Using Language Models for Classifying Startups Into the UN's 17 Sustainable Development Goals

Kfir Bar
School of Computer Science
The College of Management Academic Studies
Rishon LeZion 7570724, Israel
kfir.bar@runi.ac.il

Abstract

We propose a new computational tool for classifying startups into the United Nation's sustainable development goals (SDG), based on their main mission, products, and services. The new tool is based on a natural-language-processing classifier that automatically analyzes a description written about each company using a pretrained language model. We experiment with a number of model architectures and report on the results. The datasets that we use for training and evaluating our models are obtained from company listings that store impact-related information. Our best models achieve accuracy around 77% when assigned to classify companies for the five main SDG categories. We make our best model publicly available, to encourage the community to advance research in the relevant directions.

1 Introduction

Impact Tech Startup (ITS) is a new, rapidly developing concept, defining startups which adopt innovative strategies to tackle a variety of social and environmental problems within a for-profit framework. They are usually backed by private investments, similar to the venture capital models seen in other domains. Unlike other companies, which may focus on different areas such as security (e.g., face recognition), gaming (e.g., virtual-reality computer games) or finance (e.g., developing financial trading platforms), ITSs create social and/or environmental impact, thereby making a contribution to levels of social or environmental causes.

Examples of such startups are those dealing with loneliness of elderly people, fighting hunger and food waste, helping refugees find employment, using microalgae and nanofluids to generate energy and oxygen, or capturing airborne moisture for generating clean drinking water. This category of organizations has been recently discussed in [Gidron *et al.*, 2021]. This new hybrid form of organization has been recognized by the United Nations (UN) interagency task team (IATT) on science, technology and innovation (STI) for sustainable development goals (SDG) in 2015, as an emerging form with the potential to catalyze the business sector toward aiding SDGs [Vinuesa *et al.*, 2020].

The importance of conducting research on this new organizational category is related to the growing trend of introducing technology-based innovative solutions (products and services) in the domains of society and environment.

Traditionally, the societal and the environmental domains have been in the responsibility of the public and the non-profit sectors, and those sectors are less inclined to engage in technology-based innovations. The startup format, anchored in entrepreneurship of individuals, has demonstrated the potential of products and services based on new technologies, to be easily applied in all corners of the globe. The application of the startup format into the societal and the environmental domains that is currently taking place in the form of ITSs, is a very important development that calls for the attention of the research community [Gidron *et al.*, 2021].

One of the challenges of studying impact has been the lack of agreement and shared language on what constitutes positive social and environmental impact [Molecke and Pinkse, 2017; Perrini *et al.*, 2020]. In recent years, an important unifying factor in defining the parameters of impact have become the UN SDGs. This framework, issued in 2015, sets out 17 interlinked goals for a better and more sustainable future, along with 169 indicators. These SDGs are considered the most ambitious effort to place such goal setting at the center of global policy and governance [Biermann *et al.*, 2017]. Concomitantly, SDGs have been identified as highly beneficial for both business and investors as they present the best long-term strategic market outlook for global policy making [Pedersen, 2018; Surana *et al.*, 2020]. More specifically, SDGs are a valuable reference point for impact investors [Reisman and Olazabal, 2016; Schramade, 2017].

Measuring impact and finding the relevant organizations has been traditionally done with conventional tools such as surveys. Using such tools has always been a challenge when the intention is to collect information at a large scale worldwide. To mitigate this challenge, in this work, we present a new computational method for automatically finding ITSs in local and international business listings. To do that, we train a classifier that assigns an SDG label to a startup, using only a short description written in English about the startup's main mission. Specifically, we use customized datasets for training a natural-language-processing (NLP) classifier based on a pre-trained language model, designed in different configurations. Our base language model is BERT [Devlin *et al.*, 2018].

The methodology we devise can also serve as the basis for research on single regions, countries, or cities, as well a comparative analysis between them. That said, once an algorithm to identify ITSs in startup databases has been created, it is less time- and resource-consuming than the traditional options, and has the built-in advantage of easy comparison of variables with those of non-impact startups. We provide an example of such an analysis in Section 5. Our model design and experiments are described in Section 3. We make our best model publicly available.[1]

2 Related Work

There is a growing body of research that examines computational methods for measuring impact and SDG adherence. The ones that are most relevant to our study, use computational tools to process unstructured texts for SDG classification. In [Guisiano and Chiky, 2021], the authors build a simple BERT-based classifier for automatically labeling an input document with the SDGs that were mentioned in the text. Technically speaking, the classifier was designed to label every text representing a project or an idea submitted to the OnePlanet[2] platform. The dataset that was used for training the classifier on was relatively small, and created synthetically, and in any case, the dataset did not contain startup descriptions.

