Text Mining Domains: Sustainability

Myrthe Buckens 2705570 | Dyon van der Ende 2707850 | Eva den Uijl 2712021

March 26th 2021

Abstract

In 2015, the United Nations proposed the Sustainable Development Goals (SDGs), a set of goals that should contribute to a more sustainable future (United Nations, 2015). In this study we designed a classifier to assign an SDG to a research paper. Although our research focuses exclusively on English and a subset of the SDGs, the system is designed in such a way that it can easily be expanded for more languages and SDGs. Of several tested configurations, we found that a system with the multilingual BERT (mBERT) model and the abstracts of scientific publications as input achieved a macro f1-score of 0.912. These results can hopefully form a basis for further research in the field of SDG classification and can help contribute to a more sustainable future.

Github repository: https://github.com/dyonende/TMD

Total word count: 5223

Contents

1 Introduction

In 2015, the United Nations designed a collection of 17 interlinked global goals as a strategy to achieve a better and more sustainable future (United Nations, 2015). All these 17 Sustainable Development Goals (SDGs) are focused on a different field in which preservation of the earth, ending poverty and achieving world peace are the main objectives. These goals are intended to be achieved by the year 2030 and cover various aspects of life on earth. All these goals are defined by a total of 169 separate targets (See Appendix A for a description of all goals and their target goals).

Since the UN introduced these SDGs as one of their top priorities, numerous papers and scientific articles are published on one - or several - of these goals. For a lot of companies and organisations, the Sustainable Development Goals play an important role in strategic plans and policy making (Scott and McGill, 2019). An example of that is that universities are expected to explain how their research helps achieving these goals. Consequently, a system to automatically match each university's publications to one of the goals could greatly contribute to the main goal of matching academic excellence witch societal relevance. An organisation that is focused on this matter is the Aurora Universities Alliance. This is a collaboration between nine universities in Europe that share knowledge and have as core ambition the matching of academic excellence with societal relevance. This ambition is made concrete by different projects that are subdivided into six themes, of which one is sustainability. One of the projects within the sustainability theme is the realisation of a dashboard where research papers can be filtered by SDGs. However, there is no accurate way yet to efficiently classify research papers based on the SDGs they discuss. Therefore, there is a need for an objective and efficient way to label any publication to each of the 17 SDGs.

The goal of the Aurora Universities Alliance is to have a classification method that correctly identifies which SDGs are discussed in a paper and to which target they belong. The Aurora network of universities uses the SDGs to clarify the role that its members play in society and its position regarding global challenges[1]. Analysing research papers from the universities is one of the ways that is used to measure this. However, most papers that are published do not list the SDGs that they contribute to. This means that this data is not readily available and should be extracted from the paper, which is not only a task that takes a lot of time, but also requires knowledge on the SDGs.

To solve this problem, we designed a classifier that can predict the SDG label given a paper using state-of-the-art techniques from the field of natural language processing (NLP). The main focus of this project is to find out what the most efficient system for SDG classification is. In order to find this out, we designed two pipelines: one using mBERT and one with XML-RoBERTa.

First, Section 2 will provide a more detailed insight in the needs of our client. Then, we will provide a description of the data that is used for the current project, found in Section 3. In Section 4, we will provide an overview of the different pipelines used for our classifier, where Section 5 will describe the experiments we carried out in order to find the most efficient design of our SDG classifier. The results of these experiments can be found in Section 6. To finalize this report, we

[1]https://aurora-network.global/activity/societal-impact-and-relevance-of-research-sirr/

will provide an error analysis in Section 7 and end with a discussion and conclusion with recommendations for the client on how to continue from here in Section 8.

2 Problem definition

In this section of the report, we will provide a detailed description of the needs of Aurora and the challenges that might arise when trying to fulfill those needs. First we will focus on the client expectations, then we will discuss the technical constraints. To finalize this section, we will discuss the decisions we made based on the challenges and constraints.

As mentioned in Section 1, Aurora is an international organisation that is connected to universities in several countries. Their main goal is to use the SDG classifier to explain the role of its representatives on global challenges. In order to reach this goal, the classifier should meet the following certain criteria:

- The classifier needs to make its prediction based on input that consists of the available data and meta-data of a paper (described in Subsection 3.2);
- The paper can be written in any of the following languages: English, German, French, Italian, Spanish, Catalan, Czech, Dutch, Icelandic and Danish;
- The classifier predicts SDG labels as well as target labels based on the input.

The intuitive idea behind text classification is to classify a document under a predefined category (Ikonomakis et al., 2005). In a formal way we can explain this as labeling a document d_i in the set of documents D, with at least one label c_j out of the set of categories $\{c_1, c_2, \ldots, c_n\}$. To correctly assign SDG-labels to scientific publications, we will have to design a system for automatic text classification in which the predefined categories are the 17 SDGs.

When a document is assigned to more than one category, this is called Ranking Classification. In the scenario that a specific document gets assigned a single category, we speak of Hard Categorization (Ikonomakis et al., 2005).

The data set that was provided for this project contains, inter alia, the full text of all abstracts, the keywords and the title of the publication. To make the main goal of our system more concrete, we will illustrate the task of SDG classification with an example from our data set. We first take the text of the abstract. (for the purpose of this example we will only take a segment of the abstract):
"*Background: The incidence of gastric cancer is decreasing in Australia, yet it remains a common cause of cancer-related mortality. Surgical resection remains the cornerstone of curative treatment. High-volume specialized units have reported superior perioperative and oncological outcomes. The role of D2 lymphadenectomy has been controversial as a result of concerns over increased morbidity...*" (for full text of the abstract see Thomson et al., 2014).
Based on the content of this abstract, we can tell that this publication should be assigned to *Goal 3: Ensure healthy lives and promote well-being for all at all ages*.

2.1 Scope

The fact that there is to our best knowledge still no efficient SDG Classifier available, despite the obvious needs, already suggests that labeling SDGs is not an easy task. There are several difficulties that need to be considered regarding the task of SDG classification. These considerations are reported in this section.

The data available, both on SDGs as on the task of classifying them, has formed a great challenge in the current project. The first challenge we faced was that there currently is no benchmark annotated data set for SDG labelling available, which is even more challenging considering the fact that there is also no straightforward way of creating one. The reason behind this is that it would take a lot of time and in-depth knowledge on SDGs and the field of research.

Furthermore, Aurora provided us with a data set that they had extracted from their own English library based on queries (a more detailed description on the data and how it was collected can be found in Section 3). As a result, the available data covered only one language. The inaccessibility of multilingual data poses a threat to the goal of designing an efficient multi-lingual classifier.

Another difficulty to be taken into account is the recency of SDGs. Since SDGs are a relatively new field of research, not a lot of papers have been published specifically on the classification of SDGs. What this concretely means is that there is no information available on how difficult the task is to solve and what results to aim for.

