n 

**Muhammad Ahmed  
34716018**

**Table of Contents**

[1. Introduction 2](#_Toc166483515)

[2. Methodology 2](#_Toc166483516)

[2.1. Data Selection 2](#_Toc166483517)

[2.2. Data Preprocessing 3](#_Toc166483518)

[2.3. Modeling Approach 3](#_Toc166483519)

[3. Analysis of Topics 4](#_Toc166483520)

[3.1. Topic Overview 4](#_Toc166483521)

[3.2. Inter-topic Relationships 4](#_Toc166483522)

[3.3. Visualizations 5](#_Toc166483523)

[4. Discussions 8](#_Toc166483524)

[4.1. Interpretations of Results 8](#_Toc166483525)

[4.2. Limitations 9](#_Toc166483526)

[5. Conclusions 9](#_Toc166483527)

[5.1. Summary 9](#_Toc166483528)

[5.2. Future Work 9](#_Toc166483529)

[6. References 10](#_Toc166483530)

# 1. Introduction

Today’s generation of the social media is far advanced. It views Twitter as a central stopping point where issues that arise are discussed due to the infinite conversations and wide-ranging discourse about climate change for example. This report aims to considerably address climate priorities on Twitter by employing much advanced topic modeling combined with theme grouping to articulate and group the topics that come up in these discourses. The dataset comprised of tweets that encompassed different feelings about climate change obtained from a source that had been evaluated for a considerable time to enable the dynamic view.

Topic modelling is an essential part of text mining which guarantees breakdown of the large amount of unstructured data into smaller chunks that are easier to handle and comprehend thus facilitating a deeper understanding of the content and sentiments communicated through social media texts (Alvarez-Melis & Saveski, 2016: From the review of studies in the peer-reviewed journal "Science" by these authors, one can witness that a strong link exists in providing a home for the homeless (Curiskis et al. , 2020). The new approach called BERTopic was created by authors Gillioz et al. (Gillioz et al. , 2020) based on transformer NN structures (Gillioz et al. , 2020) to not only find the common topics occurring, but also explore how they are related. The use of this kind of an approach becomes very important as it makes LDA and NMF into a superior version of each other, thus helping to fight the big amount of tweets (Egger & Yu, 2022). This study is of significance, as it can pass through a lot of Twitter data and encodes themes that show public opinion and the direction the conversation is taking on climate change. The findings will give an informed understanding about public perception and communication on this critical global issue (Karabacak & Margetis, 2023; Laureate et al. , 2023). This report is focused on the issues under consideration and the relationship between them, thus giving a comprehensive view of the social media about climate change.

# 2. Methodology

## 2.1. Data Selection

In this study, the data is represented by a well-constructed collection of Twitter presents that points the finger at discussions centered on climate issue alone. The tweets which were made between April 26, 2015, and February 21, 2018, were mined from the publicly available repository. The total number of tweets mentioned is 43,943 and all of them were thoroughly annotated and verified for relevance and accuracy in regard to the topic of climate change. Every single tweet in this representative dataset is marked by sentiment and can belong to supportive and relatively unfavorable views on the subject. This set was chosen because it heavily covers it and also has a specific focus towards climate change; it's used to analyze people's opinions and discourse. A sample of the dataset is provided below to illustrate the nature of the data used:

Table 1: Sample of Dataset

|  |  |  |
| --- | --- | --- |
| **Sentiment** | **Message** | **Tweetid** |
| **-1** | @tiniebeany climate change is an interesting hustle as it was global warming but the planet stopped warming for 15 yes while the suv boom | 792927353886371840 |
| **1** | RT @NatGeoChannel: Watch #BeforeTheFlood right here, as @LeoDiCaprio travels the world to tackle climate change https://t.co/LkDehj3tNn httÃ¢â‚¬Â¦ | 793124211518832641 |
| **1** | Fabulous! Leonardo #DiCaprio's film on #climate change is brilliant!!! Do watch. https://t.co/7rV6BrmxjW via @youtube | 793124402388832256 |
| **1** | RT @Mick\_Fanning: Just watched this amazing documentary by leonardodicaprio on climate change. We all think thisÃ¢â‚¬Â¦ https://t.co/kNSTE8K8im | 793124635873275904 |
| **2** | RT @cnalive: Pranita Biswasi, a Lutheran from Odisha, gives testimony on effects of climate change &amp; natural disasters on the poÃ¢â‚¬Â¦ | 793125156185137153 |

## 2.2. Data Preprocessing

This preprocessing included the normalization of text, e. g. lowercasing, removing symbolic characters, URL links, and Twitter handles, which in turn, cleaned up the data for analysis. This was a must to avoid giving an idea or information with one-sided or insignificant point of view. Moreover, stopwords were removed to focus on relevant words that are the basis of topic modeling.

