Proceedings of the 7th International Conference on Economics and Social Sciences (2024), ISSN 2704-6524, pp. 526-536

The 7th International Conference on Economics and Social Sciences
Exploring Global Perspectives:
The Future of Economics and Social Sciences
June 13-14, 2024
Bucharest University of Economic Studies, Romania

The Worldwide Progress of SDGs. Depicting the Yearly Hot Topics, using Language Processing

Andreea PERNICI[1*], Stelian STANCU[2]

DOI: 10.24818/ICESS/2024/049

Abstract

The Sustainable Development Goals have become indispensable for the international agenda since their official launch in 2015. Corroborated, the publication of the yearly SDG reports represents a candid focal point for the gained progress, as long as a clear depiction of the global challenges, resources, and knowledge. Thus, in the 2016-2023 period, eight annual reports have been published, containing more than 500 content pages. Starting from those, we aim to design an explorative analysis framework, through the computation of several language processing techniques from the family of text mining, and sentiment analysis. Our objective will be to identify the top-level hot topics of every year, while also pinpointing certain temporal dynamics in terms of new or recurrent subjects, international threats, or social-economic discourse elements. Deep-diving into the methodological instruments, the analysis will be based on a mix of techniques, starting from the illustration of the most common terms and collocations. Afterward, we will apply the Term Frequency-Inverse Document Frequency algorithm to highlight the importance of several words relative to the entire collection of documents, with a focus on the evolutive dimension. Going forward, we will evaluate the sentiment scores for each report, to assess whether in certain moments the general tone of the conversation has switched to a more negative or positive valence. Throughout the applicative section, we will present the results through several graphical visualisations meant to better profile the connections between concepts. Finally, we consider the current approach to be valuable for understanding the general evolutions in terms of sustainable development, creating a summative and computational-based exploration of the progress generated worldwide in the last decade.

Keywords: Sustainable Development Goals, Language Processing, Text Mining, Sentiment Analysis, TF-IDF.

JEL Classification: Q01, Q56, F53.

[1] Bucharest University of Economic Studies, Bucharest, Romania, andreea.pernici@csie.ase.ro.
* Corresponding author.
[2] Bucharest University of Economic Studies, Bucharest, Romania; Romanian Academy, Bucharest, Romania, stelian.stancu@csie.ase.ro.

1. Introduction

The Sustainable Development Goals (SDGs) could be described as a constant hot topic on the international agenda, ever since their official launch in 2015. Designed as a fundamental instrument of the Sustainable Development Pillar, the goals have become the centrepiece for understanding the global context, with a critical focus on the multiple shortages and risks that have appeared throughout the last decade. However, assessing the international status quo is not an easy task, so along with the 17 official SDGs, a great deal of data, research, and expertise has been put forward by individual experts, representatives of the related literature, and linked institutions. To get a glimpse of the volumes, as of May 2024, according to the official SDGs website[3], there were 169 targets, almost 4.000 events, and over 1.300 publications on the subject, with the numbers increasing daily.

As part of those publications, one of the most important will be the *SDG Progress Report*, presented every year by the UN Secretary-General. The goal behind it is simple: highlight in a concise and documented way the evolution towards the targets, while also raising awareness on the urgency needed to reach a more sustainable future. However, even though these publications will be as brief as possible, when studying the entire period, the volume of information is undoubtedly increasing, making it more difficult to identify certain temporal dynamics and more subtle insights. Thus, our approach is coming as an extension of the individually done research, completing the perspective with a language processing computation, meant to extract the most relevant conclusions in terms of word patterns, overall perceived sentiment, and the potential network, and interconnections found under the *development* umbrella.

2. Problem Statement

Going forward to the research framework, as mentioned in the introduction, our main source of data will be represented by the SDG progress reports from the last eight years, more specifically the period 2016-2023. Thus, in the applicative section, we will study these publications as both individual components, as well as in conjunction, merged together to form a new text entry. However, before proceeding to the computational part, it can be useful to get a glimpse of their structure.