In another work [Hajikhani and Suominen, 2021], the authors built a classifier for tagging patents with SDG labels. They trained a probabilistic model (Naive Bayes) on a set of patents that were automatically tagged using a deterministic keyword-based algorithm which they developed.

Topic modeling is a traditional NLP task for automatically grouping words together into clusters. It has been used by [LaFleur, 2019] to classify documents, which were published by the United Nations, into SDGs.

All of these works did not analyze data about startups or any other type of organization.

In a recent work [Tiba *et al.*, 2021], 28 startup ecosystems have been automatically analyzed for measuring sustainability-related activities. The authors used a topic modeling algorithm to process the websites of the ecosystems' startups and measured the intersection of the lists of keywords (i.e., topic) they obtained by the topic modeling algorithm, with the words appear on the websites. By tagging every topic with the related SDGs, they were able to attribute SDGs to the relevant ecosystems.

Words tend to be ambiguous; their true meaning usually depends on their surrounding context. Consequently, word-matching algorithms such as the one they used, may suffer from high rate of false positives. Our language-modelling approach is less sensitive to this problem as it takes the entire text as an input. Therefore, our model is capable of labeling every startup individually and does not require the entire ecosystem to be present at processing time.

[1] https://huggingface.co/amannor/bert-base-uncased-sdg-classifier

[2] https://oneplanet.com

3 Method

We develop a new algorithm to automatically classify a startup for the relevant SDGs. We take a pure data-driven approach by training a machine-learning based classifier on digital representations of startups, which have been manually classified by domain experts. Specifically, in this study we use a short description written about every startup, either by the marketing professionals of each company (e.g., typically known as the "About Us" section in a commercial website) or by the platform curators from which we obtain most of our data for this study.

Our classifier takes as input a text describing the startup's main mission, and returns a label representing the relevant SDG. We experiment with two different label sets. In one set we have 18 labels, broken down into the 17 SDG labels in addition to "no-impact" representing companies executing with no social/environmental impact. For the second set of labels, referred to as `5Ps`, we cluster the 17 goals into 5 groups, according to the definitions provided by the UN:[3] People (SDGs 1, 2, 3, 4, and 5), Planet (SDGs 6, 12, 13, 14, and 15), Prosperity (SDGs 7, 8, 9 ,10 ,11), Peace (SDG 16), and Partnerships (SDG 17). Similarly, we add to this set the "no-impact" label.

3.1 Datasets

The first dataset we use in this study is provided by Rainmaking,[4] an online platform that groups a number of entrepreneurs who came together to use their experience to build impactful businesses. As part of their assets, they developed Compass, a database that provides information about over 2,000 impact startups from around the world addressing the 17 SDGs. Compass provides some additional information about each startup, such as the domain, maturity, founding year, and funding stage. Impact professionals had manually connected each company in the database with the relevant SDGs, focusing on its main mission and product line. With Rainmaking's permission, we use all companies from Compass that have an English description as well as an assigned SDG label, and defined under one of the following startup maturity conditions: seed, series A, as well as undisclosed. All of the chosen startups are labeled with one primary SDG, assigned by a group of experts from Rainmaking after considering a suggestion made by the startups. To decide which SDG applies for a specific startup, they had to look at the targets that were originally defined for every SDG and find indications in the company's mission for addressing those targets. Some startups are additionally labeled with SDG targets, as well as with a secondary SDG label. For lack of information on how the secondary goals were annotated, and the low frequency of target assignments, we decided not to use them in this study. The description is a short text about the company and it is typically taken from the company's website, sometimes with a few edits done by Rainmaking's professionals. The descriptions are 120.2 words long on average. The original Compass dataset, by its definition does not contain non-impact startups. We realized that non-impact startups might

[3] https://www.unescwa.org

[4] https://rainmaking.io

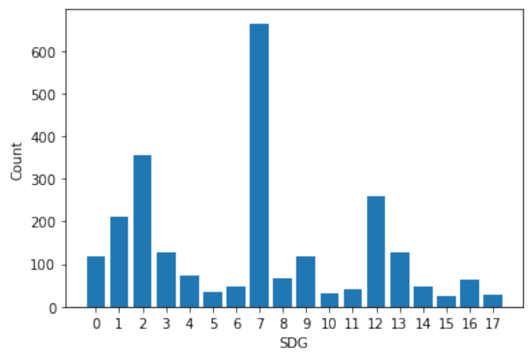

Figure 1: SDG Distribution of the Compass dataset. SDG 0 refers to "no-impact".