## 2.3. Modeling Approach

The BERTopic approach considering its popularity for natural text analysis later accompanied by a great ability to manipulate short texts such as tweets (Gillioz et al. , 2020). Different from conventional methods including LDA and NMF, BERTopic keeps up with the dynamic topic modeling different from later times, thus it is well-fitted for data streams like Twitter (Egger & Yu, 2022). From a semantic perspective, using a transformer-based technique by BERTopic is a highly sophisticated way of interpreting meaning that is significantly impactful given the complex discourse nature of social media (Zhou et al. , 2020). This enables putting forth a far attribute analysis that includes specific phrases or context which would be difficult to be picked using traditional methods to look for customized nuances like concealed ideas within a background of sheer data (Dimet Thijs, 2016; Curiskis et al. , 2020).

The approaches and the further volume of tweet dataset of Twitter data which were used to analyze the public comments about climate change on Twitter, demonstrate how the people watch and the polarization as a result in this concern.

# 3. Analysis of Topics

## 3.1. Topic Overview

The BERTopic model was able to distinguish a wide range of topics from the Tweeters with regards to climate change, which shows just how many views and concerns there are. Prominent topics include:

1. **Political Policies and Global Agreements**: This topic features to international treaties such as Paris Agreement that focus on fighting against global warming and the need for multinational solidarity.
2. **Scientific Debate and Skepticism**: Talks here revolve around whether the science of climate change is real or if it is a fake, with skeptics and proponents arguing.
3. **Impact on Wildlife and Nature**: This theme centers on the influence of climate change on biodiversity, and it highlights the issue of the endangered species and natural habitats.
4. **Economic Impact and Energy Policies**: As part of the discussions, we should talk about the finances of transition to renewable energy sources and also the price of environmental policies.
5. **Activism and Public Awareness Campaigns**: The scope of both national grassroots movements and international campaigns has been demonstrated through this subject, which shows how public is taking an active role in creating the awareness about climate justice.

## 3.2. Inter-topic Relationships

By showing the similarity matrices and hierarchical clustering, the analysis of relationships among topics has been the most important part of the topic relationships and how these topics interconnect. Comparison matrix saw that in some cases, topics like “Political Policies and World Agreements” and “Economic Impact and Energy Policies” were closely related in their focus of policy-making, but topics “Scientific Debate and Skepticism” were relatively separate from the rest of the matrix, indicating disparity in public perception and discussion that was not shared by other thematic content as well (Curiskis et al., 2020).

Hierarchical clustering also the relations among these topics, showing clusters of closely related topics such as policy and economic discussions, which often merge in the public discourse on strategic responses to climate change (Alvarez-Melis & Saveski, 2016). The segregation further reveals specific matters that are secluded, indicating that there are some critical components in the global warming discourse that may require refined communication techniques.

## 3.3. Visualizations

1. **Similarity Matrix**

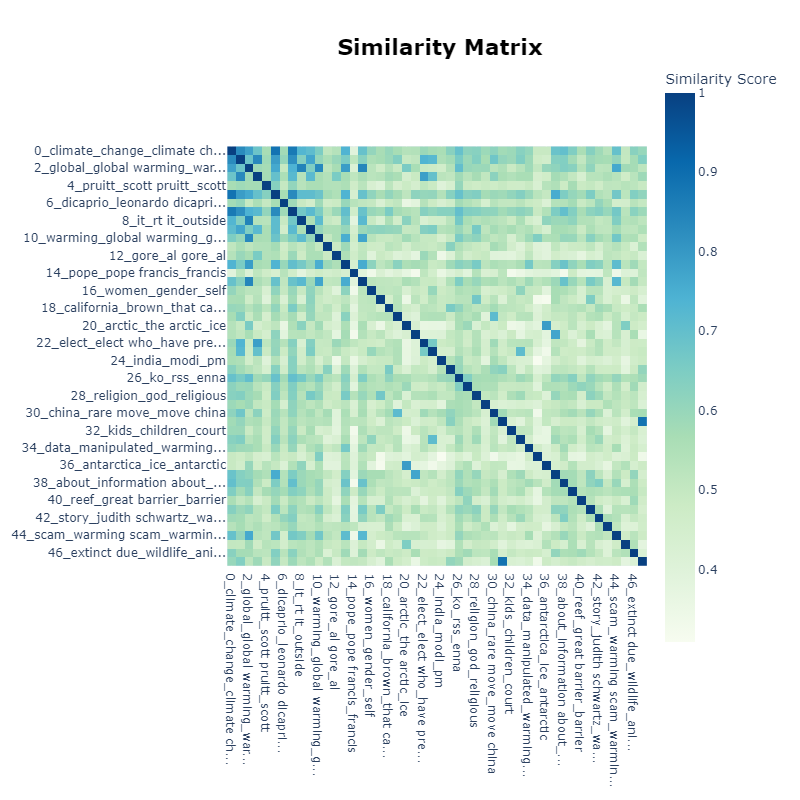


Figure 1: Similarity Matrix showing the subject cohesion between different subject, pointing out both the separate and closely-knitted areas on the topic of the climate change. This visualization helps in understanding the common topics that are being discussed in the discussion.

