THE DEVELOPMENT ASSISTANCE COMMITTEE:
ENABLING EFFECTIVE DEVELOPMENT

LINKING AID TO THE SUSTAINABLE DEVELOPMENT GOALS - A MACHINE LEARNING APPROACH

Arnaud Pincet, Shu Okabe, Martin Pawelczyk

OECD DEVELOPMENT CO-OPERATION WORKING PAPER 52

Authorised for publication by Jorge Moreira da Silva, Director, Development Co-operation Directorate

BETTER POLICIES FOR BETTER LIVES

January 2019

OECD Working Papers

Comments on the present Working Paper are welcomed and may be sent to arnaud.pincet@oecd.org — the Development Co-operation Directorate, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Please cite this paper as Pincet, A., Okabe, S. and Pawelczyk, M. (2019), “Linking Aid to Sustainable Development Goals: A machine learning approach”, OECD Development Co-operation Working Papers, No , OECD Publishing, Paris.

Keywords: Official Development Finance, Sectors, Sustainable Development Goals, Machine Learning, Text Mining, Artificial Intelligence, Credit Reporting System, Innovation

Table of contents

Tables

Figures

Boxes

Acknowledgements

The authors would like to express their deep gratitude to Olivier Cattaneo head of the Policy Analysis team and Haje Schütte, head of the Financing for Sustainable Development division for supporting the development of innovative methods inside the Development Co-operation Directorate Division. Authors would also like to thank Valérie Gaveau, Valérie Thielemans, Cecilia Piemonte, and Olivier Bouret for their technical advices, their expertise, and their constant availability during all the phases of the project.

On behalf of the Secretariat, authors are truly grateful for the data provided by the Italian Cooperation Agency, especially by its WP-Stat representative: Adolfo Morrone.

Julia Benn, Giorgio Gualberti, Rachel Morris and Konstantin Poensgen have significantly contributed to improve the working paper with advice, comments and review.

Finally, special thoughts for Sebastien Miroudot, Koen Deconinck, and Eric Bensel for their strategic insights.

Outside the Development Cooperation Directorate, authors would like to thank Ciara Begley and Mary Brophy; Ana Henche Cuesta and her colleagues from MAEC, Hubert Drolet and his team, Michael Morris and his colleagues Alice Marshall and Paddy Brock, for their thoughtful comments and insights on how this work can be relevant for the DAC and its broader community.

Abbreviations and acronyms

CRS	Creditor reporting system
CSR	Corporate social responsibility
DAC	Development Assistance Committee
DCD	Development Co-operation Directorate of the OECD
FSD	Financing for sustainable development
HAI	Human Assets Index
IDFC	The International Development Finance Club
LDC	Least developed country
LIC	Low income country
LLDC	Landlocked developing country
LMIC	Low-middle income country
ODA	Official development assistance
OECD	Organisation for Economic Co-operation and Development
SDG	Sustainable Development Goal
SDSN	Sustainable Development Solution Network
TOSSD	Total official support for sustainable development
UMIC	Upper-middle income country
UN	United Nations
USD	United States Dollars

1. Introduction

1.1. Machine learning is an emerging tool in the international development context

How is machine learning already applied to international development? In the development field, machine learning has been used in a wide range of applications, from identifying successful entrepreneurs in emerging economies (Mckenzie and Sansone, 2017[1]) to migration patterns in Bangladesh (Lu et al., 2016[2]).

Machine learning is increasingly used to track progress toward certain SDGs. For example, the World Bank used both traditional supervised methods to predict poverty rates (SDG1) and topic modelling to classify their reports into meaningful categories (Dupriez, 2018[3]). Text-as-data has also been used to track inequalities (SDG10) (Parthasarathy, Rao and Palaniswamy, 2017[4]). Other uses of machine learning research are found in poverty prediction, housing conditions, food security or over-fishing estimations (SDG1, 2, & 14) (Blumenstock, 2018[5]), (Jean et al., 2016[6]), (Goldblatt et al., 2017[7]), (Gounden, Irvine and Wood, 2015[8]), (Park et al., 2016[9]), (Parthasarathy, Rao and Palaniswamy, 2017[4])

1.2. Why do we need new tools in support of SDG and aid statistics?

The emergence of smart computing tools can help the Development Assistance Committee (DAC)[1] maintain its relevance in the Sustainable Development Goals (SDG) era, by making better use of large volumes of complex information, respond faster to changes in the policy environment, and alleviate the reporting burden of its members.

The DAC is an international platform hosted by the OECD. It gathers the largest national contributors of aid. The DAC "*promotes development co-operation and other policies so as to contribute to sustainable development, including pro-poor economic growth, poverty reduction, improvement of living standards in developing countries, and a future in which no country will depend on aid*." In this regard, the Secretariat of the DAC collects the transaction flows that corresponds to aid: the official development assistance figures (ODA).

ODA comprises "*flows to countries and territories on the DAC List of ODA Recipients and to multilateral institutions which are: provided by official agencies, including state and local governments, or by their executive agencies; and each transaction of which: is administered with the promotion of the economic development and welfare of developing countries as its main objective; and is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 per cent).*" (OECD, 2018[10]). In 2017, ODA from members of the DAC amounted USD 146.6 billion. ODA comprises multiple development projects published in a database maintained by the OECD: the Creditor Reporting System (CRS). The CRS also gathers data on development projects

[1] DAC members comprise 29 countries and the EU institutions. Please refer to the OECD DAC website to see the full list.

reported by other countries, multilateral organisations, and philanthropic foundations. It gather information on other official transaction, for example loans from aid agencies that are not concessional: Other official flows (OOF)[2]. **Overall, more than 250 000 projects are reported to the CRS each year**.

The Working Party on Development Finance Statistics (WP-STAT), a subsidiary body of the DAC, has been developing and testing methods for tracking development assistance in support of the 2030 Agenda for Sustainable Development. For example, attempts have been made to map the existing statistical classifications (purpose codes and policy markers[3]) to the Sustainable Development Goals (SDGs) (Benn and Gaveau, 2015[11]). The conclusion has been, however, that a direct link is not feasible, even if these classifications were updated to cover all SDG activity areas[4].

With existing methodologies, it is not possible to link projects to certain SDGs: some purpose codes are not matched by a corresponding SDG (1, 8, 10 or 12). Despite attempts to update the mapping since then, the coverage is still partial and gaps exist in the mapping. This does not allow policy analysts and donors to get a full picture of where ODA stands in terms of SDG financing. Moreover, current purpose codes[5] allow reporters to link each project to only one category. This does not reflect the complexity of projects as well as the multiple causes one project can serve. For example, a country financing a project in health and education would not be able to report on both categories. The members of the DAC agreed to introduce a system of multiple purpose code at the July 2016 meeting of the WP-STAT but the implementation of this system is likely to take a few years.

Such reasons have pushed the OECD Development Co-operation Directorate (DCD) to propose a method for reporting the focus of development projects in CRS: "*To properly identify the contribution of development co-operation activities to SDGs and targets it is therefore necessary to generate new information on a project-by-project basis, creating a field to store the data and common reporting guidelines*" (Benn and Gaveau, 2015[11]).[6] The WP-STAT approved the creation of a new field for collecting data on the SDG focus of aid activities at its June 2018 meeting.

This approach allows to understand how future projects will be targeting the SDGs, but it is insufficient to assess past trends. In addition, given that future reporting on the SDG field will be on a voluntary basis, the CRS may not provide a comprehensive assessment of how ODA relates to the SDGs. Several years of reporting are required in order to get an overall understanding of trends of ODA financing in support of the SDGs.

In this context, the emergence of smart computing tools can help the DAC maintain the relevance of its statistics in the SDG era, by making better use of CRS, responding

[2] Official development finance (ODF) consists of ODA and OOF.

[3] Purpose codes are seeking to "*identify the specific area of the recipient's economic or social structure the transfer is intended to foster*" (OECD, 2018[34]).

[4] See: https://one.oecd.org/document/DCD/DAC/STAT(2018)41/REV1/en/pdf

[5] The members of the DAC agreed to introduce a system of multiple purpose code at the July 2016 meeting of the Working Party on Development Finance Statistics (WP-STAT).

[6] SDG Targeting is also crucial for the new statistical measure developed at the OECD: the total official support for sustainable development (TOSSD).

faster to changes in policy environment, harnessing greater technology benefits, and alleviating the reporting burden of its members.

The paper presents a new methodology based on machine learning techniques (more specifically supervised learning techniques, see Box 1.1). It uses the CRS text descriptions as input for categorisation. With 250 000 transactions per year submitted to OECD-DCD, it would be too time consuming to attribute SDGs manually to each CRS project. The new methodology explores the potential for today's more powerful computing capacity to identify the relations between CRS text descriptions and their contribution to the SDGs. The Secretariat has not identified cases where semantic classification is used to classify CRS transactions into one of 17 predetermined SDG categories. Previous attempts to classify text to the SDGs have been conducted yet not tailored to the CRS specific vocabulary (Galsurkar et al., 2018[12])

The methodology presented can attribute none, one, or multiple SDGs to each CRS project, depending on the score of the project description allowing studying the interlinkages between SDGs. Finally, despite the fact that SDG definitions were introduced by the UN as late as 2015, the method can be easily applied to past releases of the CRS data, allowing to study both the compatibility of past CRS projects with the SDGs and global trends/shifts in SDG financing.

As the previous OECD report on sector financing explain, the development community "*do not yet have the right dashboard in place to monitor progress*" (OECD, 2018[13]). **The working paper tries to address this issue by offering a tool to monitor progress in SDG financing. Policy analysts can now estimate how much aid targets each SDG and how the adoption of the 2030 Agenda has changed the behaviour of donors.** The paper raises some policy consideration such as how to develop aid projects with SDGs in mind.

Box 1.1. What is machine learning?

Applications of machine learning programming are found in any sector from imagery detection of anomalies to disease prediction. "*Machine learning is based on algorithms that can learn from data without relying on rules-based programming*." (Pyle and San José, 2015[1]). In contrast, a machine learning programme does not have a specific instruction set and can 'tweak' its parameters to fit the new data it receives; progressively improving its performance (Rao, 2015[13]). Algorithms detect patterns and provide recommendations based on the input data and past experiences of outcomes.

Some techniques date from the 1960s but machine learning has become popular in the past decade due to:

- Massive improvement of machines' computational power
- Large volume increase of data created in many domains (images, financial data, GPS positions, …)
- Standardisation of machine learning methodology

Three main types of machine learning algorithms exist: supervised learning, when both inputs and outputs are known; unsupervised learning, when data is unlabelled and unknown patterns are looked for; reinforcement learning, when the algorithm tries to attain an objective (what is the best strategy to win a game).

Our methodology uses supervised learning methodology.

Supervised learning corresponds to algorithms using "*training data and feedback from humans to learn the relationship of given inputs to a given output*" (e.g., how the content of an email predicts whether it is spam or not) (McKinsey, 2015[14]).

Since the output of the algorithm is known (in our case the SDGs), the algorithm needs to find the steps in order to correctly link the inputs (the projects in CRS) to the output (the SDGs) (see Figure 2.1. Why use machine learning). It finds the steps using a training data set that a human has provided. The process is iterative until the algorithm finds the most optimal process to link the inputs to the outputs.

2. How to link CRS to the Sustainable Development Goals?

The following section provides a step by step guide to tailor machine learning for SDG and CRS analysis. It first gives the overview of the methodology used by the Secretariat and then explains the cleaning and translating phases that are required to obtain a clear and homogenous database. Finally, it explains which input texts are selected as teaching material and the improvement steps taken during the development of the algorithm.

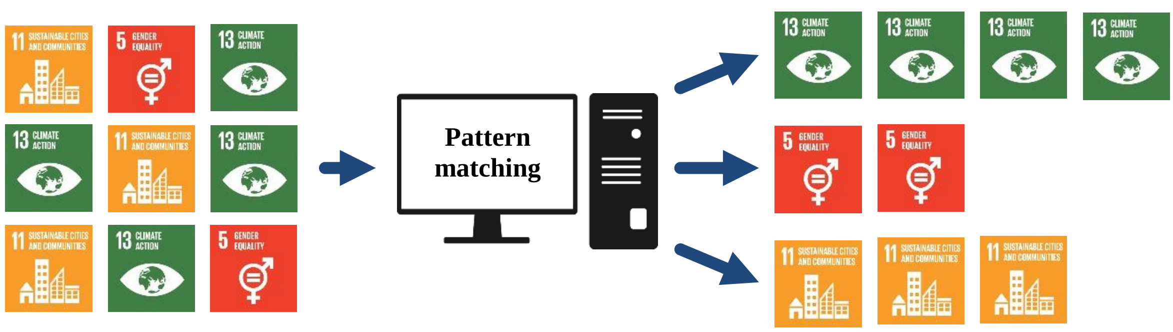
2.1. How is the machine-learning model applied to analyse the relationship between the CRS and the SDGs?

To link all the projects to SDGs, the algorithm "*reads*" the textual description of each CRS project, identifies patterns of text attributed to SDGs and links a project to zero, one or multiple SDGs (see Figure 2.1). The developed methodology uses an algorithm that draws on pre-existing text examples (e.g. SDG descriptions, reports on specific SDGs) with respective information on SDG classification attached.

The methodology uses definitions provided by the UN, classified projects and external PDF sources to learn how to distinguish between the SDGs (see 7.Annex C)**.** This information is provided as input to the algorithm that then learns how to link other text to SDGs based on observations made in the provided samples. Drawing on existing understanding of the SDGs is crucial since the CRS projects currently do not have SDG labels attributed to them.

Once this step is accomplished, it uses the built supervised model and performs out of sample predictions using the CRS project descriptions (see Figure 2.3). Due to the lack of labels, there is no direct way of evaluating the model quality. In a third step, OECD analysts validate the results comparing them to the purpose codes, examine the accuracy of the word occurrences per SDG, and conduct manual classification to evaluate the performance of our model (see 2.4 and 3.4).

Figure 2.1. Why use machine learning to unpack the SDGs?



2.2. Data preparation: cleaning, harmonising, and enhancing descriptions across countries

To ensure accurate prediction, it is important that the database is cleaned from errors, translated in English and contains as much information as possible to increase likelihood of correct predictions.

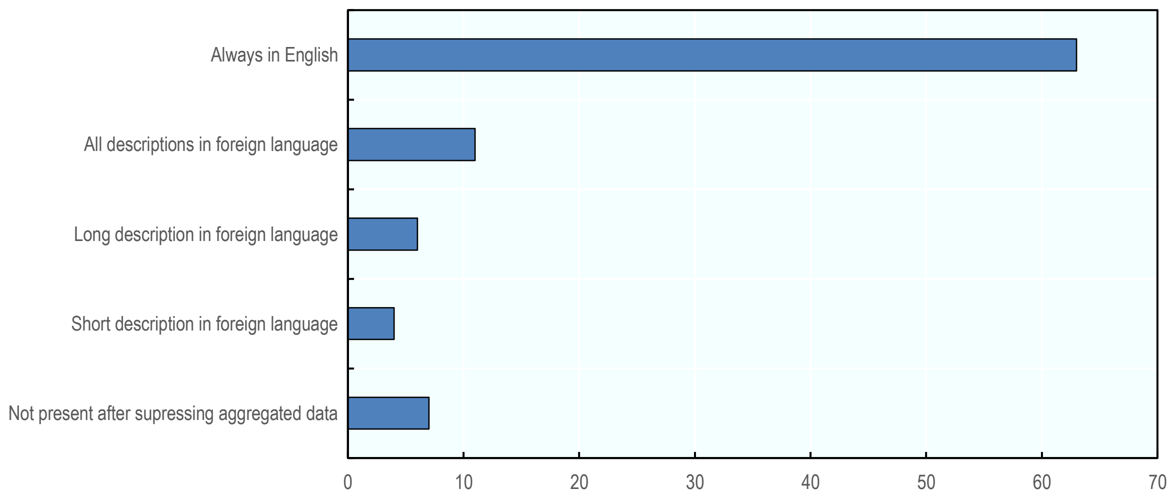
Among the many variables of CRS, the algorithm uses the textual fields as input for prediction. The database has many entries such as the donor name, the recipient, the US dollar amount committed or disbursed, a project title, and both a short and a long description for each project. Such descriptions are provided to increase the informative content of each project. Description fields were created in order to help OECD statisticians controlling for the accuracy of the projects and their eligibility to be counted as official aid. This also helps the database user understand which projects are counted as official development assistance.

Because CRS gathers financing provided by all OECD DAC providers and from other countries, multilateral organisations, and philanthropies, the descriptions are not always in English and are written with specific vocabulary, expressions, and acronyms. The algorithm first detects the acronyms and then harmonises them. For example, *HIV/AIDS, AIDS, AIDS/HIV,* is converted in *AIDS* to ensure comparability between descriptions.

The computer then automatically translates all text into English. Although the title of each project reported to the CRS must be in English, both short and long descriptions can be in the language of the reporting country. Figure 2.2 shows the number of DAC reporters that provide text in different languages. Before running the algorithm, a programme first checks whether the project descriptions are in English or not and then translates them if necessary. It replaces some abbreviations written in non-English language to their English counterparts (see Table A A.1). The projects are translated using Google Translate API (Google Inc., 2018[15]).

Figure 2.2. Harmonisation of the reporting language is key to data cleaning

Count of donors that report their long and short descriptions in English or another language



Source: Authors calculations.

To enhance the information content of the descriptions, the channel of delivery is added to the overall descriptions. The agency names that are specific to a certain SDG are also added. The channel of delivery is the first implementing partner in case of multilateral aid (OECD DAC, 2018[16]). The extending agency is the entity (central, state or local government agency or department) financing the activity from its own budget (OECD DAC, 2018[16]). Examples of channel name or agency names added are *Ministry of Health* or *Research Institution*. These names represent sources of information for the projects. It is assumed that a project financed by the 'Ministry of Health' will more likely contribute to *Good Health and Well Being* (SDG 3) whereas one from a university will contribute to knowledge creation and therefore *Quality Education* (SDG4). Additionally, names that do not directly bring informative content such as *Public Sector* are not included to avoid bias.

