

Received 20 December 2021; Revised 17 January 2022; Accepted 2 February 2022

DOI: 10.33119/EEIM.2021.62.5

Tóth, T.E. (2021). Analysis of the Twitter Discourse on Sustainability Using Natural Language Processing. *Education of Economists and Managers*, 62(4), 101–118.

Retrieved from: <https://econjournals.sgh.waw.pl/EEiM/article/view/2893>

---

# Analysis of the Twitter Discourse on Sustainability Using Natural Language Processing

Tímea Emese Tóth

*Department of Social Research Methodology, Faculty of Social Sciences,  
Eötvös Loránd University*

## Abstract

This publication aims to map the environmental sustainability discourse on Twitter. This will be achieved through two commonly used methods of natural language processing; topic modelling, which is used to uncover hidden themes in the document collection, and sentiment analysis, which is used to detect the attitudes of the authors of the text towards a particular attitude. The exploration of communication can provide an opportunity to find a solution to a multifaceted problem in order to protect our common future.

**Keywords:** sustainability, sustainable development goals, SGDs, natural language processing, topic modelling, sentiment analysis, NLP, discourse analysis, Python, Twitter

**JEL Classification Code:** C1

## Introduction

The aim of this article is to use computational text analytics to explore what typical attitudes can be detected in sustainability-related communication on Twitter. According to some experts, we are currently facing risks that make it impossible for people between the ages of 20 and 30 to live up to a peaceful old age on this planet without natural disasters, forest and bush fires, poverty, famine, epidemics, or wars over largely intact territories. The question is, why do we not act?

The period under study, 2007–2020, is not only of historical importance for environmental sustainability. With the rapid development of technology, the Internet, and social media, the everyday user has become a content producer, generating billions of pieces of textual content that reveal a lot about people's opinions, mindsets, attitudes, and actions. To research the textual content produced in this way, new disciplines have developed and come to the fore, such as the field of quantitative text analysis (Németh, Katona, & Kmetty, 2020), which allows the process, the discourse, to be studied as an external participant, without the respondent's compliance constraint. Those who use classical survey methods are increasingly running into obstacles that hinder the proper use of the method, for example, the tendency of non-response to increase or the aforementioned compliance constraint (Németh, Katona, & Kmetty, 2020) As I mentioned earlier, the method of quantitative text analysis, even if it is certain that it will never be equivalent to the survey method and will not replace it, makes discourse analysable as an external observer.

The analytical database consists of tweets that were generated on Earth Overshoot Day when we use up our annual energy reserves. An additional criterion was that only posts in English that contain at least one of the terms 'sustainable' and 'sustainability' were analysed. Two methods are used to explore the discourse: topic modelling using Latent Dirichlet Allocation to identify sub-topics in the discourse, and analysis using the VADER sentiment lexicon to identify attitudes and attitudes presented by posters of relevant tweets. If this exploration of communication is carried out, it may be possible to see the gaps in communication, the paths that have not yet been taken, and the paths that can be taken towards sustainable living.

## Theoretical background overview

Before presenting the current situation on environmental sustainability in detail, I reckon it is important to relate the Internet-based revolution discussed earlier to the professional context, thus validating the relevance of Internet sources in the next chapter.

At the level of everyday experience, we can observe that when we need information, we turn to the Internet first and foremost rather than going to a library or a newsagent for our journal. We simply often lack the time for anything else. So how can we expect organisations to deal with global problems to keep up with the rapidly changing situation, even on a daily basis, with new official publications on the state of play?

The primary aim of the Internet resources used in this chapter is, therefore, to show where more in-depth information and up-to-date data can be found nowadays (official websites are constantly updated between consecutive reports).

The second aim is to highlight that the scientific discourse on sustainability is not always available in the form of publications. Science also has to adapt to the information habits of an accelerated world, especially if it is to make a real difference. And this is an issue so central from the viewpoint of sustainability that it cannot be ignored.

## Description of the current situation

A mass extinction event is defined as a large-scale and sudden loss of a species, trait, or characteristic of one of the billions of species around us. According to experts on the history of the Earth, humanity is on the verge of the sixth extinction event.

The Intergovernmental Panel on Climate Change (IPCC)<sup>1</sup> reports in its Fifth Assessment, published in 2014, that the human impact on climate is clear and that, if it continues to increase, there will be severe and irreversible impacts (IPCC, 2020). Since then, special reports<sup>2</sup> have shown that biodiversity is on an accelerating path to extinction as the planet's resilience declines,<sup>3</sup> despite a series of agreements to avoid climate catastrophe.<sup>4</sup>

In their research summary book *Big World, Small Planet*, published in 2015, Johan Rockström and Matthias Klum describe the components that influence the Earth's ecosystems.