Each of the SDG reports will start by presenting a foreword section, aimed at summarising certain trends and dynamics in the current international state of affairs. In the more recent reports, an emphasis on several crucial and urgent pillars will also be present, as well as an acknowledgment of the severe risks faced by humanity. Nevertheless, each report will focus on a summary of all 17 goals, based on the army of indicators behind them, ending with a closing statement, strengthening the main actions that need to be taken. Finally, each publication will be completed with visual summaries, valuable for a rapid skimming of the document.

Going back to our application, the first step we endeavoured was to extract the main idea from each report, an initial milestone in the process. The result can be

[3] Retrieved from: https://sdgs.un.org/goals.

observed in Table 1, where we can find the reports' references, number of pages, and yearly simplified key points. As a methodological mention, the main ideas have been generated empirically, through observation.

Table 1. Description of the reports

Reference	Pag	Main Ideas
(UN DESA, 2016)	56	A starting point, with a focus on presenting each goal and the need for key milestones and frequent evaluation going forward.
(UN DESA, 2017)	64	Progress is visible, however not equitable, being uneven among different demographics. A bolder vision is needed to meet the SDG targets.
(UN DESA, 2018)	40	Emphasising the need for immediate and accelerated action, with a reflection on the global challenges. A focus on the disadvantaged groups.
(UN DESA, 2019)	64	Climate change is defined as the most urgent area for action, along with inequality, poverty hunger, and diseases. The exploration of the interlinkages across goals and the opportunities to accelerate progress.
(UN DESA, 2020)	68	The effects of the COVID-19 pandemic, which could generate transformative pathways in turbulent times. A need for data innovations.
(UN DESA, 2021)	68	The continuation of the pandemic and its toll on the society and economy. A focus on vaccines, global solidarity, and a commitment to a multidimensional, inclusive recovery.
(UN DESA, 2022)	68	Interlinked global crises: the pandemic, a fragile economic recovery, the war in Ukraine, and climate change. A focus on a sustainable and urgent rescue effort.
(UN DESA, 2023)	80	The age of poly-crisis. A special edition that presents a vision of hope, highlighting the progress and the potential for further advancements.

Source: authors' own processing, based on the mentioned references.

After a quick understanding of the general message of each report, the next step was to turn to the related literature, in the effort of finding similar papers that employed language processing tools. Thus, we have found several articles that used text-mining methods to analyse specific goals or processes, for example, climate change (Hwang et al., 2021) or the global sustainable development concept (Roy et al., 2022). Regarding sentiment analysis, an interesting approach has been designed by Shen et al. (2021) in the exploration of social media for SDGs, however, the comparability with our approach will be limited.

Therefore, we can consider the current paper as a valuable contribution to the related field, adding an element of novelty through the methodology and the correlation with the SDG reports. At the same time, it can be seen as a continuation of previous authors' work (Pernici et al., 2023).

3. Research Methods

Relating to the research methodology, we will split our analysis into two sections. The first one will be text mining, through the use of several instruments, such as identifying the most common words, collocations, and measuring the TF-IDF frequencies. The second one will focus on sentiment analysis, through the computation of the sentiment scores and exploring certain word valences. Both of the sections will be described theoretically going forward.

Before proceeding, there is one last step that needs to be described, namely the preprocessing stage. Therefore, we will first ensure the curation of our datasets, through the removal of the numbers, punctuations, special characters, and stopwords from each of the eight individual texts. Next, the data frames will be tokenised, or in other words, reduced to an easy-to-process format made out of singular elements. With this new format, we will go forward with the application of the language processing methods, using a Python environment.

3.1 Text Mining

For the first family of methods, we will start with the illustration of *the most common words.* In this stage, we will also focus on graphical representations, especially through the form of *word clouds*, for each of the eight SDG reports. Afterward, we will continue with the exploration of the *collocations*, or the combinations of words that are frequently found together throughout the text. In this step, we will design a network graph meant to better highlight the conceptual connections. To complete the picture, we will also employ a frequently used measure in language processing, the TF-IDF, explained in detail below.