Given the available time and resources to build the classifier, we need to narrow down the scope of the research. With the aforementioned limitations in mind, we decided to focus on English and to consider only two out of seventeen SDG goals (Goal 3 and Goal 14) with hard categorization. Although we will mainly focus on English in the current project, we will keep the possibilities of implementing the application in other languages in mind while designing the system. To measure the results of our classifier, we will compare it to a baseline classifier that is described in Section 5.

3 Data

In the current section, we will provide an overview of how the data was collected. Then, we will present a detailed description of the data that was used to train and test our systems.

3.1 Collection of the Data

The data was collected by using search queries for the different SDGs. These queries are published by the Aurora Universities Alliance, in order to obtain domain specific output associated with the Sustainable Development Goals (Vanderfeesten et al., 2020). These queries form a construction with keyword combinations and Boolean operators, which form a unique phrase every different target within all SDGs.

An example of a query for SDG 3, *3.1: By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births* is the following:

```
   TITLE-ABS-KEY
( ("reduce" OR "end" OR "ending" OR "ratio")
W/3
( ( "maternal" W/3 ("mortality" OR "death*") ) ) )
```

This query will find all papers that in the title, abstract or keywords have one or more of the terms *reduce, end, ending* or *ratio* within a window of 3 of the term *maternal*, which in its terms also has to be within a window of 3 within the terms *mortality* or *death*.

When there is a match for the query, the target label and SDG label are assigned to the article.

3.2 Data Description

The data consists of abstracts from research articles on SDGs from the Scopus database crawled by Aurora Universities Alliance. The data is enriched with meta data on the publication, such as id, title, SDG goal, SDG target, and keywords for scopus index, see Table 1. In total there are 8223 rows of data.

Table 1: Data format before preprocessing

Property	Description
scopus_abstract_retrieval.coverDate	The date that the data was retrieved
_id	Unique identification number for each paper
query_title	The SDG label that was assigned to the paper
query_id	The SDG targets that were assigned to the paper
scopus_abstract_retrieval.title	The title of the paper
scopus_abstract_retrieval.abstract	The abstract of the paper
scopus_abstract_retrieval.idxterms	Relevant keywords of the paper
scopus_abstract_retrieval.authkeywords	Keywords that were provided by the author of the paper
scopus_abstract_retrieval.doi	Unique digital object identifier of the paper
scopus_abstract_retrieval.url	Link to Elsevier api with request for the paper
unpaywall_response.free_fulltext_url	Link to the full text of the paper without paywall

3.3 Train and Test Data

In order to train and test our model, a few decisions regarding the data set had to be made. This section will cover the decisions made, as well as the reasoning behind those decisions.

The first step we took was to create a negative class. The main intuition behind this is that all articles that do not cover the topic of SDG 3 or SDG 14, are not taken into account. To do so, we converted all labels except SDG 3 or SDG 14 to 0. The reason we picked those two SDGs is that the subject they cover is rather dissimilar

(Good health and well-being against Life below water). Therefore, we expect the chance that a classifier is able to make the distinction to be higher compared to SDGs that are more similar (like SDG 15: Life on land).

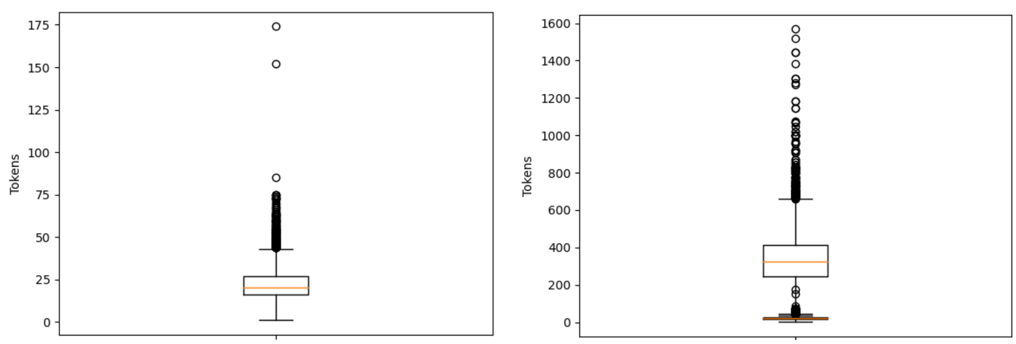
To continue, we decided to ignore all data that covered more than one SDG. More concrete, this means that all data covering multiple SDGs was removed from the data set. The reason behind this is that we focus on hard categorization in this project, resulting in the fact that our classifier is not fit to learn multiple labels per data entry. Since we expect titles and abstracts to be the most valuable aspects in the data to predict SDG on, we decided to leave out the instances for which the abstracts are missing.

For the training and evaluation of our model, we will need training data and test data. To achieve this, we split up the cleaned data with a ratio of 0.8. The set containing 80% of the collected data will form our training set, and 20% of the data will form the test set. The distribution of the labels between the training and the test data are kept the same. To obtain a clearer view of the data, Table 2 provides an overview on the statistics of the training data.

Table 2: Overview of statistics for the training data.

	Training Data
# Articles	5220
Language	English
Label distribution	*0*: 59,2%
	3: 20,8%
	14: 20,0%

As stated earlier, we think that the features that carry the most relevant information for this task are the title and the abstract. To tune the classifier in Section 4, the length of these features are of importance. The plots in Figure 1 show the distribution of the number of tokens of the title (Figure 1a) and the abstract (Figure 1b). As expected the title contains less tokens than the abstract, with an average of 22,1 compared to 341,0 tokens per abstract. It is safe to say that most papers do not have a title that contains more than 50 tokens or an abstract that contains more than 800 tokens.



(a) Tokens per title (b) Tokens per abstract

Figure 1: Number of tokens in the training data, tokenized by mBERT Tokenizer.

3.4 Data Quality

When looking at the data, we cannot draw a conclusive conclusion on the quality of it. The advantage of the available data set is that it is very rich in information. We have access to the title, abstracts and we can in fact retrieve the full texts from the urls and dois provided. However, since collection of the data is based on queries, we do have some concerns for the data. An article is assigned an SDG label when the words in the search query match with the words in the title, keywords or abstract. This method comes with the following three limitations:

1. There are articles in the data that do contain the words in the search query, but do not cover the topic of the SDG they are labeled with.

2. There are articles not labeled with an SDG because the words in the search query do occur in the articles, but the proximity requirements are not met.

3. There are articles not labeled with an SDG because the words in the search query do not occur in the article, but the article does cover an SDG topic.

The first limitation is of influence on our model its performance because the model learns based on the input that it is given. The two latter are of concern for the real world application and validity of the model. More on what this means for our research will be discussed in Section 8.

4 System Design

In this section of the current report, different models for the task of text classification will be discussed. We will shortly describe BERT, and its multilingual adaptions; mBERT and XLM-RoBERTa and finalize this section with the different systems we designed using the examined language models and the rationale behind them.