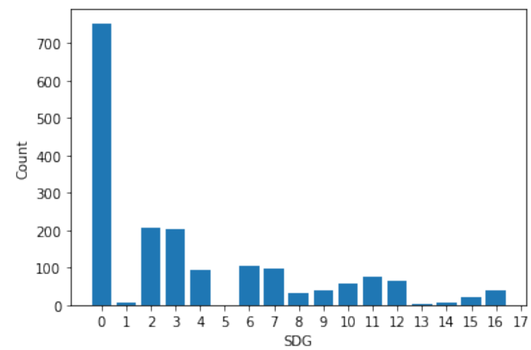

Figure 2: SDG Distribution of the SNC dataset. SDG 0 refers to "no-impact".

get classified as ITS by the classifier, only based on a topical similarity with ITSs from the training set that define their mission in the same domain. For example, we observe this ambiguity with non-impact financial technology (FinTech) startups that might get labeled as SDG 1 (No Poverty), since there are a substantial amount of ITSs under SDG 1 in the dataset which use blockchain technologies to achieve their goal. Therefore, a human impact expert (one of the authors) had followed an annotation methodology similar to what was done to create Compass, to add 119 non-impact startups to the dataset, which are active in the relevant fields. The non-impact startups were chosen from another dataset, which we obtained from Start-up Nation Central (see the description below). Figure 1 shows the distribution of the Compass dataset for SDGs, including the "no impact" label, represented by 0. Overall the dataset has 2,448 startups, which we randomly split into train and test sets.

Additionally, we collaborate with Start-up Nation Central (SNC),[5] an independent non-profit organization that builds and maintains Start-Up Nation Finder, an online search platform for thousands of startups founded by Israeli entrepreneurs. A team of experts from SNC annotated startups with SDG labels following a slightly different annotation approach. The startups were randomly selected from a database of more than 6,000 companies, not necessarily considered as ITS. We refer to this dataset as SNC. Overall, there are originally about 2,000 startups in the dataset; some were labeled as uncertain by the experts. By removing those companies from the dataset, we were left with 1,799 startups; their SDG distribution is shown in Figure 2. Their description length is 120 words long on average.

It is clear from the data that the two datasets have different SDG distributions. This difference stems from the different annotation guidelines, as well as from the different geographical distribution of the startups. While SNC contains only Israeli startups, the Compass dataset contains startups from all over the world.

As seen, both datasets are not distributed equally for the SDGs. To handle this unbalance, we combine SNC and Compass into one dataset. We run our experiments on each dataset separately, as well as on the combined one and report on results in the following section. Table 1 summarizes the sizes of the three datasets.

[5]https://startupnationcentral.org/

Dataset	Train	Test
Compass	1,934	514
SNC	1,447	352
Combined	3,381	866

Table 1: Dataset sizes

As mentioned above, we cluster the labels into 5 categories, the `5Ps`; the distribution of both datasets for the 6 labels (0 indicates "no-impact") is provided in Figure 3.

3.2 SDG Classifiers

Our classifiers are all based on BERT, a masked language model which utilizes large amounts of free English texts taken from unpublished books as well as from Wikipedia.

We take three different fine-tuning approaches, and compare their results.

Document Classification

In the first approach, referred to as BERT-Classifier, we follow a common practice of fine tuning BERT for document classification [Adhikari *et al.*, 2019], using the Hugging Face's transformers library [Wolf *et al.*, 2019]. We use the description of the startup as an input, and configure the model to classify it for either 18 or 6 labels, depending on the label set we experiment with.

