ANALYZING PUBLIC SENTIMENT ON CLIMATE CHANGE USING NATURAL LANGUAGE PROCESSING

Arish Morshed Zobeyer*[1] and Tamjidul Islam Illin [2]

[1] *Research Intern, Technovative Solutions Limited, UK, e-mail: arishzobeyer@gmail.com*

[2]*Undergraduate Student, Bangladesh University of Engineering and Technology, Bangladesh, e-mail: 1816018@wre.buet.ac.bd*

***Corresponding Author**

Abstract
Climate change has emerged as one of the most prominent topics of the decade, resulting in severe environmental disruptions such as global warming, extreme weather events, sea-level rise, and loss of biodiversity, all of which pose threats to ecosystems and human livelihoods. Given the increasing environmental concerns associated with climate change in recent years, understanding public sentiment on this matter is essential for formulating effective policies and practices. Sentiment analysis is a valuable method for researchers to gauge public perception and foster a social consciousness regarding this issue. This research investigates the effectiveness of sentiment analysis in mining social media data to assess public opinion on climate change. Although extensive research has been conducted in this area—primarily focusing on social media platforms due to the vast amounts of data generated during climate discussions—there remains a gap in studies that examine public sentiment from YouTube. To address this, the current study examines YouTube comments that feature term "Climate Change," using TextBlob that offers an easy-to-use API for conducting Natural Language Processing (NLP) tasks, including sentiment analysis to assess public sentiment. Textblob categorizes comments as positive, negative, or neutral, assigning sentiment scores to determine their polarity. The findings indicate that public attitudes towards climate change are generally neutral to mildly positive, reflecting a sense of mild optimism and constructive dialogue about potential solutions. While this sentiment is not overwhelmingly positive, it suggests an openness to engaging with climate-related issues. Insights derived from this analysis can offer policymakers valuable information to guide more informed and impactful decisions regarding climate change adaptation and mitigation strategies.

Keywords
Climate Change, global warming, Sentiment analysis, Natural Language Processing,

1. Introduction

One of the most urgent issues of the 21st century is climate change, driven by greenhouse gas emissions, such as carbon dioxide, released into the atmosphere. Since the 1800s, human activities—primarily the burning of fossil fuels like coal, oil, and gas—have been the main contributors to this change [1]. Climate change impacts the environment in various ways, including rising temperatures, sea-level rise, droughts, flooding, and more. It presents a significant challenge and concern at global, regional, and national levels [1-2]. According to the World Meteorological Organization [3], 2020 was one of the hottest years ever recorded, even with a temporary decrease in greenhouse gas emissions due to COVID-19 measures. Additionally, 2020 witnessed a notable increase in extreme events such as hurricanes, heatwaves, severe droughts, and wildfires, resulting in population displacement and fatalities [3]. Therefore, it is crucial for us to work together to mitigate further climate-related damage, which can be informed by examining public sentiment and implementing policy changes.

The impacts of climate change are recognized globally [3], and the topic has been a prominent part of public discourse through media events such as UN climate conferences, the release of IPCC reports, and civil protests since 2018. The increase in user engagement on social media in recent years demonstrates a growing public interest in climate change and highlights heightened concern for the issue [4]. Platforms like YouTube, Instagram, Twitter, and Facebook have allowed users to express their views on climate change more than ever. As these users play a vital role in spreading information and raising awareness, social media has become a crucial medium for communicating climate-related issues [5].

Sentiment analysis, a branch of Natural Language Processing (NLP), focuses on developing techniques to uncover the sentiments expressed in online comments, reviews, or opinions [6]. Its primary aim is to study and analyze perceptions and viewpoints regarding specific entities within discussions. Among various discussion mediums, analyzing natural language to extract sentiments remains the most popular application in this field [3]. Additionally, by examining the tone of words in a text, sentiment analysis can reveal people's reactions to a given topic, categorizing them as negative, neutral, or positive. This insight can assist decision-makers in understanding public behavior and addressing pressing issues like global warming and climate change [7].