For each of these, a planetary limit has been set, which, if exceeded, will upset the planet's life-support balance and threaten the very existence of our own species, but if we stay within the limits, our existence can remain sustainable (Rockström & Klum, 2015).

---

<sup>1</sup> In 1988 it was established by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) to provide political leaders with regular scientific assessments.

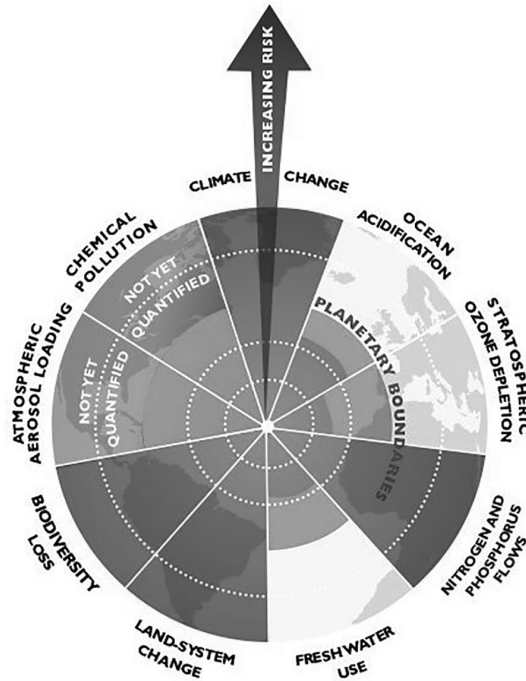
<sup>2</sup> <https://www.ipcc.ch/reports/>

<sup>3</sup> <https://www.overshootday.org/>

<sup>4</sup> <https://www.consilium.europa.eu/en/policies/climate-change/paris-agreement/>

The elements included were climate change, ocean acidification, stratospheric ozone depletion, nitrogen and phosphorus flows, fresh-water use, land-system change, biodiversity loss, atmospheric aerosol loading chemical pollution.

Figure 1. Planetary boundaries



Source: Finnish Environment Institute.

## Opportunities for change

Fortunately, a sustainable future is not lost yet, as a group of researchers and scientists have outlined a number of possible ways to achieve sustainability goals.

Economist Kate Raworth, in her book *The Doughnut Economics: 7 Ways to Think Like a 21<sup>st</sup> Century Economist* (2017), has created a model based on the work of Rockström and Klum, which includes the steps needed to live sustainably. The model is called 'The Doughnut of social and planetary boundaries'. The goals and instruments set out in this model are very similar to the UN Sustainable Development Goals (SDGs), which are based on the 2030 targets adopted in 2015.<sup>5</sup> In the paper,

<sup>5</sup> <https://sdg-tracker.org/>

Raworth presents seven new ways to change fundamentally economic thinking that is key to the future, also because the actions of its actors have a significant weight in both energy use and the choice of how it is delivered (Raworth, 2017). A key question in examining the process of human population is whether we will reach the plateau of sustainable living referred to in the demographic context, where we can simultaneously achieve profit, human well-being, and the good of the Earth. To achieve the goals, we would have to maintain the living standards of rich countries by reducing their ecological footprint and increase those of poor countries, thus accelerating their demographic transition by maintaining their ecological footprint. To tackle these challenges, we must both question the indicators that express the well-being of a country (Dasgupta & McKenzie, 2020), and consider whether growth is indeed a fundamental element of economic and political success. The question is: can economic growth be green? The answer is to be found in a multivariate equation, which is currently being solved (Drews & Antal, 2016; Drews & Reese, 2018).

The IEA analysis<sup>6</sup> published on 2 March 2021 shows that global carbon dioxide emissions in December 2020 were 2%, or 60 million tonnes, higher than in the same month the previous year. On these trends, we can expect an average temperature increase of 2.5–2.8°C by 2100, well above the 1.5–2°C limit set in the Paris Agreement.<sup>7</sup> Achieving this objective would require, first and foremost, a change in energy sources towards clean energy. This could reduce further greenhouse gas emissions and slow down ocean acidification. While the issue of energy storage remains unresolved, a series of efforts to promote the use of solar, wind, hydro, and geothermal energy is a sure way forward (Attenborough & Hughes, 2020).