The algorithm then lemmatises the text, converting it to its "dictionary form". This step enhances the capacity of the machine to detect similar patterns. It reduces the dimensionality of the problem. For example, "*educated*", "*educates*", and "*educate*", convey similar meaning; converting all of them to "*educate*" enables the computer to deal with fewer words to learn from and therefore provide better predictions.

Finally, the programme then removes numbers, special characters and punctuation. It also converts some words and abbreviations specific to the database to make them more explicit and removes the stop words such as "*the*", "*he*", "*on*", which do not hold informative content to the *bag of words*[7] created.

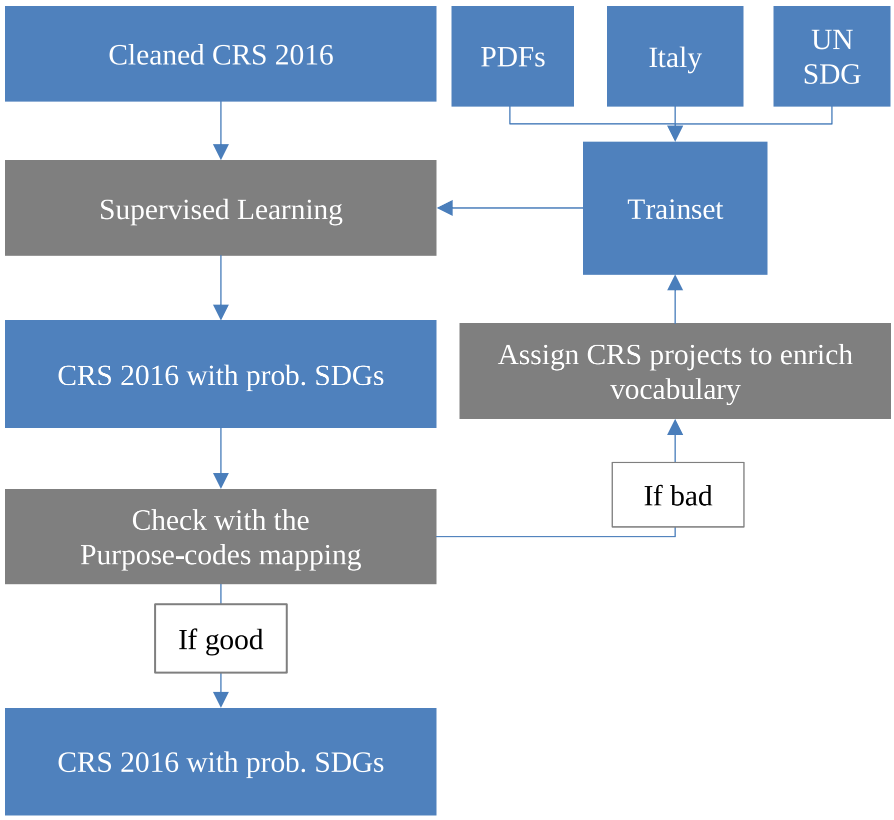
[7] This methodology uses a bag of words approach meaning that the algorithm disregard the grammatical representation of sentences or the order of words. It rather considers each sentences as a "bag of word" i.e. a specific list of words for each text. The words that do not convey any important meaning such as "the", "he", "she", are therefore disregarded to simplify the problem.

2.3. Trainset creation: preparing the teaching material for the computer

This section provides an overview of how to create a *trainset* that serves as teaching material for the computer. More concretely, a trainset contains texts where the algorithm already knows the categories it needs to link with. In this case, the trainset contains text that have already been linked to SDGs. The algorithm will train on these texts to accurately predict the SDGs and reproduce it to new text that do not have pre-defined classification.

In the current project, the algorithm learns from texts (inputs) that are assigned to a specific SDG (outputs) to predict the SDG score of each project. The pre-assigned texts constitute the trainset. It is based on these texts that the algorithms will train and determine a relationships between the words (embedded in the text) and the expected SDGs. Different supervised learning methods exist; the current methodology uses off-the-shelf supervised learning techniques namely XGBoost (Chen and Guestrin,(n.d.)[17]) and Random Forests (Breiman and Leo, 2001[18]).

Figure 2.3. Summary of the methodology



One of the difficulties is to create a training set that is able to capture the complexity of the SDGs. In fact, experts have different views on what should be included in the SDGs (OECD, 2017[20]) and SDGs are often interlinked (Le Blanc, 2015[20]). For CRS, this would require manually assigning a large number of projects to each SDG. Such classification would be very time consuming and gives rise to potential biases introduced via the annotator. "*Labelled examples are often, however, very time consuming and expensive to* *obtain, as they require the efforts of human annotators, who must often be quite skilled"* (Zhu, Ghahramani and Lafferty,(n.d.)[21]).

To address this complexity while decreasing the likelihood of human bias, the methodology relies on external sources to build the specific vocabulary of each SDGs. The training database extracts text from:

- More than 200 reports that each target a specific SDG. Such reports come from academics, public sector, NGOs, and multilateral organisations (see full list in 7.Annex C). It can be, for example, for SDG 2 the *Zero Hunger Challenge National Action Plan of Nepal*, a report of the World Food Organisation, or a report from PwC to businesses on how to engage with the goals.
- The description of the SDGs on the UN Website (Progress and Info of 2016-17), and description of the targets and indicators.
- Assigned CRS projects from Italy[8].

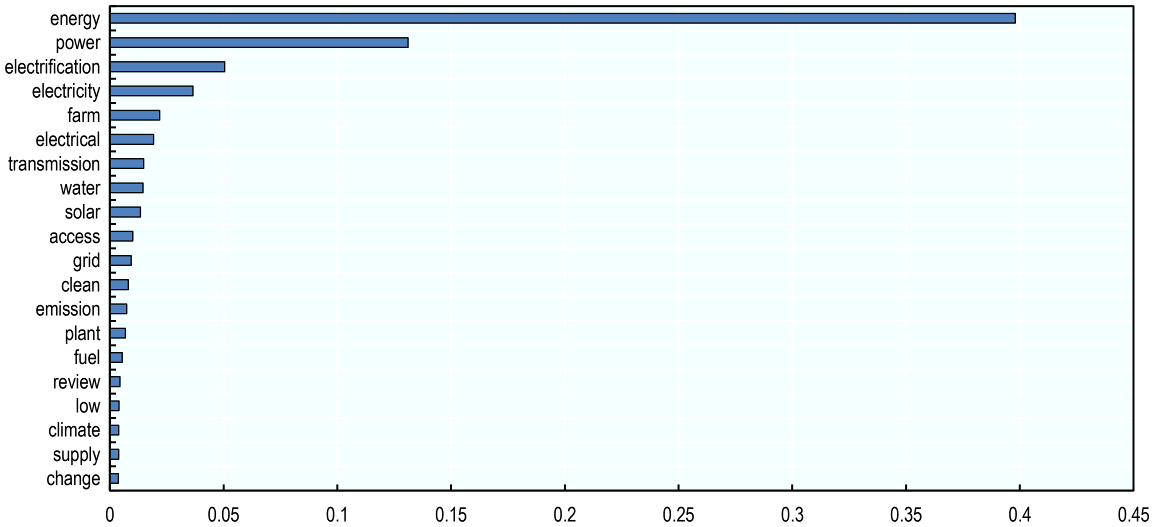
Combining the UN descriptions, the PDFs, and assigned projects from Italy gives a broader understanding of what vocabulary should be associated to each SDG, raising the level of heterogeneity of texts in the training set. In fact, each paragraph of each text represents a source of vocabulary that the machine will learn from. Such methodology also reduces likelihood of bias by avoiding that a single analyst defines itself what should be included in a SDG category rather than letting multiple experts add their understanding of the SDGs.

The algorithm bases its training on the word frequency. It derives the relative importance of words for each SDG. For example, the three most important words it has

[8] The DAC delegation of Italy and the Italian Agency for Development Cooperation (AICS) have provided the Secretariat with some CRS projects they have manually assigned to SDGs for 2017. This represents 380 projects after removing projects that have similar text descriptions.

found to attribute a project to the energy SDG are *energy*, *power* and *electrification*. Weighting of word for each SDG is in 0.

Figure 2.4. Weight of words for SDG 7 prediction



Source: Authors' calculations, preliminary results.

2.4. Trainset improvement: improving the teaching material to tailor it to CRS

If used only with external text to learn from, the algorithm will not take into account the specificities of the CRS database. **The following section explains how projects from CRS are added to increase the information content of the training set.**

The vocabulary used in reports can be different from the one found in the project descriptions of DAC members. This can decrease the likelihood of the algorithm to detect certain patterns in the CRS texts. Therefore, there is a need to validate the current training set by comparing the predictions obtained with other methodologies that link CRS projects and SDGs. The machine learning approach is therefore compared with the matrix that partially links purpose-codes and SDGs (Benn and Gaveau, 2015[11]), and with CRS Markers[9] (*Biodiversity*, *Climate Mitigation*, or *Gender*). These comparisons enables to spot three type of errors:

- **Missing words in our algorithm**
- **Error in reporting the purpose codes from reporting donors**
- **Lack of an accurate description**

[9] The Rio Conventions were established in 1992 on Climate Change, Biological Diversity and Desertification. Developed country parties committed to assist developing countries in the implementation of these Conventions. A scoring system of three values is used, in which development co-operation activities are "marked" as targeting the environment or the Rio Conventions as the "principal" objective or a "significant" objective, or as not targeting the objective. Rio markers cover Biodiversity, Climate Change Adaptation, Climate Change Mitigation, Desertification (OECD, 2018[35]).

Missing words are usually specific words that are not found in reports with overlooking view. For example, projects with specific descriptions are added to our trainset: reports might mention *disease* but not specifically *polio* or *malaria*.

Errors from reporting countries can be found; for example, a text description consisted of "*Tunisia second natural resources management project"*. It is currently linked to Road Transport (purpose code 21020) which would correspond to SDG 9: Industry, Infrastructure, and Innovation. Our algorithm will categorise it as SDG 12 Responsible Consumption and Production, as it is directly linked to management natural resources. The implication of detecting reporting errors is discussed in the section 6.2.

Finally, OECD analysts have systematically controlled the top 100 projects for each donor in terms of disbursement. This provides a benchmark for the performance of our algorithm and enables us to spot once again areas that might not have been covered by the reports used as training. This process is iterated until overall accuracy has reached its plateau (see 7.Annex B for accuracy for each DAC donor).

3. Results

The algorithm attributes 76% of the CRS database to at least one SDG. Controlling the largest 100 projects per DAC donor in terms of disbursement gives an accuracy of 87% (from 79% for Slovenia to 96% for Greece and the United States). This methodology identifies funding for all SDGs especially ones that could not be identified before such as Inequalities (SDG10), Cities (SDG11), or Peace and Justice (SDG16). The results of the project demonstrate that while machine learning delivers a high level of accuracy, further fine-tuning is needed to achieve higher accuracy and classification of SDG labels.

Despite the further work required, the project has revealed several important initial findings presented in the section below: which SDGs are most frequently associated to CRS transactions? Are there some SDG darlings? Along which SDG do recipients receive the most aid?

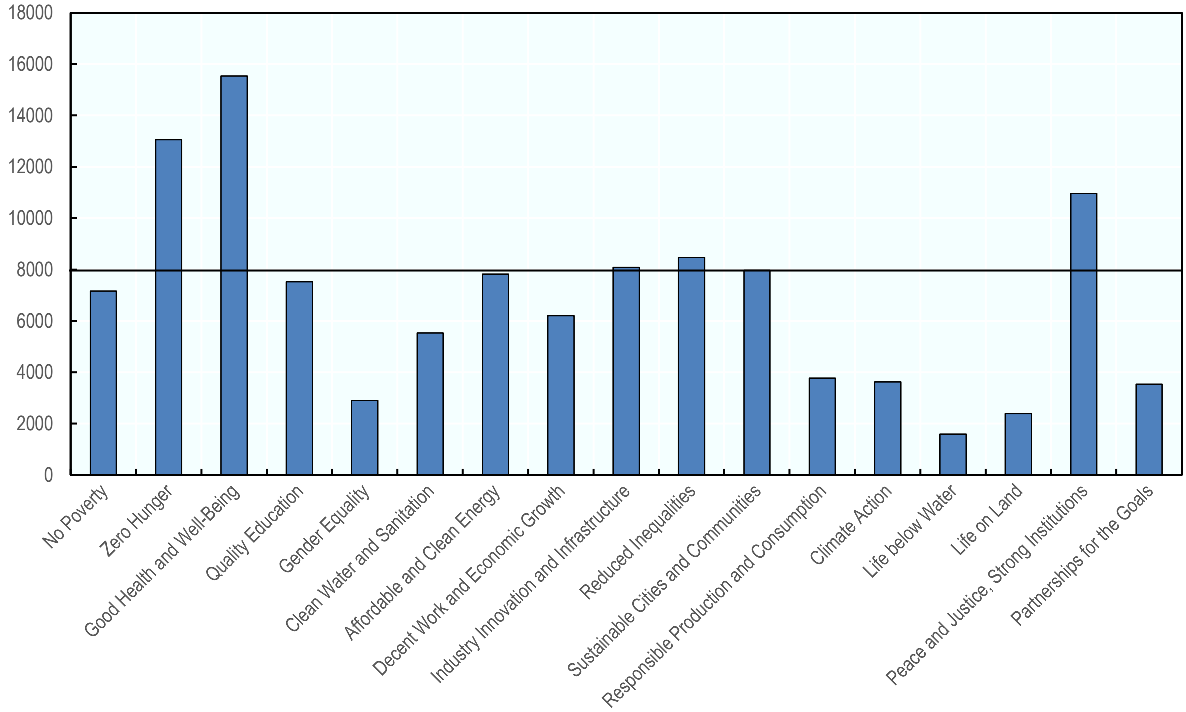
The final part of the section validates the results by comparing them with other existing methodologies.

3.1. Machine learning can reveal which SDGs are most often targeted

The results provide a first glance at how ODA and development finance more broadly target the SDGs. Overall, the most targeted SDG is Good Health and Well Being. At the DAC level (Figure 3.1), the top tier SDGs also include Good Health and Well-Being, Affordable and Clean Energy, Industry Innovation and Infrastructure, Zero Hunger, and Sustainable Cities and Communities (respectively SDG 3, 7, 9, 2, and 11). Future analysis is required to understand the implications of such targeting. Is there a risk of SDG darlings or orphans? How can this information help DAC members identify the synergies between these goals?

Figure 3.1. Aid per SDG for all DAC donors

(USD million)

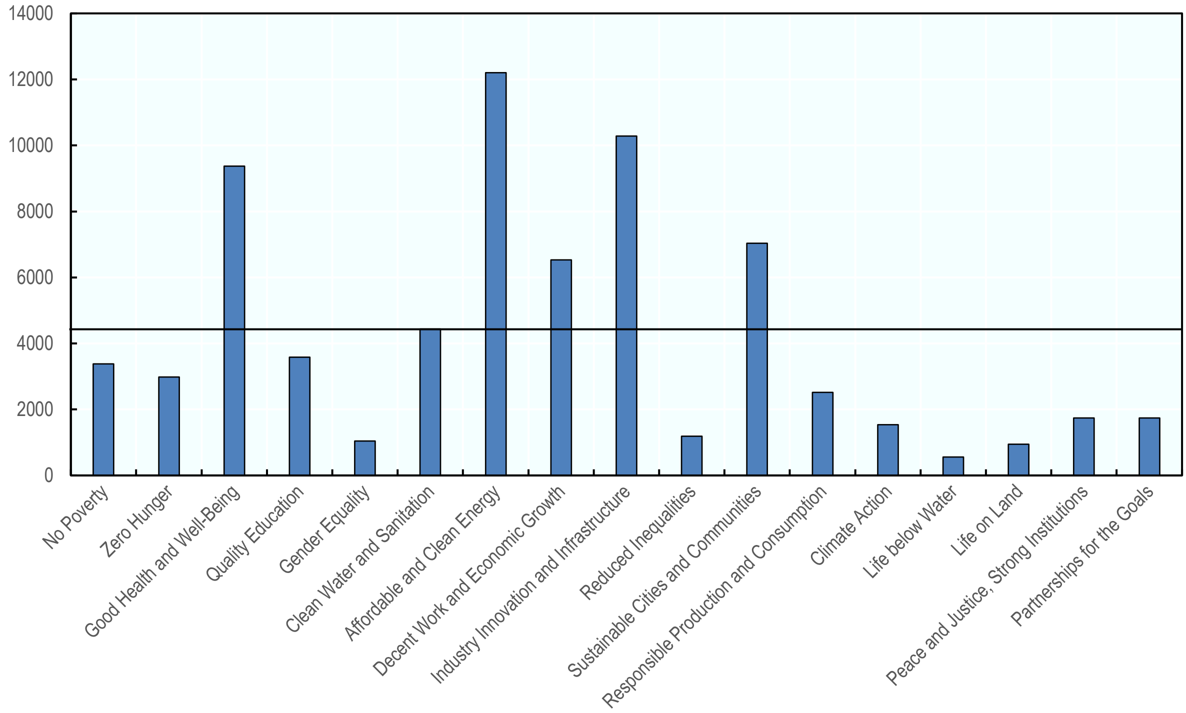


Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated.
Source: CRS 2016, Authors 'calculations. Figures are preliminary.