Term Frequency-Inverse Document Frequency (TF-IDF)

The $TF - IDF$ is a popular technique used to determine the relative contribution of one word in correlation with an entire corpus of documents. Thus, there will be three main elements that together compose the $TF - IDF$ scores. The first one will be the *term frequency* (TF), or the number of instances of a given word t, in a document d, while the second one will be the *document frequency* (DF), which will count the number of occurrences of the word t in the document set N (1). Going forward, the third element will be the *inverse document frequency* (IDF), which will assess the number of documents in the corpus separated by the frequency of the text. Regarding this last element, it is important to mention that the IDF scores will be processed with a base 2 logarithm to reach the final values (2).

$$TF(t,d) = \frac{count\ of\ t\ in\ d}{number\ of\ words\ in\ d}\ ,\ \ DF(t) = occurance\ of\ t\ in\ N = N(t) \quad (1)$$

$$IDF(t) = \frac{N}{DF(t)} => IDF(t) = \log\left(\frac{N}{DF(t)}\right) \quad (2)$$

Thus, the method will ultimately be a weighting system that assigns a weight to each word based on its term frequency (t) and the reciprocal document frequency. As a last step in the computation, the $TF - IDF$ scores will be calculated (3).

$$TF - IDF(t,d) = TF(t,d) * IDF(t) \quad (3)$$

3.2 Sentiment Analysis

Reaching our second part of the paper, the sentiment analysis, here the focus will be on extracting the sentiment scores, at both a document and word level. Therefore,

using the *VADER* lexicon in Python, we will be able to identify the general tone of voice in each SDG report, assess whether in time these publications have become more positive or negative, and extract the most significant results at an individual term level. In terms of methodology, the application will start from the tokenised format, search, and find each token in the specific lexicon and their adjacent sentiment value, and finally aggregate the total sentiment scores.

4. Findings

4.1 Text Mining

Most common words

Discussing now the first results, namely the illustration of the most common words in each year, it is important to mention that before arriving at a conclusive illustration, we have eliminated plenty of words from the tokenised dataset. This decision was made considering their significance, many of them being usual words that do not bring extra value (for example *percent, dollars, many*), or terms that are evident for the subject at hand (for example *sustainable, development, action).*

Therefore, the first method in which we chose to illustrate the most common words is shown in Table 2, where we can find the top 25 frequent terms for the period 2016-2019 (after the curation step), along with the absolute and relative frequencies. What is important to mention here is the fact that we have kept the geographical terms as well, considering them valuable for understanding several dynamics.