A priority to protect our waters would be to further increase the extent of marine protected areas and to develop a sustainable fisheries strategy, which the UN and the World Trade Organisation are currently working on.<sup>8/9</sup>

The protection of forests and land should also aim at restoring the biodiversity of the population of the area, i.e., reforesting the area. The main means of achieving this are to maintain existing forests,<sup>10</sup> reduce logging, afforest agricultural land (Alexandratos & Bruinsma, 2020; FABLE, 2020). It would also be important to reduce soil-damaging operations, reform agricultural practices,<sup>11/12</sup> and change nutritional habits.<sup>13</sup>

---

<sup>6</sup> <https://www.iea.org/news/after-steep-drop-in-early-2020-global-carbon-dioxide-emissions-have-rebounded-strongly>

<sup>7</sup> <https://ourworldindata.org/future-emissions>

<sup>8</sup> <https://www.thebluecarboninitiative.org/>

<sup>9</sup> <https://www.protectedplanet.net/en/thematic-areas/marine-protected-areas>

<sup>10</sup> <https://www.un-redd.org/>

<sup>11</sup> <https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Programmes/Green-Cities/Urban-food-systems.htm>

<sup>12</sup> <https://www.rethinkx.com/food-and-agriculture-executive-summary>

<sup>13</sup> <https://www.pnas.org/content/early/2016/03/16/1523119113>

## Theoretical overview of the methodology

### Text data generation and the amount of information

With the rapid development of technology and the emergence of the Internet and social media, the ordinary user has become a content producer themselves, with the opportunity to actively comment, share, and react to content on the Internet. As a result, there are now billions of such data, which tell us a lot about people's opinions, mindsets, and attitudes (Evans-Acaves, 2016). In numerical terms, thanks to the Big Data revolution, 4.5 billion people will have used the Internet by the beginning of 2020, 3.8 billion of whom will have been members of a social media site, according to a report published by Digital Reports.<sup>14</sup> One of the most popular platforms is Twitter, where an average of 6,000 tweets are posted every second, which means 350,000 tweets per minute and 500 million per day.<sup>15</sup>

Processing texts and extracting information from them is not too challenging for the average user, given that he or she only deals with a small detail. However, comprehensively reviewing and processing such a large amount of data can be time-consuming and challenging with manual processing methods, but with algorithmic solutions this is no longer an obstacle.

### Topic modelling

The method aims to identify the themes of a collection of documents. The method assumes that there are words with strong semantic information and that documents on similar topics use similar groups of words. One of the best-known topic modelling techniques is the Latent Dirichlet Allocation (LDA), first described by Blei et al. in 2003 (Blei et al, 2003). LDA performance can be measured by, among other things, log likelihood, or log likelihood-based metrics such as perplexity, most often on the part of the corpus retained for testing, the test corpus (Blei et al, 2003). With the topical coherence indicator, the interpretability of the topics, the internal coherence, can be examined by considering the semantic similarity of the terms that characterise the topics (Stevens et al., 2012). Among the coherence indicators, I use the *c\_v* indicator, which is based on the co-occurrence of word pairs and can take a value between 0 and 1. Roughly speaking, we can say that the coherence of a model

---

<sup>14</sup> <https://datareportal.com/reports/digital-2020-global-digital-overview>

<sup>15</sup> <https://www.internetlivestats.com/twitter-statistics/>

is considered poor up to 0.0–0.3, low between 0.31–0.4, acceptable between 0.41–0.6, good between 0.61–0.8, and poor again above that. In practice, I perform the topology modelling using the Gensim package<sup>16</sup> implemented in the Python library.

## Sentiment analysis

The aim of sentiment analysis is to find out people's opinions, attitudes, and even feelings about a particular issue, based on texts alone. In practice, I carried out sentiment analysis using the VADER programme. It is a special lexicon containing 9,000 lexical features, including words, emoticons, abbreviations, and slang expressing emotion. VADER does not calculate the polarity and intensity of words separately but aggregates the results in an indicator called a normalised 'valence-score'. For text fragments, it calculates the proportion of negative/neutral/positive terms in the fragment, considering the intensities, and then normalises it to a range from -1 to 1. VADER performs quite accurately when compared to other competitor tools and by itself, coming first in the evaluation of tweets, both when considering human annotators and other software. In the other tests, only the human annotator scored higher in accuracy (Hutto & Gilbert, 2015).

It is important to note that I do not treat the results of the sentiment analysis as an indication of attitudes towards the subject, but I interpret them according to the typical attitudes (positive-negative-neutral) presented by the texts on the subject.

