# Climate crises are real: People's perspectives changing over time

s1073016
Radboud University

## ABSTRACT

The topic of discussion regarding climate change has always been limited to the environmentalist or climate activist. However, with the increase in temperature and unannounced disastrous events in past few years have made people alert about the the consequences of climate change. People, today have started showing their concerns about these crises on various social media platforms. Sentiment analysis is the best way to understand the opinions of people through text and Topic Modeling helps in knowing people are concerned about what. In this paper, tweets related to climate change were extracted and sentiments were generated accordingly. Furthermore, using LDA model, more insights about what kind of tweets people are sharing were studied. The analytical output of above mention approach will give more knowledge about how people's perspective regarding climatic change has been evolved and what factors alleviated these changes.

## 1 INTRODUCTION

Climate change has been making the planet warmer with each passing day causing natural events such as floods and cyclones which were not anticipated. Scientist and researchers has already foreseen the effect of global warming and has proposed a lot of work and models to minimise the adverse effect of it. However, even with the scientific findings, people have always overlooked this vital subject of matter, considering it as politically polarised topic. An opinion article in New York Times back [1] in 2014 claimed a bold statement: "We're much more likely to believe that there are signs that aliens have visited Earth (77 percent) than that humans are causing climate change (44 percent).". Nonetheless, people have started raising concerns towards climate change issues in recent years. Therefore, many work has been proposed for tracking the perspectives of people by either designing a manual surveys or using social media data. While, designing and implementing manual surveys cost immense amount of resources and time, a lot researchers prefer social media analysis.

Micro-blogging sites like Twitter and Reddit has turned the Web into a substantial ledger of comments on a huge variety of topics, making them a potential source of information for all kinds of research areas. This paper focuses on Twitter data, since its allows its users to generate tweets that are short messages upto 140 to 280 words. Also, Twitter is one of best platform, where people can share their emotions and thoughts in short sentences which helps in performing sentiment analysis.

Being a part of social media analysis, sentiment analysis helps in identifying and extracting the subjective of data along with assessing it. Adding to this, sentiment analysis also provides the ability to understand and interpret the perspectives and feelings people. With the help of sentiment analysis, public awareness and empathy towards climate change can be studied. Another important part of social media analysis is topic modeling, which provides more perceptive about the topics, that people are discussing about more. Using both sentiment analysis and topic modeling, researchers and policy maker can understand, what kind of events or things people are concerned about the most

This paper performs sentiment analysis on the Twitter data that focuses on climate change tweets, to answer the research question: "How people's opinion about climate change and global warming has been changed over the time?". Along with this, the proposed approach also gives insights about the topics people are discussing when they post positive or negative tweets about climate change. Lastly, the paper also depicts the factors that triggered the changes in people's perspectives.

## 2 RELATED WORK

Many prior work has been done by researchers to tackle climate change using machine learning models and AI systems. One such work by David Rolnick and his team [7] where they have discussed potentials of using machine learning techniques to mitigate climate change. However, scientific approaches alone are not enough, since even though it is a natural effect, we humans play a major role. Therefore, understanding people's opinion is equally important and social media analysis is way to understand their thoughts. Russell and his team [8] performed social media analysis, specifically Twitter and found that, people's emotions, attitude and behaviour can be studied by analysing their tweets. This opened a new field of interest Sentiment Analysis for the researchers. Patrick Baylis [2] in his study drew a correlation between the effects of change in temperature and sentiments of Twitter users. Another work has been proposed by Allison Koenecke and her team[5] where they examined the change in climate change sentiments in response to five different natrual disasters in United States, by studying Twitter data. Along with sentiment analysis, researcher have also tried searching for trends regarding climate changes from Twitter discussions. One such work has been proposed in this [9] paper, where the team perform topic modeling on to analyse the common topic about climate change that people discuss on Twitter. Another interesting work by D. Maynard, K. Bontcheva[6], whose aim was to help organisation to create better campaigns by analysing and understanding the societal engagement regarding climate using Twitter platform.

## 3 APPROACH

The entire approach consists of collection of data, pre processing of data, generating sentiment labels, implementing topic modeling

[1] https://www.nytimes.com/2014/01/19/opinion/sunday/kristof-neglected-topic-winner-climate-change.html


*TxMM'21-'22, Radboud University, Nijmegen, Netherlands*

and performing exploratory analysis of data to gain more insights about the tweets. Reference code can be found here [2].

### 3.1 Data Collection

Around 5M tweets were extracted using snscrape [3] which is a scraper for social networking services. These tweets were collected under three major constraints. One being that the only those tweets were collected which had climate change or global warming as one of their hashtags, second being that tweets no older than 2015 where collected and lastly only those tweets were extracted whose language is English. Since performing natrual language processing on other languages would be difficult, so tweets written in English were collected. The dataset consist of Date, User ID, Tweets and Hashtags. The dataset can be found on Kaggle [4].