Table 2. Most common words – absolute and relative frequency – 2016-2019

2016		2017		2018		2019	
Asia	219 - 1.6%	**Asia**	242 – 1.1%	**Water**	91– 0.7%	**Asia**	167 – 0.9%
Africa	116 - 0.9%	**Africa**	210 – 1%	**Asia**	77 – 0.6%	**Africa**	151 – 0.8%
Children	61 – 0.5%	**America**	158 – 0.7%	**Africa**	70 – 0.6%	**America**	94 – 0.5%
Subsaharan	57 – 0.4%	**Subsaharan**	129 – 0.6%	**Access**	62 – 0.5%	**Subsaharan**	93 – 0.5%
Women	52 – 0.4%	**Women**	104 – 0.5%	**Energy**	50 – 0.4%	**Water**	81 – 0.4%
Water	51 – 0.4%	**Children**	92 – 0.4%	**America**	44 – 0.3%	**Women**	75 – 0.4%
America	51 – 0.4%	**Europe**	90 – 0.4%	**Subsaharan**	43 – 0.3%	**Children**	72 – 0.4%
Growth	48 – 0.4%	**Latin**	76 – 0.4%	**Sanitation**	38 – 0.3%	**Access**	67 – 0.4%
Caribbean	48 – 0.4%	**Caribbean**	74 – 0.3%	**ODA**	37 – 0.3%	**Health**	60 – 0.3%
Latin	47 – 0.2%	**Zealand**	71 – 0.3%	**Women**	34 – 0.3%	**Poverty**	55 – 0.3%
Access	43 – 0.3%	**Water**	70 – 0.3%	**Use**	33 – 0.3%	**Climate**	55 – 0.3%
Oceania	43 – 0.3%	**Growth**	70 – 0.3%	**Urban**	32 – 0.3%	**Growth**	50 – 0.3%
Urban	41– 0.3%	**Australia**	70 – 0.3%	**Forest**	32 – 0.3%	**Resources**	50 – 0.3%
Energy	40 – 0.3%	**Oceania**	70 – 0.3%	**Consumption**	29 – 0.2%	**Australia**	49 – 0.3%
GDP	33 – 0.2%	**Access**	65 – 0.3%	**Growth**	29 – 0.2%	**Zealand**	49 – 0.3%
Caucasus	33 – 0.2%	**Poverty**	61 – 0.3%	**Implementation**	28 – 0.2%	**Energy**	47 – 0.3%
Health	30 – 0.2%	**Health**	59 – 0.3%	**Management**	28 – 0.2%	**Latin**	46 – 0.3%
Risk	29 – 0.2%	**Urban**	56 – 0.3%	**Agenda**	27 – 0.2%	**Europe**	45 – 0.2%
Deaths	29 – 0.2%	**Deaths**	53 – 0.2%	**Policies**	27 – 0.2%	**Caribbean**	42 – 0.2%
Marine	29 – 0.2%	**Energy**	43 – 0.2%	**Resources**	27 – 0.2%	**Oceania**	39 – 0.2%
Poorest	28 – 0.2%	**Marine**	42 – 0.2%	**Living**	25 – 0.2%	**Use**	37 – 0.2%
LDCS	28 – 0.2%	**Climate**	40 – 0.2%	**Health**	25 – 0.2%	**Income**	35 – 0.2%
Resources	27 – 0.2%	**Education**	38 – 0.2%	**Material**	25 – 0.2%	**Public**	34 – 0.2%
Education	26 – 0.2%	**Girls**	38 – 0.2%	**LDCS**	24 – 0.2%	**Policies**	33 – 0.2%
Food	26 – 0.2%	**Risk**	37 – 0.2%	**Support**	24 – 0.2%	**GDP**	33 – 0.2%

Source: authors' own processing.

Thus, a first insight that prevails is the high volume of *Asia* and *Africa* occurrences, a fact, however, that is due to the way the UN is mapping the regional groupings, with a higher segmentation in these two continents. Regarding the conceptual words, we can observe that in the first years, the focus was put on *children*, *women*, and *water access*, while some important other concepts also make an appearance: *sanitation*, *ODA*, *energy*, *marine*, and *education*.

When reaching the 2020-2023 period (Table 3), we can determine that *COVID* and the *pandemic* terms have naturally become one of the most used notions. Corroborated, we can see the *health* dimension as being prevalent, along with the *crisis* term. Interestingly, in 2023, the most used concept has become *climate*, while *energy* has also climbed some steps, a fact that could be interpreted as a new priority at the global level. Last but not least, *Ukraine* has been identified as entering the top 25 most common words in 2022, due to the aggression happening on its territory.