1. **Hierarchical Clustering Diagram**

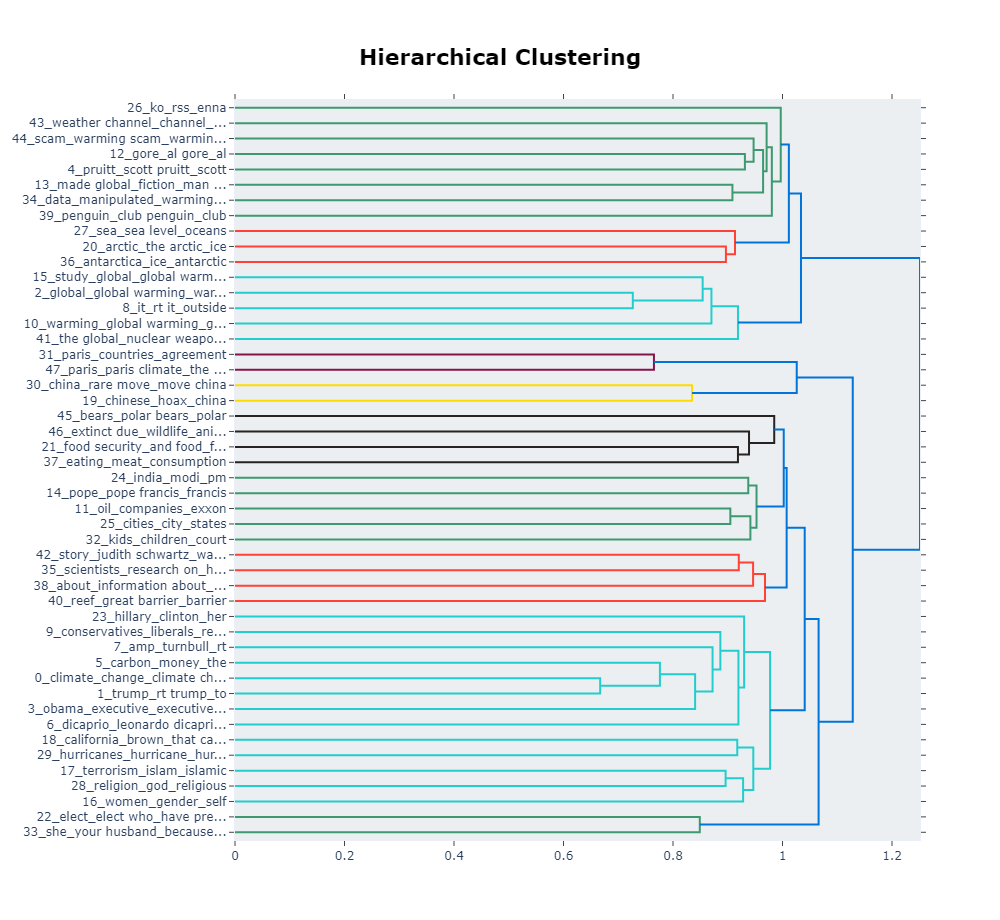


Figure 2: Hierarchical Clustering of topics derived from Twitter data regarding climate change to show the relationship, along with distance, of different discourse streams. Clusters point to hot topics which are often discussed together on the internet and social media.