4.1 Language Models

Over the last few years, language models gained popular ground for text classification tasks. In 2018, BERT, short for Bidirectional Encoder Representations from Transformers, was introduced as the first bidirectional language model (Devlin et al., 2019). It was different from existing language models because it uses masked language models (MLM), meaning that random input tokens get masked resulting in the system having to predict the correct token based on the context (Liu et al., 2019). With the use of MLM, it became possible to take the context on the left and the right of the masked token at the same time as input for the pre-training of the deep bidirectional Transformer (Devlin et al., 2019). Until then, existing models like ELMo were only able to process language unidirectional for learning general language representation. Based on the structure of BERT, new language models were introduced that involved multilingual processing like mBERT, and different architectures that were trained on more data such as RoBERTa.

4.2 Multilingual BERT

The Multilingual BERT (mBERT) language model yields a sentence representation generalized over 104 different languages and is trained on Wikipedia data (Libovick et al., 2019). The great coverage of languages makes mBERT useful for multi-lingual tasks. Since the main objective is to design a system for SDG classification in 11 languages, using a multi-lingual model is of great importance. Research has shown that the performance of mBERT for low resource language can be poor (Wu and Dredze, 2020). For the languages that Aurora is interested in, only Icelandic is a low resource language.

4.3 XLM-RoBERTa

Another model we used in order to execute the current text classification task, is the pre-trained model XLM-RoBERTa (XLM-R). Conneau and colleagues (2019, 2020) describe how the performance of cross-lingual transfer tasks can be improved by using a multi-lingual pre-trained model: The combination of XLM and RoBERTa. The RoBERTa model is build on BERT and modifies key hyperparameters, removing the next-sentence pretraining objective and training with much larger mini-batches and learning rates (Wolf et al., 2020). The improvement of XLM-R over RoBERTa is that XLM-R was trained on 2.5TB filtered CommonCrawl data (Conneau et al., 2020).

We use XLM-R in this project since training of the model for the given task must take the full sentence into account to extract semantic features. On top of that, Conneau et al. (2019) show that XLM-R outperforms mBERT significantly on various cross-lingual benchmarks.

4.4 Classifier implementation

With the language models described in the previous sections, we designed two types of pipelines: a fined-tuned mBERT classifier that uses a neural classification layer; and an XLM-RoBERTa model with an SGD classifier.

The first step for both implementations is the same: preparing the data. As mBERT and XLM-R have a maximum input length of 512 tokens and long inputs put more strain on the hardware that is used, it is crucial to deal with input that is too long. The way we have dealt with this, is by removing tokens from the end of the input text, until the desired length has been reached.

With the mBERT model, the next step is to convert the labels, as it can only deal with integer labels that are continuous. In our case, negative class and SDG labels 3 and 14 would be converted to 0, 1 and 2 respectively. The training data is then used to train the mBERT model with an extra neural layer for classification. Finally, this fine-tuned and trained model can be used to make predictions on the prepared test data. A schema for this pipeline can be found in Figure 2. We chose to fine-tune mBERT as there is good documentation[2] available for BERT models on how to do this and the performance benefits that it can offer (Sun et al., 2019).

The pipeline for XLM-RoBERTa has more steps, as it uses the language model to vectorize the prepared input text, but not for the classification. These vectors are then used as features for the classifier, which is an SVM classifier with stochastic

[2]https://colab.research.google.com/drive/1ywsvwO6thOVOrfagjjfuxEf6xVRxbUNO?usp=sharing

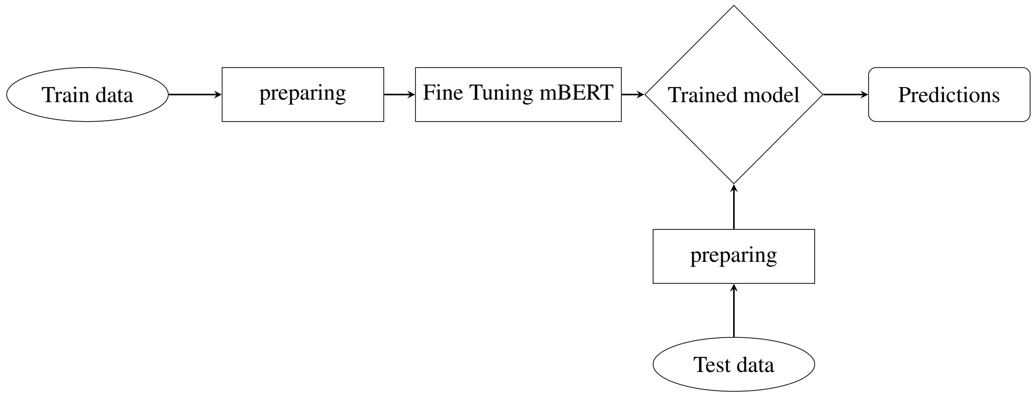


Figure 2: mBERT architecture.

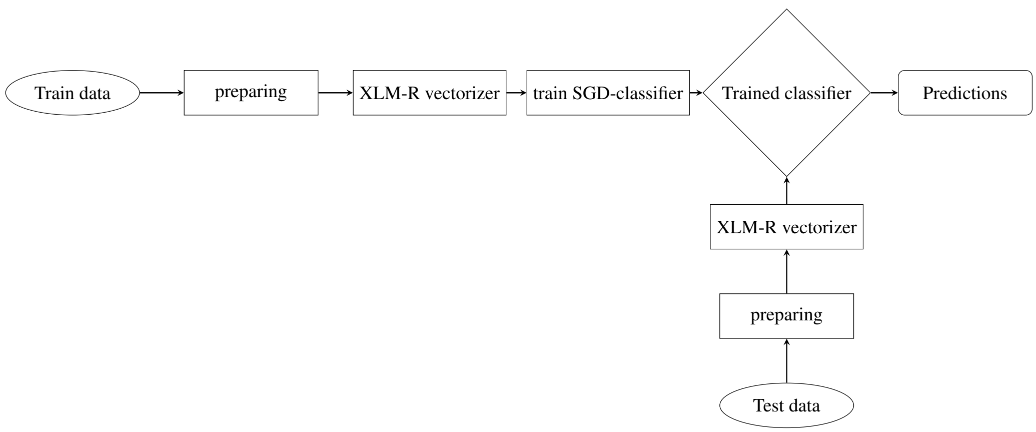


Figure 3: XLM-RoBERTa SGD architecture.

gradient descent (SGD) implemented. After the classifier has been trained on the training data, it can make predictions for the test data. See Figure 3 for a schematic representation.

5 Experiments

The main objective of this project is to design a system that is able to take scientific publications as input and then determine how the particular article is related to one of the seventeen Sustainable Development Goals. In order to pursue this objective, we tested several configurations of the systems that were described in the previous section. In the current section, we will describe these different configurations. For simplicity, these systems are refered to by the language model they use: mBERT or XLM-R.

In this experimental setup, the configurations differ for the type of input that is used by the system. This input can either be the title, the abstract or a combination of both. Furthermore, the maximum number of input tokens is different for each configuration. Using mBERT, we tested three different configurations: titles and up to 128 tokens (mBERT-t), abstract and up to 400 tokens (mBERT-a) and titles and abstract concatenated up to 400 tokens (mBERT-ta). For XLM-R we tested two configurations: one using title and up to 75 tokens (XLMR-t) and one using abstracts and up to 300 tokens. (XLMR-a). An overview of these configurations is provided in Table 3.