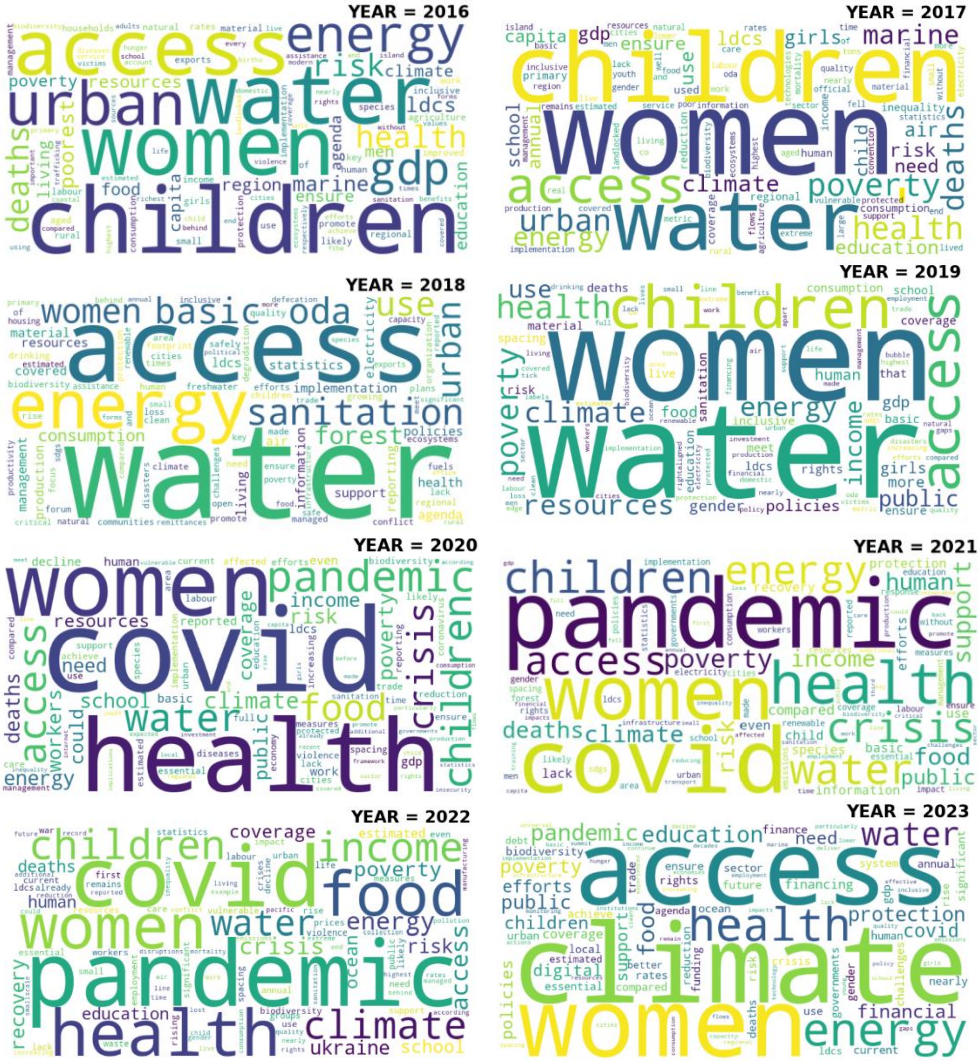
Table 3. Most common words – absolute and relative frequency – 2020-2023

2020		2021		2022		2023	
COVID	158 – 0.7%	**Pandemic**	168 – 0.8%	**Pandemic**	140 – 0.6%	**Asia**	158 - 0.6%
Asia	149 – 0.7%	**COVID**	152 – 0.7%	**Asia**	139 – 0.6%	**Africa**	131 - 0.5%
Africa	139 – 0.6%	**Asia**	143 – 0.7%	**COVID**	118 – 0.5%	**Climate**	110 - 0.4%
Health	117 – 0.5%	**Africa**	120 – 0.6%	**Africa**	104 – 0.5%	**Access**	102 - 0.4%
Women	92 – 0.4%	**Women**	103 – 0.5%	**America**	95 – 0.4%	**America**	99 - 0.4%
Pandemic	91 – 0.4%	**Health**	98 – 0.5%	**Health**	93 – 0.4%	**Women**	91 - 0.3%
America	88 – 0.4%	**America**	97 – 0.5%	**Women**	86 – 0.4%	**Energy**	87 - 0.3%
Subsaharan	79 – 0.4%	**Children**	72 – 0.3%	**Food**	83 – 0.4%	**Health**	81 - 0.3%
Children	78 – 0.4%	**Water**	66 – 0.3%	**Children**	77 – 0.4%	**Water**	76 - 0.3%
Water	73 – 0.3%	**Crisis**	66 – 0.3%	**Income**	77 – 0.4%	**Growth**	71 - 0.3%
Access	73 – 0.3%	**Subsaharan**	63 – 0.3%	**Climate**	71 – 0.3%	**Pandemic**	70 - 0.2%
Food	65 – 0.3%	**Energy**	60 – 0.3%	**Water**	59 – 0.3%	**Food**	69 - 0.2%