### 3.2 Data Preprocessing

Data preprocessing helps in normalizing text to perform further analysis and would help in getting more insights. This process is also known as feature extraction process. Following where the stages involved in preprocessing data:

- Removing all the repeated tweets. Extracting tweets using a scrapper can lead to have duplicate tweets.
- Extracting all hashtags from tweets and store as a separate field. Hashtags can be used in identifying the trend and emotions of the tweets.
- Removing handles (names mentioned using '@') from the tweets, since it can be considered as noise when generating senitments for the tweets.
- Removing punctuaiton marks, emojis, numeric values, urls and extra white space. Theses are the noise entities and does not provide meaningful knowledge.
- Removing stopwords from the tweets as they are the most frequently appearing words. Also they do not express any sorts of emotions hence its presence might affect the performance of the model.
- Generating tokens from the tweets, which help in topic analysis.

The processes of removing stopwords and tokenisation is perform after generating sentiments, since structured sentences are required having both syntactic and semantic meanings.

### 3.3 Generating Sentiments

Sentiments are the base for the sentiment analysis. Since the tweets were directly extracted using a scrapper, sentiments needs to be generated. SentimentIntensityAnalyzer [5] which is a Vadar sentiment analysis tool [4] that is provided by nltk packages has been use to generate sentiment intensity scores. This tools uses rule based model to perform general sentiment analysis. The scores, also known as polarity, is further divided into positive, neutral and negative sentiments by giving boundaries.

The boundaries can be set according to the requirements. If the individual is interested in knowing how strong the sentiments are then setting the score greater than some value as Strong Positive followed by Weak Positive and so on.

**Table 1: Sentiment Labels Scale**

| Polarity > 0 | Polarity = 0 | Polarity < 0 |
|---|---|---|
| Positive | Neutral | Negative |

### 3.4 Topic Modeling

Further, I applied topic modeling to draw the general outline about the discussion related to climate changes. I used Latent Dirichlet Allocation model which is based on generative probabilistic model, that calculates the probability of the word appearing in each topic. The word with the highest probability in a particular topic, defines that topic. For performing LDA, I have used gensim [6] which is a python open source library for performing topic modeling. The number of topic was set to 6 to avoid overlapping of the topics and helps in easy generalising.

## 4 RESULTS AND ANALYSIS

### 4.1 Sentiment Analysis

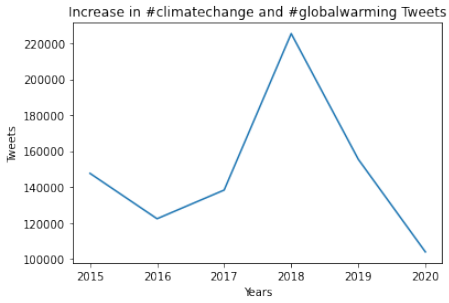


**Figure 1: No. of Tweets about global warming**

The fig shows the number of tweets about climate change or global warming being posted everyday. It can be inferred from the above figure that more people tweeted the most about climate change and global warming between the year 2017 to 2020. One factor that cause such a huge spike is the Paris Climate Agreement 2018, when United States of America, withdrew the agreement [10]. A lot of people have criticised the former president Donald Trump for his actions over Twitter. Another climate event was the heat waves in Europe during 2019.

Furthermore, detailed analysis has been performed, where individual sentiments have been plotted over the 5 years. There is not much difference between number of negative tweets and positive tweets. Over time period, all the three show similar trend, however, it is expected that during major climate events, the negative sentiment should rise rapidly and exceed positive sentiment which is not seen here. One of the reason could be, that during these events,

[2]https://github.com/Yashika-Sorathia/Sentiment-Analysis
[3]https://github.com/JustAnotherArchivist/snscrape
[4]https://www.kaggle.com/yashikasorathia/global-warming-tweets
[5]https://www.nltk.org/api/nltk.sentiment.vader.html
[6]https://radimrehurek.com/gensim/models/ldamodel.html

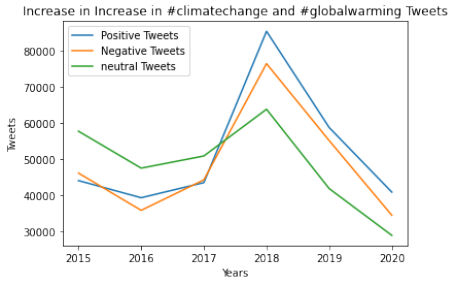


**Figure 2: Sentiments Throughout 5 Years**

people are showing more concerns and sympathies towards them and simultaneous people are showing more anger and grief over the lose of life and lack of services provided to aided regions. Due to this, even during major events, positive and negative sentiments are showing similar trends.

## 4.2 Topic Analysis

Topic modeling using LDA was performed on dataset to gain insights of topics of discussion people are having about climate change and global warming on Twitter platform.