Name	**Model**	**Input**	**Max Tokens**
mBERT-t	mBERT	titles	128
mBERT-a	mBERT	abstracts	400
mBERT-ta	mBERT	titles, abstracts	400
XLMR-t	XLM-RoBERTa	titles	75
XLMR-a	XLM-RoBERTa	abstrats	300

Table 3: Overview of the different configurations.

The systems will be compared to each other and to a baseline. The implementation of the baseline is very straightforward since it will always predict a random label that is either the negative class, SDG 3 or SDG 14.

In the evaluation, we will focus on the macro precision, recall and f1-score. We chose for macro, as all categories are equally important for the functioning of the classifier and with macro, the categories are weighted equally, despite class imbalance.

The best performing classifier was also tested with all 17 SDG labels (and without a negative class). The results can be found in Section 6.

6 Results

In total we have classified SDGs with a baseline, three mBERT configurations and two XLM-R configurations. An overview of the results for these classifiers is listed in Table 4. In Appendix B more detailed results for each classifier can be found.

Table 4: Performance Based on Macro Scores for All Systems.

	precision	**recall**	**f1-score**
baseline	0.346	0.344	0.322
mBERT-t	0.828	0.824	0.826
mBERT-a	**0.916**	**0.911**	**0.912**
mBERT-ta	0.905	0.893	0.899
XLMR-t	0.783	0.704	0.734
XLMR-a	0.799	0.829	0.810

As shown in the table, we found moderate results for the baseline reaching a f1-score of only 0.322, which we expected. All the other more sophisticated systems show much better performance. Of these, XLMR-t achieves the lowest f1-score of 0.734. The highest f1-score was achieved by mBERT-a, reaching an f1-score of 0.912. Interesting about these results is that for all classifiers the difference between precision and recall is very small.

Another notable thing is that titles perform worse than abstracts when used as features. The combination of the two in mBERT-ta also performs slightly worse than mBERT-a, probably because the use of the title shortens the part of the abstract that can be used.

In Table 5 the results for mBERT-a are shown with more detail on the performance on individual classes.

We find that SDG 14 scores comparable to the negative class, both having

Table 5: Results for mBERT-a. SDG 0 is the negative class.

Class	**precision**	**recall**	**f1-score**	**support**
SDG 0	0.933	0.942	0.938	773
SDG 3	0.898	0.838	0.867	272
SDG 14	0.916	0.954	0.935	262
macro avg	0.916	0.911	0.913	1307

an f1-score of 0.93 and a slightly higher precision than recall, despite the class imbalance. For SDG 3, the recall is lower than the precision, 0.838 compared to 0.898.

As mBERT-a was the best performing classifier, we trained and tested it with all SDGs and no negative class. The total number of instance is lower than with the previous test set, because more duplicates were removed. The results are found in Table 6.

Not surprisingly, the performance of the system with all the SDG labels is worse than for just the two SDGs. However, the difference between the macro precision and recall is again very small. The macro f1-score is only 0.542 and some individual classes have an f1-score of 0. As these classes are the smallest classes in the data, this is somewhat expected. More interesting is that for some of the smaller classes, the results are actually quite decent. SDG 10 counts only 24 instances in the test data, but its f1-score is 0.545. SDG 7 achieves an even more impressively 0.710 with a support of 34. That is higher than SDG 15, which scores just 0.635 with a support of 107. What this shows is that with more data the results get better, but also that some SDGs are harder to classify than others.

Table 6: Results for mBERT-a on a test set with 17 SDG labels.

	precision	**recall**	**f1-score**	**support**
SDG 1	0.627	0.875	0.730	96
SDG 2	0.000	0.000	0.000	24
SDG 3	0.752	0.846	0.796	136
SDG 4	0.762	0.800	0.780	60
SDG 5	0.273	0.429	0.333	21
SDG 6	0.787	0.778	0.783	81
SDG 7	0.786	0.647	0.710	34
SDG 8	0.000	0.000	0.000	23
SDG 9	0.000	0.000	0.000	19
SDG 10	0.600	0.500	0.545	24
SDG 11	0.848	0.808	0.828	104
SDG 12	0.486	0.425	0.453	80
SDG 13	0.853	0.725	0.784	40
SDG 14	0.943	0.943	0.943	262
SDG 15	0.587	0.692	0.635	107
SDG 16	0.568	0.658	0.610	76
SDG 17	0.303	0.278	0.290	36
macro avg	0.540	0.553	0.542	1223

7 Error Analysis

In this section, we will try to gain insight in what our system can and cannot do by discussing the errors that are made by our best performing system. The model that will be discussed in this error analysis is the mBERT model that was fine-tuned up to 400 tokens of the abstract as input. We chose to analyse this configuration as it achieved the most promising results and further insights in the types of mistakes that it makes could help increase performance in further research. Table 7 shows the confusion matrix for this system.

Table 7: Confusion Matrix for mBERT-a predictions on the test data.

Predicted Gold	0	3	14
0	**729**	15	29
3	44	**224**	4
14	8	1	**253**

When looking at the confusion matrix for mBERT-a, we can see that our initial hypothesis that SDG 3 and SDG 14 are easy to distinguish, holds. The matrix show that SDG 3 and 14 are only mixed up by the system five times. SDG 3 is wrongly predicted to be SDG 14 for 4 articles, while SDG 14 is mistaken for SDG 3 only once. Because of this low number of mix-ups, it made sense to do a manual analysis for these articles.

In Appendix C the title and the abstract for each of these mislabelled papers is found. Table 14 contains the data for the case where our model predicted SDG 3, while the gold label was SDG 14. The title tells us that this article is about anti-fouling, a chemical substance that is used for the maintenance of boats, and its impact on marine life. The abstract of this paper is very technical with chemical formulas and terms which can be hard to understand even for non-expert humans. The confusion seems to come from the fact that the anti-fouling is toxic, but in this case, the toxicity is more harmful for live below water. In practice anti-fouling is also toxic for humans, so although SDG 14 is indeed the correct label for this abstract, SDG 3 is not that strange of a prediction.

By looking at the instances that were labelled with SDG 14 while the gold label was SDG 3, we expected four different instances, but to our surprise we found that it were only two different instances that were duplicated in the data. Upon further inspection, it turned out that, despite our efforts to prevent this, there were multiple duplicate entries in our train and test data. The implications of this are discussed in Section 8. For the analysis, it means that only two papers have to be analysed. Both can be found in Table 15 and Table 16 and appear to be health related and again are very technical. The first talks about vaccination and there are no clear clues that this paper has any connection to life below water. The second paper makes this even clearer by stating that its research is focused on concentrations of toxic chemicals in *humans*. Our only explanation for this confusion of SDGs could come from the fact that both papers talk about toxicity of some kind, but in both cases SDG 14 is clearly not applicable.

