of the growers and their families there would be no beans for the coffee enjoyed daily by one billion coffee lovers around the world. In the last three years drought and crop pests like coffee rust have caused losses worth billions of dollars to coffee farmers in Latin America alone.</p> <p>2. How Starbucks and Green Mountain serve up shared value A shocking dichotomy exists in our food system. The more than 70 percent of the global population that is food insecure are the very people we rely on to produce about 70 percent (PDF) of the world's food supply. They are typically small farmers living in rural areas across the globe. Today is World Food Day, the opportunity to highlight the call for action to end hunger for the 1 billion currently hungry (PDF) and to fulfill our future population's food needs. This year's theme of family farming acknowledges the crucial role these farms play. Building the capacity of smallholder producers is necessary to empower them to move out of poverty and to eradicate hunger.</p> <p>3. Costa Coffee owner to raise prices in response to national living wage Whitbread, the owner of Premier Inn hotels and Costa Coffee, said it expected to cut spending and increase some prices to counter the a substantialâ€ cost of the national living wage, as it posted slowing second-quarter sales. The chancellor, George Osborne, announced a bumper pay increase in July, with the current 6.50 minimum wage set to rise to 7.20 for those aged over 25 from next April. The rate will grow steadily over the following four years to about 9.35 an hour."</p>	<p>Gold Positive</p> <p>SVM Negative</p>

Possible explanation for misclassifications

Text 1) The gold label here is wrong and the classifier worked correctly

Text 2) The classifier was incorrect due to the negative background description of the problem that the overall positive connotation of World Food Day would have towards SDG 1.

Text 3) Unclear misclassification

Text:	misclassified as:
<p>1. Slave labor found at Starbucks-certified Brazil coffee plantation At first sight, the C�lrego das Almas farm in Plumihi, in rural Minas Gerais state, seems to be a model property. No slave or forced labor is allowed, reads one of several signs that display international certifications including one linked to the U.S.-based company Starbucks corporation.But investigators have found that laborers on the farm coffee plantations were working under degrading conditions and living in substandard housing without sewerage or drinking water. A Ministry of Labor team inspection conducted at the site rescued 18 rural workers in conditions analogous to slavery.The farm, locally known as Fatura (Portuguese for Abundance), also boasts the UTZ seal a Netherlands-based sustainable farming certificate considered one of the most prestigious in the coffee industry.</p>	Positive
<p>2. There is no money in coffee any more: Guatemalans driven to US migration From his wooden hut in the foothills of the Sierra Madre, Rodrigo Carrillo can see the product of his life savings: a vast green sea of coffee plants, sprouting red berries like tiny Christmas ornaments. Those plants once seemed like a life-changing investment. Carrillo joined a cooperative that sells beans to Starbucks and several certified fair-trade organisations. In Guatemala��'s fertile highlands, there was no faster way out of poverty than to supply American coffee drinkers. But in recent years, the price of coffee has crashed, leaving Carrillo, 48, with a choice to make.</p>	
<p>3. Children as young as eight picked coffee beans on farms supplying Starbucks High street coffee shop giant Starbucks has been caught up in a child labour row after an investigation revealed that children under 13 were working on farms in Guatemala that supply the chain with its beans.Channel 4 Dispatches filmed the children working 40-hour weeks in gruelling conditions, picking coffee for a daily wage little more than the price of a latte. The beans are also supplied to Nespresso, owned by Nestle. Last week, actor George Clooney, the advertising face of Nespresso, praised the investigation and said he was saddened by its findings. Channel 4 Dispatches filmed the children working 40-hour weeks in gruelling conditions, picking coffee for a daily wage little more than the price of a latte. The beans are also supplied to Nespresso, owned by Nestle. Last week, actor George Clooney, the advertising face of Nespresso, praised the investigation and said he was saddened by its findings.</p>	

Possible explanation for misclassifications

Text 1) The idiomatic phrase “at first sight” giving the description as first found is the logical reason for this error.

Text 2) Only the last line is negative and the introduction describes a situation that no longer exists. The full story although being negative uses positive language to describe the past situation.

Text 3) Quite obviously negative the inclusion of the celebrity George Clooney should not have impacted the classifier. Difficulty to explain misclassification in this instance.