Fine-tuning with Masked Keywords

In the second approach, which we referred to as BERT-MLM, we take inspiration from other works [Kawintiranon and Singh, 2021; Tian *et al.*, 2020], which showed that putting the focus of the language model on domain relevant words prior to the fine tuning process on the downstream task, performs better than the standard pre-trained BERT model. Instead of finding those relevant words automatically, we manually compile a list of SDG keywords from relevant UN literature. Overall, 393 words were included in this list (see Appendix A). During pretraining, BERT was fitted to correctly guess words that were intentionally masked in the in-

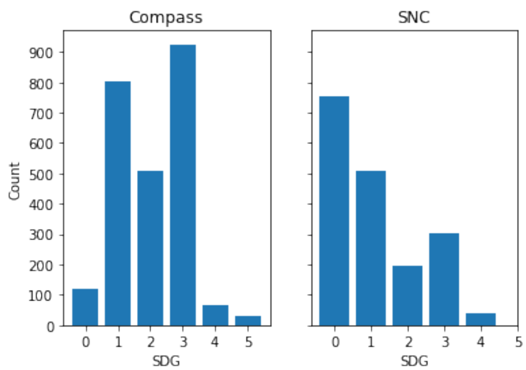

Figure 3: 5Ps distribution of both the Compass and the SNC datasets. Label 0 refers to “no-impact”.

put.[6] Originally, 15% randomly chosen words from the training data were modified; some were masked and some were replaced by a random word. This training task is called masked language-modelling (MLM). In this approach, we start by fine-tuning the pre-trained BERT model for the MLM task on a collection of company descriptions [Qader *et al.*, 2018] in which we masked all instances of the 393 relevant keywords. That gave us 4% of masked words in the training set. In order to get to the recommended rate of 15%, we masked some additional random words. The collection contains about 25 Million words. We run this MLM fine tuning for 3 epochs.

As a second step, we fine tune this model for document classification, following the same process as described in our first approach.

Prompt-based Learning

In our third approach, referred to as BERT-Prompt, we exploit the natural BERT's MLM characteristic for performing classification. In this technique, which is called prompt-based learning, we modify the input by adding a prompt to it for triggering the model to guess a label-related masked word. For example, given the input text ”The movie was great!” for sentiment analysis, we concatenate it with ”This is a `[MASK]` review”, requesting the model to guess the masked word; ideally, the model will prefer the word ”positive” rather than the word ”negative”. This topic is adequately covered by a recently published review [Liu *et al.*, 2021]. We experiment with two prompt templates: (1) `The mission of the company is related to SDG [MASK]. [SEP] <DESC>`, and a shorter version (2) `Sustainable goal (SDG) [MASK]. [SEP] <DESC>`. In both templates, `<DESC>` is a placeholder for the input textual description, `[MASK]` is the special mask token, and `[SEP]` is a special separator token. During fine tuning, we train the model to predict the number of the relevant SDG, including 0 for the non-impact startups. During testing, the model predicts a distribution of the entire vocabulary. We consider only the tokens representing the 18 labels; the token with the highest probably is chosen as model's prediction.

[6]Masking is done by replacing the word with a special token `[MASK]`

Model	Compass	SNC	Combined
Baseline (chance)	13.2	22.2	13.1
BERT-Classifier	**79.0**	**68.4**	73.6
BERT-MLM	**79.0**	66.4	73.5
BERT-Prompt (1)	78.5	65.2	**74.0**
BERT-Prompt (2)	77.2	66.0	73.0

Table 2: Evaluation results on the 18-labels set. Results are reported for all classification approaches described in the text. BERT-Prompt (1) and BERT-Prompt (2) refer to the first and second prompts, respectively.

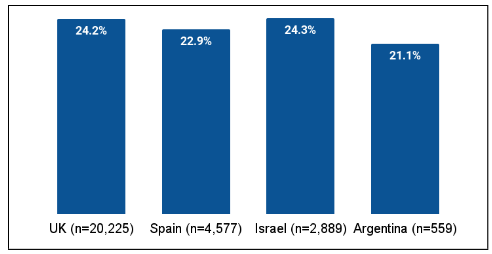

Figure 4: ITS percentage in each examined country.

4 Experimental Results

All our experiments are evaluated with weighted-average F1. We run every experiment 5 times, each with a different seed value, and take their average results. Unless written specifically, we fine tune every model for 5 epochs, with learning-rate value of $5e-5$. We train and run our models on the Galax GeForce RTX 3090 GPU with 24GB of memory.