The project further shows the composition of actors by SDG. Outside the DAC (Figure 3.2), top tier projects are in Good Health and Well Being, Zero Hunger, Peace and Justice, Strong Institutions, Reduced Inequalities, and Industry Innovation and Infrastructure (respectively SDG 3, 2, 16, 10, and 9). This confirms the intuition that multilateral development banks are mostly financing infrastructure projects, the large involvement of the Bill and Melinda Gates Foundation in health development finance, and the importance of the World Food Programme.

Similarities in top SDG (Good Health and Well-Being) from DAC donors and non-DAC donors raises policy questions. Is health the SDG that needs the most funding? And who are the recipients? Is there a complementarity or overlap between DAC donors and non-DAC donors? Should DAC donors keep investing in SDGs where multilateral partners are the most present?

Figure 3.2. Aid per SDG for non DAC donors (USD millions)



Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated. Non DAC Donors corresponds donors that are not member of the DAC but still report their projects that are then integrated into CRS.
Source: CRS 2016, Authors 'calculations. Figures are preliminary.

3.2. Mapping the SDGs by donor

The section aims to provide an overview of the top donors in volume terms for each SDG. **While some well-known trends are confirmed** (eg Bill & Melinda Gates Foundation on health or Japan on Water, energy, etc), **other results are more surprising.** For example, Germany on Poverty reduction and Inequalities or Japan as major contributors to cities.

Table 2.1, Figure 3.3, and Figure 3.4 exemplify the specificities of donors' profiles. This confirms existing knowledge of donors such as the Bill and Melinda Gates Foundation focusing on SDG 3, Health (Figure 3.3), or Japan (Figure 3.4), focusing on Water, Energy, Infrastructure, and Cities (SDG 6, 7, 9, and 11). Table 2.1 might not reflect countries' actual effort to target a specific goal because of lack of accurate reporting or too succinct descriptions (see section 4).

Table 2.1. Top 5 Donors per SDG

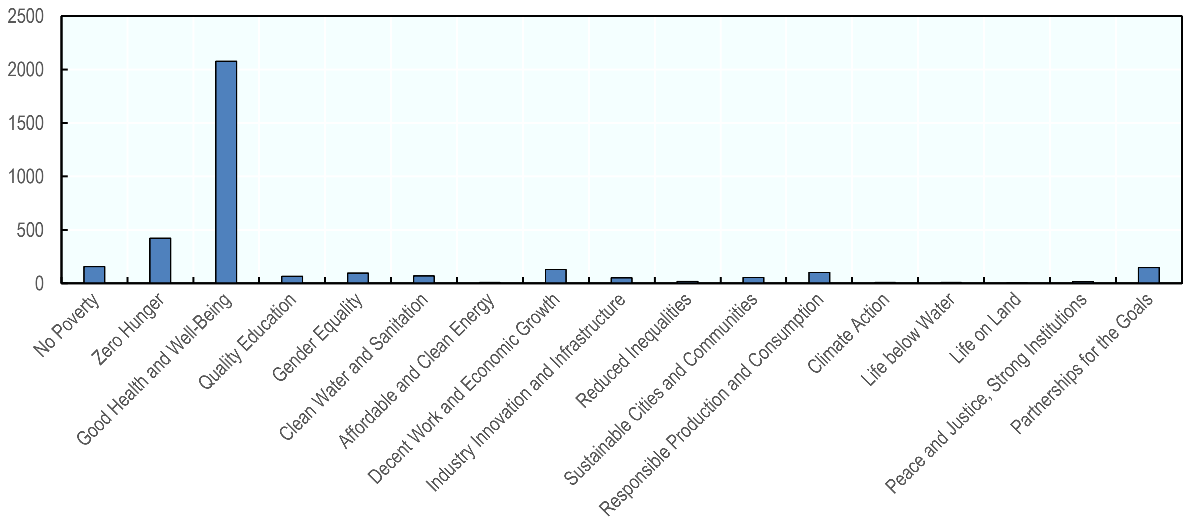
Top 5 donors per SDG in USD disbursement in CRS 2016.

SDG	Donor 1	Donor 2	Donor 3	Donor 4	Donor 5
Poverty	Germany	United Kingdom	United States	EU Institutions	Japan
Hunger	United States	EU Institutions	United Kingdom	Germany	Canada

Health	United States	United Kingdom	EU Institutions	Germany	Canada
Education	United Kingdom	Germany	United States	EU Institutions	France
Gender	United States	United Kingdom	EU Institutions	Sweden	Norway
Water	Japan	EU Institutions	Germany	France	United Kingdom
Energy	Germany	EU Institutions	Japan	France	United States
Eco. Growth	EU Institutions	Germany	United States	United Kingdom	Norway
Ind. & Inf.	EU Institutions	Japan	France	United Kingdom	Germany
Inequalities	Germany	United States	United Kingdom	EU Institutions	Japan
Cities	EU Institutions	Japan	France	Germany	United Kingdom
Cons. & Prod.	United States	Germany	Japan	EU Institutions	France
Life b.Water	EU Institutions	Japan	United States	United Kingdom	France
Life on Land	Germany	United States	Norway	EU Institutions	Japan
Peace & Just.	United States	EU Institutions	United Kingdom	Germany	Sweden
Partner.	United States	Germany	Netherlands	EU Institutions	Norway

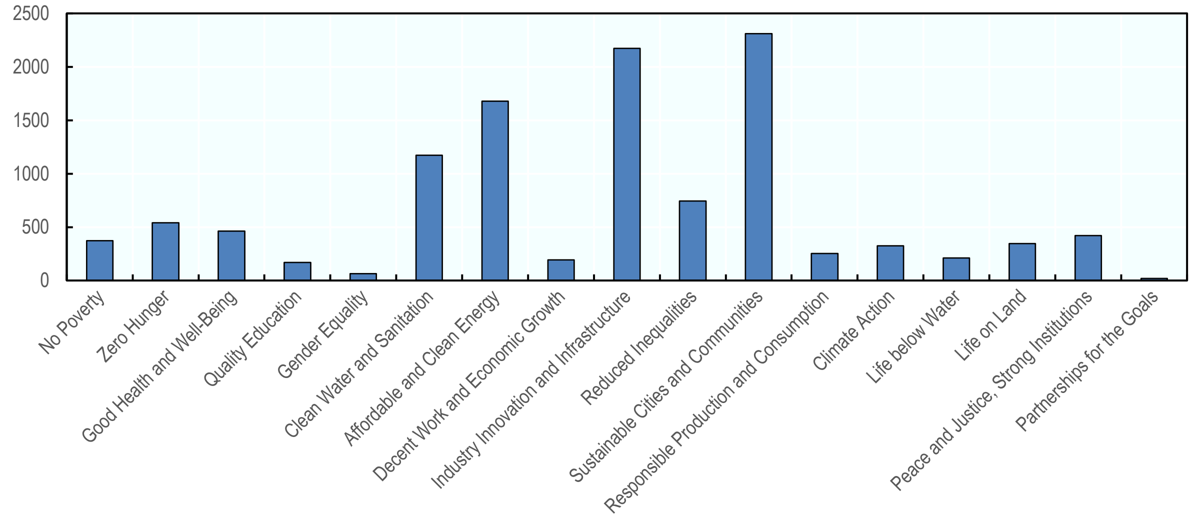
Source: Authors calculations. Figures are preliminary.

Figure 3.3. Aid per SDG for the Bill and Melinda Gates Foundation (USD millions)



Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated.
Source: CRS 2016, Authors 'calculations. Figures are preliminary.

Figure 3.4. Aid per SDG for Japan (USD millions)



Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated.
Source: CRS 2016, Authors 'calculations. Figures are preliminary.

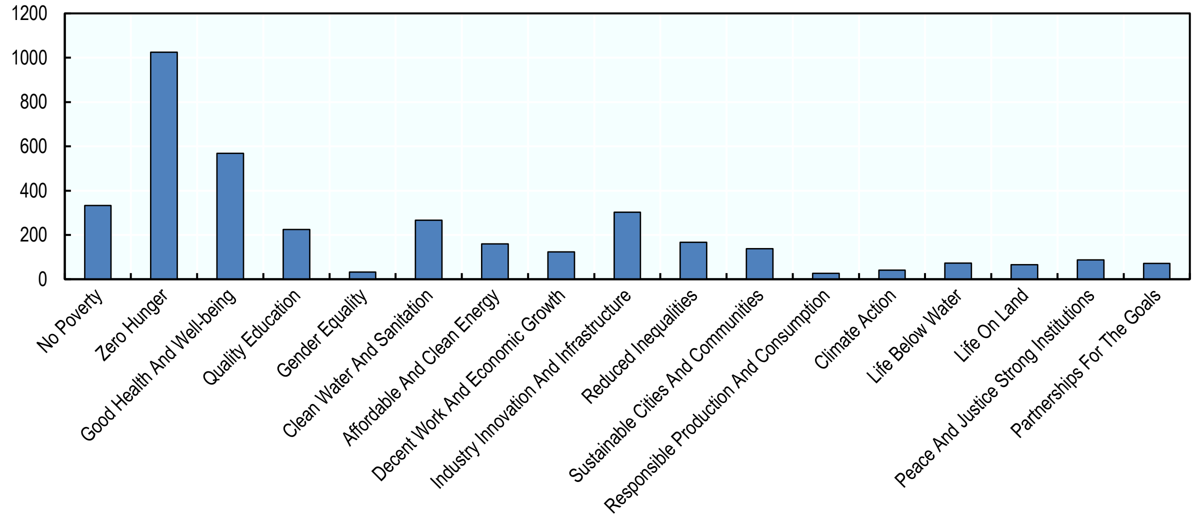
3.3. Mapping the SDGs by recipient

Because of the CRS structure, similar analysis is possible at the recipient level. The SDG mapping allows to see the different aid profiles that recipient receives. For example, Figure 3.5 shows the SDG sectoral allocation of aid flows in Ethiopia. Zero Hunger and Health (SDG 2 & 3) are the most targeted SDG which is in line with the LDC status of the country where basic social needs still need to be addressed[10].

Looking at the recipient level enables differentiating aid by country categories: income (LIC, LMIC, UMIC) or by countries most in need (SIDS, LDCs, LLDCs). Further work needs to be conducted to link the financial inputs listed by SDG to the SDG indicators developed by the UN. The policy implications of the sectoral allocation of aid at the recipient level is detailed in section 4.1.

[10] The Human Assets Index (HAI) of Ethiopia is 45.3, below LDC average of 53.1 and graduation threshold of 66. The HAI is a composite indicator including, health, education, and nutrition indicators and reflecting a country's development in terms of human capital.

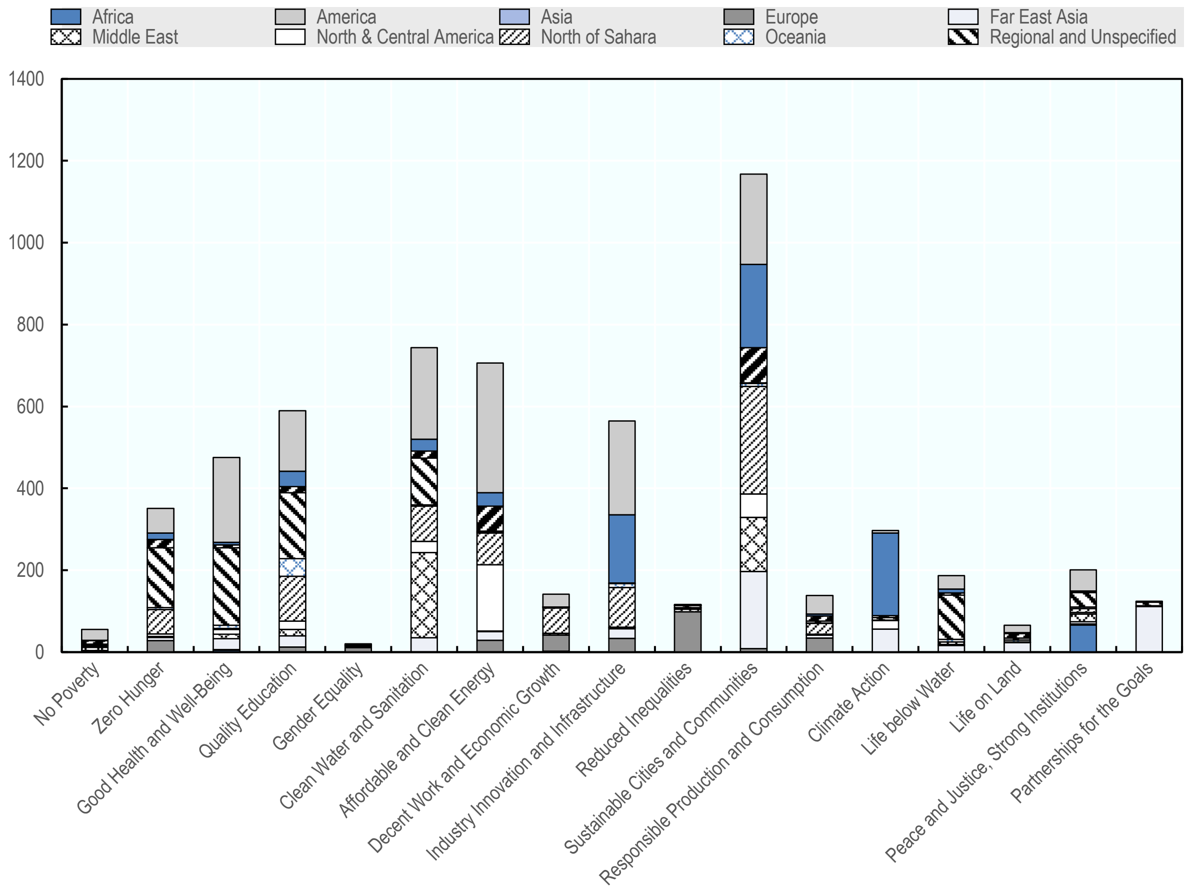
Figure 3.5. Aid per SDG for Ethiopia (USD millions)



Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated.
Source: CRS 2016, Authors 'calculations. Figures are preliminary.

Finally, it is possible to link donors, recipients and SDGs to understand how donors tailor their aid by regions or countries. For example, France's contribution varies largely across regions. Figure 3.6 shows that France's support to infrastructure and cities development largely takes place in Africa (SDG 9 and SDG 11) whereas the alleviation of the refugee burden mainly takes place in Europe, more specifically in Turkey (SDG 10).

Figure 3.6. Aid per SDG for France (USD millions)



Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated.
Source: CRS 2016, Authors 'calculations. Figures are preliminary.

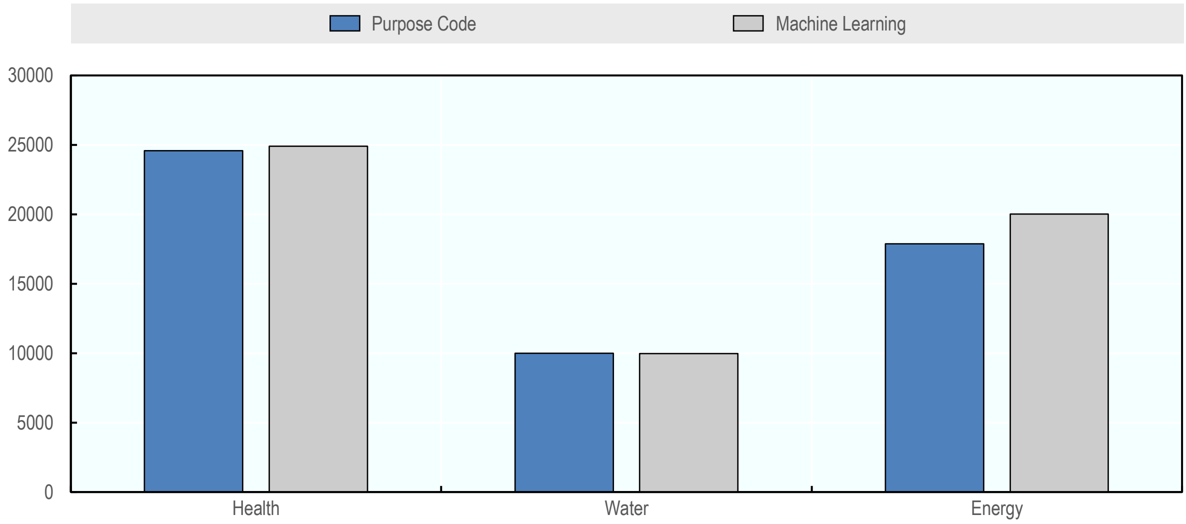
The projects that are not classified or that fall into unspecified regions are usually general budget support to a specific fund or to an NGO. In these cases, there is not an explicit mention of what the projects are trying to achieve. For example, a "*general contribution to the United Nations International Children's Emergency Fund*" cannot be attributed. Children are explicitly mentioned in many of the SDGs (SDG 1, 2, 3, 4, 5, 8, 11, and 16), but such description does not allow the algorithm which specific target is addressed. Therefore, projects non-attributed are not necessarily not contributing to any of the SDGs.

3.4. Comparing with purpose code classification can help to validate results

The machine learning methodology is in line with other methods to assess aid flows along the SDGs. It confirms the pertinence of machine learning in correctly linking projects to the SDGs. As seen in Figure 3.7, the results are aggregated and compared with purpose codes that can be clearly attributed to certain SDGs. Figure 3.7 shows the comparison with the two methodologies resulting in small differences (from 0.22% for the water sector to 10.8% for the energy one).

Figure 3.7. Disbursement allocation with purpose code and algorithm for specific SDGs

Disbursement (USD millions)



Source: Authors' calculations.

4. Future policy research – assessing aid in the SDG era

The methodology opens large array of future research. Some SDGs such as Health (SDG 3), Education (SDG 4), or Energy (SDG 7) are easy to capture with the purpose codes. Others such as Decent Work and Economic Growth (SDG 8), Sustainable Cities (SDG 11), Life below Water (SDG 14) or Justice (SDG 16) provided more difficulty to identify projects related to them. Providing descriptive figures on these goals and explaining which kind of projects they contain will be of interest.