Crisis	61 – 0.3%	**Access**	60 – 0.3%	**Energy**	55 – 0.3%	**Subsaharan**	69 - 0.2%
Growth	60 – 0.3%	**Zealand**	51 – 0.2%	**Access**	53 – 0.2%	**Education**	67 - 0.2%
Climate	57 – 0.3%	**Poverty**	50 – 0.2%	**Crisis**	51 – 0.2%	**Protection**	67 - 0.2%
Poverty	48 – 0.2%	**Europe**	50 – 0.2%	**Growth**	51 – 0.2%	**Public**	62 - 0.2%
Latin	48 – 0.2%	**Australia**	49 – 0.2%	**Latin**	50 – 0.2%	**Poverty**	61 - 0.2%
Australia	47 – 0.2%	**Support**	46 – 0.2%	**Europe**	48 – 0.2%	**Financial**	57 - 0.2%
Zealand	46 – 0.2%	**Latin**	46 – 0.2%	**Caribbean**	48 – 0.2%	**Europe**	57 - 0.2%
Energy	45 – 0.2%	**Caribbean**	44 – 0.2%	**Subsaharan**	46 – 0.2%	**COVID**	55 - 0.2%
Risk	44 – 0.2%	**Climate**	42 – 0.2%	**Poverty**	44 – 0.2%	**Children**	53 - 0.2%
School	43 – 0.2%	**Income**	41 – 0.2%	**Zealand**	41 – 0.2%	**Efforts**	50 - 0.2%
Deaths	43 – 0.2%	**Food**	41 – 0.2%	**Australia**	40 – 0.2%	**Caribbean**	49 - 0.2%
Europe	43 – 0.2%	**Human**	39 – 0.2%	**Ukraine**	38 – 0.2%	**Latin**	48 - 0.2%
Caribbean	42 – 0.2%	**Risk**	37 – 0.2%	**Risk**	38 – 0.2%	**Policies**	47 - 0.2%