## Advantages and disadvantages of the method used

Obviously, as with all methods, there are advantages and disadvantages. I mentioned the advantages in the introduction and referred to them before, but briefly, once again I would like to point out that it is that algorithmic processing of the text allows us to analyse large amounts of data quickly without direct intervention. Its drawbacks include, among others, as compared to classical surveys, coverage errors and representativeness issues. Not everyone is represented in each community space, which further leads to the fact that we cannot analyse everyone's opinion based on different inclusion probabilities. Several reasons for the lack of representation of subjects in a given space can be listed, which can be explored either by examining classical sociodemographic variables or by examining the temporal dynamics of community surfaces. Another important methodological issue is

---

<sup>16</sup> <https://pypi.org/project/gensim/>

the comparability of the results of different publications and the extent to which the results can be generalised. For such research, again in a similar way to survey methods, a sampling design should be developed. Here, too, we need to select the platforms, the texts that we consider relevant (this is often filtered by search terms), we need to consider the temporal aspects, we can examine the relative trends in social media use and the possibilities for analysis, which social media platforms are more popular in which period, which ones have the most privacy and data protection and data request possibilities. In addition, there is still the issue of tuning the parameterisation of the algorithms to fit the research topic, which can show quite a lot of variability. With all this, I do not want to emphasise that the results of the method are not suitable for comparison with the results of other studies, but only to point out the circumstances in which they occur. In this research, I have taken as a basis several studies on similar topics, e.g., Kirilenko & Stepchenkova, 2014; Cody et al., 2015; Nik-Bakht & El-Diraby 2016; Dahal & Kumar, Li, 2019; Ballestar, Cuervo-Mir, & Freire-Rubio, 2020; Radi & Shokouhyar, 2020; however, I do not aim to compare the results due to its multiple aspects.

## Description of the research

### Data collection

My analytical database consists of scraped tweets from Earth Overshoot Day, from the creation of Twitter. Criteria included that the tweets contain at least one of the words: ‘sustainable’ and ‘sustainability’. The scraping was done using the Twint package, which can be implemented in Python. Both in selecting the time period and in finding relevant search terms I took into account the results of previous research on a similar topic (Kirilenko & Stepchenkova, 2014; Cody et al., 2015; Nik-Bakht & El-Diraby 2016; Dahal, Kumar, & Li, 2019; Ballestar, Cuervo-Mir, & Freire-Rubio, 2020; Radi & Shokouhyar, 2020). Retweets are not included in the database, which contains all the relevant posts of the whole day, 61,083 in total.

**Table 1. Database Structure**

Year	Word Overshoot Day	Number of tweets	Unit of analysis
2006	2006.08.19	no data	-
2007	2007.08.17	3	1
2008	2008.08.14	49	1
2009	2009.08.18	1,587	1



Year	Word Overshoot Day	Number of tweets	Unit of analysis
2010	2010.08.07	1,329	1
2011	2011.08.04	4,457	2
2012	2012.08.04	2,923	2
2013	2013.08.03	2,537	2
2014	2014.08.04	5,920	2
2015	2015.08.05	8,491	2
2016	2016.08.05	6,168	3
2017	2017.08.01	6,538	3
2018	2018.07.29	4,183	3
2019	2019.07.29	9,543	3
2020	2020.08.22	7,355	3

Source: own work.

The categorisation is based on David Attenborough’s book *A Life on Our Planet* and the Paris Agreement. Attenborough identified an important phase from 1997–2011 and then another phase from 2011–2020, but in order to carry out the analysis with a more balanced number of cases at even intervals, I split the second phase into two parts based on the Paris Agreement’s declaration of 2015.

### Steps of the analysis

In the first step, to make the data usable, I perform a data cleansing process, removing variables that I will not use later but that ‘go with’ the tweets while scraping, and the text that is otherwise inappropriate, such as only spaces. On the other hand, I convert the appropriate texts to the appropriate form so that the algorithm can interpret them (Németh, Katona, & Kmetty, 2020; Denny & Spirling, 2018). Then, in the second part of the analysis, I use the Gensim package to find the best fitting and meaningful topics for each section, using both qualitative text processing and algorithmic methods. For the algorithmic method, I optimise the parameters between 3 and 11 using the literature. In the third section, I carry out the sentiment analysis of the subsamples. My hypothesis is that I will find predominantly neutral and positive texts, on the one hand, because emphasising negatives in the sustainability discourse can lead to counterproductivity, and on the other hand, because social networking sites are characterised by the positivist approach. The results will be presented section by section, using a combination of methods.