The methodology also has to be applied to past years to get an understanding of past trends in development finance and track policy shifts of donor countries. Such work can be interesting to understand whether the adoption of SDGs and the call to make ODA compatible with the AAAA has influenced donors' behaviour, avoiding "SDG washing"

The following section details potential areas for policy analysis beyond descriptive statistics. It puts the ODA figures in a broader development context by linking aid flows to development gaps. It then looks at the different approaches donors are taking to tackle the SDGs by looking at the different connections between SDGs among donors.

Finally, this section shows that the machine learning allows expansion beyond the CRS database. It is possible to look at textual information from other sources: Chinese foreign aid flows or private sector reports.

4.1. Identifying the SDG gaps and opportunities for aid flows

The current methodology allows linking financial aid inputs with development outcomes in the SDG era. SDGs have a set of 232 indicators[11] that offer tangible measurement to monitor progress and ensure accountability. Beyond this set of indicators and official statistics aggregated by the United Nations, other organisations have worked in designing tools to track progress along the SDGs: the World Bank SDG Atlas or the SDSN SDG Index.

Linking with outcome data allows identifying financing opportunities and potential inefficiencies in donors' aid strategies. A recipient with high level of aid in education (SDG 4) despite high scores in the SDG 4 indicators raises questions of whether aid flows need to be diverted to a country in greater education need. Similarly, a recipient with persistently high level of aid in a certain SDG despite low progress along the SDG indicators could indicate low aid efficiency questioning donor's strategy.

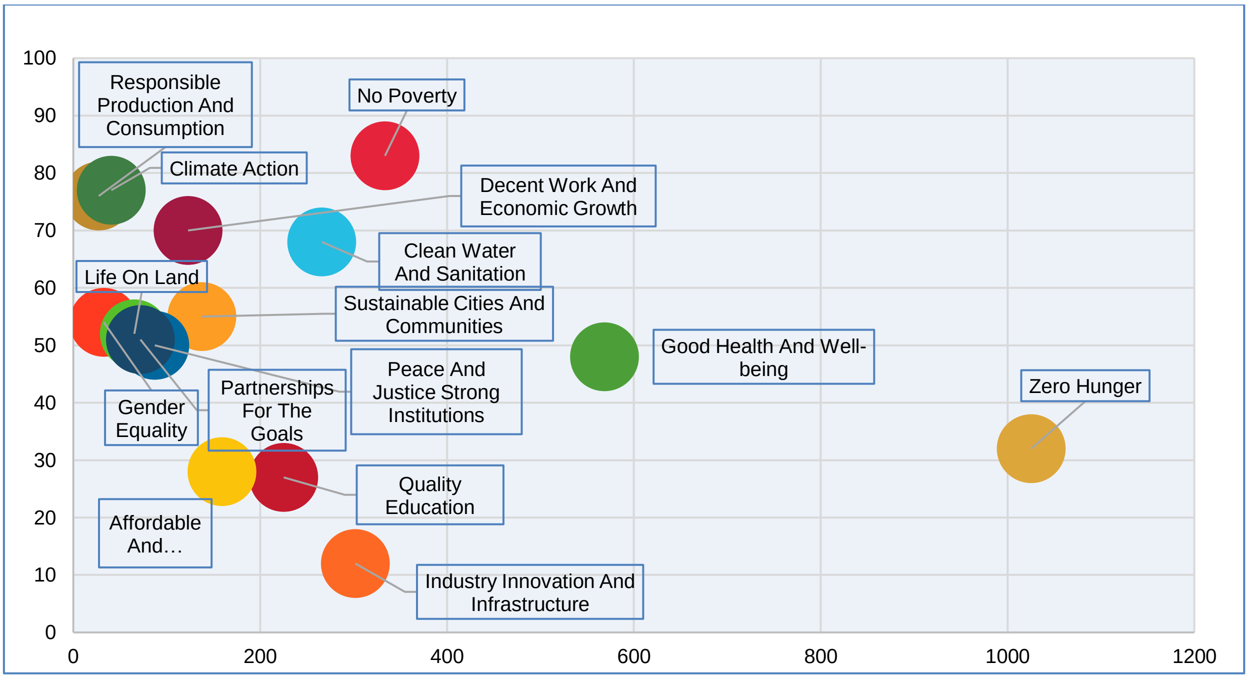
For example, SDSN identifies seven SDGs for which Ethiopia faces major challenges (see Figure 4.1). If donors target SDG 2, 3, and 9 (Hunger, Health, and Industry Innovation and Infrastructure), SDG 7 and 16 (Energy, Peace and Justice) receives low funding compare to the assessed needs (see Figure 3.5). Discrepancies between donors funding and the needs raise crucial policy questions:

- **Are donors' aligning priorities with recipients' biggest needs?**
- **Does domestic expenditure address the SDGs that are not targeted by donors?**

[11] Indicators are metrics linked to SDG indicators in order to measure progress in reaching the goal.

- **Will focusing on specific SDGs will allow for synergies, improvement of other SDGs? Which SDGs can be enablers?**
- **When a country still relies on aid despite improved SDG indicators, does this imply too high reliance on aid? Should domestic expenditure take over?**

Figure 4.1. Ethiopia SDSN Score and Aid Flows per SDG



Note: SDG Index scores indicates how likely a country is to attain the SDG by 2030. The lower the score, the less likely it is to reach the goal.
Source: SDSN, 2018 SDG Index and Dashboard Report, Country Profiles, Ethiopia

Policy analysts should remain cautious when analysing the gaps between funding and indicators. SDGs should be consider as interlinked and diminution of aid for a certain goal might diminish chances to attain other goals. For example, poverty reduction might pass by better infrastructure and higher level of education (see (World Bank, 2017[22])).

4.2. Exploring SDG interlinkages and policy approaches to SDG tackling

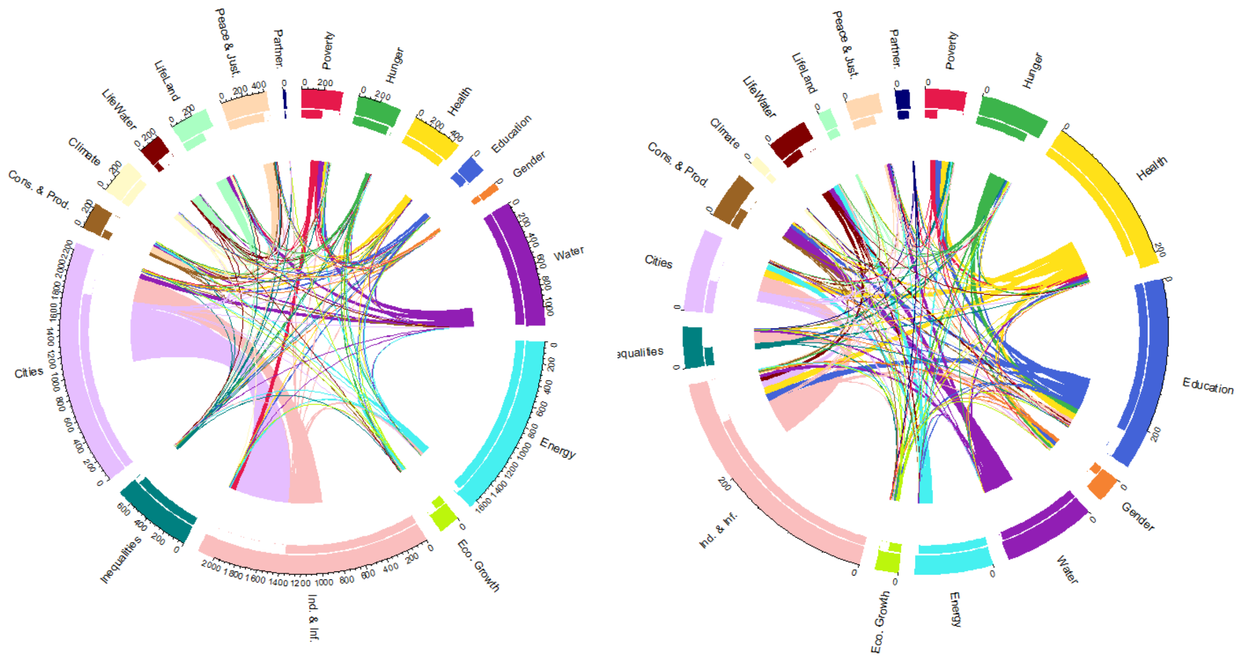
Because the algorithm can attribute multiple SDGs, it draws linkages between SDGs for projects that are targeting more than one goal. A donor can focus on gender equality through better education of young population when the other donor will ensure that women are better protected by law enforcement authorities. In the first case, Gender Equalities (SDG 5) will be linked with Education (SDG 4) whereas the second case will be linked with Peace Justice and Strong institutions (SDG 16).

Drawing these linkages identifies different *SDG nexuses* and projects with potential synergies. How many projects target reproductive health? What type of Decent Work and Economic Growth projects explicitly aim at reducing poverty? Which infrastructure is built in cities? Are some SDGs more standalone than others? This approach gives a quantitative understanding of the integrated SDG system and allows identifying projects that lie at the intersection of SDGs.

Interlinkages allow greater policy comparison between donors by further detailing their areas of specialisation. Figure 4.2 compares the SDG interlinkages of Japanese and Korean aid. While both countries have large industry and infrastructure financing (pink area), Japan infrastructure financing mostly target cities (e.g. financing construction of metro in large emerging Asian metropolitan areas). Is it because it is an explicit goal of Japanese policy? Are countries using different financing mechanisms or instruments?

Figure 4.2. Aid by SDG, Japan and Korea,

USD millions



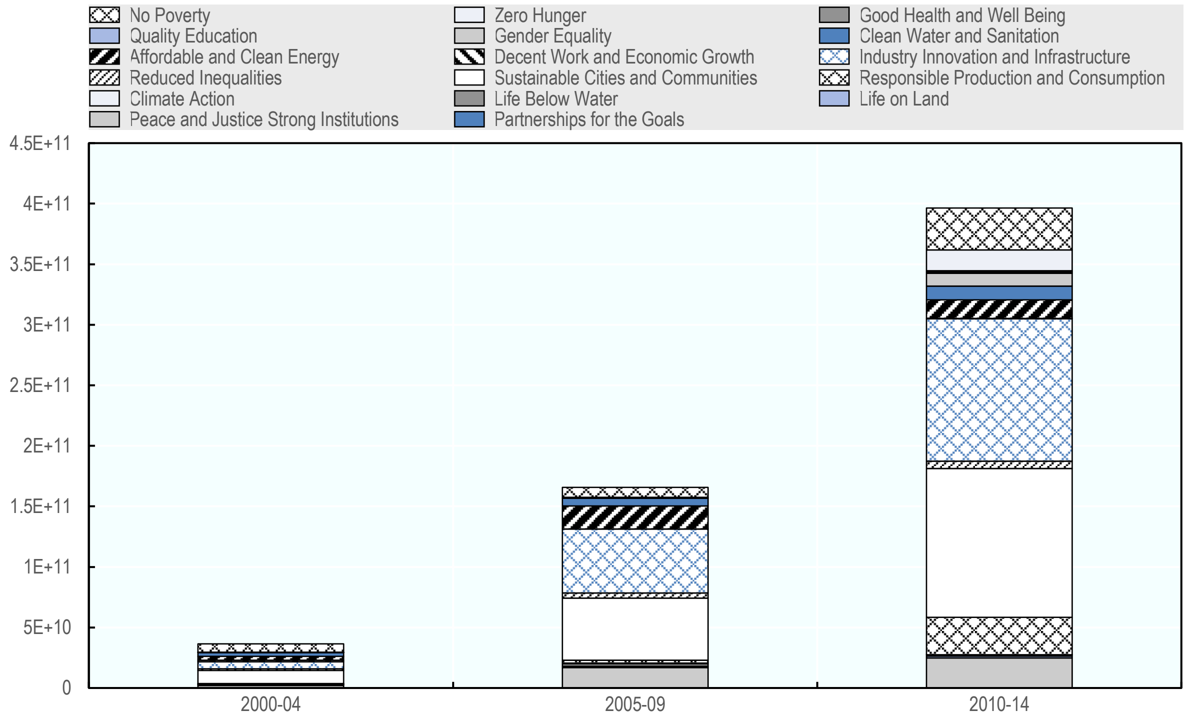
Note: Inner circle represents the primary SDG i.e. the SDG with the highest score for a given project. The width of sections corresponds to the amount in million USD. The line from one SDG to the other represents the projects that have two SDGs. The second circle represent the USD amount of primary and secondary SDGs. For example, a project that has primary SDG in Cities and secondary in infrastructure will contribute to the primary and secondary circle of Cities, draw a line from Cities to Industry, and contribute to the secondary circle of Industry.
Source: Authors' calculation. Preliminary results.

4.3. Expanding the scope: the new donors

Although China does not report publicly its aid contributions, research institutes have compiled records of aid and non-concessional official financing and provide estimates of China's development footprint (AidData, 2017[23]) (CARI, 2018[24]). The AidData database comprise more than 4'000 records representing $350 billion of Chinese investments. Each record has a text description attached providing information on the project financed. This allows the algorithm to retroactively classify them along the different SDG categories (see Figure 4.3).

Figure 4.3. Looking at Chinese Aid with SDG lenses

Chinese official financing, concessional and non-concessional, USD 2014 deflated.



Note: The sum of each SDG does not correspond to total aid flows since some projects can be double counted as they might be allocated to more than one SDG and some other projects are not allocated.
Source: AidData, Authors' calculation.

Political will and greater involvement of non-DAC providers such as China will be critical to yielding greater effectiveness of aid[12]. This stylised exercise shows the potential of the tool in assessing non DAC flows. AidData will publish the latest aid flows from China in the second semester of 2019. Assessing Chinese aid in the SDG era will allow finding potential areas for collaboration and synergies with DAC donors.

Similar analysis needs to be conducted for all multilateral development banks and for other non-DAC providers. For example, the International Development Finance Club (IDFC), a club of multiple national development banks can play a significant role in pushing its member to align their strategies to the SDG agenda (Morris, 2018[25]). If the projects financed directly by IDFC organisations might not directly be compared with aid transactions, similar attempts could be done to assess the potential of its members. Expanding to non-DAC reporters allows a more holistic approach to map aid flows drawing lessons and enabling to understand respective roles for the different actors.

[12] See for example: Transition Finance Challenges for Commodity-Based Least Developed Countries, the Example of Zambia; OECD Development Co-Operation Working Paper 49.

4.4. Expanding the scope: involving the private sector to shift the trillions

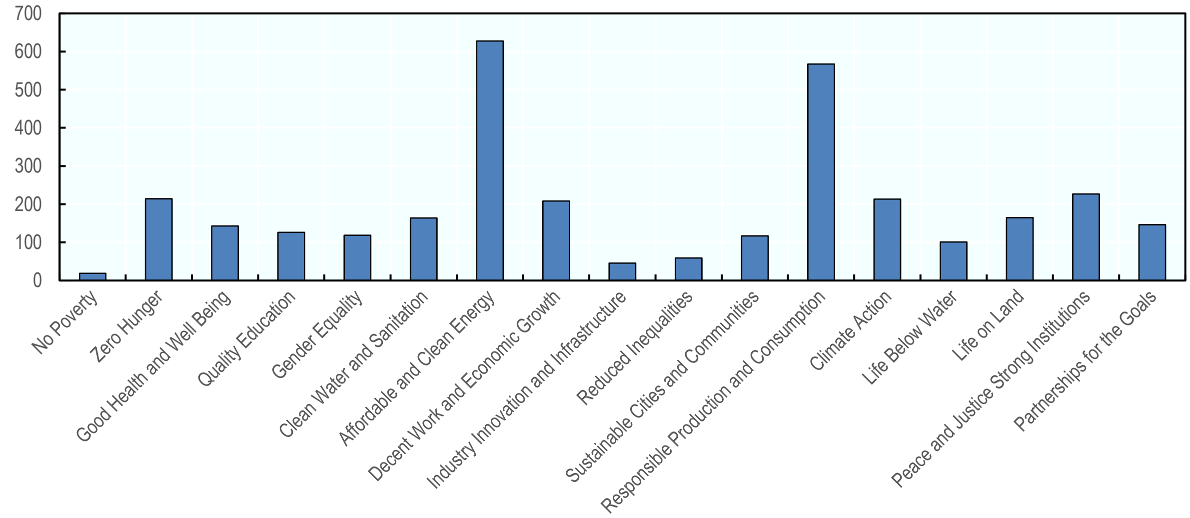
The Corporate Social Responsibility (CSR) reports from large companies offer vast sources of textual information that can be analysed using the SDG lenses. In the case of sustainability reports, the algorithm identifies sentences which vocabulary pertain to a certain SDG. By counting the frequency of SDGs appearing in the text, policy analysts can estimate the primaries area of focus of the companies. Unlike CRS, sustainability reports do not offer direct quantification to SDG financing, rather indicating trends in SDG targeting.

The OECD has run a pilot on Corporate Social Responsibility reports submitted by the Global Reporting Initiative (GRI). The algorithm identified the most mentioned SDG for 100 companies, representative of the Fortune 500 (see Figure 4.4). The two most mentioned SDGs are Affordable and Clean Energy (SDG 7) and Responsible Production and Consumption (SDG 12); two goals where private sector has a large influence due to its carbon footprint and manufacturing capacities. Decent Work and Economic Growth does not appear to be mentioned on a frequent basis raising questions on private sector role in economic development. This work follows the work analysing reporting patterns by MNEs (Winkler, 2017[26]).