In the recent years, the widespread access to the Internet has led to a surge in social media platforms, attracting a vast number of users who interact and share their thoughts or opinions on various topics. This activity has produced an enormous volume of both structured and unstructured data, which can be analyzed for valuable insights. The popularity of these platforms continues to grow annually, as they facilitate faster dissemination of information compared to traditional news outlets. Among these, YouTube has gained significant traction over the years. Furthermore, the opinions and sentiments expressed by users have demonstrated a considerable impact on society [7]. Individuals from various backgrounds—including NGOs, activists, celebrities, politicians, and the general public—have voiced their views on climate change issues on YouTube. Key topics in these discussions often revolve around the reality of climate change and potential mitigation strategies [7]. Thus, by leveraging the wealth of opinions available on this platform, researchers can explore the evolution of sentiment regarding social issues like climate change.

Sentiment analysis includes two main approaches: lexicon-based methods and machine learning methods. Lexicon-based approaches utilize a pre-built dictionary that categorizes words as positive, negative, or neutral, often including sentiment intensity for each word. In contrast, machine learning approaches rely on statistical techniques that use feature extraction methods, where each word or phrase is vectorized to serve as independent features during training [7]. One advantage of lexicon-based approaches is that they do not require large training datasets, making them a suitable choice when a substantial training corpus is unavailable [8].

Research on public comments regarding social issues has a long history. Mandel et al. [9] examined public responses to Hurricane Irene, pioneering a trend in disaster response research through comments. Subsequent studies analyzed public opinions during different disaster stages, including preparedness, emergency response, impact, and recovery [10]. Additionally, researchers assessed the intensity of collective emotions on various climate change topics, such as heat waves and floods, by analyzing comment content [9]. Twitter has emerged as a platform that fosters social and political action to tackle climate change, with its diverse user base enabling a wide range of participants—including NGOs, politicians, celebrities, and grassroots movements—to engage in these discussions [11].

While discussions on climate change comments have occurred, they have not yet been effectively linked to attitudes about climate change as reflected in those discussions [11]. This study aims to highlight the sentiments and opinions of YouTube users regarding climate change by collecting a corpus of comments from YouTube videos, preprocessing the text, conducting sentiment analysis and topic modeling, and calculating the mean sentiment score. The structure of the paper is as follows: Section "Data Collection and Preprocessing" describes the data and its source, Section "Methodology" outlines the study's methodology, Section "Results" presents the implementation and findings, and the paper concludes in Section "Conclusion."

2. Data Collection and Preprocessing

In this study, we begin by outlining the data collection process from YouTube using specific keywords. We start with the search term "Impact of climate change." Videos with disabled comment sections are excluded from our analysis. The collected comments are stored in a data frame. We set a limit of 20,000 comments to extract, while the maximum number of YouTube videos searched is capped at 1,000,000. During the next step, the obtained data are processed to prepare it for the following stage. Several stages are included in this stage:

- Punctuation was removed to eliminate marks from the text data, a standardization method that allows 'happy' and 'happy!' to be treated identically. The list of punctuation to exclude must be carefully selected based on the specific use case.
- HTML breaks were eliminated, and whitespace and newlines were normalized.
- The next step in the preparation process involved removing any URLs, DOI references, and academic citations from the data. In comment analysis, it's common for comments to include URLs, which typically need to be removed for further research.
- Unwanted whitespace, which increases text size without adding value, was eliminated. Thus, removing extra whitespace is a simple yet crucial text preprocessing step.
- Emojis were removed using a regex (Regular Expression) pattern.
- Tokenization, the process of breaking down the text into individual tokens, was performed as this is necessary for many NLP procedures. Word tokens were converted to lowercase and filtered to retain only alphabetic tokens.
- Stop words, which are common words that carry little significant meaning, were removed.
- Finally, the modified data frame containing the cleaned comments was printed.

3. Methodology

In Python, there are many packages available for sentiment analysis, each employing different methods. This research utilized the TextBlob natural language processing algorithm for analyzing textual data. Designed for both Python 2 and 3 [12], TextBlob offers a straightforward API for various NLP tasks, including part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and translation. It provides a polarity score that indicates the emotional sentiment of a statement, ranging from −1 (negative sentiment) to +1 (positive sentiment). Additionally, subjectivity scores, which range from 0 to 1, reflect the author's perspective on the topic: a score of 0 signifies a very objective comment based on factual information, while a score of 1 indicates a highly subjective opinion. NLTK (Natural Language Toolkit), a comprehensive library, was also used for tokenization and stop word management.