Although SDG 3 and SDG 14 were only mixed up with each other in few occasions, Table 7 does show some other cases in which the system misclassifies these SDGs with the negative class. To find out what the problem is in these cases, we performed a systematic investigation of the abstracts of all publications based on the most frequent lemmas. For informativeness, we filtered out stop words and punctuation.

Table 8: Most frequent words in abstracts for each of the combination clusters of gold and predicted label. Note that the frequencies are from the aggregation of all abstracts and that a word does not have to appear frequently in all abstracts.

Cluster (#)	**Most frequent words (#)**
3 predicted as 0 (44)	study (112), health (86), patient (66), cost (64), year (54), treatment (46), intervention (42), relationship (40), mother (36), adolescent (36)
0 predicted as 3 (15)	patient (25), health (23), adolescent (21), level (21), study (19), woman (18), mortality (17), care (17), girl (17), use (16)
3 correct (224)	patient (674), mortality (576), study (290), rate (274), high (254), increase (252), risk (230), year (230), cause (226), death (204)
0 predicted as 14 (29)	study (22), specie (22), population (21), change (21), high (18), site (17), result (15), water (15), region (14), area (14)
14 predicted as 0 (8)	year (13), emission (12), production (10), selectivity (9), tourist (9), study (9), patch (9), exploitation (8), transport (8), water (8)
14 correct (253)	high (237), marine (201), study (189), specie (181), change (157), result (155), model (153), area (150), water (143)

When reviewing the results of the current error analysis, the most remarkable facts as shown in Table 8 are that both clusters of the SDGs contain the same frequent words. For both SDG 3 its false positives and true negatives and its correctly classified cases, the same or similar tokens are found in the list of most frequent words. An example of this is the occurrence of *mortality* in the false positives in comparison to *death* in the true positives. For SDG 14 we see a similar pattern. The word *study* occurs in the list of frequent words over the whole cluster.

For the cases in which abstracts of the negative class were predicted to be SDG 3, we found that *patient* (25), *health* (23), *sanitation* (16) and *mortality* (17) were frequents words in the data set. These findings are interesting because for the words mentioned we could intuitively say that they are indeed related to *Good Health and Well-Being*. For the cases in which abstracts of the negative class were predicted to be SDG 14, we found a similar pattern. The most frequent words were, among others, *site* (17), *water* (15) and *mussel* (12). The occurrence of these words in abstracts that are incorrectly classified to be on *Life Below Water* is interesting since the words mentioned evoke the feeling that they are related to sea life in some extent.

It is hard to tell from this analysis, whether our classifier wrongly connected these papers to the predicted, because of certain keywords, or that these papers actually discussed the predicted SDG but that the gold labels were maybe incorrect. We can say, however, that when the classifier makes mistakes, they are at least explainable.

8 Discussion and Conclusion

In this section, we will discuss the applicability of our system, the points of improvement and recommendations for further development.

Our goal was to design a classifier for Aurora Network that could assign an SDG label to research papers that could eventually be extended to cover multiple languages. Due to limited resources we focused on creating a system that can classify two SDGs in English. The results of our research are very promising, with our best performing system reaching a macro f1-score of 0.912 on our data. We tried to apply our model to a data set that contained all 17 SDGs, but due to the class imbalance for the SDGs in the data that was available, the classifier is not optimized for classifying all SDGs. Only when there were 30 or more articles available, the f1-score for that SDG raised above 0.7. For optimizing and training the classifier to be able to classify all SDGs, more data is needed.

A consequence of our design of the pipelines is that the configurations with mBERT and XLM-R did use a different maximum token input and classifier, which is not ideal for making a definitive comparison between the two.

The data that we had available had several limitation which impacted the performance of our model, apart from class imbalance. Since the collection of data was executed by using search queries for string matches, the system learns based on these specific words occurring in the data. This is a limitation of the system for when it encounters new data that does not contain any of those search query words.

During the error analysis, we found out that the training and test data contains duplicates. Although in the code we used for preprocessing the data we dropped the duplicates based on the id number of the article, in the training and test data there were still articles in there with the same id number. This led to some SDGs seemingly having more data than others. We expect that this has no real consequences for the reliability of the model, but without the duplicates, the test results probably would have been different.

8.1 Recommendations

Our best performing model is the multilingual BERT model trained and fine-tuned on the abstracts of the articles. As described in section 4, recent studies showed that XLM-RoBERTa outperforms mBERT on various cross lingual benchmarks. With the available resources at the time of this project, it was not possible to fine-tune XLM-RoBERTa to our data to see if this claim also applies to our project. We recommend looking in the use of XLM-RoBERTa at a later moment again to see if there are improvements that allow for fine-tuning on our data.

Our best configuration used a maximum token input of 400 tokens because of the available hardware that was used. BERT models have a maximum input length of 512 tokens. According to our data, most abstract have a maximum length of under 800 tokens. Since it is not possible to run the whole abstract through BERT, research has shown that a combination of the first 128 tokens and the last 384 tokens can achieve better results (Sun et al., 2019).

Lastly, we were not able to test our system on data that was in a different language than English, as we did not have data available for this. Our design is such that is should work for data that is in different language, but the model is currently fine-tuned on the data that we had, which was in English. There are two ways to deal with this: Fine-tune a model specifically for each desired language, or create a data set that consists off all desired languages and fine-tune the model on this data.

References

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2020). Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv:2006.13979*.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186.

Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS transactions on computers*, *4*(8), 966–974.

Libovick, J., Rosa, R., & Fraser, A. (2019). How language-neutral is multilingual bert? *arXiv preprint arXiv:1911.03310*.

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*.

Scott, L., & McGill, A. (2019). PWC SDG Challenge 2019.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune bert for text classification? *China National Conference on Chinese Computational Linguistics*, 194–206.

Thomson, I. G., Gotley, D. C., Barbour, A. P., Martin, I., Jayasuria, N., Thomas, J., & Smithers, B. M. (2014). Treatment results of curative gastric resection from a specialist australian unit: Low volume with satisfactory outcomes. *Gastric Cancer*, *17*(1), 152–160.

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development.

Vanderfeesten, M., Otten, R., & Spielberg, E. (2020). *Search Queries for ”Mapping Research Output to the Sustainable Development Goals (SDGs)” v5.0* (Version 5.0). Zenodo. https://doi.org/10.5281/zenodo.3817445

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6

Wu, S., & Dredze, M. (2020). Are all languages created equal in multilingual BERT? *Proceedings of the 5th Workshop on Representation Learning for NLP*, 120–130. https://doi.org/10.18653/v1/2020.repl4nlp-1.16

A SDG Goals

The United Nations designed 17 Global Goals and a total of 169 targets divided over all of these goals (United Nations, 2015). Below, a short overview will be provided for all 17 SDGs and their targets.

SDG 1 **No Poverty**

The first goal is to end poverty worldwide. This goal consists of the following targets:

Target 1 Establishing a minimum income of $1.25 per day for all people everywhere on earth.