Analysing private sector reporting permits to identify potential synergies between public institutions and the private sector. For example, Responsible Production and Consumption (SDG 12) is largely mentioned in the sustainability reports yet few aid flows target this SDG. On the other hand, both aid flows and private companies are largely mentioning or financing energy.

- **Is it because of the climate imperative? Does the private sector have a competitive advantage in the production goal (SDG 12)?**
- **Are firms only concerned of carbon neutrality and better production?**
- **Why do companies report so few on Decent Work and Economic Growth? Why is Industry Innovation and Infrastructure so low? Is it because CSR reporting do not focus on the core business strategies of firms?**
- **How can public institutions incentivise the private sector to address other goals? Should it be sector specific?**
- **At the algorithm level, is the algorithm missing the main messages? Is companies' vocabulary too specific?**

Figure 4.4. SDG in the CSR reports of selected Fortune 500 companies



Note: Count of text mentioning a particular SDG
Source: GRI database, authors' calculations, preliminary results

Further analysis is necessary to map SDG reporting by firms' sector. Preliminary results indicate that large discrepancies exist between sectors in SDG reporting: Healthcare companies primarily focuses on Health (SDG 3) whereas financial companies are mainly mentioning Peace Justice and Strong Institutions (SDG 16) and Responsible Production and Consumption (SDG 12).

Finally, limitations of the algorithm for the sustainability reports are similar to the ones regarding the CRS. The following section provides an overview of the limitations of the methodology.

5. Limitations of the algorithm

The method is powerful and has a high accuracy; nonetheless, it has its limits. As seen in section 3, it links between 67 to 96% of the database depending on the respective DAC donors. Moreover, the complexity of the SDGs, the specificity of the database, and the algorithm used create some limitations.

5.1. At the SDG level: the complexity of the goals presents important limitations

One of the main difficulties in linking projects to SDGs is due to the complexity of the Goals and the fact that they combine means and ends. The creation of solar panels and of road activity will contribute to economic growth and therefore to poverty reduction. Taking into account a crucial question is how far it needed to take into account the spill over effects of the aid projects.

The interlinkage of SDGs is also a pressing issue. As Le Blanc demonstrates, SDGs are unequally connected between each other at the target level (Le Blanc, 2015[20]). Moreover, some targets that would be perceived as belonging to a certain SDG can be found in other ones. As an example: "*Create sound policy frameworks […], based on pro-poor and gender-sensitive development strategies, to support […] poverty eradication actions*" is included in SDG1, Ending Poverty, although it has a gender specific content (SDG5) (United Nations, 2016[27]).

Some targets are also similar from one SDG to the other: "*By 2030, devise and implement policies to promote sustainable tourism that creates jobs and promotes local culture and products*" (Target 8.9) and "*Develop and implement tools to monitor sustainable development impacts for sustainable tourism which creates jobs, promotes local culture and products*" (Target 12.b) (United Nations, 2016[27]). This creates confusion for the analyst on where to manually allocate some projects in the training set. It also creates confusion at the machine level since the vocabulary the machine is learning from is based on the UN SDG description. The vocabulary related to tourism will therefore be found in 8 and 12 at the same time.

Box 5.1. Why is Poverty not the most targeted SDG?

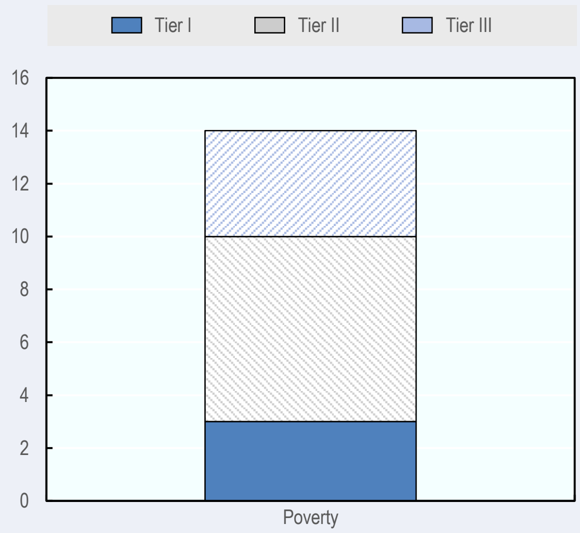
SDGs consist of a mix of means and ends

In an aid context, one could think that No Poverty (SDG 1) would be the most targeted SDG. As mentioned in (OECD, 2017[20]), poverty reduction is cross cutting by nature. The goal is also unclear and hard to track. Only 20% of the indicators are conceptually clear with an internationally established methodology, see Figure 5.1. Poverty Indicators Classification (UNSTATS, 2018[28]). This exemplifies the difficulty to reach a consensual definition of what is needed to be included in Poverty Reduction or in SDGs in general.

If most of the aid projects strive to reduce poverty, it is rather a result of action taken along the other SDGs. For example, an education program will reduce poverty through job creation of teacher and by offering better economic opportunities to students. The result will however be diffuse in time.

Therefore, projects counted in SDG1 are the ones targeting explicitly the poorest populations or providing basic services. Projects related to economic growth in general or businesses are attributed to SDG 8.

Figure 5.1. Poverty Indicators Classification



Note: Tier 1: Indicator is conceptually clear, has an internationally established methodology and standards are available, and data are regularly produced by countries for at least 50 per cent of countries and of the population in every region where the indicator is relevant. Tier 2: Indicator is conceptually clear, has an internationally established methodology and standards are available, but data are not regularly produced by countries. Tier 3: No internationally established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested. (UNSTATS, 2018[28])
Source: IAEG-SDGs, Tier Classification for Global SDG Indicators.

5.2. At the CRS level: project descriptions are heterogeneous

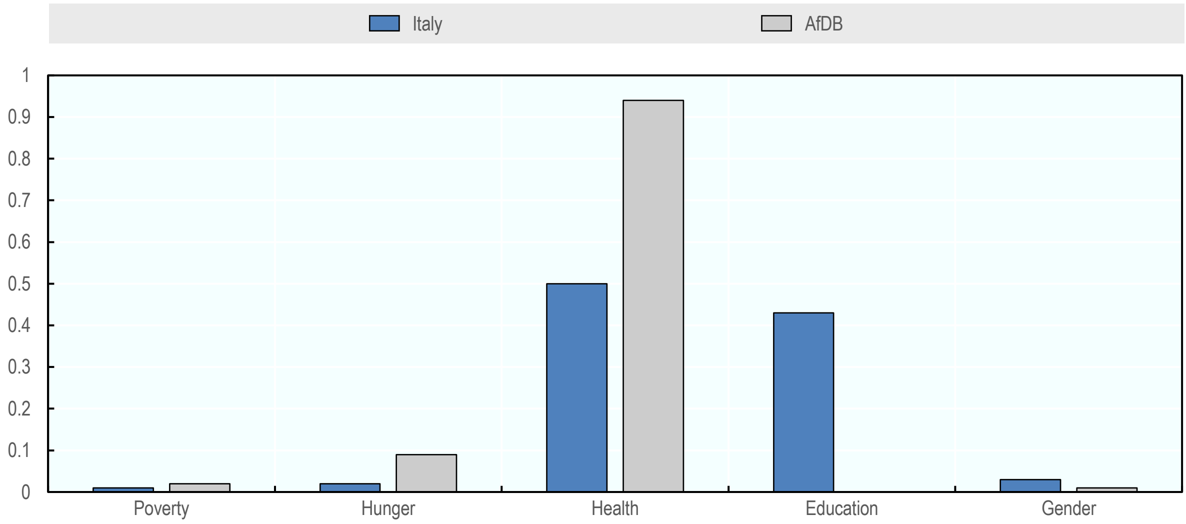
In the CRS database, descriptions from donors are heterogeneous, long enough descriptions will be more likely to have more than one SDG attributed. For example, the average description of Canada is 1145 characters while it is 217 for Czech Republic. Some descriptions also include elements that are not directly linked to the specificity of the project such as detail of the political situation of the country. Since the algorithm functions on the relative importance of words, a project with a long description but only one sentence mentioning explicitly the purpose has a lower chance of being attributed to the corresponding SDG. Unrelated text will also interfere by creating noise and making it harder for the algorithm to detect the correct goal.

On the other hand, very short descriptions do not provide the full understanding of aid projects. This refrains the algorithm from finding multiple SDG although they might exist. For example, the first of the following descriptions provides a good description of the goal of the project whereas the second one does not provide sufficient details:

- Project 1: "*Education and health policies for the [...] youth in Peru enhancing education and health rights for local children, through the implementation of an* *education plan dealing with tutorial, creative, recreational and health-protection activities.*" Attributed to: SDG 3 (Health); SDG 4 (Education).

- Project 2: "*Government support development of the health system [...] sector: social. [...] sub-sector: health*" Attributed to: SDG 3 (Health).

Figure 5.2. Scores per SDG of two descriptions with different informational content



Source: CRS 2016, Authors 'calculations.

Another limitation is that the CRS database is limited to input rather than impact-driven data. The CRS database is a database that lists the donor's contribution for official development assistance. If verification is made by the OECD to make sure each project complies with the official assistance definition, the OECD does not assess the efficiency or the effectiveness of such projects. Therefore, the methodology does not prejudge from how well the money is spent.

A higher score in attribution for a project compared to another does not indicate that the first project contributes more to this SDG. It indicates that the description was more in line with the definition the algorithm uses to assess each project along each SDG. As an example, projects helping fisheries are allocated to Life below the Ocean. If the donor does not track whether the fisheries it assists are fishing in a sustainable way, the ODA counted as contribution to SDG 14 (Life below Water) might actually have a harmful contribution to this particular SDG.

Given the importance of the quality of project descriptions, the section 6.1 provides some recommendation to DAC donors on how to best report the description to ensure better machine learning results.

5.3. At the machine level: the inherent complexity of human language

Projects with very specific/rare words cannot be detected. The machine needs to identify words into a certain amount of projects to attribute words to a certain SDG. A description that will detail a very specific disease (e.g. *Herpesviral encephalitis*) without mentioning *health* or *doctors* might not be linked to *Health*. Words need to be present at least in 30 projects. This limitation is the main reason why projects that are incorrectly non-allocated. Further work is required to create SDG taxonomies that will help addressing this issue.

The machine is not able to recognise negation. It bases its scores on the weight of each word present in the description (see 2.3). A project described as a project "*not in the water sector*" is still attributed to SDG 6 due to the presence of the word *water*. However, in practice this limitation rarely appears. Among the 3 000 projects controlled, only two contain negative descriptions.

Finally, the specificity of context in which the vocabulary is employed has an influence in the allocation of SDG. For example, most of the projects that mention *shelter* belong to the refugee corpus and are attributed to SDG 10 (construction of shelters for refugees). There are nonetheless projects where *shelter* refers to a building built as a consequence of the Chernobyl accident. The projects that are in minority for a specific confusing word face the risk to be misallocated.

Despite the limitations mention above, preliminary results prove to be robust and reliable as shown in Figure 3.7. **Access to more data such as reports from programme managers will improve the accuracy of the tool.**

6. Implications for the DAC

The methodology developed at the OECD has multiple implications for the DAC donors. It helps better reporting from the country by controlling potential errors. It provides a comprehensive overview of donor's financing along the SDGs. The following section provides recommendation to DAC members on how to pursue better reporting in a machine-learning era. It also explains how the stream of work relates with other DAC initiatives: TOSSD and the SDG purpose field.

6.1. Improved CRS Reporting 2.0

By assigning scores to each project, the algorithm enables cross-comparison with the purpose codes. This enables countries' statisticians to track reporting errors from grant managers on the field by checking the discrepancies between a given purpose code and low scores attached to it. Similarly, it helps OECD statisticians to focus on projects that might have potential issues and be sure that the ones with the highest scores are correctly attributed.

Such approach will only be efficient for purpose codes that are linked with SDGs but will not work for the whole database (see DCD/DAC/STAT(2015)9).

In general, the Secretariat recommends limiting the length of the descriptions. Long descriptions should be around 4-6 sentences. Half of the text should focus on an objective description of the project itself:

- Example 1: "*The FarSolar project is building 30MW of solar panels in the Hafar region. It will consist of three distinct plants that will be connected to local minigrids.* SDG 7, Energy

The second half of the description should focus on the expected outcomes, the intentions:

- Example 1.a: "*This project aims at alleviating travelling distances for woman to charge their phones.*" SDG 5, Gender
- Example 1.b: "*This project aims at reinforcing the local capacities of small textile producers by powering their machines. It will increase economic development in the region.*" SDG 8, Economic Growth; SDG 9, Industry, Innovation, and Infrastructure.

By doing so, the donor can report on SDGs that are more input-oriented such as infrastructure and provide detail on what outcomes are targeted.

The above-mentioned recommendations could be translated in reporting guidelines and shared with members. Technical assistance and deployment of the tool in the relevant organisations could be provided.

6.2. TOSSD and SDG focus field – using machine learning as a controlling tool

This methodology is a complement to the work on TOSSD and the SDG focus field in CRS. In fact, by analysing the text description, the algorithm can only assess what has been provided in the system. Therefore, it cannot replace the knowledge of the managers that are running the aid programmes. It is a useful tool to control the reporting of the grant managers and other reporters. By having managers and statisticians reporting along the SDGs, the CRS database will be enriched by direct intelligence coming from the field.

Since reporting on SDG only started, the methodology is useful to create a first assessment of where countries stand. This might create conflicting results with countries engagement and this is why the OECD-DAC recommends donors to report along the SDGs. Moreover, further reporting on SDG by DAC donors will provide material where the computer can train and improve its accuracy in attributing the SDGs.

The Secretariat proposes that the methodology be used as a controlling tool for statisticians within the OECD and DAC members' organisations.

7. Conclusion

Finally, future research should focus on linking output of the SDGs with the financial inputs in order to find the gaps and necessary areas for improvement. By linking the SDG sectoral allocation of ODA to the SDG indicators (Tier 1 and tier 2), policy analysts will be able to identify countries that might be lacking of specific funding for a particular SDG.

The Secretariat aims to extend the methodology to previous years to assess changes in donors' behaviour when entering in an SDG era. This work will also enable to explore the interlinkages between the SDGs and how policies have changed synergies between goals. To strengthen current methodology, the Secretariat has created a scientific expert group consisting of machine learning and development experts to present improvement and validate hypothesis.

References

AidData (2017), *AidData's Global Chinese Official Finance Dataset, 2000-2014, Version 1.0*, AidData, https://www.aiddata.org/data/chinese-global-official-finance-dataset (accessed on 10 January 2019). [24]

Development Co-operation Directorate (ed.) (2015), *Comprehensive Review of Purpose Codes and Policy Markers in Light of the SDGs and Post-2015 Agenda*, http://dx.doi.org/DCD/DAC/STAT(2015)9. [11]

Blumenstock, J. (2018), "Don't forget people in the use of big data for development", *Nature*, Vol. 561/7722, pp. 170-172, http://dx.doi.org/10.1038/d41586-018-06215-5. [5]

Breiman, L. and Leo (2001), "Random Forests", *Machine Learning*, Vol. 45/1, pp. 5-32, http://dx.doi.org/10.1023/A:1010933404324. [19]

CARI (2018), *Data — China Africa Research Initiative*, http://www.sais-cari.org/data/ (accessed on 10 January 2019). [25]

Chen, T. and C. Guestrin((n.d.)), "XGBoost: A Scalable Tree Boosting System", http://dx.doi.org/10.1145/2939672.2939785. [18]

Dupriez, O. (2018), *An empirical comparison of machine learning classification algorithms*, World Bank, http://pubdocs.worldbank.org/en/666731519844418182/PRT-OD-presentation-V2.pdf (accessed on 20 September 2018). [3]

Galsurkar, J. et al.((n.d.)), *Assessing National Development Plans for Alignment with Sustainable Development Goals via Semantic Search*, http://www.aaai.org (accessed on 21 January 2019). [38]

Galsurkar, J. et al. (2018), *Semantic Searching for Efficient Assessment of Sustainable Development in National Plans Under guidance of In Partnership with the United Nations Development Programme: Devika Iyer*, https://dssg.uchicago.edu/wp-content/uploads/2017/09/galsurkar.pdf (accessed on 24 January 2019). [12]

Global Fishing Watch (2017), *Global Fishing Watch*, http://globalfishingwatch.org/ (accessed on 20 September 2018). [33]

Goldblatt, R. et al. (2017), "Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover", *Remote Sensing of Environment*, Vol. 205, pp. 253-275, http://dx.doi.org/10.1016/j.rse.2017.11.026. [7]

Google Inc. (2018), *Translation API | Google Cloud*, https://cloud.google.com/translate/docs/apis (accessed on 18 September 2018). [16]

Gounden, C., J. Irvine and R. Wood (2015), "Promoting Food Security through Improved Analytics", *Procedia Engineering*, Vol. 107, pp. 335-336, http://dx.doi.org/10.1016/J.PROENG.2015.06.089. [8]

Hausmann, R. et al. (2018), *The Atlas of Economic Complexity*, https://s3.amazonaws.com/academia.edu.documents/30678659/HarvardMIT_AtlasOfEconomicComplexity_Part_I.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1537525021&Signature=7%2BrghudU%2Bop%2FusCxAaOsKR8foSw%3D&response-content-disposition=inline%3B%20filenam (accessed on 21 September 2018). [34]

Jean, N. et al. (2016), "Combining satellite imagery and machine learning to predict poverty.", *Science (New York, N.Y.)*, Vol. 353/6301, pp. 790-4, http://dx.doi.org/10.1126/science.aaf7894. [6]

Le Blanc, D. (2015), *Towards integration at last? The sustainable development goals as a network of targets*, UN-DESA, http://dx.doi.org/ST/ESA/2015/DWP/141. [21]

Lu, X. et al. (2016), "Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh", *Global Environmental Change*, Vol. 38, pp. 1-7, http://dx.doi.org/10.1016/J.GLOENVCHA.2016.02.002. [2]