Source: authors' own processing.

Next, in order to better illustrate the most frequent sustainability topics, we have chosen to design several word clouds, one for each year of analysis. However, this time we chose to eliminate the geographical terms, so the representation could better profile the status quo. Thus, throughout Figure 1, we can see the confirmation of the insights generated before, namely the clear focus on *children*, *women*, and *water* elements in the period 2016-2019, followed by the appearance of *COVID*, *pandemic*, *health*, *energy*, and *climate* dimensions starting in 2020.

Figure 1. Wordcloud 2016 (left) – 2017 (right)



Source: authors' own processing.

Term Frequency-Inverse Document Frequency (TF-IDF)

Going forward, to study more in-depth the evolution of certain concepts throughout the years, we will proceed with the computation of the TF-IDF scores. Therefore, in Table 4 we can find the final scores for a selection of concepts related to sustainability. From an interpretation point of view, the words that show a higher value will be the more frequent ones in that year's report, however, rare across the entire corpus. The colored cells will show the peaks for each concept. Therefore, we can note that more recently, words such as *food*, *climate*, *education*, *digital*, *technology*, *crisis*, *recovery*, or *war* have increased their values, gaining more significance in the overall context.

Table 4. TF-IDF Scores – selection of words - 2016-2023

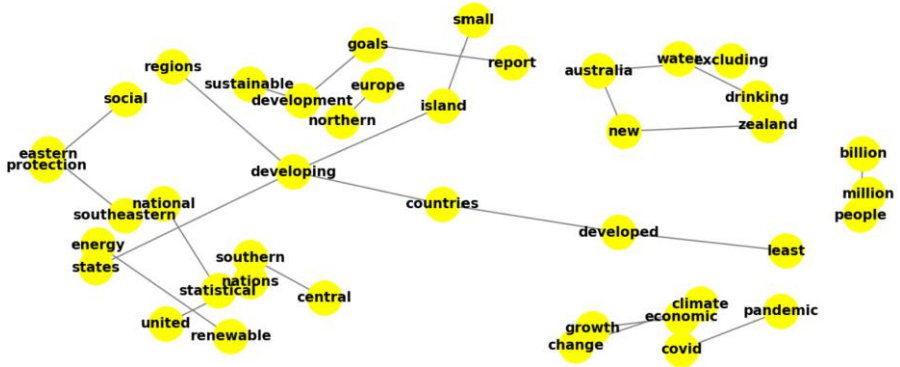
Term/Year	2016	2017	2018	2019	2020	2021	2022	2023
Children	0.14	**0.15**	0.06	0.14	0.12	0.11	0.12	0.07
Women	0.12	**0.17**	0.10	0.15	0.15	0.16	0.14	0.13
Water	0.12	0.11	**0.28**	0.16	0.12	0.10	0.09	0.11
Energy	0.09	0.07	**0.15**	0.09	0.07	0.09	0.09	0.12
Health	0.07	0.09	0.08	0.12	**0.19**	0.15	0.15	0.11
Deaths	0.07	**0.09**	0.02	0.05	0.07	0.06	0.06	0.05
Marine	**0.07**	**0.07**	0.03	0.03	0.03	0.02	0.03	0.04
Food	0.06	0.03	0.04	0.06	0.10	0.06	**0.13**	0.10
Pandemic	0.00	0.00	0.00	0.00	**0.23**	**0.41**	**0.36**	0.16
Climate	0.06	0.06	0.05	0.11	0.09	0.06	0.11	**0.16**
Education	0.06	0.06	0.03	0.05	0.04	0.04	0.06	**0.09**
Poverty	0.06	0.10	0.04	**0.11**	0.08	0.08	0.07	0.09
Digital	0.01	0.00	0.00	0.00	0.01	0.01	0.01	**0.07**
Technology	0.01	0.02	0.01	0.01	0.00	0.01	0.02	**0.04**
Ukraine	0.00	0.00	0.00	0.00	0.00	0.00	**0.13**	0.04
Income	0.04	0.04	0.03	0.07	0.06	0.06	**0.12**	0.04
Crisis	0.01	0.00	0.00	0.01	**0.11**	**0.11**	0.09	0.06
Recovery	0.00	0.00	0.00	0.00	0.03	0.06	**0.07**	0.04
War	0.00	0.00	0.00	0.00	0.01	0.00	**0.06**	0.02
Economy	0.01	0.00	0.01	0.01	**0.04**	0.03	0.01	0.01

Source: authors' own processing.

Collocations

Regarding the collocations, we decided to illustrate a Bigrams network graph, aimed at highlighting the most common linkages of words. Therefore, the computation has been done using the entire corpus of documents, since the yearly differences did not vary considerably. The result is visible in Figure 2.

Figure 2. Collocation Network Graph, the entire corpus



Source: authors' own processing.

In terms of interpretation, one can observe several clusters of words, for example, the geographical one, the one that links the development adjectives to certain spatial units (*states*, *countries*, *regions*), as well as several individual collocations (*covid-pandemic*, *renewable-energy*, *drinking-water*, *united-nations*).

4.2 Sentiment Analysis

Finally, regarding the computation of the sentiment analysis application, we have used the *VADER* lexicon in Python to identify the overall scores for each progress report. Thus, in Table 5 we can see sentiment scores for each year, along with a classification into positive, negative, and neutral words.

Table 5. Sentiment Scores and distribution of words

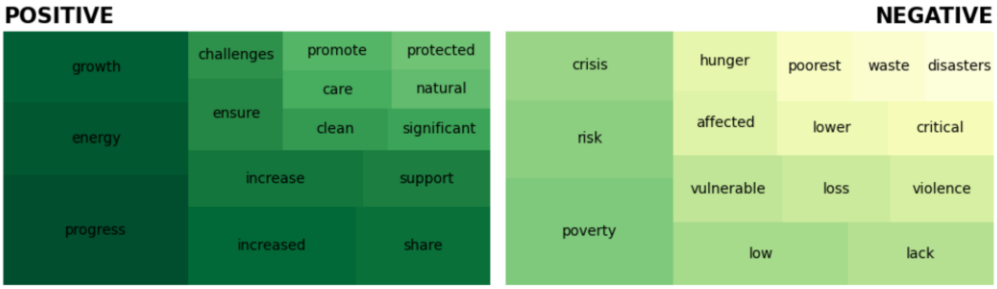
Year	Sentiment Scores	Number of positive words	Number of negative words	Number of neutral words
2023	314.5	2.161	1.096	23.415
2022	-23.2	1.319	1.156	18.031
2021	96.3	1.376	937	18.013
2020	61.5	1.421	1.063	17.716
2019	175.9	1.342	686	15.271
2018	176	939	377	10.580
2017	194.6	1.424	773	17.785
2016	138.7	915	442	11.227

Source: authors' own processing.