Neural Network classifier

- 19/40 in total misclassified
- 3 misclassified as negative which were positive
- 3 misclassified as positive which were negative
- 6 misclassified as neutral which were positive
- 7 misclassified as neutral which were negative

- 13 in total classified as false neutral

Text:	misclassified as:
<p>text 1 Costa Coffee owner to raise prices in response to national living wage Whitbread, the owner of Premier Inn hotels and Costa Coffee, said it expected to cut spending and increase some prices to counter the substantial cost of the national living wage' as it posted slowing second-quarter sales. The chancellor, George Osborne, announced a bumper pay increase in July, with the current 6.50 minimum wage set to rise to 7.20 for those aged over 25 from next April. The rate will grow steadily over the following four years to about 9.35 an hour."</p> <p>Text 2 Whitbread commits to 100% renewable energy Whitbread, the UK's largest hospitality company, has switched its entire energy demand to renewable sources. From 1st April 2017 all Whitbread brands, including Costa Coffee and Premier Inn, have been sourcing their purchased electricity in the UK from renewable sources. The decision means that the energy used to power 68,000 Premier Inn rooms and 2,200 Costa Coffee stores serving more than 28 million customers every month across the UK, is now coming from 100% renewable sources.</p> <p>Text 3 If You Love Coffee... You Can Help Eradicate Poverty Every day Americans enjoy more than 580 million cups of coffee, while worldwide, more than 3 billion cups are consumed. Coffee is produced by more than 26 million growers, with over 250 million people in coffee producing countries depending directly or indirectly on coffee production for their daily survival. Sadly, most of them live in poverty or in extreme poverty. Without the daily struggle of the growers and their families there would be no beans for the coffee enjoyed daily by one billion coffee lovers around the world. In the last three years drought and crop pests like coffee rust have caused losses worth billions of dollars to coffee farmers in Latin America alone.</p>	<p>Gold positive</p> <p>NN Negative</p> <p>confidence NN</p> <p>text 1 92 % <i>neg</i> 1 % <i>pos</i></p> <p>text 2 46 % <i>neg</i> 44% <i>pos</i></p> <p>text 3 41 % <i>neg</i> 39 % <i>pos</i></p>

Possible explanation for misclassifications

Text 1) The text mentions; raising prices, cost of living increase and bumper pay increase. The classifier is very confidently labeling it negative which does not really make sense as the increase in worker wages is a positive contribution to reducing poverty and the text overall should be labeled positive

Text 2) As the text should in first instance not be marked related as it does not overlap with SDG 1 but more with SDG7. Also our own manually annotated label is wrong, it should have been neutral which the classifier sort of indicates with a similar negative as positive score.

Text 3) Understandably the classifier marked the text negative through; extreme poverty, losses of billions, living in poverty and daily survival. But the conveyed meaning should have been positive as the text states from the start that drinking coffee will eradicate poverty

Text:	misclassified as:
<p>text 1</p> <p>Slave labor found at Starbucks-certified Brazil coffee plantation At first sight, the C��rrego das Almas farm in Piumhi, in rural Minas Gerais state, seems to be a model property. No slave or forced labor is allowed reads one of several signs that display international certifications" including one linked to the U.S.-based company Starbucks corporation. But investigators have found that laborers on the farms coffee plantations were working under degrading conditions and living in substandard housing without sewerage or drinking water. A Ministry of Labor team inspection conducted at the site rescued 18 rural workers in conditions analogous to slavery. The farm, locally known as Fartura (Portuguese for Abundance), also boasts the UTZ seal a Netherlands-based sustainable farming certificate considered one of the most prestigious in the coffee industry.</p> <p>Text 2</p> <p>Theres no money in coffee any more: Guatemalans driven to US migration From his wooden hut in the foothills of the Sierra Madre, Rodrigo Carrillo can see the product of his life savings: a vast green sea of coffee plants, sprouting red berries like tiny Christmas ornaments. Those plants once seemed like a life-changing investment. Carrillo joined a cooperative that sells beans to Starbucks and several certified fair-trade organisations. In Guatemala����s fertile highlands, there was no faster way out of poverty than to supply American coffee drinkers. But in recent years, the price of coffee has crashed, leaving Carrillo, 48, with a choice to make.</p> <p>Text 3</p> <p>Slave Labor Discovered At Starbucks, Verified Coffee Plantation In Brazil The farm appears to be a model property. Brazil, and the Carego das Almas farm in Piumhi, in rural Minas Gerais state has signs posted that slave labor is not allowed. These signs also display international certifications as testament to the fact that the labor practices and production values on the farm are above reproach. Mongabay reports on the possible illegal conditions uncovered at the farm. Brazil Labor Ministry officials raided the farm and rescued 18 workers who were living and working there under conditions akin to slavery. The farm holds the C.A.F.E. Practices certification, which is owned by Starbucks in a partnership with SCS Global Services.</p>	<p>Gold negative</p> <p>NN positive</p> <p>confidence NN</p> <p>text 1 0.2 % neg 96 % pos</p> <p>text 2 2 % neg 87% pos</p> <p>text 3 8 % neg 63 % pos</p>

Possible explanation for misclassifications

text 1) The idiomatic phrase "at first sight" giving the description as first found is the logical reason for this error. (same as SVM mis-classification)

text 2) Only the last line is negative and the introduction describes a situation that no longer exists. The full story although being negative uses positive language to describe the past situation. (same as SVM mis-classification)

text 3) Mentions the same scandal as text 1, also the key idiomatic phrase "at first sight" explains the misclassification.