## Results

### The first analysis phase (2007–2010)

After completing the pre-processing, a total of 2,938 tweets were included in this section. The most frequently used words, regardless of the topic, were *'green'*, *'energy'*, *'design'*, *'pioneer'*, *'live'*, *'food'*, *'lifestyle'*, *'new'*, *'business'*, *'make'*, which are mostly related to the general lifestyle in line with the environment. For this section, **50.27% neutral, 42.95% positive, and 6.78% negative** sentiment tweets were produced. For the tweets in the 2007–2010 range, a total of 6 different and distinct topics were found, with a good fit.

Topics three and six are quite close, which may be justified by the theory that well-chosen, quality products result in significant energy savings due to longer life. For all the themes, tweets reflecting negative attitudes were the least represented, with only the third and sixth, i.e., the connected themes, showing examples of tweets with positive attitudes being more represented than neutral ones. The most positive topics are related to agriculture and food, and the most negative are related to sustainable homes. In general, the mood of the topics is no different. The most common words that make up the topics tend to refer more to the general lifestyle of individuals and less to reflection on the global situation.

### The second analysis phase (2011–2015)

A total of 24,090 tweets were included in the analysis phase from 2011 to 2015, which is almost ten times the number of tweets included in the first part. A total of 5 topics were created, similar to the previous method, also with acceptable topic coherence. The 10 most frequent words, regardless of the topic were: *'development'*, *'AMP'*,<sup>17</sup> *'food'*, *'make'*, *'business'*, *'CSR'*,<sup>18</sup> *'green'*, *'need'*, *'environment'*, *'future'*. In contrast to the previous survey, social responsibility and common sustainability efforts were given greater emphasis, with **48.45% neutral and 44.54% positive, 7% negative**.

---

<sup>17</sup> AMP is a company dedicated to the development and management of renewable energy infrastructure, based in Canada, founded in 2009.

<sup>18</sup> CSR is corporate social responsibility.

**Table 2. Subtopics and sentiment analysis of tweets from 2007–2010**

Topic number	1 <sup>st</sup> topic	2 <sup>nd</sup> topic	3 <sup>rd</sup> topic	4 <sup>th</sup> topic	5 <sup>th</sup> topic	6 <sup>th</sup> topic
Item number	503	513	933	383	231	375
Most common words	green, live, building, come, future, great, event, use, eat, market	food, lifestyle, technology, agriculture, local, way, world, open, city, project	energy, new, job, word, find, global, solution, become, producer, blog, post	business, water, development, com, get, source, management, work, look, help	design, make, good, build, home, need, report, take, programme, think	pioneer, community, organic, go, check, product, time, consider, life, social
Main topic	Green buildings, markets	Agriculture, food	Energy using	Business development, management	Green, sustainable home	Organic, own products, thoughtful shopping
Senti-ment of the topic	Negative: 2.98% Neutral: 46.72% Positive: 50.30%	Negative: 3.70% Neutral: 43.85% Positive: 52.45%	Negative: 9.75% Neutral: 54.34% Positive: 35.91%	Negative: 4.44% Neutral: 49.87% Positive: 45.69%	Negative: 9.96% Neutral: 50.22% Positive: 39.82%	Negative: 9.07% Neutral: 54.13% Positive: 36.8%

Source: own work.

**Table 3. Subtopics and sentiment analysis of tweets from 2011–2015**

Topic number	1 <sup>st</sup> topic	2 <sup>nd</sup> topic	3 <sup>rd</sup> topic	4 <sup>th</sup> topic	5 <sup>th</sup> topic
Item number	5,328	4,594	4,447	4,721	5,000
Most common words	make, environment, goal, work, use, way, thank, good, solution, live	development, amp, need, take, future, late, build, great, help, seafood	green, energy, global, water, go, report, step, city, building, new	csr, change, world, target, new, lead, think, fashion, today, products_ mainstream	food, business, get, education, cattle, RTTC,19 job, company, travel, value
Main topic	Environmental solutions	Development needs	Greening, development	Mindset	Consumption, travel
Senti-ment of the topic	Negative: 7.02% Neutral: 44.01% Positive: 48.97%	Negative: 8.16% Neutral: 49.96% Positive: 41.88%	Negative: 5.87% Neutral: 46.32% Positive: 47.81%	Negative: 7.46% Neutral: 49.88% Positive: 42.66%	Negative: 6.5% Neutral: 55.54% Positive: 37.96%

Source: own work.