BERT comes in two main configurations. The most common one is `bert-base-uncased`, consisted of 12 Transformer [Vaswani *et al.*, 2017] layers, assuming input texts written in lower case. A larger configuration is `bert-large-uncased`, which is almost double in size. As a first step we try both configurations, specifically arranged for our first classification approach, BERT-Classifier, on the combined dataset. The results show no significant improvement by the large model (76.5 vs. 76.2 F1); therefore, in all that follows we use the `bert-base-uncased` model.

Table 2 summarizes the evaluation results of all models and datasets on the 18-labels set. For each dataset, we use the train/test split described in the previous section. The first row relates to the F1 score we obtain by guessing the label for every company, using the label distribution.

We see no significant difference between the three approaches; if at all, it seems like the most standard approach works best in most of our experiments. The best results are obtained on the Compass testing set, when the models are fine-tuned on the Compass's training set. That is despite the smaller chance to guess the correct label in Compass than in SNC. Therefore, a possible explanation for this might be the fact that Compass was designed specifically to store impact-tech companies. Their descriptions may include clear indications for impact, as opposed to SNC, which was designed for a different goal, unnecessarily related to impact tech. On the other hand, SNC is focused on one country, Israel, and thereby might be biased toward the SDGs mostly

SNC model, SDG 6 (Clean Water and Sanitation)
water quality monitoring solutions . cywat technologies develops cyber - like physical water protection systems for the water - monitoring industry and water authorities , with the endorsement of the israeli government . the company offers water suppliers across the globe a solution for real - time water quality monitoring , preventing water pollution events and ensuring public health and safety .
Compass model, SDG 6 (Clean Water and Sanitation)
water quality monitoring solutions . cywat technologies develops cyber - like physical water protection systems for the water - monitoring industry and water authorities , with the endorsement of the israeli government . the company offers water suppliers across the globe a solution for real - time water quality monitoring , preventing water pollution events and ensuring public health and safety .
SNC model, SDG 7 (Affordable and Clean Energy)
automatic meter reading for electrical systems . unique technology develops automatic meter - reading solutions designed to address challenges including power theft , conservation , and grid stability . the company ' s technology enables two - way communication over power lines . unique technology also makes systems for automated street lighting , electrical power for marinas , and electric vehicle charging . working hand - in - hand with local utilities and businesses worldwide , the company develops tailor - made solutions to meet individual needs and budget constraints , providing technical support and after - sale service across the globe .
Compass model, SDG 7 (Affordable and Clean Energy)
automatic meter reading for electrical systems . unique technology develops automatic meter - reading solutions designed to address challenges including power theft , conservation , and grid stability . the company ' s technology enables two - way communication over power lines . unique technology also makes systems for automated street lighting , electrical power for marinas , and electric vehicle charging . working hand - in - hand with local utilities and businesses worldwide , the company develops tailor - made solutions to meet individual needs and budget constraints , providing technical support and after - sale service across the globe .

Table 3: Example of two startups which were correctly classified as SDG 6 and 7, respectively. For each startup, the most important text segments for classifying the startup with the correct label are highlighted in Red. Each description is shown twice. Once in the eyes of the model that was trained on SNC, and in the other in the eyes of the model that was trained on Compass. Segments highlighted in Blue are the less important ones for the same label. The highlights were automatically created using the SHAP library (https://shap.readthedocs.io)

Model	Compass	SNC	Combined
Baseline (chance)	25.0	26.1	26.2
BERT-Classifier	**83.6**	**72.0**	**77.7**
BERT-MLM	82.6	70.1	77.1
BERT-Prompt (1)	83.5	69.0	76.6

Table 4: Evaluation results on the `5Ps` label set. Results are reported for all classification approaches described in the text. BERT-Prompt (1) refers to the first prompt mentioned in the text.

addressed by Israeli companies. Our ultimate goal is to build a model that can successfully classify companies taken from any repository. Therefore, our final model, which we release with this paper, is trained on the combined dataset.

In both datasets the label distribution is extremely unbalanced; therefore, we cluster the labels as described in Section 3 and run the same experiments with the set of clusters, also known as the `5Ps`. As before, label 0 indicates "no-impact". Table 4 summarizes the evaluation results of all models and datasets on the `5Ps` label set.