Therefore, what we can note is that most of the documents will have a generally positive tone of voice, with the exception of 2022, which will register the only negative value. This is explained by a variety of factors, correlated with the leitmotiv of the report, namely, the *road map out of crisis.* Thus, multiple factors have played a role in the overall negative valence, such as the pandemic, natural disasters, pollution, climate change, or the war context. What is interesting, however, is the fact that the number of negative words will be lower than the positive ones, so the explanation is coming from the intensity of the gloomy sentiment, with words such as *apocalyptic*, *catastrophe*, or *devastating* being between the most frequent terms. Nevertheless, the good news is coming from the 2023 publication, where we can observe the most positive sentiment score, a fact that is confirmed by the general objective of the report, namely the focus on the progress gained until now, and a hopeful plan for the future. To complete this insight, *success*, *rescue*, or *recovery* will be between the more positive words prevalent throughout the document.

Lastly, the sentiment exploration could be concluded by highlighting the most frequent positive and negative words in the collection of documents (Figure 3).

Figure 3. The most frequent positive and negative words in the entire corpus



Source: authors' own processing.

As a result, in Figure 3 we can see the top 15 positive and negative words, where the dimension of the figure represents the absolute frequency. Thus, in the positive category, we can see words such as *progress, increase, support,* or *care*, while in the negative group, we have found terms such as *poverty, risk, crisis,* and *vulnerable.*

5. Conclusions

In conclusion, we can consider the current approach as a starting point in understanding the dynamics shown in the SDG progress reports, one of the most conclusive publications for the international scene and the centrepiece of the sustainable development UN agenda. Through the use of both text mining and sentiment analysis, we have been able to explore some of the most important concepts of each year, starting from the children and women focus, and reaching towards climate, economic recovery, and building a more inclusive global community. At the same time, the urgency of action remains prevalent, with a general negative sentiment, however, recently sweetened by the progress registered at the half of the path towards 2030.

Bibliography

[1] Hwang, H., An, S., Lee, E., Han, S., Lee, C.H. (2021). Cross-societal analysis of climate change awareness and its relation to SDG 13: A knowledge synthesis from text mining. Sustainability, 13(10), 5596.

[2] Pernici, A., Stancu, S., Vulpe, M.I. (2023). Exploring green energy in economics: conceptual evolution. A literature review based on text mining and sentiment analysis. Revista Romana de Economie [Romanian Economics Journal], 57.

[3] Roy, A., Basu, A., Su, Y., Li, Y., Dong, X. (2022). Understanding recent trends in global sustainable development goal 6 research: scientometric, text mining and an improved framework for future research. Sustainability, 14(4), 2208.

[4] Shen, C.W., Luong, T.H., Pham, T. (2021). Exploration of social media opinions on innovation for sustainable development goals by topic modeling and sentiment analysis. In Research and Innovation Forum 2020: Disruptive Technologies in Times of Change (pp. 459-471). Springer International Publishing.

[5] UN DESA (2016). The Sustainable Development Goals Report 2016. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2016/.

[6] UN DESA (2017). The Sustainable Development Goals Report 2017. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2017/.

[7] UN DESA (2018). The Sustainable Development Goals Report 2018. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2018/.

[8] UN DESA (2019). The Sustainable Development Goals Report 2019. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2019/.

[9] UN DESA (2020). The Sustainable Development Goals Report 2020. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2020/.

[10] UN DESA (2021). The Sustainable Development Goals Report 2021. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2021/.

[11] UN DESA (2022). The Sustainable Development Goals Report 2022. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2022/.

[12] UN DESA (2023). The Sustainable Development Goals Report 2023. New York, USA. Retrieved from: https://unstats.un.org/sdgs/report/2023/.