Target 2 Reduce number of people living in poverty by factor 2, based on the national definitions.

Target 3 Apply social protection systems for the poor and the vulnerable.

Target 4 Establish equal rights and access to economic resources.

Target 5 Reduce the exposure and vulnerability to climate events and economic, social and environmental disasters for the poor and the vulnerable.

Target a Ensure a substantial mobilization from different sources in order to provide sufficient and predictable means for developing countries. In particular for least developed countries, to adopt programs and policies to end poverty in all its aspects.

Target b Create a framework at regional, national and international level to support investment in plans to overcome poverty.

SDG 2 **Zero Hunger**

The second goal of the SDGs is to end hunger, to achieve food security and improvement of nutrition and to promote sustainable agriculture. This goal consists of the following targets:

Target 1 End hunger and make nutritious and sufficient food accessibility for everyone worldwide, all year round.

Target 2 End all forms of malnutrition, including ending prevalence of both underweight and overweight in children.

Target 3 Double the agricultural productivity and incomes for small scale producers of food by providing equal access to land, productive resources and inputs, knowledge and financial services.

Target 4 Establish sustainable food production systems; increase productivity, maintain ecosystems, strengthen capacity to adapt to climate change and other weather conditions and disasters.

Target 5 Maintain the diversity of seeds, plants and farmed and domesticated animals and their related wild species.

Target a Increase investment in rural infrastructure, agricultural research and plant and livestock banks.

Target b Avoid and correct restrictions on trade in agricultural markets.

Target c Limit extreme food price volatility by facilitating access to market information.

SDG 3 **Good Health and Well-being**

The third Global Goal is to ensure healthy lives and to promote the well-being for everyone at all ages. This goal consists of the following targets:

Target 1	Reduce the global maternal mortality ratio to less than 70 per 100,000 births.
Target 2	Reduce neonatal mortality (preventable death of newborns and children under the age of 5) to 12 per 1,000 births.
Target 3	End epidemics of AIDS, tuberculosis, malaria and tropical diseases.
Target 4	Reduce the premature mortality by one third by prevention and treatment of non-communicable (chronic) diseases.
Target 5	Reinforce the prevention and treatment of narcotic drug and alcohol abuse.
Target 6	Reduce the number of deaths from road traffic accidents with the factor 2.
Target 7	Make universal sexual and reproductive health-care services accessible.
Target 8	Achieve global health coverage including financial risk protection, access to health-care and access to safe medicine and vaccines for all.
Target 9	Reduce the number of deaths and illnesses from chemicals in air, water and soil.
Target a	Strengthen the WHO its convention on tobacco control in all countries.
Target b	Support research and development of vaccines and medicines for diseases.
Target c	Increase health financing and the recruitment, development and training of the health workforce in developing countries.
Target d	Strengthen global capacity for risk reduction and management and global health risks.

SDG 4 **Quality Education**

The fourth goal on the global agenda of the UN is to ensure inclusive and qualitative education for all, and to promote life-long learning. This goal consists of the following targets:

Target 1 Make sure all children complete free and qualitative primary and secondary education for effective learning outcomes.

Target 2 Ensure children have access to care and pre-primary education so that they are ready for primary education.

Target 3 Make education accessible for all men and women.

Target 4 Increase the number of youths and adults who are relevantly skilled.

Target 5 Eliminate disparity in education based on gender, race and disability.

Target 6 Achieve literacy and numeracy for a substantial portion of people.

Target 7 Make sure that all learners achieve a level of knowledge to promote development on - among others - human rights, culture, gender equality.

Target a Upgrade education facilities to make them non-violent and inclusive for all.

Target b Expand the number of available scholarships for developing counties.

Target c Increase the supply of qualified teachers.

SDG 5 **Gender Equality**

SDG 5 is to achieve gender equality and to empower all women and girls. This goal consists of the following targets:

Target 1 End all forms of gender discrimination

Target 2 End all forms of violence against females in public and private spheres.

Target 3 End all harmful practices against women, such as forced marriage and genital mutilation.

Target 4 Provide public services for unpaid care and domestic work.

Target 5 Ensure participation of women for leadership at all levels.

Target 6 Ensure accessibility to sexual and reproductive health.

Target a Give women equal rights to economic resources.

Target b Enhance the use of technology by women.

Target c Strengthen and reform policies for the promotion of gender equality.

SDG 6 **Clean Water and Sanitation**

Goal 6 aims at ensuring availability and sustainable management of water and sanitation. This goal consists of the following targets:

Target 1 Achieve global access to safe drinking water for all.

Target 2 Achieve global access to sanitation and hygiene for all.

Target 3 Improve water quality.

Target 4 Increase water-use efficiency.

Target 5 Implement water resource management at all levels.

Target 6 Protect and restore water-related ecosystems.

Target a Expand global cooperation and capacity-building to developing countries in water related activities.

Target b Strengthen the participation of local communities in water and sanitation management.

SDG 7 **Affordable and Clean Energy**

The seventh goal of the seventeen SDGs is to ensure access to affordable, reliable, modern and sustainable energy for everyone. This goal consists of the following targets:

Target 1 Establish global access to modern energy services.

Target 2 Increase the share of renewable energy.

Target 3 Double the rate of improvement in efficient energy usage.

Target a Intensify the international cooperation to facilitate access to clean energy.

Target b Expand infrastructure and upgrade technology for energy supply in developing countries.

SDG 8 **Decent Work and Economic Growth**

Goal 8 is to promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all. The following targets are part of this goal:

Target 1 Support economic growth per capita; at least 7% gross domestic product growth.

Target 2 Reach higher levels of economic productivity.

Target 3 Promote development policies for productive activities as job creation and entrepreneurship.

Target 4 Improve progressively global resource efficiency in consumption.

Target 5 Achieve full and productive employment and work for all regardless of gender, race and disability.

Target 6 Reduce the proportion of unemployed youth.

Target 7 Take measures to end forced labor, modern slavery and human trafficking.

Target 8 Protect labor rights and promote safe and secure environments for all employers.

Target 9 implement policies to promote sustainable tourism.

Target 10 Strengthen capacity of domestic financial institutions.

Target a Increase Aid for Trade for developing countries.

Target b Develop a global strategy for youth employment and implement the Global Jobs Pact of the International Labour Organization.

SDG 9 **Industry, Innovation and Infrastructure**

Goal 9 is to build infrastructure, to promote inclusive and sustainable industrialization and to foster innovation. The following 8 targets are part of the ninth SDG:

Target 1 Develop infrastructure that is qualitative, reliable and sustainable to support economic development and human well-being.

Target 2 Promote inclusive and sustainable industrialization.

Target 3 Increase accessibility of small-scale industries and enterprises.

Target 4 Upgrade infrastructure and industries to make them more sustainable with increased resource-use efficiency.

Target 5 Upgrade technological capabilities of industrial sectors in all countries and encourage scientific research.