Overall, we see an improvement in performance on the clustered label set than on the 18-labels set. The task is relatively easier than classifying startups for the 18-labels set. The chance performance (first row in Table 4) provides the evidence for that. As can be seen in Figure 3, labels 4 and 5 have a relatively small support, representing the Peace and Partnership categories, respectively. Each of those two categories includes only one SDG, 16 and 17, while the other categories group more than one SDG together. Therefore, naturally our classifiers are more focused on classifying companies for the first 0-4 labels. At a higher level, we believe that it is more difficult to build innovative technologies that address SDGs 16 and 17, especially for startups. We provide two examples of startup descriptions in Table 3, along with highlights of the most important text segments for two models.

5 Practical Use Case: Mapping Impact in Different Countries

To demonstrate how this model can be used in practice, we use our BERT-Classifier model which was trained on the combined dataset, to process some unlabeled descriptions of startups from different countries. Thankfully, we have been given a research-level access to Crunchbase[7] which has live data about private companies from all over the world, built and maintained by a community of contributors, partners, and in-house data experts. Each company is indexed with some additional relevant information, such as the founding date, the company's business fields, location identifiers, as well as a description of its main mission, written in English. The Crunchbase's programming interface enabled us to retrieve lists of startups, organized by their country of origin. We use a similar definition of startup as in [Gidron *et al.*, 2021]; specifically, we query Crunchbase for active companies with up to 250 employees which reported on overall revenue that does not exceed 50 million dollar. We exclude all companies

[7]https://www.crunchbase.com

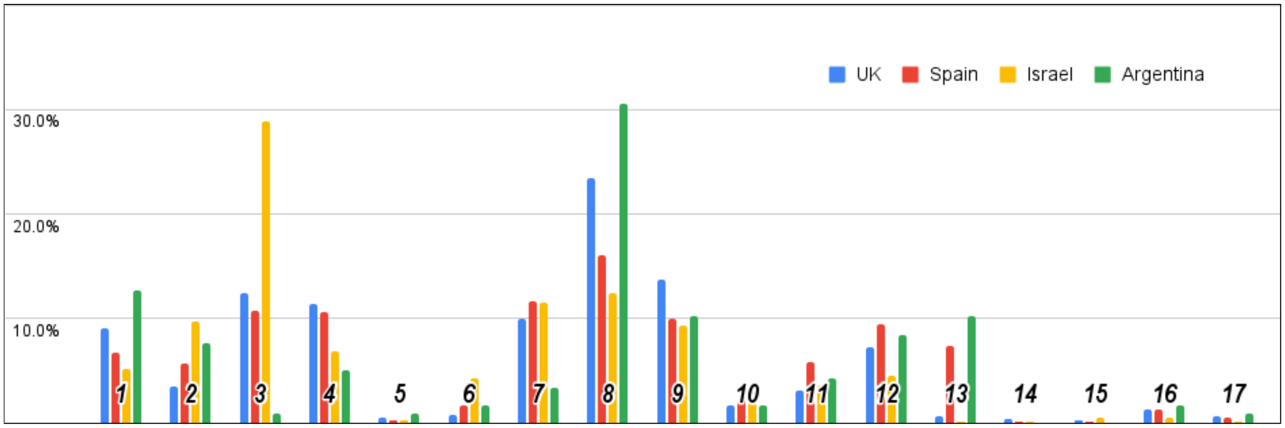

Figure 5: SDG Distribution per country.

that do not have a textual description written in English. We ran all queries during January, 2022. For the remaining companies, we use our 18-labels model to automatically assign an appropriate SDG label which we aggregate per country, and calculate the percentage of dominance of each SDG among the ITSs in each country.

We randomly pick 4 countries to demonstrate the potential of using our model to gain impact-related insights. However, this analysis can be done with all other countries for which Crunchbase has data. Figure 4 shows the percentage of startups that are labeled as ITS (assigned with SDG 1-17) per country. The number of startups that we analyze in each country is mentioned next to its name.

While the differences are insignificant, we still learn from Figure 4 that the UK and Israel have the highest rate of ITSs among the startups that we analyzed from Crunchbase. We provide another visualization of the same analysis in Figure 5, which summarizes the percentage of dominance of each SDG among the ITSs in each country. We can learn from this figure that in Israel the percentage of ITSs that focus on SDG 3 (Good Health and Well-being) is significantly higher than in the other 3 countries. On the other hand, in Argentina, the UK, and to some extent Spain, the most dominant SDG is 8 (Decent Work and Economic Growth). More studies can be carried out using our model; for example, focusing only on startups that have recently raised funding, as well as monitoring the focus of ITSs on specific SDGs over time per country.