In this case, the 3/4/5 topics are very close to each other, and their keywords include reflecting pairs of words, such as SCR (corporate social responsibility) or RTTC (responsible travel and tourism collective). Collective action and thinking are also reflected in the texts, which at this stage are mainly focused on the improvement of buildings and other areas, responsible travel, and the discovery of local values. More emphasis is given to reports on the environment, on specific objectives. Environmental and sustainability values are also infiltrating the life of companies and institutions, and consumption trends are also starting to shift towards more sustainable options. The period ends with the Paris Agreement, which further reinforces the importance of collective responsibility and global trends towards cooperation.

Taking sentiment into account, tweets with negative attitudes are also the smallest proportion, but the overall proportion is smaller than in the previous section. In two out of five cases, the majority of positive tweets outweigh the neutral ones, as can be seen in the case of the environmental solutions and the greening/enhancement of buildings. The positive-neutral categories are also more distinct. The most negative topics are related to development and the most positive ones to environmental solutions.

<sup>19</sup> RTTC stands for responsible travel and tourism collective.

### The third analysis phase (2016–2020)

In this section, I examined a total of 33,464 tweets. I created 7 topics with an acceptable topic coherence of 0.42. The ten most frequent terms in the corpus are: ‘make’, ‘solution’, ‘AMP’, ‘need’, ‘go’, ‘learn’, ‘would’, ‘think’, ‘development’, ‘future’. These terms describe the phenomenon whereby environmentally conscious actors act with their own life path and future in mind and encourage others to do the same. This is also supported by the topics. The period covered is one in which people are beginning to realise that our lack of action on sustainability is putting our own future at risk. The proportions of sentiments across the whole period were **51.7% positive, 38.98% neutral and 9.32% negative**. As in the past, there are also some topics that are close to each other. There are two such groups: one is a group of 1–2 topics, the other is a grouping of 4/5/6/7 members.

Let us now pay more attention to the separate topic with 1,950 tweets. The importance of this topic lies in the fact that it summarises a hitherto unrepresented part of the environmental movement, namely ethical clothing and packaging and related issues. Its relevance is underlined by the renaissance of the second-hand clothing market, the thrift shop, observed in the last few years. There are more and more posts explaining that wearing second-hand goods is not about a lack of style, but about the fact that it is possible to dress in an environmentally responsible and fashionable way. The choice of apparel nowadays also takes into account the place of production, the use of water and organic ingredients.

Another major part of this is packaging. Consumers prefer environmentally friendly packaging (paper, glass, metal) or packaging-free alternatives to the plastic packaging they used to use, and not just in fashion. Importantly, this phase has seen the emergence of explicitly topical terms that have been identified by researchers as key concepts, such as ‘climate change’ or ‘move the date’, a key term in the movement to delay the Beyond Day. In addition, there are other topics related to food, environmental movements, economics, and mindsets. The most positive topics are those related to environmental communities and projects, and the most negative are those related to mindsets.

**Table 4. Subtopics and sentiment analysis of tweets from 2016–2020**

Topic number	1 <sup>st</sup> topic	2 <sup>nd</sup> topic	3 <sup>rd</sup> topic	4 <sup>th</sup> topic	5 <sup>th</sup> topic	6 <sup>th</sup> topic	7 <sup>th</sup> topic
Item number	11,399	13,296	2,646	1,950	1,773	1,113	1,287
Most common words	need, make, go, would, way, think, people, good, get, work	amp, solution, learn, business, help, new, build, environmental, community, project	check, product, fashion, ethical, available, packaging, thank, live_good, organic, live good	energy, water, reduce, plant, climate change, power, clean, renewable, farmer, tree	development, future, global, week, role, shri_discuss, login_detail, goal, pm_list, economic	year, world, day, planet, today, earth, resource, nature, movethedate, together	food, system, enter, (sweepstakes_ beautyofcarpet), farm, production, win, feed, learn, course
Main topic	Mindset	Environmental communities, projects	Sustainable fashion, environmentally conscious packaging	Climate change, renewable energy	Global economy, development	Overshoot day	Food, nutrition
Sentiment of the topic	Negative: 17.43% Neutral: 30.75% Positive: 51.82%	Negative: 3.97% Neutral: 39.1% Positive: 56.93%	Negative: 5.82% Neutral: 39.27% Positive: 54.91%	Negative: 9.64% Neutral: 47.85% Positive: 42.51%	Negative: 4.91% Neutral: 58.32% Positive: 36.77%	Negative: 6.74% Neutral: 58.31% Positive: 34.95%	Negative: 7.69% Neutral: 53.38% Positive: 38.93%

Source: own work.