Mckenzie, D. and D. Sansone (2017), *Man vs. machine in predicting successful entrepreneurs : evidence from a business plan competition in Nigeria*, http://documents.worldbank.org/curated/en/968231513116778571/Man-vs-machine-in-predicting-successful-entrepreneurs-evidence-from-a-business-plan-competition-in-Nigeria (accessed on 20 September 2018). [1]

McKinsey (2015), *An executive's guide to AI*, Mc Kinsey Insights, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/An%20executives%20guide%20to%20AI/An-executives-guide-to-AI.ashx (accessed on 03 October 2018). [15]

Morris, S. (2018), *The International Development Finance Club and the Sustainable Development Goals | Center For Global Development*, Center for Global Development, https://www.cgdev.org/publication/international-development-finance-club-and-sustainable-development-goals (accessed on 24 January 2019). [26]

OECD (2018), *DAC Glossary of Key Terms and Concepts - OECD*, http://www.oecd.org/dac/dac-glossary.htm#Purpose_Code (accessed on 03 October 2018). [36]

OECD (2018), *Development finance standards - OECD*, OECD-DAC, http://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/ (accessed on 18 September 2018). [30]

OECD (2018), *OECD Statistics on External Development Finance Targeting Environmental Objectives Including the Rio Conventions - OECD*, http://www.oecd.org/dac/environment-development/rioconventions.htm (accessed on 09 October 2018). [37]

OECD (2018), *Official Development Assistance (ODA); What is ODA?*, OECD, http://oe.cd/oda-eligibility-database (accessed on 30 January 2019). [10]

OECD (2018), *Sector Financing in the SDG Era*, The Development Dimension, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264307711-en. [13]

OECD (2017), “Data Collection on SDG Focus of Development Co-operation Activities”, OECD DAC, Paris, http://dx.doi.org/DCD/DAC/STAT(2017)28. [20]

OECD (2003), *OECD Glossary of Statistical Terms - Official development assistance (ODA) Definition*, https://stats.oecd.org/glossary/detail.asp?ID=6043 (accessed on 03 October 2018). [35]

OECD DAC (2018), *Converged Statistical Reporting Directives for the Creditor Reporting System (CRS) and the Annual DAC Questionnaire*, OECD, Paris, https://one.oecd.org/document/DCD/DAC/STAT(2018)9/FINAL/en/pdf (accessed on 19 September 2018). [17]

OECD DAC (2010), *Reporting Directives for the Creditor Reporting System*, OECD, http://dx.doi.org/DCD/DAC(2007)39/FINAL/CORR5. [31]

Park, S. et al. (2016), “Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions”, *Agricultural and Forest Meteorology*, Vol. 216, pp. 157-169, http://dx.doi.org/10.1016/J.AGRFORMET.2015.10.011. [9]

Parthasarathy, R., V. Rao and N. Palaniswamy (2017), *Deliberative inequality : a text-as-data study of Tamil Nadu's village assemblies*, http://documents.worldbank.org/curated/en/582551498568606865/Deliberative-inequality-a-text-as-data-study-of-Tamil-Nadus-village-assemblies (accessed on 20 September 2018). [4]

Pyle, D. and C. San José (2015), “An executive’s guide to machine learning | McKinsey”, *McKinsey Quarterly*, https://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning (accessed on 20 September 2018). [32]

Rao, A. (2015), *Demystifying machine learning*, PWC - Next in Tech, http://usblogs.pwc.com/emerging-technology/demystifying-machine-learning/?_ga=2.132073356.2086681082.1537427163-629790633.1536587932 (accessed on 20 September 2018). [14]

United Nations (2016), *Final list of proposed Sustainable Development Goal indicators*, Inter-Agency and Expert Group on Sustainable Development Goal Indicators, http://dx.doi.org/E/CN.3/2016/2/Rev.1. [28]

UNSTATS (2018), *IAEG-SDGs — SDG Indicators*, IAEG-SDGs, https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/ (accessed on 20 September 2018). [29]

Winkler, D. (2017), “How Do Multinationals Report Their Economic, Social, and Environmental Impacts? Evidence from Global Reporting Initiative Data”, No. 8274, World Bank, http://econ.worldbank.org. (accessed on 08 January 2019). [27]

World Bank (2017), *Creating Markets in Ghana Country Private Sector Diagnostic*, World Bank, https://www.ifc.org/wps/wcm/connect/ac42c20a-c82c-48b7-8432-221c0e066e2a/CPSD-Creating-Markets-in-Ghana-Nov-2017_v1.pdf?MOD=AJPERES (accessed on 08 January 2019). [23]

Zhu, X., Z. Ghahramani and J. Lafferty((n.d.)), *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, http://www.aaai.org/Papers/ICML/2003/ICML03-118.pdf (accessed on 17 September 2018). [22]

Annex A. Translated Donors

Table A A.1. Language of Long and Short Description in the CRS database for each donors

Long Description and Short Description are in English	United States, Japan, International Development Association, Canada, United Kingdom, Sweden, AsDB Special Funds, Norway, International Bank for Reconstruction and Development, Korea, Italy, UNFPA, Bill & Melinda Gates Foundation, Australia, Finland, Ireland, IDB Special Fund, World Health Organisation, Asian Development Bank, Global Environment Facility, United Arab Emirates, UNAIDS, OPEC Fund for International Development, Portugal, European Bank for Reconstruction and Development, New Zealand, Global Alliance for Vaccines and Immunization, Global Fund, IFAD, Kuwait, Islamic Development Bank, Slovak Republic, WFP, Arab Fund (AFESD), Estonia, Lithuania, Hungary, Council of Europe Development Bank, Romania, Caribbean Development Bank, Greece, Thailand, Iceland, Dutch Postcode Lottery, OSCE, Climate Investment Funds, Green Climate Fund, Nordic Development Fund, UNECE, Latvia, Malta, Russia, Kazakhstan, Global Green Growth Institute, Azerbaijan, Swedish Postcode Lottery, UNEP, Adaptation Fund, People's Postcode Lottery, UNRWA, Chinese Taipei, Croatia, Liechtenstein
Long Description and Short Description are in non-English language	EU Institutions, France, Spain, UNICEF, Switzerland, UNDP, Belgium, Netherlands, International Labour Organisation, Luxembourg, World Tourism Organisation
Long Description in non-English language	Germany, Austria, Inter-American Development Bank, Czech Republic, Poland, Slovenia
Short Description in non-English language	Denmark, African Development Fund, African Development Bank, UN Peacebuilding Fund

Annex B. Accuracy and Performance for each DAC country

Table A B.1. Summary of classification for each DAC Donor

Country	Correctly classified	Not Correctly classified	Accuracy	Assigned	Not Assigned	Assignment Rate	False Positive	False Negative
Australia	90	10	**90%**	81	19	**81%**	5	5
Austria	93	12	**89%**	70	35	**67%**	4	8
Belgium	83	17	**83%**	74	26	**74%**	7	10
Canada	90	11	**89%**	89	12	**88%**	7	4
Czech Republic	89	11	**89%**	83	17	**83%**	0	11
Denmark	90	10	**90%**	81	19	**81%**	5	5
EU Institutions	92	8	**92%**	90	10	**90%**	3	5
Finland	95	18	**84%**	91	22	**81%**	4	14
France	90	10	**90%**	84	16	**84%**	6	4
Germany	83	17	**83%**	85	15	**85%**	8	9
Greece	96	4	**96%**	97	3	**97%**	4	0
Hungary	97	3	**97%**	93	7	**93%**	0	3
Iceland	82	12	**87%**	75	19	**80%**	5	7
Ireland	92	8	**92%**	68	32	**68%**	5	3
Italy	96	22	**81%**	90	28	**76%**	9	13
Japan	84	16	**84%**	88	12	**88%**	8	8
Korea	86	14	**86%**	81	19	**81%**	6	8
Luxembourg	109	8	**93%**	87	30	**74%**	1	7
Netherlands	85	15	**85%**	77	23	**77%**	13	2
New Zealand	84	16	**84%**	79	21	**79%**	9	7
Norway	90	12	**88%**	81	21	**79%**	7	5
Poland	81	19	**81%**	76	24	**76%**	8	11
Portugal	86	14	**86%**	88	12	**88%**	9	5
Slovak Republic	80	20	**80%**	82	18	**82%**	9	11
Slovenia	79	21	**79%**	79	21	**79%**	9	12
Spain	98	12	**89%**	86	24	**78%**	4	8
Sweden	84	19	**82%**	81	22	**79%**	10	9
Switzerland	92	8	**92%**	83	17	**83%**	6	2
United Kingdom	91	9	**91%**	90	10	**90%**	5	4
United States	96	4	**96%**	96	4	**96%**	3	1

Note: Authors have manually classified 100 top projects in USD disbursement for each country and compared it with algorithm's predictions. Project is considered as correctly classified if primary SDG is hit. Project is not considered as correct if secondary SDG is hit but not the first one or if the algorithm does not predict any of the SDG.
Source: Authors 'calculation

Annex C. List of PDFs

Table A C.1. List of PDFs used for training

SDG	Title of the Publication	Name or Institution	Year
1	2017 HLPF Thematic Review of SDG 1: End Poverty in All its Forms Everywhere	UN - HLPF	2017
1	Ending extreme poverty by 2030	Devinit	2014
1	No Poverty, Educational Resource for Teachers and Facilitators	Concern Active Citizenship	2017
1	Statistical Yearbook for Asia and the Pacific - No Poverty	UN-ESCAP	2017
1	End poverty in all its forms everywhere	SDG Asia Pacific	2015
1	A Critical Note on Poverty Eradication Target of Sustainable Development Goals	Dr. Palash Kamruzzaman	2016
1	SDG 1: No poverty, End poverty in all its forms everywhere	PWC	2016
2	SDG 2 End Hunger, Achieve Food Security and Improved Nutrition and Promote Sustainable Agriculture	Mollier, Seyler, Chotte, Ringler	2017
2	Towards zero hunger and sustainability	FAO	2017
2	Zero Hunger Challenge	UNDP	2016
2	SDG 2 (Zero Hunger) in the Context of the German Sustainable Development Strategy: Are We Leaving the Starving Behind?	German Development Institute	2017
2	Toward Zero Hunger, a Strategic Review of Sustainable Development Goal 2 in Uganda	The Government of Uganda	2017
2	Zero hunger by 2030: The not-so-impossible dream	OFID	2016
2	International Society and Sustainable Development Goals	Lalaguna, Diaz Barrado, Fernandez Liesa	2016
2	Namibia Zero Hunger	Republic of Namibia	2017
2	Nepal: Zero Hunger Challenge National Action Plan (2016 - 2025)	Nepalese Ministry of Agricultural Development	2016
2	Working for Zero Hunger	World Food Programme	2017
2	Cooperatives for Zero Hunger in Africa	International Co-operative Alliance	2018
2	Zero Hunger Strategic Review	Republic of Liberia	2017
2	SDG 2: Zero Hunger End hunger, achieve food security and improved nutrition and promote sustainable agriculture	PWC	2017
3	Health in the Sustainable Development Goals	WHO	2016
3	SDG Health and Health Related Targets	WHO	2016
3	Achieving Health SDG 3 in Africa through NGO Capacity Building - Insights from the Gates Foundation Investment in Partnership in Advocacy for Child and Family Health (PACFaH) Project	Judith-Ann Walker	2016
3	2017 HLPF Thematic Review of SDG3: Ensure healthy lives and promote well-being for all at all ages	UN - HLPF	2017
3	Health in the Framework of Sustainable Development, Technical Report for the Post-2015 Development Agenda	SDSN	2014
3	From MDGs to SDGs: Implications for Maternal Newborn Health in Africa	Hodin & al	2016
3	Sustainable Health Agenda for the Americas 2018-2030: a Call to Action for Health and Wellbeing in the Americas	Pan American Health Organisation; WHO	2017
3	Health innovation and the Sustainable Development Goals	Global Health Technologies Coalition	2015
3	SDG3: Good health and well-being. Ensure healthy lives and promote well-being for all at all ages	PWC	2017
4	Enhancing Collaboration in Pursuit of SDG 4: Literacy and Lifelong Learning	KPMG	2017
4	Case Studies: Case Study SDG 4 "Education"	Credit Suisse	2017
4	Sustainable Development Goal 4 and Refugee Education	UNHCR	2015
4	PISA for Development and the Sustainable Development Goals	OECD	2017
4	Cashing in on SDG 4	Antonia Wulff	2017

4	SDG 4 South Africa - Country Profile	UNESCO	
4	Sustainable Development Goal 4 - Discussion Summary	The Commonwealth Education Hub	2015
4	SDG4 - Education 2030 in Asia and the Pacific	Maki Hayashikawa	2017
4	SDG Target 4.7 and the importance of monitoring learning materials	Global Education Monitoring Report	2016
4	SDG 4: Quality education - Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all	PWC	2017
5	Goal 5 Achieve gender equality and empower all women and girls	UNSTATS	2016
5	SDGs and Gender Equality: UN Interagency Guidance Note for the Europe and Central Asia Region	UN Europe and Central Asia Issue-Based Coalition on Gender	2017
5	Statistical Yearbook for Asia and the Pacific - Gender Equality	UN - ESCAP	2016
5	Case Study: SDG5: Gender Equality	Toshiba	
5	Achieving SDG 5 Achieve gender equality, and empower women and girls everywhere	Bina Agarwal - NITI	2017
5	Gender equality and sustainable development: Achieving the twin development goals in Africa	GSDR	2015
5	SDG 5 Achieve gender equality and empower all women and girls	Sr. Libania Fernandes	
5	SDG5 in Complex Fragile Settings: Participation, Peace and Politics	Alberoth & al	2016
5	Gender Equality and the Sustainable Development Goals - Toward a More Feminist United Nations	ICRW	2014
5	Achieving Gender Equality and Empowering Women and Girls: Is SDG 5 Missing Something?	Gita Sen	2015
5	Advancing SDG 5 through Inclusive Sourcing	GSC - UN GC	2016
5	SDG 5: Gender equality - Achieve gender equality and empower all women and girls	PWC	2017
6	SDG 6 in-depth review: UN-Water Synthesis Report 2018 on Water and Sanitation	UN Water	2018
6	SDG 6: Clean water and sanitation - Ensure availability and sustainable management of water and sanitation for all	PWC	2017
6	Goal 6 Ensure availability and sustainable management of water and sanitation for all	UNSTATS	2016
6	SDG 6 along the Water and Nutrient Cycles	AGUASAN	2017
6	SDG 6: Ensure Availability and Sustainable Management of Water and Sanitation for all -India's performance in terms of SDG 6	NITI	2018
6	Integrated approaches for Sustainable Development Goals planning: The case of Goal 6 on water and sanitation	UN ESCAP	2017
6	The SDGs in practice: Measuring and managing sustainable development water targets	Lisa-Maria Rebelo	2015
6	Water & Sanitation - A People's Guide to SDG 6 - A rights-based approach to implementation		2017
6	Water and Sustainable Development Opportunities and Challenges in Africa Region - SDGs for Water & The African Water Vision 2025	Stephen Max Donkor	2013
6	Synthesis Report SDG 6 Water and Sanitation	Angela Ortigara	2018
6	ROAD TO SDG 6: IMPACT X 20 - Water For People | Strategy Summary 2017–2021	Water for People	2017
6	SDG 6: Clean water and sanitation - Ensure availability and sustainable management of water and sanitation for all	PWC	2017
7	SDG7 Ensure Access to Affordable, Reliable, Sustainable and Modern Energy for All	McCollum, Echeverri, Riahi, Parkinson	2017
7	SDG 7 as an enabling factor for sustainable development: the role of technology innovation in the electricity sector	Alloisio & al	2018
7	SDG 7: Affordable and clean energy - Ensure access to affordable, reliable, sustainable and modern energy for all	PWC	2017
7	2030 Agenda: Review Process of SDG7 on Energy	European Union Energy Initiative	2018
7	Tracking SDG7: The Energy Progress Report 2018	IEA & al	2018

7	Sustainable Energy for the Implementation of the SDGs and the Paris Agreement - Vienna Energy Forum	UNIDO & al	2018
7	Progress on SDG 7 and its Interlinkages with other SDGs in support of First Review of SDG7 during the 2018 High-Level Political Forum	UN	2017
7	Abstract Book of the Conference - Sustainable Energy for Africa	KAOWARSOM	2018
7	Statistical Yearbook for Asia and the Pacific - SDG7	UN - ESCAP	2015
7	The Energy Transition and Disruptive Technologies - Asia Pacific will Usher in a Sustainable Energy Future	AMER7	2017
7	Energy Sector Strategy: Sustainable Energy for Asia	AIIB	2017
7	Accelerating Progress toward SDG 7: UN System Contributions	HLPF 2016	2016
8	SDG 8: Decent work and economic growth - Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	PWC	2017
8	SDG 8: Decent Work and Economic Growth	Toshiba	
8	Statistical Yearbook for Asia and the Pacific - SDG 8	UN ESCAP	2015
8	Tracking the SDGs in Canadian Cities: SDG 8	IISD	2018
8	Goal 8 Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	UNSTATS	2016
8	India’s roadmap for SDG 8 – A Brief Introduction	NITI	2016
8	Goal 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.	UUA	
8	Case Study SDG 8 “Sustainable Economic Growth”	Credit Suisse	2017
8	Building Partnerships to Localize SDG8	ILO - UCLG	2016
8	SDG 8: Promote Sustained, Inclusive and Sustainable Economic Growth, full and productive employment and decent work for all.	NIC	
8	SDG Note: Engaging the Private Sector on Decent Work-Business Operations and Investments	ILO	2017
8	Decent work for sustainable development	Governing Body - ILO	2016
9	SDG 9: Industry, innovation and infrastructure - Build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation	PWC	2017
9	2017 HLFP Thematic Review of SDG-9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	UN - HLPF	2017
9	Statistical Yearbook for Asia and the Pacific - SDG9	UN - ESCAP	2015
9	SDG 9: Build Resilient Infrastructure; Promote Inclusive and Sustainable Industrialization and Foster Innovation	NITI	2016
9	How to develop resilient infrastructure (SDG 9)	Interreg & al.	2017
9	Build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation	Nature Counts	2016
9	Goal 9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	The Danish Institute for Human Rights	
9	Goal 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	Grete Faremo	2015
9	Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation	Atlas of Sustainable Development Goals	2017
9	Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	SDG Compass	2016
10	SDG 10: Reduced inequalities - Reduce inequality within and among countries	PWC	2017
10	Goal 10 Reduce inequality within and among countries	UNSTATS	2016
10	SDG 10: Reducing inequalities – Concepts and approaches for development cooperation	KFW	2016
10	Statistical Yearbook for Asia and the Pacific - SDG 10	UN - ESCAP	2015
10	Sustainable Development Goal 10 - Reduced inequalities. Where does Portugal stand?	Silva & al.	2016
10	Goal 10: Reduce inequality within and among nations.	UUA	
10	Inequality SDGs: Countries Still Not Ready	actionaid	2016