One of the authors, who is a domain expert in the field of impact tech startups, have manually analyzed the results of randomly chosen 85 startups, reporting on about 70% agreement with the algorithm decisions.

6 Conclusions

In this paper we made an attempt to identify a new organizational category (ITS) through a computational approach. This category is important for the world's sustainable goals as it groups together organizational entities that engage in protecting the planet and promoting society, and do so by introducing technologically based innovations. The methodology we developed enables their identification and categorization, studying them in depth and eventually supporting them through investments, mentoring and policy.

We took a pure data-driven approach by training a natural-language-processing classifier to process short descriptions written about startups, and classify them into the SDGs. We experimented with different model architectures, based on an English pre-trained language model, which we evaluated on datasets obtained from two different sources.

We believe that we have demonstrated the potential of using such a tool for processing large databases of companies originated in different countries. We make our model publicly available for the community, to encourage research in the areas of interest for advancing social and environmental causes which are related to all the SDGs.

Our methodology has some limitations. The focus on startups plays an important role in our work. Usually, a startup has one main mission, focusing on building a product or providing a service for achieving a specific goal, for which we measure impact. In this work we take a simple deterministic approach for identifying startups in the databases that we work with, typically by limiting the number of employees and/or capping revenue. However, some small number of non-startup organizations such as consultancy firms as well as co-working hubs may still be included in our startup lists.

Another limitation lies in the design of our models to predict only one SDG for a given startup. Some startups may be related to several SDGs, as was seen in both the Compass and the SNC datasets. We plan to work on supporting multilabel classification as one of our future directions.

We realize that tagging a startup with a specific SDG label sometimes can be difficult even for a human expert due to the relatively high ambiguity in the definition of some of the SDGs, as well as insufficient information in the description written about the company. Additionally, companies may represent themselves through those descriptions in an overly positive light with respect to social impact. It has been previously referred to as *greenwashing* [Lyon and Montgomery, 2015]. Therefore, we plan to explore better ways in which we could take into consideration the targets and indicators of the SDGs and find evidence for their application in daily activities of the company.

References

[Adhikari *et al.*, 2019] Ashutosh Adhikari, Achyudh Ram, Raphael Tang, and Jimmy Lin. Docbert: BERT for document classification. *arXiv preprint arXiv:1904.08398*, 2019.

[Biermann *et al.*, 2017] Frank Biermann, Norichika Kanie, and Rakhyun E Kim. Global governance by goal-setting: the novel approach of the UN Sustainable Development Goals. *Current Opinion in Environmental Sustainability*, 26:26–31, 2017.

[Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Gidron *et al.*, 2021] Benjamin Gidron, Yael Israel-Cohen, Kfir Bar, Dalia Silberstein, Michael Lustig, and Daniela Kandel. Impact tech startups: A conceptual framework, machine-learning-based methodology and future research directions. *Sustainability*, 13(18):10048, 2021.

[Guisiano and Chiky, 2021] Jade Guisiano and Raja Chiky. Automatic classification of multilabel texts related to Sustainable Development Goals (SDGs). In *TECHENV EGC2021*, 2021.

[Hajikhani and Suominen, 2021] Arash Hajikhani and Arho Suominen. The interrelation of Sustainable Development Goals in publications and patents: A machine learning approach. In *CEUR Workshop Proceedings*, volume 2871, pages 183–193, 2021.

[Kawintiranon and Singh, 2021] Kornraphop Kawintiranon and Lisa Singh. Knowledge enhanced masked language model for stance detection. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4725–4735, 2021.

[LaFleur, 2019] Marcelo LaFleur. Art is long, life is short: An SDG classification system for DESA publications. 2019.

[Liu *et al.*, 2021] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*, 2021.

[Lyon and Montgomery, 2015] Thomas P Lyon and A Wren Montgomery. The means and end of greenwash. *Organization & Environment*, 28(2):223–249, 2015.

[Molecke and Pinkse, 2017] Greg Molecke and Jonatan Pinkse. Accountability for social impact: A bricolage perspective on impact measurement in social enterprises. *Journal of Business Venturing*, 32(5):550–568, 2017.

[Pedersen, 2018] Claus Stig Pedersen. The UN Sustainable development goals (SDGs) are a great gift to business! *Procedia Cirp*, 69:21–24, 2018.