Target a Facilitate infrastructure development in developing countries.

Target b Support technology development in developing countries.

Target c Increase accessibility of information and communication technology.

SDG 10 **Reduced Inequality**

The 10th SDG focuses on reducing inequality within and among countries. This goal consists of the following targets:

Target 1 Accomplish and support pay development of the last 40% of the populace at a rate higher than the national average.

Target 2 Enable and advance the social, financial and political incorporation of all, regardless of age, gender, disability, race, nationality, origin, religion or other status.

Target 3 Guarantee equivalent freedom and lessen imbalances of result, including taking out prejudicial laws, and advancing proper enactment, strategies and activity in such manner.

Target 4 Make arrangements, particularly fiscal, wage and social security approaches, and continuously accomplish more equality.

Target 5 Improve the guidelines and monitoring of worldwide business sectors and foundations and reinforce the usage of such guidelines.

Target 6 Guarantee upgraded portrayal and voice for agricultural nations in dynamic in global international economic and financial institutions.

Target 7 Facilitate safe and responsible migration and mobility for all.

Target a Execute the rule of uncommon and differential treatment for developing countries, specifically least developed nations, as per World Trade

Organization arrangements.

Target b Enhance official development assistance and financial flows.

Target c Reduce the transaction costs of migrant remittances to less than 3 per cent.

SDG 11 **Sustainable Cities and Communities**

SDG 11 aims at making cites and human settlements inclusive, safe and sustainable. The following targets form this goal:

Target 1 Make housing and basic services affordable, safe and adequate for all.

Target 2 Provide safe, accessible and sustainable transport systems.

Target 3 Increase inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning.

Target 4 Protect the world's cultural and natural heritage.

Target 5 Reduce the number of deaths and number of people affected by disasters directly related to economic losses.

Target 6 Improve the environmental impact of cities by monitoring air quality and waste management.

Target 7 Make public spaces safe, accessible and green.

Target a Support links between urban, per-urban and rural areas based on economic, social and environmental impact.

Target b Increase the number of cities and human settlements that have plans for inclusion, resources efficiency and mitigation.

Target c Support developing countries in building sustainable buildings through financial and technical assistance.

SDG 12 **Responsible Consumption and Production**

The main focus of goal 12 is to ensure the sustainable consumption and production patterns. This goal includes the following targets:

Target 1 Conduct a ten-year framework for programs on sustainable consumption and production.

Target 2 Achieve sustainable management and efficiency on the use of natural resources.

Target 3 Reduce the global food waste at retail and consumer level by factor 2.

Target 4 Achieve environmental management of chemicals and waste throughout the human life cycle.

Target 5 Reduce waste generation.

Target 6 Encourage companies to adopt sustainable programs.

Target 7 Promote sustainable public procurement practices.

Target 8 Create awareness and provide relevant information on sustainable development and lifestyles in harmony with nature.

Target a Support developing countries towards more sustainable patterns of consumption and production by strengthening their scientific and technological capacity.

Target b Develop and implement tools to monitor sustainable development.

Target c End inefficient use of fossil-fuels.

SDG 13 **Climate Action**

Goal 13 of the SDGs is to take urgent action to combat climate change and the impacts of climate change. The following targets are documented to achieve this goal:

Target 1 Strengthen the adaptive capacity for climate related hazards and disasters in all countries.

Target 2 Integrate climate change measures into policies, strategies and planning nation wide.

Target 3 Improve education and human and institutional capacity on climate change.

Target a Commit to the United Nations Framework Convention on Climate Change to a goal of mobilizing jointly $100 billion annually by 2020 for developing countries.

Target b Advance systems for raising limit with regards to compelling environmental change-related arranging and the executives in least created nations and little island creating States, remembering centering for ladies, youth and nearby and minority networks.

SDG 14 **Life Below Water**

Goal 14 aims at conserving and sustainable using the oceans, seas and marine resources for sustainable development. The targets of this goal are the following:

Target 1 Prevent and significantly reduce marine pollution of all kinds.

Target 2 Sustainable manage and protect marine and coastal ecosystems.

Target 3 Minimize the impact on ocean acidification.

Target 4 Regulate harvesting and over-fishing and end illegal, unreported and unregulated fishing.

Target 5 Conserve 10 percent of coastal and marine areas.

Target 6 Prohibit the subsidisation of fisheries that contribute to over-fishing and illegal, unreported and unregulated fishing.

Target 7 Increase the benefits of Small Island developing States and developing countries from sustainable use of marine resources.

Target a Increase scientific knowledge, research capacity and marine technology to improve ocean health.

Target b Provide access for small-scale fishers to marine sources and markets.

Target c Improve the protection and manageable conservation of seas and their assets by executing global law as reflected in UNCLOS. To give a lawful structure to the preservation and practical conservation of seas and their assets

SDG 15 **Life on Land**

Goal 15 is to protect, restore and promote the sustainable use of terrestrial ecosystems. To sustainably manage forests, combat desertification and halt and reserve land degredation and to prevent biodiversity loss. The targets of which this goal consists are the following:

Target 1 Conserve and sustainably use the terrestrial and inland freshwater ecosystems, forests, mountains, drylands and wetlands.

Target 2 Promote the implementation of sustainable management of forests. Halt deforestation, restore degrades forests and sustainable increase afforestation and reforestation globally.

Target 3 Combat desertification and restore land and soil.

Target 4 Ensure the conservation of mountain ecosystems.

Target 5 Take action to reduce degradation of natural habitats.

Target 6 Promote fair share of benefits from genetic resources.

Target 7 End trading and trafficking of protected species of flora and fauna.

Target 8 Introduce measures to prevent the impact of invasive species in various ecosystems.

Target 9 Implement biodiversity and ecosystem values into national and local planning.

Target a Mobilize and increase financial resources to conserve biodiversity and ecosystems.

Target b Mobilize and increase resources to conserve biodiversity and ecosystems for developing countries.

Target c Globally support efforts to combat trafficking of protected species.

SDG 16 **Peace and Justice Strong Institutions**

Goal 16 is to promote peaceful and inclusive societies for sustainable development. As well as providing access to justice for all, and build effective, accountable and inclusive institutions at all levels. The targets of this goal are:

Target 1 Reduce violence and violence related death rates.

Target 2 End trafficking, abuse, exploitation and all other forms of violence against children.

Target 3 Ensure equal access to justice for all.

Target 4 Reduce all forms of organized crimes

Target 5 Reduce bribery and corruption of all forms.

Target 6 Develop institutions that are transparent, accountable and effective at all levels.

Target 7 Enhance inclusive and representative decision-making at all levels.

Target 8 Increase the participation of developing countries in global governance.

Target 9 Make legal identity and birth registration accessible for all.

Target 10 Make public information accessible and ensure fundamental freedoms.

Target a Make the national institutions stronger through international cooperation.

Target b Enforce non-discriminatory laws.

SDG 17 **Partnerships to achieve the Goal**

The last goal of the 17 SDGs is to strengthen the means of implementation and revitalize the global partnership for sustainable development.