## Conclusion

The aim of this paper was to map the environmental sustainability discourse on Twitter using two commonly used tools of natural language processing: topic modelling and sentiment analysis. The analysis dataset consisted of tweets that were posted on Earth Overshoot Day.

One of the methods used to carry out the research was topical modelling. In this thesis, a model with appropriate topical coherence was fitted to each of the three sub-databases. Based on the results, it can be said that the theme of the tweets is mostly focused on green economy, energy, and sustainable agriculture options for global issues. The texts dealing with issues at the individual level are organised around the themes of environmental communities, responsible product purchase and use, and mindsets. The other method used was sentiment analysis. By examining the sentiment of tweets it can be concluded that over time the sentiment of tweets undergoes a significant polarisation. The proportion of positive sentiments decreases by almost 10% and the proportion of negative sentiments increases by 3%. This is partly due to the worsening situation and partly due to the erosion of the positivist outlook on social media platforms. The most positive topics are about environmental communities, projects, and solutions, while the most negative ones are about development and mindsets. Unfortunately, the tweets under review do not deal at all or only sporadically with the situation of our salt and fresh waters, forests, air and chemical pollution, and biodiversity loss in general. The SDGs are numbered 6, 14, 15. So these are the areas where education is still needed, in addition to informing people as widely as possible about the interrelationships between the sub-areas. Emphasising the general context is also important because improvements in some areas, particularly technological ones, may reduce the threat of climate change but exacerbate other problems. What we clearly need now is to find comprehensive solutions together to protect our common future. In the future it would certainly be worthwhile to carry out similar research on other platforms, either by defining search terms more specifically or by selecting a longer, continuous period. It may also be interesting to repeat previous research while ensuring comparability.

## Aknowledgements

I would like to take this opportunity to thank my former supervisor, Márton Rakovics, without whom the basis for this work would not have been possible. I would also like to thank Renáta Németh, for her current professional support I cannot be grateful enough.

This paper also supported by the ÚNKP-21–3 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund and non-financed supported by NKFIH K-134428.

## References

- Attenborough, D., & Hughes, J. (2020). *A Life on Our Planet: My Witness Statement and a Vision for the Future*. London: Ebury Publishing.
- Alexandratos, N., & Bruinsma, J. (2012). *World agriculture towards 2030/2050: the 2012 revision*. ESA Working paper No. 12–03. Rome, FAO. Retrieved from: <http://www.fao.org/3/ap106e/ap106e.pdf> (accessed: 29.03.2021).
- Ballestar, M.T., Cuerdo-Mir, M., & Freire-Rubio, M.T. (2020). The Concept of Sustainability on Social Media: A Social Listening Approach. *Sustainability*, 12(5), 2122. MDPI AG. <http://dx.doi.org/10.3390/su12052122>
- Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Cody, E.M., Reagan, A.J., Mitchell, L., Dodds, P.S., & Danforth, C.M. (2015). Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll. *PLoS ONE*, 10(8), e0136092. DOI:10.1371/journal.pone.0136092
- Dahal, B., Kumar, S.A. P., & Li, Z. (2019). Topic modelling and sentiment analysis of global climate change tweets. *Soc. Netw. Anal. Min.*, 9(24) <https://doi.org/10.1007/s13278-019-0568-8>
- Dasgupta, P., & McKenzie, E. (2020). *The Dasgupta Review – Independent Review on the Economics of Biodiversity Interim Report*. Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/882222/The\\_Economics\\_of\\_Biodiversity\\_The\\_Dasgupta\\_Review\\_Interim\\_Report.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/882222/The_Economics_of_Biodiversity_The_Dasgupta_Review_Interim_Report.pdf) (accessed: 28.01.2021).
- Denny, M., & Spirling, A. (2018). Text Preprocessing or Unsupervised Learning: Why It Matters, When It Misleads, And What to Do About It. *Political Analysis*, 26(2), 168–189. DOI:10.1017/pan.2017.44
- Drews, S., & Miklós, A. (2016). Degrowth: A “Missile Word” That Backfires? *Ecological Economics*, 126, 182–187. DOI: 10.1016/j.ecolecon.2016.04.001