[Perrini *et al.*, 2020] Francesco Perrini, Laura A Costanzo, and Mine Karatas-Ozkan. Measuring impact and creating change: a comparison of the main methods for social enterprises. *Corporate Governance: The International Journal of Business in Society*, 2020.

[Qader *et al.*, 2018] Raheel Qader, Khoder Jneid, François Portet, and Cyril Labbé. Generation of company descriptions using concept-to-text and text-to-text deep models: dataset collection and systems evaluation. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 254–263. Association for Computational Linguistics, 2018.

[Reisman and Olazabal, 2016] Jane Reisman and Veronica Olazabal. *Situating the next generation of impact measurement and evaluation for impact investing*. Rockefeller Foundation New York, NY, USA, 2016.

[Schramade, 2017] Willem Schramade. Investing in the un sustainable development goals: opportunities for companies and investors. *Journal of Applied Corporate Finance*, 29(2):87–99, 2017.

[Surana *et al.*, 2020] Kavita Surana, Anuraag Singh, and Ambuj D Sagar. Strengthening science, technology, and innovation-based incubators to help achieve Sustainable Development Goals: Lessons from india. *Technological Forecasting and Social Change*, 157:120057, 2020.

[Tian *et al.*, 2020] Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4067–4076, Online, July 2020. Association for Computational Linguistics.

[Tiba *et al.*, 2021] Sarah Tiba, Frank J van Rijnsoever, and Marko P Hekkert. Sustainability startups and where to find them: Investigating the share of sustainability startups across entrepreneurial ecosystems and the causal drivers of differences. *Journal of Cleaner Production*, 306:127054, 2021.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[Vinuesa *et al.*, 2020] Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications*, 11(1):1–10, 2020.

[Wolf *et al.*, 2019] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.

A Appendix: List of SDG-related Keywords

circular organized vulnerable crime school dioxide city ocean detention illegal national tax animals birth exploitation work medicine hydroelectric capitalism life resources psychological infected genetics indigenous accountable fishermen irrigation impact plants internet coverage parks town consumption societies aids disease international air carbon public vaccines disasters age literacy degradation inclusive wetlands greenhouse water migrant heritage women inclusion population river contamination service investments children sanitation consume road wave electricity power law sexual supply labour extinct autonomy infrastructure inadequate governance reef materials species environmental distortion drought opportunity alcohol community land development technological scholarship biomedical alternative charcoal free support education torture migrants clean illicit human empowerment fossil sharing production living qualified society habitats creativity economy refugees services weather fisher social information opportunities transfer pollution arms girl threats religion chain aid vehicles mobile transparency statistics green science race conflict change technologies adaptation cooperation displaced sea fuel chains tourism warming slavery identity maize cop care fauna training standards climate health disadvantaged malaria communication violence diversity skills wildlife child abuse temperature food reproductive infection coral dept institutions biodiversity agricultural hunger maternal roads accident sustainability census affordable network crimes rights domestic citizenship developing female retail sustainable civil energy assets data tobacco patterns medical management equal technology disaster solid procurement paris discrimination market natural coast security teachers phone girls oceans stability industry vaccine wealth improving migration foreign harvesting job mortality trade mental dignity suburban transportation therapy weapon investment hygiene regional disappearance invasive economic bribery conserve hate fisheries agriculture accidents dental reduction seizures forest culture access wash peaceful toilets alien ice lifelong coal plant loss stolen commercial goods genetic income decent partnership sex disabilities youth financial ict universal unpaid pattern capture insurance homeless warning tree finance hungry legal justice trafficking rivers cultural emission pay stocks conversion bees reduce accountability marriage settlements peace poor wood recycling rural childhood productivity soldiers fisherman organisms emissions hazards poverty tuberculosis resource fishing conflicts wind responsible trees conservation organism safe studying cities transport feminism workers hiv renewable nutritional wastewater knowledge healthcare coastal quality theft cleaner soil fuels cars resolution diseases enterprises turbine disability needs jobs global equality corruption secure gas safety hepatitis paid entrepreneurship teacher housing medicines gender protection healthy fish marine solar drinking waste ethnicity industrial equity urban growth electrical flora treatment detainees produce skill unemployment wasting capacity extreme terrorism gdp battery smart doha nutrition efficiency product ecological planning environment innovation ecosystems employment freedom banking networks learning who forests.