Text:	misclassified as:
<p>text 1</p> <p>Starbucks, home of the \$4 latte, is moving into poor areas Starbucks has a point to prove: There's more to the company than selling \$4 lattes to rich people. The Seattle-based coffee giant that has cultivated a reputation for being socially responsible said Thursday it is expanding its effort to put more coffee shops and create more jobs in poor neighborhoods. Starbucks plans to open or remodel 85 stores by 2025 in rural and urban communities across the U.S. Each store will hire local staff, including construction crews and artists, and have community event spaces. The company will also work with local United Way chapters to develop programs at each shop, such as youth job training classes and mentoring. The effort will bring to 100 the number of community stores Starbucks has opened since it announced the program in 2015.</p>	<p>Gold</p> <p>text 1 positive</p> <p>text 2 negative</p> <p>text 3 negative</p>
<p>Text 2</p> <p>Slave Labor Found at Second Starbucks-Certified Coffee Farm In July of last year, Brazilian labor inspectors found six employees on the Cedro II farm in Minas Gerais state were subjected to slave labor. The coffee plantation had been certified by both Starbucks and Nestle-affiliated brand Nespresso. Yet when the farm landed on Brazil's Dirty List, a government-backed listing of companies and employers in Brazil linked to slavery, Starbucks and Nespresso announced they would no longer source from the farm. The newest Dirty List, published in April 2019, adds 48 new employers. Among them is a second coffee producer, Helvacio Sebastiao Batista, which had been certified with Nespresso and Starbucks quality seals and used to provide coffee for both brands.</p>	<p>NN</p> <p>Neutral</p> <p>confidence NN</p>
<p>Text 3</p> <p>Starbucks pays 18.3m tax but 348m in dividends Starbucks UK-based European business paid just 18.3m in tax last year, while paying the coffee giants parent company in Seattle 348m in dividends collected from licensing its brand. The company, which has faced years of heated criticism for paying very little tax in the UK, reported on Thursday that its European, Middle Eastern and Asian business (EMEA) paid \$23.6m (18.3m) in tax on \$230m of royalty payment earnings in the year to the end of September 2018. Starbucks EMEA, which collects royalty income from 38 countries from Norway to Morocco, said it paid an effective tax rate of 23.7% above the UK's corporate tax rate of 19%. At the same time, Starbucks EMEA sent \$448m to its global headquarters in Seattle. The money is being brought back to the US following president Donald Trump's tax holiday on overseas earnings.</p>	<p>text 1 1% <i>neg</i> 59% <i>neu</i> 40% <i>pos</i></p> <p>text 2 18% <i>neg</i> 80% <i>neu</i> 2% <i>pos</i></p> <p>text 3 0% <i>neg</i> 97% <i>neu</i> 3% <i>pos</i></p>

Possible explanation for misclassifications

An overall explanation could be that in the gold data no neutral labels were assigned where the neural network classifier does assign neutral labels

Text 1) The classifier picked up a large fraction of the positive connotations (40% positive) but marked the overall text as neutral. This is understandable as the text states facts about Starbucks development. However, the overall text should be marked positive as all developments have a positive effect on reaching SDG 1

Text 2) As the text explains events already happened it is understandable it is marked neutral, however with the slave labour found at plantations it should have been marked negative

Text 3) As the text does not make use of very negative words like "tax evasion" or "slave labor" the overall sentiment points in the direction of text evasion through comparison of dividend payments to tax payments which is negative for achieving SDG 1. It is understandable the classifier missed this underlying meaning.

Appendix I: Error analysis document classifier

6 articles that were marked related but are unrelated to SDG 1

- These articles met the threshold for the Keyword lookup

1) 1st Ld-Writethru: Six hundred wartime forced Chinese laborers receive aid A ceremony was held here on Sunday to grant nongovernmental financial aid to 586 survivors of forced Chinese laborers in World War II, most of whom were reduced to slavery-like labor by the Japanese army in the 1930s and '40s. Ten representatives of survivors, mainly from China's northern provinces of Hebei and Shanxi, central Henan and eastern Shandong, attended the ceremony that was held by the China Red Cross Foundation (CRCF) and Chinese Federation of Demanding Compensation From Japan (CFDC). Each of the survivors was given 1,000 yuan (131.6 U.S. dollars), and one who was "extremely poor" was given 3,000 yuan. They obtained domestic aid of the kind for the first time. Tong Zeng, head of the CFDC, said the financial aid offered by Hong Kong Philanthropist Zhong Huiming, will be mainly used for helping WWII victims with "special difficulties" and women forced to be sex slaves in the war.