- Drews, S., & Reese, G. (2018). “Degrowth” vs. Other Types of Growth: Labelling Affects Emotions but Not Attitudes”. *Environmental Communication*, 12(6), 763–772. DOI: 10.1080/17524032.2018.1472127
- Evans, J.A., & Aceves, P. (2016). Machine translation: mining text for social theory. *Annu.Rev. Sociol.*, 42(1), 21–50. <https://doi.org/10.1146/annurev-soc-081715-074206>
- FABLE (2020). *Pathways to Sustainable Land-Use and Food Systems*. 2020 Report of the FABLE Consortium. Laxenburg and Paris: International Institute for Applied Systems Analysis (IIASA) and Sustainable Development Solutions Network (SDSN). <https://doi.org/10.22022/ESM/12-2020.16896>
- Furman, E., Häyhä, T., & Hirvilammi, T. (2018) A future the planet can accommodate. *SKYKE & SITRA*, 2. ISBN: 978-952-11-4938-2. Retrieved from: [https://helda.helsinki.fi/bitstream/handle/10138/235418/PB\\_A-future-the-planet-can-accommodate.pdf?sequence=1](https://helda.helsinki.fi/bitstream/handle/10138/235418/PB_A-future-the-planet-can-accommodate.pdf?sequence=1) (accessed: 08.03.2021).
- Hutto, C.J., & Gilbert, E. (2015). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. Proceedings of the 8<sup>th</sup> International Conference on Weblogs and Social Media, ICWSM 2014.
- Intergovernmental Panel on Climate Change – The Fifth and the Sixth Assessment Cycle (2020). Retrieved from: [https://www.ipcc.ch/site/assets/uploads/2020/05/2020-AC6\\_en.pdf](https://www.ipcc.ch/site/assets/uploads/2020/05/2020-AC6_en.pdf) (accessed: 06.03.2021).
- Kirilenko, A.P., & Stepchenkova, S.O. (2014). Public microblogging on climate change: One Year off Twitter worldwide. *Global Environ. Change*. <https://dx.doi.org/10.1016/j.gloenvcha.2014.02.008>
- Németh, R., Katona, E.R., & Kmetty, Z. (2020). Az automatizált szövegelemzés perspektívája a társadalomtudományokban [Perspectives on automated text analysis in the social sciences]. *SZOCIOLÓGIAI SZEMLE*, 30(1), 44–62. ISSN 1216–2051.
- Nik-Bakht, M., & El-Diraby, T.E. (2016). Sus-tweet-ability: Exposing public communities perspective on sustainability of urban infrastructure through online social media. *International Journal of Human-Computer Studies*, 89, 54–72, ISSN 1071–5819. <https://doi.org/10.1016/j.ijhcs.2015.11.002>
- Radi, S.A., & Shokouhyar, S. (2021). Toward consumer perception of cellphones sustainability: A social media analytics. *Sustainable Production and Consumption*, 25, 217–233, ISSN 2352–5509. <https://doi.org/10.1016/j.spc.2020.08.012>
- Raworth, K. (2017). *Doughnut economics: seven ways to think like a 21<sup>st</sup>-century economist*. London: Random House.
- Rockström, J., & Klum, M. (2015). *Big world, small planet: Abundance within planetary boundaries*. USA: Yale University Press.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). *Exploring topic coherence over many models and many topics*. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL ‘12). Association for Computational Linguistics, USA, 952–961.

**Online sources:**

<https://www.consilium.europa.eu/en/policies/climate-change/paris-agreement/>  
<https://datareportal.com/reports/digital-2020-global-digital-overview>  
<https://www.iea.org/news/after-steep-drop-in-early-2020-global-carbon-dioxide-emissions>-<https://www.internetlivestats.com/twitter-statistics/>  
<https://www.ipcc.ch/reports/>  
<https://ourworldindata.org/future-emissions>  
<https://www.overshootday.org/>  
<https://www.pnas.org/content/early/2016/03/16/1523119113>  
<https://www.protectedplanet.net/en/thematic-areas/marine-protected-areas>  
<https://pypi.org/project/gensim/have-rebounded-strongly>  
<Research/Programmes/Green-Cities/Urban-food-systems.htm>  
<https://www.rethinkx.com/food-and-agriculture-executive-summary>  
<https://sdg-tracker.org/>  
<https://www.thebluecarboninitiative.org/>  
<https://www.un-redd.org/>  
<https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental->

**Tímea Emese Tóth**

She studied Sociology at the Károli Gáspár University of the Reformed Church in Budapest, and completed her Master's degree in Survey Statistics at the Eötvös Loránd University (ELTE). She is currently a lecturer at the Department of Social Research Methodology (ELTE) and a researcher at the Research Center for Computational Social Science (ELTE). Her research is closely related to environmental sustainability, using survey data analysis, and natural language processing as methodological approaches.

e-mail: [toth.emese@tatk.elte.hu](mailto:toth.emese@tatk.elte.hu)

ORCID: 0000-0002-3584-118X