WordCloud is a visualization tool that displays the most frequently occurring words in a text dataset, highlighting the terms that are significant in sentiment analysis. By creating a word cloud, users can easily identify prominent themes and sentiments in the text, making it a valuable resource for understanding public opinion and trends. The size of a word indicates how important it is in terms of how often it appears in the text. In this research, WordCloud was used to enhance the interpretation of textual data. The primary python libraries used for this purpose are: WordCloud (for generating word clouds), Matplotlib (for visualizing the word cloud), NumPy (used for handling arrays and numerical data), Pandas (for data manipulation). This study also utilized N-grams which are sequences of n items, typically words or characters, used in NLP to analyze text patterns and context. N-grams, which can be unigrams (single words), bigrams (two-word combinations), or trigrams (three-word combinations), allow for a better understanding of the sentiment expressed by considering not just individual words but also their combinations. For instance, "not good" has a different sentiment than "good." The steps used for N-grams in sentiment analysis included the following: First, text preprocessing was conducted to clean the text data by removing punctuation and special characters, as well as converting all text to lowercase. Next, the cleaned text was tokenized to break it into individual words or tokens. Afterward, N-grams were generated by selecting sequences of words, such as unigrams, bigrams, or trigrams from the text. Then, the frequency of each N-gram was counted to identify common phrases and patterns.

4. Results

4.1 Sentiment Score

The aim of conducting sentiment analysis is to evaluate global public opinion on climate change. The findings derived from the sentiment scores are presented through various charts. For instance, Figure 1 illustrates the sentiments gathered from multiple comments, categorized into three labels: positive (≥ 0), neutral ($=0$), and negative (≤ 0). Overall, the sentiments expressed in comments during the pandemic ranged from neutral to mildly positive, suggesting that the public maintained an optimistic outlook despite the challenges posed by global climate change. Comments about "climate change" conveyed negative sentiments with words like "bad, fail, crazy, afraid, and catastrophe," neutral sentiments through terms such as "can't say, okay, normal, and fair," and positive sentiments with words like "good, joyful, wonderful, and happy." A total of 9,986 comments related to climate change were gathered from 1,000,000 YouTube videos. Of these comments, 44.26% conveyed positive sentiments, 22.87% expressed negative feelings, and 32.85% were categorized as neutral.

In this study, the mean sentiment score was calculated to be 0.06. This mildly positive score from TextBlob regarding climate change suggests that the text reflects a slightly optimistic or favorable viewpoint, potentially highlighting discussions about solutions, progress in addressing climate issues, or positive actions being undertaken. Conversely, a neutral or negative score would indicate a more critical or pessimistic perspective.

```
0       everybody wants money question certainty valid...    0.166667
1       rich countries way green china wants wait anot...    0.125000
2       lol asking quisling sellout planetkillers care...    0.700000
3       still care carnivores arnivores teeth tells us...   -0.070000
4       funding funny let face humanity good cooperati...    0.475000
...                                                   ...         ...
9982                                           guess done    0.000000
9983    going talk massive deforestation causing massi...    0.121429
9984                             well dry know worth water    0.116667
9985                   really still people believe nonsense    0.200000
9986                                nature give us back gave    0.000000
```

Figure 1: Example of computed values of sentiments

4.2 Analysis of frequent words using WordCloud

The extracted texts were tokenized, and the most frequently used words were visualized as a word cloud (Figures 2 and 3). Word frequency analysis was conducted to identify common terms, combinations, and semantic associations that could serve as additional markers of relevance to climate change, as well as to explore variations across comments. When discussing climate change, commenters frequently used terms such as "world, human, planet, people, earth, need, year, and think," among others. Figure 3 shows the top ten high frequency words.

Figure 2: Word Cloud representing high frequency word

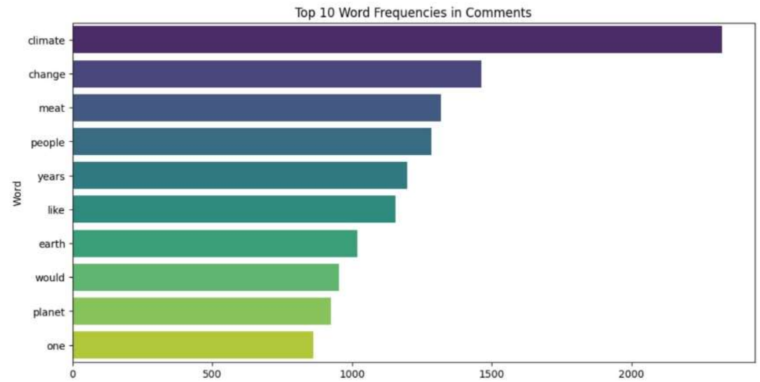

Figure 3: Bar chart of high frequency word

4.3 N-gram Analysis

N-gram in Natural Language Processing is a sequence of N words or characters that appear together in a text. In this study, 2-grams and 3-grams were calculated to identify the most commonly used pairs and triplets of words, respectively.