B Results

Table 9: Results Baseline System.

	precision	**recall**	**f1-score**	**support**
SDG 0	0.617	0.348	0.445	773
SDG 3	0.217	0.353	0.269	272
SDG 14	0.203	0.332	0.252	262
macro avg	0.346	0.344	0.322	1307

Table 10: Results for mBERT-t.

	precision	**recall**	**f1-score**	**support**
SDG 0	0.870	0.877	0.874	773
SDG 3	0.809	0.779	0.794	272
SDG 14	0.805	0.817	0.811	262
macro avg	0.828	0.824	0.826	1307

Table 11: Results for mBERT-ta.

	precision	**recall**	**f1-score**	**support**
SDG 0	0.916	0.935	0.926	773
SDG 3	0.841	0.838	0.840	272
SDG 14	0.960	0.905	0.931	262
macro avg	0.905	0.893	0.899	1307

Table 12: Results for XLMR-t.

	precision	**recall**	**f1-score**	**support**
SDG 0	0.774	0.895	0.830	773
SDG 3	0.810	0.596	0.686	272
SDG 14	0.765	0.662	0.686	226
macro avg	0.783	0.704	0.734	1307

Table 13: Results for XLMR-a.

	precision	**recall**	**f1-score**	**support**
SDG 0	0.887	0.819	0.851	773
SDG 3	0.797	0.779	0.788	272
SDG 14	0.713	0.889	0.791	262
macro avg	0.799	0.829	0.810	1307

C Mislabelled Abstracts

Table 14: The only case for which mBERT-a predicted SDG 3 while the gold label was SDG 14.

id 2-s2.0-84961213300 **gold label** 14
Title
The toxicity of the three antifouling biocides DCOIT, TPBP and medetomidine to the marine pelagic copepod Acartia tonsa
Abstract
© 2016, Springer Science+Business Media New York.Copepods, the largest group of pelagic grazers, are at risk from exposure to antifouling biocides. This study investigated the toxicity of the antifouling biocides 4,5-dichloro-2-octyl-1,2-thiazol-3(2H)-one (DCOIT), triphenylborane pyridine (TPBP) and 4-[1-(2,3-dimethylphenyl)ethyl]-1H-imidazole (medetomidine) to the copepod Acartia tonsa, using mortality and egg production as endpoints. The toxicity ranking for mortality was as follows: DCOIT (LC50 57 nmol l-1) = TPBP (LC50 56 nmol l-1) > medetomidine (LC50 241 nmol l-1). Egg production was more sensitive than mortality to TPBP (EC50 3.2 nmol l-1), while DCOIT and medetomidine inhibited egg production at roughly the same concentrations (72 and 186 nmol l-1 respectively). Furthermore, TPBP seems to affect egg hatching directly which was not the case for DCOIT and medetomidine. DCOIT and medetomidine might pose an environmental risk as they have been reported to occur in different exposure scenarios or analytical surveys at concentrations only 2-3 times lower than the respective EC10. Reported environmental concentrations of TPBP are few but clearly lower than the EC10 values reported here, suggesting current risk of TPBP to copepods to be moderate.

Table 15: The first of the two cases for which mBERT-a predicted SDG 14 while the gold label was SDG 3.

id	2-s2.0-84997785075	**gold label**	3

Title

Induction of mucosal immune responses against Helicobacter pylori infection after sublingual and intragastric route of immunization

Abstract

© 2016 John Wiley & Sons LtdThere is a current lack of effective mucosal vaccines against major gastroenteric pathogens and particularly against Helicobacter pylori, which causes a chronic infection that can lead to peptic ulcers and gastric cancer in a subpopulation of infected individuals. Mucosal CD4+ T-cell responses have been shown to be essential for vaccine-induced protection against H. pylori infection. The current study addresses the influence of the adjuvant and site of mucosal immunization on early CD4+ T-cell priming to H. pylori antigens. The vaccine formulation consisted of H. pylori lysate antigens and mucosal adjuvants, cholera toxin (CT) or a detoxified double-mutant heat-labile enterotoxin from Escherichia coli (dmLT), which were administered by either the sublingual or intragastric route. We report that in vitro, adjuvants CT and dmLT induce up-regulation of pro-inflammatory gene expression in purified dendritic cells and enhance the H. pylori-specific CD4+ T-cell response including interleukin-17A (IL-17A), interferon-γ (IFN-γ) and tumour necrosis factor-α (TNF-α) secretion. In vivo, sublingual immunization led to an increased frequency of IL-17A+, IFN-γ+ and TNF-α+ secreting CD4+ T cells in the cervical lymph nodes compared with in the mesenteric lymph nodes after intragastric immunization. Subsequently, IL-17A+ cells were visualized in the stomach of sublingually immunized and challenged mice. In summary, our results suggest that addition of an adjuvant to the vaccine clearly activated dendritic cells, which in turn, enhanced CD4+ T-cell cytokines IL-17A, IFN-γ and TNF-α responses, particularly in the cervical lymph nodes after sublingual vaccination.

Table 16: The second of the two cases for which mBERT-a predicted SDG 14 while the gold label was SDG 3.

id 2-s2.0-84978997791 **gold label** 3	
Title	
Demographic, Reproductive, and Dietary Determinants of Perfluorooctane Sulfonic (PFOS) and Perfluorooctanoic Acid (PFOA) Concentrations in Human Colostrum	
Abstract	
© 2016 American Chemical Society.To determine demographic, reproductive, and maternal dietary factors that predict perfluoroalkyl substance (PFAS) concentrations in breast milk, we measured perfluorooctane sulfonic (PFOS) and perfluorooctanoic acid (PFOA) concentrations, using liquid chromatography-mass spectrometry, in 184 colostrum samples collected from women participating in a cohort study in Eastern Slovakia between 2002 and 2004. During their hospital delivery stay, mothers completed a food frequency questionnaire, and demographic and reproductive data were also collected. PFOS and PFOA predictors were identified by optimizing multiple linear regression models using Akaike's information criterion (AIC). The geometric mean concentration in colostrum was 35.3 pg/mL for PFOS and 32.8 pg/mL for PFOA. In multivariable models, parous women had 40% lower PFOS (95% CI: -56 to -17%) and 40% lower PFOA (95% CI: -54 to -23%) concentrations compared with nulliparous women. Moreover, fresh/frozen fish consumption, longer birth intervals, and Slovak ethnicity were associated with higher PFOS and PFOA concentrations in colostrum. These results will help guide the design of future epidemiologic studies examining milk PFAS concentrations in relation to health end points in children.	

D Distribution of the work

For this project we tried to distribute the workload as evenly as possible. In the first stages, we often had meetings where we all attended, to discuss the direction of our research. Later on in the project, when we had to code, we simultaneously worked on it at the same time until someone had it working. For writing of the report we did it roughly the same. First we discussed what had to be done and then we individually incorporated this and provided feedback on the work that had been done by the others.

The presentation was completely done by Myrthe.