10	‘SDG No.10: Reduced Inequalities – Challenges and Perspectives’	giz	2016
10	Reduce Inequality within and among Countries	Toshiba	
10	Will inequality get left behind in the 2030 Agenda?	Kate Donald	2018
10	What does it mean to reduce inequality between countries? Goal 10: “Reduce inequality within and among countries”	Andrew Walton	2015
10	Reducing economic inequality as a Sustainable Development Goal - Measuring up the options for beyond 2015	nef	2014
10	Reduce inequality within and among countries	Nature Counts	2016
10	SDG-10: Reduce inequalities within the States	NIC	2010
11	SDG 11: Sustainable cities and communities - Make cities and human settlements inclusive, safe, resilient and sustainable		
11	SDG Goal 11 Monitoring Framework	UN-HABITAT TECHNICAL SUPPORT	2016
11	Goal 11 Make cities and human settlements inclusive, safe, resilient and sustainable	UNSTATS	2016
11	Case Study SDG 11 “Sustainable Cities”	Credit Suisse	2017
11	Statistical Yearbook for Asia and the Pacific - SDG 11	UN ESCAP	2015
11	SUSTAINABLE DEVELOPMENT GOAL 11 - Make cities and human settlements inclusive, safe, resilient and sustainable	UN-HABITAT TECHNICAL SUPPORT	2016
11	Cities and the Sustainable Development Goals	ICLEI	2015
11	Implementing Sustainable Development Goal 11 by connecting sustainability policies and urban planning practices through ICTs	CBD & al.	2017
11	SDG 11 - Make cities and human settlements inclusive, safe, resilient and sustainable	Daria Cibrario	2018
11	Make Cities and Human Settlements Inclusive, Safe, Resilient and Sustainable	World Sustainability Fund	2016
11	SDG 11: Supporting the delivery of cities that work for all	Shabana Shiraz	2015
11	Goal 11: Sustainable Cities and Communities	Christina Kamp	2017
12	SDG12: Responsible consumption and production - Ensure responsible consumption and production patterns	PWC	2017
12	Case Study // Round Table SDG 12: Successful approaches to delivering on Sustainable Consumption and Production by 2030	UNECE	2018
12	Sustainable Consumption and Production Indicators for the Future SDGs	UNEP	2015
12	Binding rules on business and human rights – a critical prerequisite to ensure sustainable consumption and production patterns	Martens & al.	2017
12	SDG TARGET 12.3 on Food Loss and Waste: 2017 Progress Report	Champions 12.3	2017
12	Statistical Yearbook for Asia and the Pacific - SDG 12	UN ESCAP	2015
12	Goal 12: Responsible Consumption and Production	KATE	2017
12	PURSUING SDG TARGET 12.3: Guidance on interpreting Sustainable Development Goal Target 12.3	Craig Hanson	2017
12	ACHIEVING SUSTAINABLE DEVELOPMENT GOAL 12: an exploratory study on sustainable consumption in Lima, Peru	Irene HOFMEIJER	2017
12	SDG 12: Ensure sustainable consumption and production patterns	McLarty Associates	
12	Pakistan National Action Plan on SDG 12 - Sustainable Consumption and Production	Government of Pakistan	2017
12	SDG 12 interventions - UNECE Regional Forum on Sustainable Development 2018	UNECE	2018
12	Ensure sustainable production and consumption (SDG 12): What role for trade?	giz-ICTSD	2017
12	7th Economic and Social Council Youth Forum - CONCEPT NOTE	Youth2030	2018
13	SDG 13: Climate action - Take urgent action to combat climate change and its impacts	PWC	2017
13	SDG 13: Climate Action	Metz & Gupta	2016
13	Statistical Yearbook for Asia and the Pacific - SDG 13	UN ESCAP	2015
13	Feature—Brief on Sustainable Development Goal 13 on Taking Action on Climate Change and Its Impacts: Contributions of International Law, Policy and Governance	Lofts & al.	2016

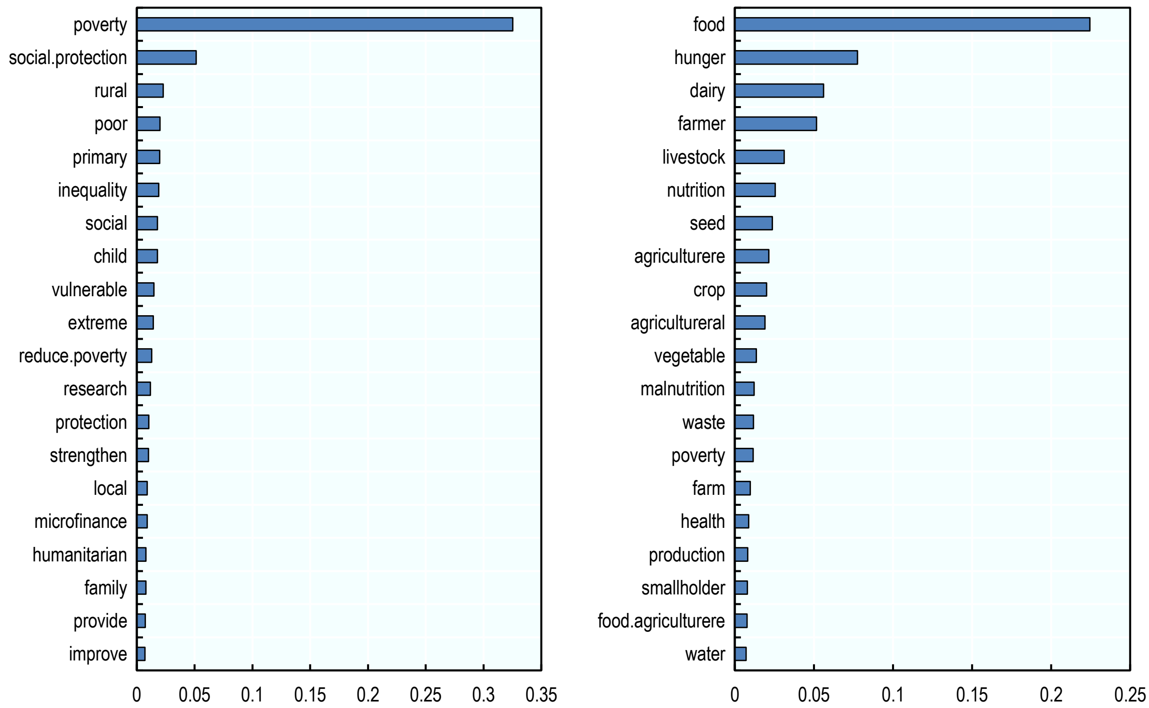
13	The pivot point: realizing Sustainable Development Goals by ending corporate capture of climate policy	Lawrence-Samuel & al.	2017
13	Goal 13 Take urgent action to combat climate change and its impacts	UNSTATS	2016
13	Opportunities and options for integrating climate change adaptation with the Sustainable Development Goals and the Sendai Framework for Disaster Risk Reduction 2015–2030	UNCCS	2017
13	Trade and SDG 13 - Action on Climate Change - ADBI Working Paper Series	ADBI	2017
13	SDG 13, the SEEA and New Zealand's missing carbon tax	Jeremy Webb	2017
13	SDG 13 Take Urgent Action to Combat Climate Change and It's Impact	Ajay K. Jha	2017
13	SDG 13: Take urgent action to combat climate change and its impacts	Global Pulse Confederation	
14	SDG 14: Life below water - Conserve and sustainably use the oceans, seas and marine resources for sustainable development	PWC	2017
14	The Large Marine Ecosystem Approach - An Engine for Achieving SDG 14	LME Learn	2017
14	Goal 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development.	UUA	
14	2017 HLPF Thematic review of SDG 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development	UN HLPF	2017
14	FAO Working for SDG 14: Healthy oceans for food security, nutrition and resilient communities	FAO	2017
14	SDG 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development	Schmidt & al.	2017
14	A preliminary assessment of indicators for SDG 14 on "Oceans "	Laura Recuero Virto	2017
14	MEETING BRIEF - Innovation in Partnerships - SDG 14: Life below Water	one Earth - IPI	2017
14	Outcomes of the UN SDG 14 Conference	Intergovernmental Oceanographic Commission	2017
14	Statistical Yearbook for Asia and the Pacific - SDG 14	UN ESCAP	2015
14	IUCN's contribution to Transforming Our World - Goal 14: Conserve and sustainably use the oceans, seas and marine resources	IUCN	2017
15	SDG 15: Life on land - Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation and halt biodiversity loss	PWC	2017
15	Transforming Tourism - Goal 15: life on land	Naturefriends International	2017
15	Land Degradation Neutrality: The Target Setting Programme	The Global Mechanism	2016
15	SDG 15 interventions - UNECE Regional Forum on Sustainable Development 2018	UNECE	2018
15	SDG 15 Life on Land: Monitoring and maximizing global environmental benefits of drylands	GEF	2018
15	Goal 15 Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	Mahmoud Mohieldin and Paula Caballero	2015
15	Implementing SDG 15: Can large-scale public programs help deliver biodiversity conservation, restoration and management, while assisting human development?	Bridgewater & al.	2015
15	SDG 15: Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	Syngenta	
15	INTEGRATE – A European-wide network to enhance nature conservation in sustainably managed forests	European Network Integrate	2018
15	Trade and SDG 15: Promoting "Life on Land" through Mandatory and Voluntary Approaches - ADBI Working Paper Series	ADBI	2017
15	A journey through the value chain of wood: the case of Austria.	Federal Ministry of Sustainability and Tourism, Austria	2018
15	Case Study SDG 15 - "Terrestrial Ecosystems"	Credit Suisse	2017

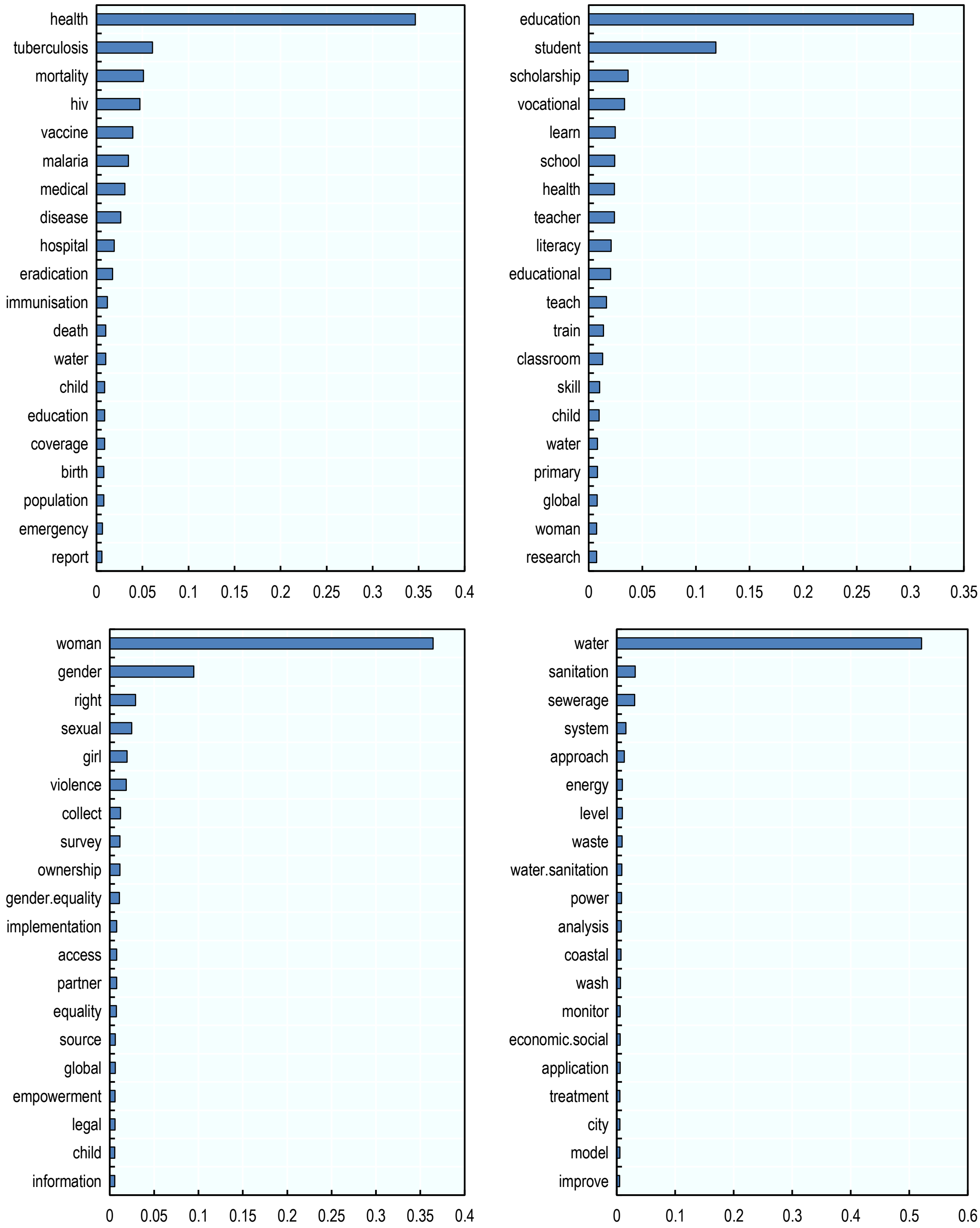
15	Protect, restore and promote sustainable use of terrestrial ecosystem, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	Christine von Weizsaecker	2017
15	Report of Roundtable on SDG 15 on Life on Land - Asia Pacific Forum on Sustainable Development	UN ESCAP	2018
15	Statistical Yearbook for Asia and the Pacific - SDG 15	UN ESCAP	2015
15	Keeping an eye on SDG 15 - Working with countries to measure indicators for forests and mountains	FAO	2013
15	Sustainable Development Goal 15: Progress and Prospects - An expert group meeting in preparation for HLPF 2018: Transformation towards sustainable and resilient societies	UN-DESA	2018
16	SDG 16: Peace and justice; strong institutions - Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	PWC	2017
16	SDG 16 on Governance and its measurement: Africa in the Lead	Cling & al.	2018
16	The Solutions Forum on SDG 16+: Towards Implementation	the Graduate Institute	2017
16	SDG 16 Innovation Challenge #SDG16Innovation	Government of Pakistan	2018
16	Sustainable Development Goal 16: The Challenge of Sustaining Peace in Places of Crisis	African Policy Circle	2017
16	GOAL 16 - Advocacy Toolkit	TAP Network	2018
16	Goal 16 Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	UNSTATS	2016
16	Business and SDG 16: Contributing to Peaceful, Just and Inclusive Societies	SDGF	2017
16	Monitoring to Implement Peaceful, Just and Inclusive Societies	UNDP	2017
16	Promote Peaceful, Just and Inclusive Societies	UNDP	2017
16	Achieving the Impossible: Can we be SDG 16 Believers?	Alan Whaites	2016
16	Statistical Yearbook for Asia and the Pacific - SDG 16	UN ESCAP	2015
16	SDG16 PROGRESS REPORT - A comprehensive global audit of progress on available SDG16 indicators	Institute for Economics & Peace	2017
16	SDG 16 - Progressive implementation of the 2030 Agenda depends on achieving sustainable peace	Ziad Abdel Samad	2017
16	How can the New Deal and SDG 16+ be Achieved?	Williams & al.	2017
16	There can be no sustainable development without peace and no peace without sustainable development	Robert Zuber	2017
16	Measuring progress towards SDG 16: Data needs and resources in Nepal	Royal Norwegian Embassy, Kathmandu	2017
16	National Workshop on SDG 16 - Proceeding Report	NGO Federation of Nepal	2017
17	SDG 17: Partnerships for the goals Strengthen the means of implementation and revitalise Global Partnership for Sustainable Development		
17	SDG 17: It's not about transactional partnerships	Philippa Smales	2017
17	Enabling a Sustainable Future through the Joint Action of Countries and Communities: A Revitalized Global Partnership for Sustainable Development	Nikhil Seth	2015
17	Partnerships for the goals - Achieving the United Nations' Sustainable Development Goals	UBS	2018
17	A Project Summary on SDG 17: Partnerships for the Goals	The Open Door Project	2016
17	Partnerships for the goals	UNESCAP	2017
17	SPACE for SDGs a Global Partnership	Giovanni Rum	2017
17	Transforming Tourism - Goal 17: Partnerships for the goals	Monshausen & al.	2017
17	A legacy review towards realizing the 2030 Agenda	UNSD	2015
17	SDG 17 - Means of implementation or means of appropriation?	Stefano Prato	2017
17	Statistical Yearbook for Asia and the Pacific - SDG 17	UN ESCAP	2015
17	Expert Group Meeting on Sustainable Development Goal 17	UNHQ	2018