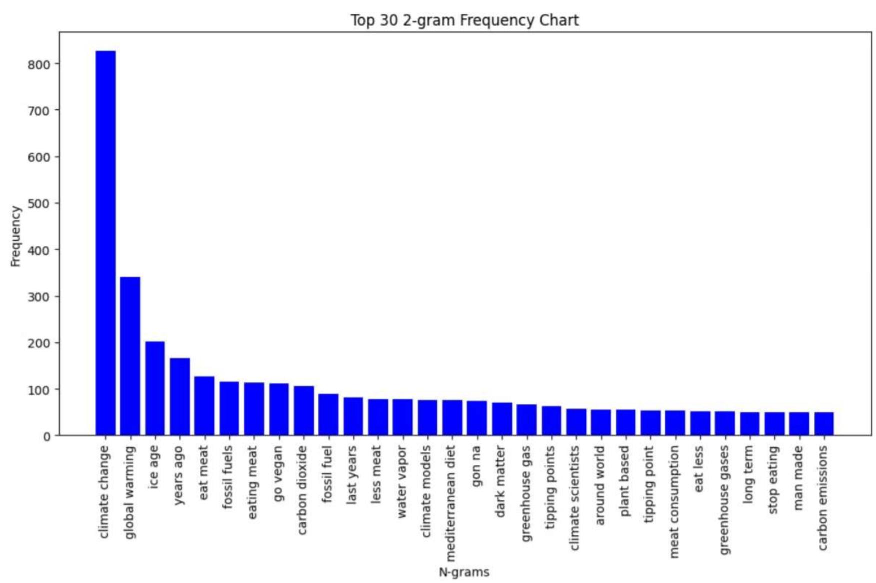

Figure 4: Most frequent two-word phrases (2-grams) associated with climate change

It is understandable that "climate change" and "global warming" are the most frequently used terms to express people's perceptions of climate issues, as shown by the N-gram analysis (Figure 4). These N-grams offer more contextual information than merely looking at single-word probabilities, enhancing our understanding of commonly used words and phrases. For example, the word “emissions” may appear often in discussions about climate change, but without context, it remains unclear whether the reference is to “reducing emissions” or “increasing emissions.” Figure 4 presents the top 30 two-word N-grams. In comments regarding climate change, users frequently mentioned associated phrases such as “global warming,” “ice age,” “years ago,” “fossil fuel,” and “carbon dioxide,” highlighting these as the predominant two-word N-grams.

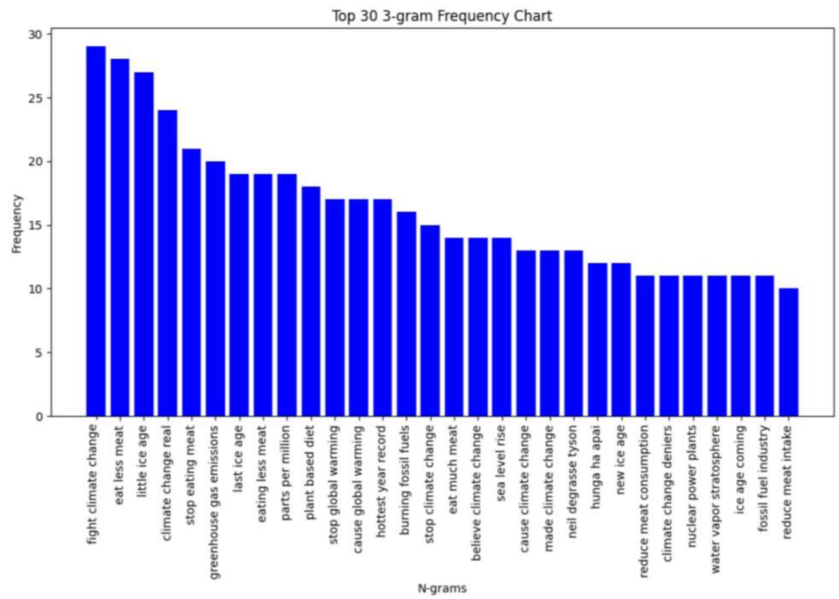

Figure 5: Most frequent three-word phrases (3-gram) associated with climate change