1. **Topic Probability Distributions**

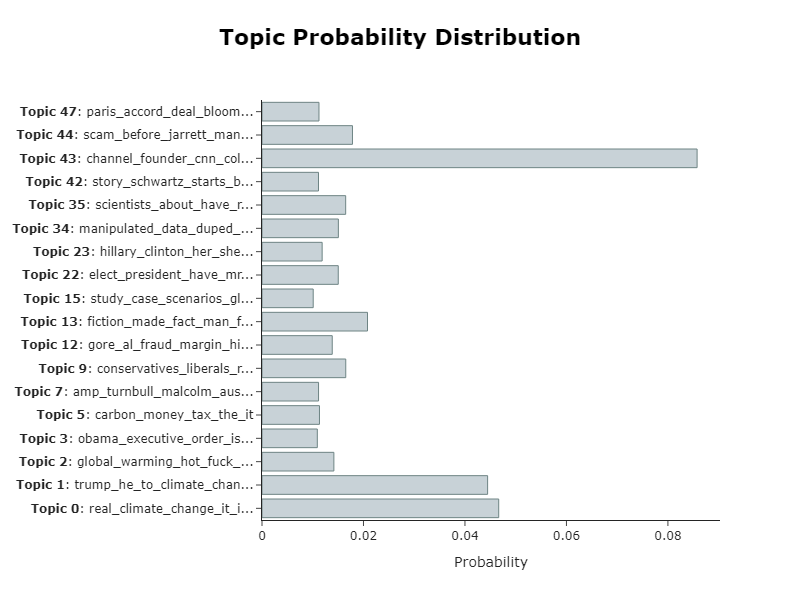


Figure 3: Bar chart showing the probability distribution for various topics that Twitter data generated on climate change. This chart is used to show which subjects are the most common in the dataset, thus it is clear which aspects of climate change are talked about the most. The visualization is a summed up visual brief for analysis sake aimed at showing more dominant topics in discussions mentioned.

# 4. Discussions

## 4.1. Interpretations of Results

Topics and interrelations between them support deepened understanding of how complex they are and how they intertwine, not to mention which factors influence them. Themes such as political policies and scientific scepticism are linked intimately via this show, they mirror ongoing debates and the discussions that often inform policies and trends in other countries. The interplay between the economic impacts and the political policies points to a growing public debate on the cost-benefit analysis of climate actions which is manifested in different social platforms (Curiskis et al. , 2020). We got a rise astounding discovery the huge talkabout biodiversity; that is a sign of high regard to the public about the wild-life and natural habitats. These might be disregarded topics in conventional media.

## 4.2. Limitations

The way of collecting data from social media may reflect people's integrated opinion, this is not always true (Karami et al. , 2020). Besides the language and mood on Twitter is (also) swayed by what is going on at the time, therefore, it could distort the stable attitude about climate change. The English-language tweets' dependence may be also excluding the non-English speaking discourse, possibly biasing the topics towards the English-speaking regions' concerns (Laureate et al. , 2023).

# 5. Conclusions

## 5.1. Summary

With this work done we have made a clear map in social media terms that divides the world of climate change conversation into many areas of such a nature as policy debate, taking personal actions into account and science based one. BERTopic becomes parts of discussions in the public sphere helping to see the effect of social media and digital media on the interconnection and development of these discussions (Egger & Yu, 2022). The results of the study show that social media has a very important role to play in the formation and reflection of public opinion on global issues like climate change.

## 5.2. Future Work

The next step could be to include multi-lingual datasets to widen the coverage of the climate debate and get a more global understanding of it. Implementing adaptive topic modelling methods instead would add the opportunity to observe how public opinions develop throughout time with the general data going through this resolution as a result of new scientific discoveries or changes in the state of affairs (Gillioz et al. , 2020). By and large, the incorporation of sentimental analysis will bring out the emotions involved in the deliberations and provide a more round picture of the public attitude toward climate change (the study of Rodrigues et al, 2022). By applying these approaches the awareness of climate discussion would be increased, thus helping to develop more focused and effective communication strategies.

# 6. References

1. Alvarez-Melis, D. and Saveski, M., 2016. Topic modeling in twitter: Aggregating tweets by conversations. In *Proceedings of the international AAAI conference on web and social media* (Vol. 10, No. 1, pp. 519-522).
2. Curiskis, S.A., Drake, B., Osborn, T.R. and Kennedy, P.J., 2020. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management*, *57*(2), p.102034.
3. Egger, R. and Yu, J., 2022. A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7*, p.886498.
4. Gillioz, A., Casas, J., Mugellini, E. and Abou Khaled, O., 2020, September. Overview of the Transformer-based Models for NLP Tasks. In *2020 15th Conference on Computer Science and Information Systems (FedCSIS)* (pp. 179-183). IEEE.
5. Karabacak, M. and Margetis, K., 2023. Natural Language Processing Reveals Research Trends and Topics in The Spine Journal Over Two Decades: A Topic Modeling Study. *The Spine Journal*.
6. Karami, A., Lundy, M., Webb, F. and Dwivedi, Y.K., 2020. Twitter and research: A systematic literature review through text mining. *IEEE access*, *8*, pp.67698-67717.
7. Laureate, C.D.P., Buntine, W. and Linger, H., 2023. A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*, *56*(12), pp.14223-14255.
8. Rodrigues, A.P., Fernandes, R., Shetty, A., Lakshmanna, K. and Shafi, R.M., 2022. Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques. *Computational Intelligence and Neuroscience*, *2022*.
9. Zhou, M., Duan, N., Liu, S. and Shum, H.Y., 2020. Progress in neural NLP: modeling, learning, and reasoning. *Engineering*, *6*(3), pp.275-290.