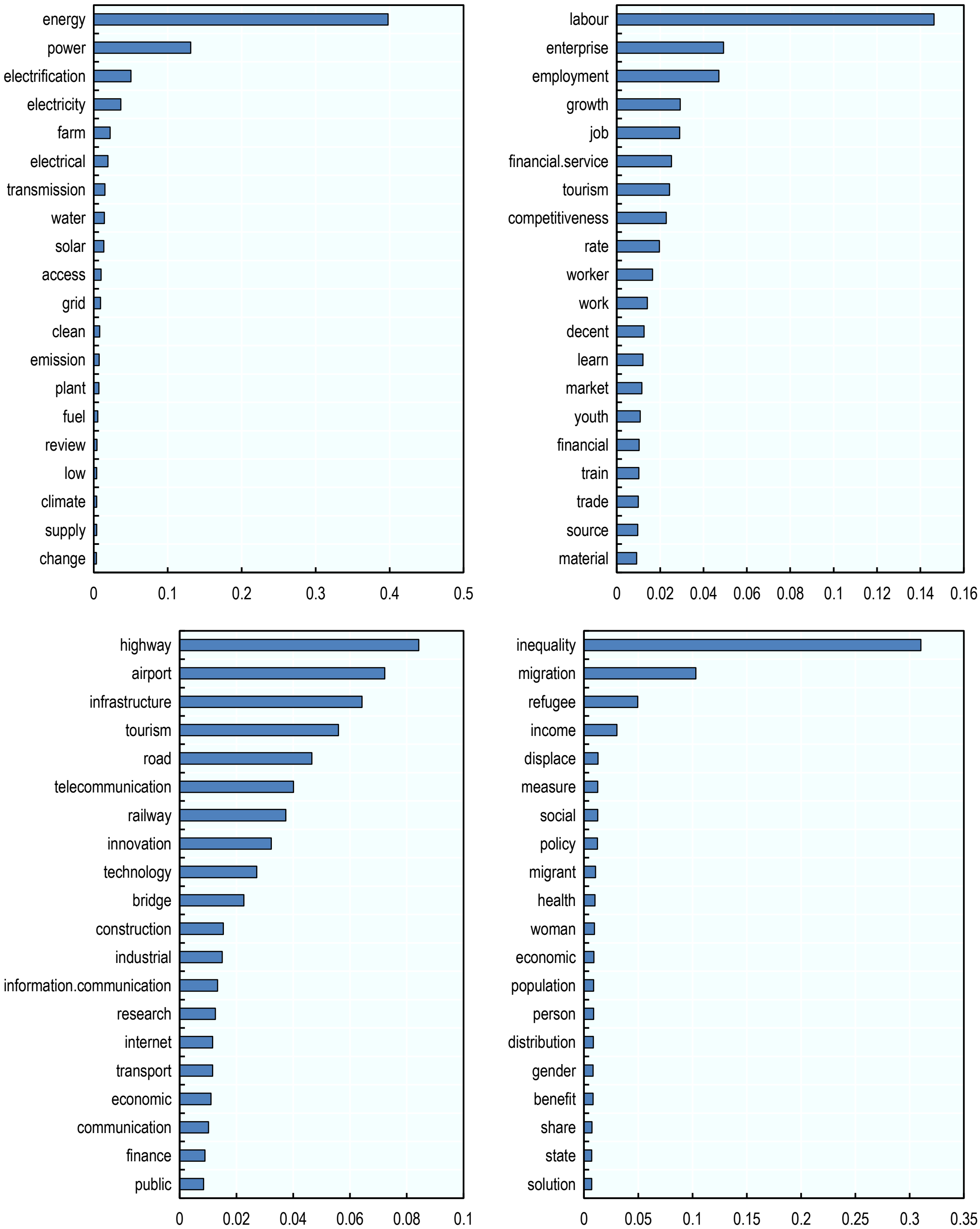
Annex D. Weight of Words per SDG

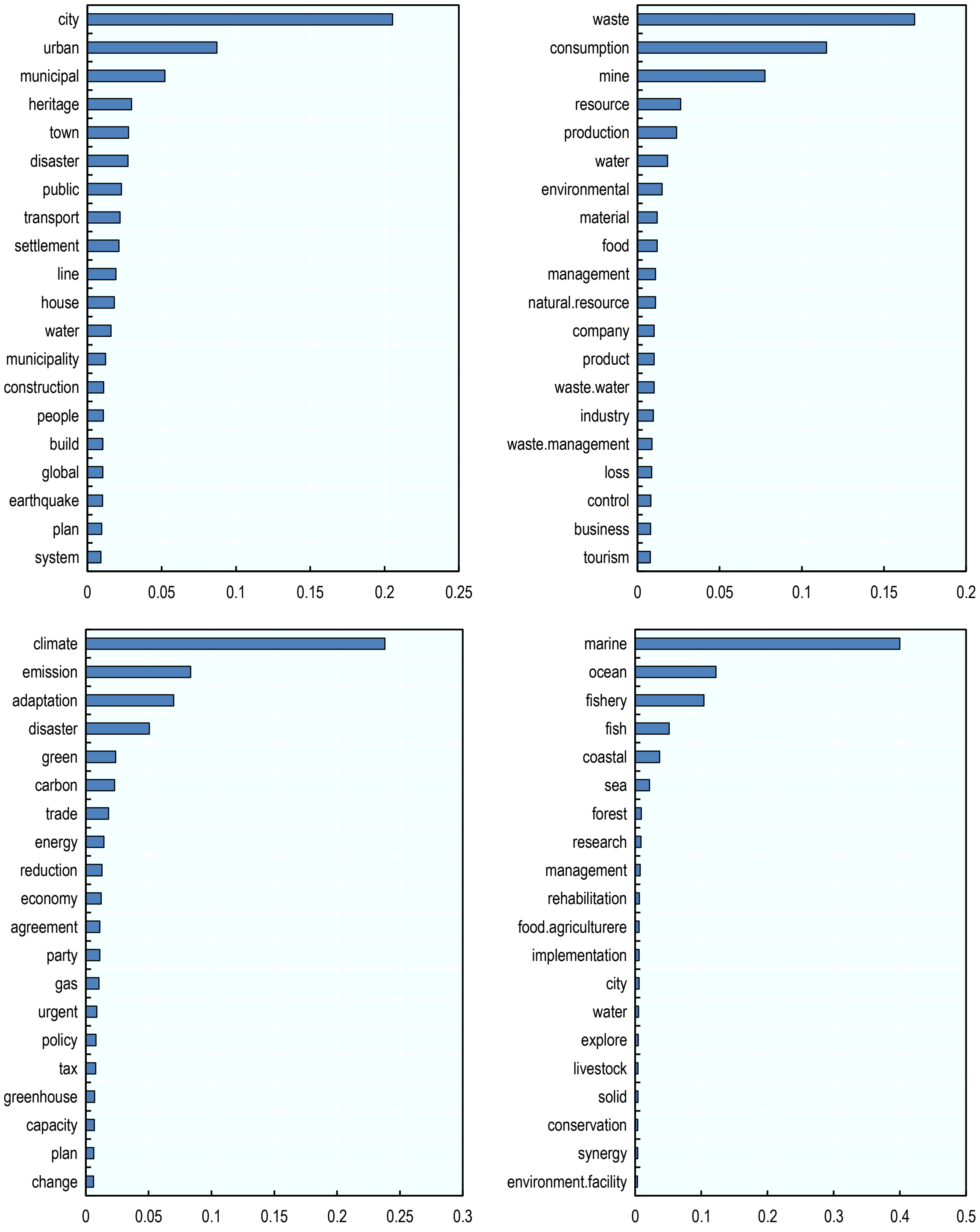
Figure A D.1. Words' weight for each SDG

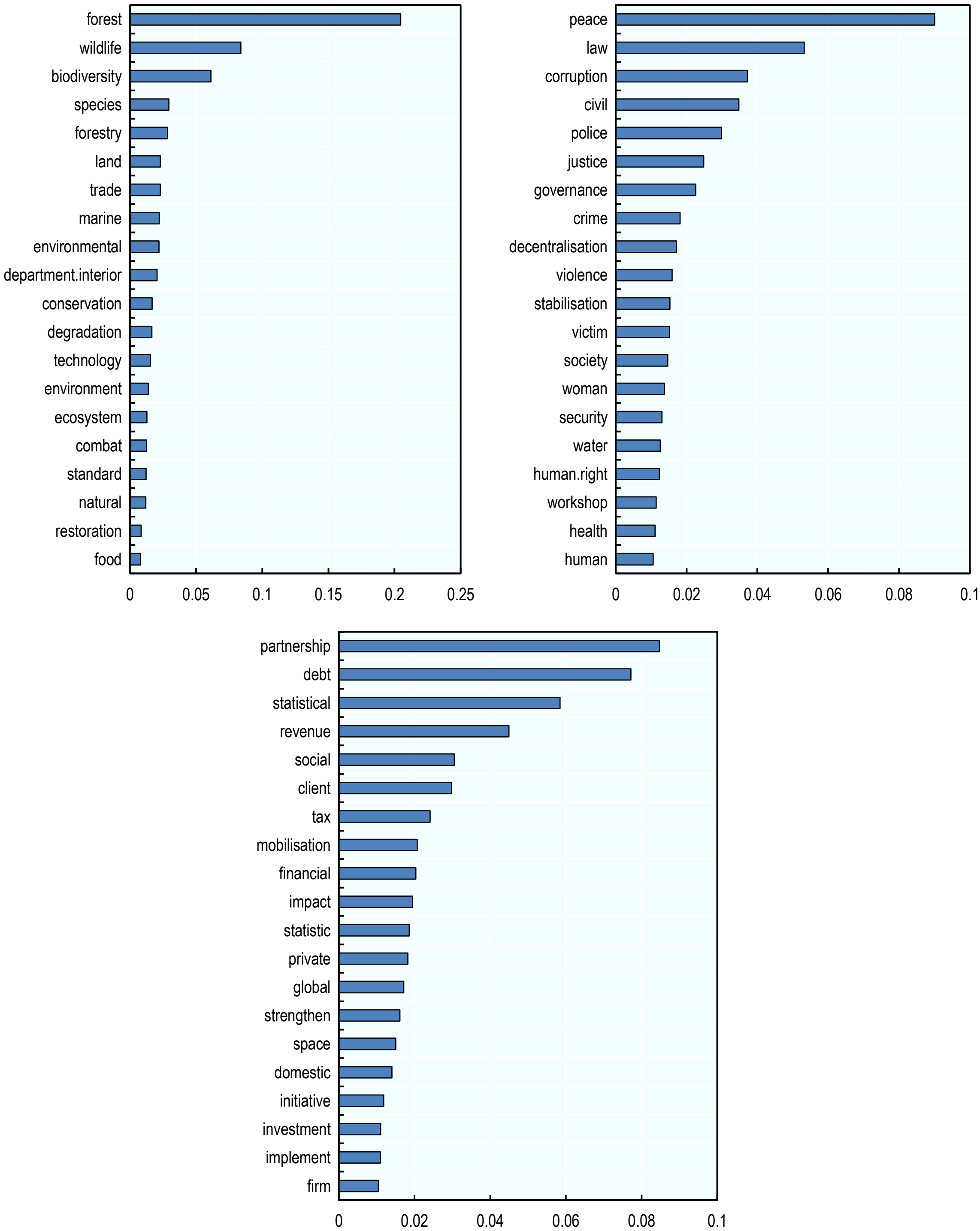
Weight of the top 20 words attributed by the algorithm to predict categorisation of each SDG




health
tuberculosis
mortality
hiv
vaccine
malaria
medical
disease
hospital
eradication
immunisation
death
water
child
education
coverage
birth
population
emergency
report
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4
education
student
scholarship
vocational
learn
school
health
teacher
literacy
educational
teach
train
classroom
skill
child
water
primary
global
woman
research
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35
woman
gender
right
sexual
girl
violence
collect
survey
ownership
gender.equality
implementation
access
partner
equality
source
global
empowerment
legal
child
information
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4
water
sanitation
sewerage
system
approach
energy
level
waste
water.sanitation
power
analysis
coastal
wash
monitor
economic.social
application
treatment
city
model
improve
0 0.1 0.2 0.3 0.4 0.5 0.6


energy
power
electrification
electricity
farm
electrical
transmission
water
solar
access
grid
clean
emission
plant
fuel
review
low
climate
supply
change
0 0.1 0.2 0.3 0.4 0.5
labour
enterprise
employment
growth
job
financial.service
tourism
competitiveness
rate
worker
work
decent
learn
market
youth
financial
train
trade
source
material
0 0.02 0.04 0.06 0.08 0.1 0.12 0.14 0.16
highway
airport
infrastructure
tourism
road
telecommunication
railway
innovation
technology
bridge
construction
industrial
information.communication
research
internet
transport
economic
communication
finance
public
0 0.02 0.04 0.06 0.08 0.1
inequality
migration
refugee
income
displace
measure
social
policy
migrant
health
woman
economic
population
person
distribution
gender
benefit
share
state
solution
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35


city
urban
municipal
heritage
town
disaster
public
transport
settlement
line
house
water
municipality
construction
people
build
global
earthquake
plan
system
0
0.05
0.1
0.15
0.2
0.25
waste
consumption
mine
resource
production
water
environmental
material
food
management
natural.resource
company
product
waste.water
industry
waste.management
loss
control
business
tourism
0
0.05
0.1
0.15
0.2
climate
emission
adaptation
disaster
green
carbon
trade
energy
reduction
economy
agreement
party
gas
urgent
policy
tax
greenhouse
capacity
plan
change
0
0.05
0.1
0.15
0.2
0.25
0.3
marine
ocean
fishery
fish
coastal
sea
forest
research
management
rehabilitation
food.agriculturere
implementation
city
water
explore
livestock
solid
conservation
synergy
environment.facility
0
0.1
0.2
0.3
0.4
0.5

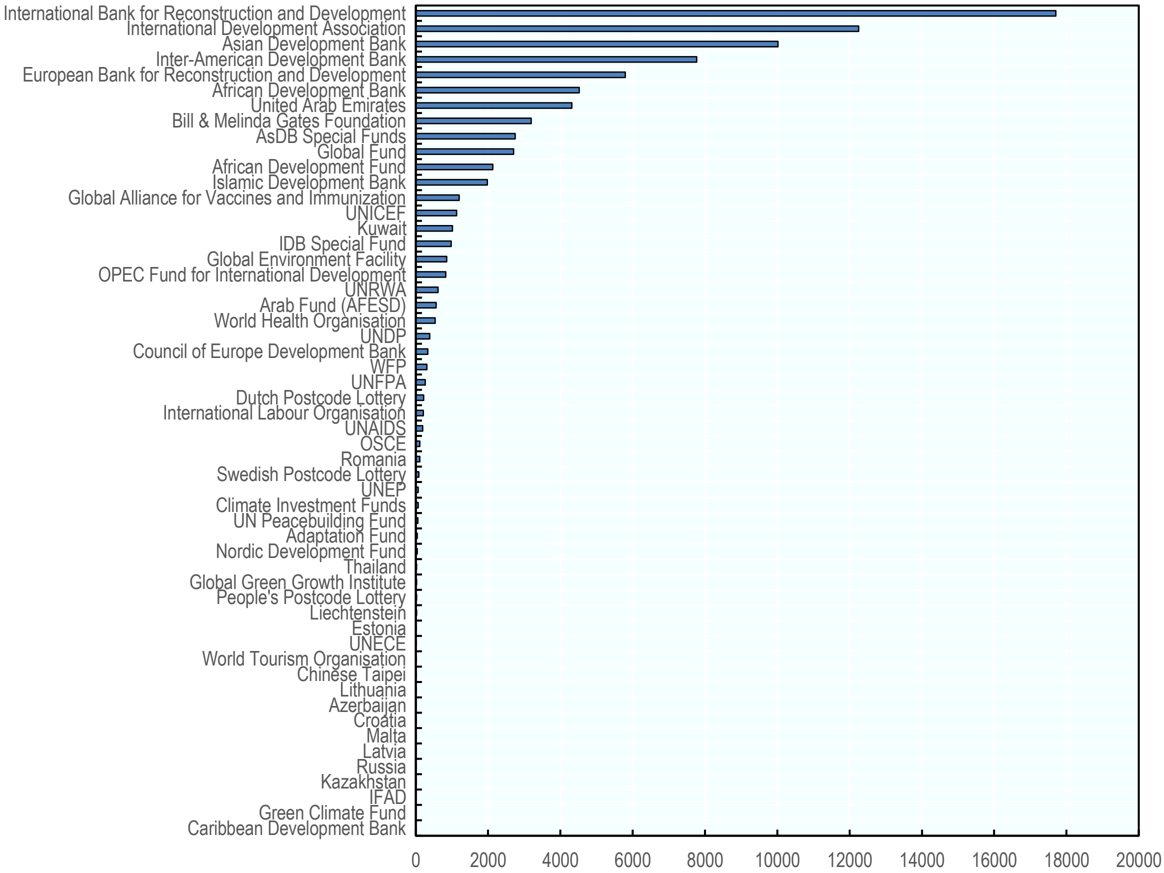


Note: The weight of words as well as which word to include is not decided by the analyst. It is based on pattern matching by the algorithm hence-why some words might belong to other categories or a not related to the SDG. Weight of word can change in the future due to algorithm improvement.

Annex E. List of non-DAC donors present in CRS

Figure A E.1. Non-DAC Reporting Donors

Non-DAC reporting donors classified by disbursement (USD million).



Source: CRS 2016.

BETTER POLICIES FOR BETTER LIVES