- Possible reason due to words "aid", "poor", "slavery"

2) BRADFORD MORSE, FORMER CONGRESSMAN AND U.N. OFFICIAL, DIES AT 73) Bradford Morse, a former Republican congressman from Massachusetts who rose to become a top-ranking U.N. official, died Sunday at his home in Naples, Fla. He was 73. The cause was heart failure, his family said. A lawyer who was first elected to the House of Representatives in 1960, Morse served six terms before resigning to join the United Nations. In his 13-year career there, Morse supervised a diverse series of relief efforts that ranged from job-training for two million Colombians to a program that built hundreds of miles of secondary roads in five African nations. As head of the U.N. Development Program, which helps finance and administer development projects in more than 150 nations, Morse became an advocate for aid to poor countries and for victims of famine in Africa.

- Possible reason for error due to "relief" and possible reference to the United Nations

3) NKorean leader Kim Jong Il sends greetings to Putin for Day of Russia holiday North Korean leader Kim Jong Il praised his country's relations with Russia in a message sent Monday to President Vladimir Putin on the occasion of an independence holiday there. "I express conviction that the long traditional friendly and cooperative relations between the two countries would continue to develop in depth in different fields in conformity with the common aspiration and desire of the two peoples in the future," Kim said in his message to Putin for the Day of Russia holiday, according to a report by the North's official Korean Central News Agency. The Russian holiday formerly known as Independence Day had marked the Russian parliament's 1990 declaration of sovereignty from the Soviet Union. However, Putin has called the Soviet breakup a national tragedy and the holiday was renamed Day of Russia in 2002. The Soviet breakup also left North Korea without its major benefactor, a development that led to famine in the 1990s also caused by outdated farming methods and natural disasters.

4) Number of NKorean defectors to SKorea tops 20,000 The number of North Koreans defecting to South Korea has surged in recent years because of economic suffering in the North, with more than 10,000 defections over the past three years, South Korea's Unification Ministry said Monday. About as many North Koreans have defected to the South since the end of 2007 as the number who had fled over the entire previous period since the 1950-53 Korean War ended with an armistice, the ministry said in a statement. The overall total stands now at 20,050. Ministry official Han Dong-ki said the rise in defections reflects North Korea's worsening economy. North Korea has relied on outside food aid since natural disasters and mismanagement wrecked its economy in the mid-1990s, when an estimated 2 million people died of famine. The North's economic troubles are thought to have worsened following a botched attempt at currency reform last year.

- Articles 3 and 4 concerned North Korea so possible that the classifier connects this noun with poverty and slavery.

5) Number of NKorean defectors to SKorea tops 20,000 The number of North Koreans defecting to South Korea has surged in recent years because of economic suffering in the North, with more than 10,000 defections over the past three years, South Korea's Unification Ministry said Monday. About as many North Koreans have defected to the South since the end of 2007 as the number who had fled over the entire previous period since the 1950-53 Korean War ended with an armistice, the ministry said in a statement. The overall total stands now at 20,050. Ministry official Han Dong-ki said the rise in defections reflects North Korea's worsening economy. North Korea has relied on outside food aid since natural disasters and mismanagement wrecked its economy in the mid-1990s, when an estimated 2 million people died of famine. The North's economic troubles are thought to have worsened following a botched attempt at currency reform last year.

- unrelated but probably marked related through economic suffering, aid, died of famine, economic troubles

6) 2nd LD Writethru: Financial crisis turning into human development crisis, say World Bank, IMF The global financial crisis, the most severe since the Great Depression in 1930s, is rapidly turning into a human and development crisis, warned an IMF- World Bank report released Friday. The crisis originated in the developed world, but it has spread quickly and inexorably to the developing world, sparing no countries, said the Global Monitoring Report 2009: A Development Emergency. The crisis is imperiling attainment of the Millennium Development Goals (MDGs) and creating an emergency for development as the poor countries are especially vulnerable, for they "lack the resources to respond with ameliorative actions." Most of the eight MDGs are unlikely to be met by 2015, including those related to hunger, child and maternal mortality, education, and progress in combating HIV/AIDS, malaria and other major diseases, the report warned.

- 52. unrelated, found due to MDG's set, with a lot of overlap of key words. Article concerned after crisis aid and the Millennium Development Goals, which have great overlap with the SDG's at hand.