**Table 2: Topic distribution coverage**

| | |
|---|---|
| Topic 1 | ["like", "people", "world", "think", "right"] |
| Topic 2 | ["climate", "change", "world", "report", "science"] |
| Topic 3 | ["ice", "sea", "melting", "arctic","rise"] |
| Topic 4 | ["global", "warming", "real", "dont", "believe"] |
| Topic 5 | ["year", "day", "record","earth", "temperature"] |
| Topic 6 | ["climate", "change", "news", "latest", "report" |

Total, 6 topics where generated, in which people have shown concerned about the rising of sea level due the melting of glaciers in arctic which what topic 3 is about. From the tweets of topic 1, it can be infered that people are being skeptical about climate change and global warming and used humor in their tweets. People have also expressed their disbelief about the climate change and global warming issues (Topic 4). However, there are people and other media agencies who have showed their concerns on the Twitter platform (Topic 2 and Topic 6). Lastly, topic 5 describes about the discussions people had over the temperature being recorded highest in the day or in the year.

Xiaoran An and the team [1], in their paper, tries to understand the whether social media analysis such as Twitter can complement and supplement the insights about the perceptions of climate change and how its been changing over time. Another work [9] has shown that there is dependency of number to tweets about climate change and global warming over the major climate events or the summits or agreements that are being held.

# 5 DISCUSSION

Sentiment Analysis and Topic Modeling together, provides lots of research scope in understanding human perspectives on climate change. One such scope is to perform analysis based on geographic location of the user, helps in understanding the landscape and if it is a highly affect areas, long term measures can be take. Furthermore, with topic modeling, deep analysis about the type of tweets can be performed where insights about what people's emotions towards climate change are, can be captured. Jennifer R. Fownes [3] suggested various future scope about how climate change is being discussed on Twitter. Adding to this, people's reactions and emotion can be studies with respect to the natural disasters as suggested by Allison Koenecke and her team[5] in their paper.

There are few limitation of this approach. This paper has used Twitter data to illustrate how the perspectives of people about climate change has changed over time. However, Twitter data does not represent all the social groups. Therefore, this approach cannot be generalise based on Twitter data alone. Also only those tweets were extracted which in English, which resulted into loss of tweets and the final findings are also affected. The sentiments generated are not true labels, which mean if the data were manually labeled, the findings might have differ. Adding to this, performing manual labelling requires time and huge amount of resources. Lastly, the Twitter data was extracted using a scrapper, which the questions ethics of this approach.

# 6 CONCLUSION

From the analysis of discussion over climate change and global warming on Twitter, the findings of this paper says that, over period of time, there is some change in people's opinion over climate change. Along with this, external factors such as, major climatic events or Global Summits, or any political influences, triggers the spike, where people discuss about these topics a lot. Adding to this, with the help of topic modeling, more knowledge about people's perspectives can be achieved, such as people are concern about the rise in sea level due to melting of ice. The findings of this paper is consists with [1], [9] and [5].

# REFERENCES

[1] Xiaoran An, Auroop R Ganguly, Yi Fang, Steven B Scyphers, Ann M Hunter, and Jennifer G Dy. 2014. Tracking climate change opinions from twitter data. In *Workshop on Data Science for Social Good*. 1–6.

[2] Patrick Baylis. 2015. Temperature and temperament: Evidence from a billion tweets. *Energy Institute at HAAS working paper* 12, 1 (2015), 1–21.

[3] Jennifer R Fownes, Chao Yu, and Drew B Margolin. 2018. Twitter and climate change. *Sociology Compass* 12, 6 (2018), e12587.

[4] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8.

[5] Allison Koenecke and Jordi Feliu-Fabà. 2019. Learning Twitter User Sentiments on Climate Change with Limited Labeled Data. *arXiv preprint arXiv:1904.07342* (2019).

[6] Diana Maynard and Kalina Bontcheva. 2015. Understanding climate change tweets: an open source toolkit for social media analysis.. In *EnviroInfo/ICT4S (1)*. 242–250.

[7] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. 2019. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433* (2019).

[8] Martha G Russell, June Flora, Markus Strohmaier, Jan Pöschko, Rafael Perez, Jiafeng Yu, Marc A Smith, and Neil Rubens. 2015. Semantic analysis of energy-related conversations in social media: A Twitter case study. In *Communicating Sustainability for the Green Economy*. Routledge, 233–251.

[9] Zhongkai Shangguan, Zihe Zheng, and Lei Lin. 2021. Trend and Thoughts: Understanding Climate Change Concern using Machine Learning and Social Media Data. *arXiv preprint arXiv:2111.14929* (2021).

[10] Yong-Xiang Zhang, Qing-Chen Chao, Qiu-Hong Zheng, and Lei Huang. 2017. The withdrawal of the US from the Paris Agreement and its impact on global climate change governance. *Advances in Climate Change Research* 8, 4 (2017), 213–219.