Figure 5 displays the top 30 three-word N-grams. In comments about climate change, users often highlighted phrases such as “fight climate change,” “eat less meat,” “little ice age,” “climate change real,” and “stop eating meat,” identifying these as the most prominent three-word N-grams. One commonly recurring phrase, “stop eating meat,” reflects people's concerns about how to reduce emissions, as the food system is responsible for 19% to 29% of global greenhouse gas emissions, with 80% to 86% stemming from agricultural production [13]. Within this agricultural sector, meat production accounts for 72% to 78% of greenhouse gas emissions as animals like cows, sheep, and goats produce methane as part of their digestive process [13].

5. Conclusion

Sentiment analysis serves as a valuable tool for gauging public opinion on various topics, including climate change. In this study, TextBlob sentiment analyzers were employed to assess people’s perception on climate change, revealing that 32.85% of comments were neutral, 44.26% positive, and 22.87% negative, indicating a general trend toward neutrality and mild positivity among users. This suggests a leaning toward optimism, likely reflecting positive actions being taken to address climate issues. The research highlights the potential of sentiment analysis to enhance understanding of public attitudes toward climate change, providing crucial insights for policymakers, businesses, and organizations focused on sustainability.

Additionally, the findings reveal the complex nature of public sentiment, emphasizing diverse perspectives that shape discourse. This understanding is essential for developing effective communication strategies and interventions that encourage dialogue among stakeholders with differing views. The study underscores the importance of social media as a critical data source during challenging times. It serves as a reference for researchers examining similar topics or utilizing social media data, offering a benchmark for comparative studies across different platforms or linguistic contexts. However, the research acknowledges limitations related to time and computational resources, which may have affected the depth of analysis. Focusing solely on English-language comments could result in an incomplete understanding of global sentiment, especially in non-English-speaking regions. Expanding the study to include other platforms like Facebook, Instagram, and Reddit, as well as additional languages, could reveal unique trends and enhance the robustness of the findings.

References

[1] Phillips M, Dickie J (2015) Climate change, carbon dependency and narratives of transition and stasis in four English rural communities. Geoforum.; 67:93–109.

[2] Cody, EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM (2015) Climate change sentiment on Twitter: An unsolicited public opinion poll. PLOS ONE **10**, e0136092, https://doi.org/10.1371/journal.pone.0136092.

[3] World Meteorological Organization. (2020) State of the Global Climate 2020; World Meteorological Organization: Geneva, Switzerland.

[4] Gómez-Casillas A, Gómez MV (2023) The effect of social network sites usage in climate change awareness in Latin America. Popul Environ. 45:7.

[5] Singh P, Kaur S, Baabdullah AM, Dwivedi YK, Sharma S, Sawhney RS, et al (2022) Is #SDG13 trending online? Insights from climate change discussions on twitter. Inf Syst Front 25:1–21.

[6] Khan, MY, Junejo, KN (2020) Exerting 2D-Space of Sentiment Lexicons with Machine Learning Techniques: A Hybrid Approach for Sentiment Analysis. Int. J. Adv. Comput. Sci. Appl. (IJACSA), *11*, 599–608.

[7] Sham NM, Mohamed A. (2022) Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches. Sustainability, *14*, 4723. https://doi.org/10.3390/su14084723.

1. [8] Gupta I, Joshi N (2020) Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic. J. Intell. Syst. *29*, 1611–1625.

[9] Mi Z, Zhan H (2023) Text Mining Attitudes toward Climate Change: Emotion and Sentiment Analysis of the Twitter Corpus. Wea. Climate Soc., 15, 277–287, https://doi.org/10.1175/WCAS-D-22-0123.1.

[10] Huang Q, Xiao Y (2015) Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery. ISPRS Int. J. Geo-Inf 4, 1549–1568, https://doi.org/10.3390/ijgi4031549.

[11] Fownes J, Yu RC, Margolin DB (2018) Twitter and climate change. Sociol. Compass, 12, e12587, https://doi.org/10.1111/soc4.12587.

[12] Anderson T, Sarkar S, Kelley R (2024) Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques. Natural Language Processing Journal, Volume 8, 100097, ISSN 2949-7191, https://doi.org/10.1016/j.nlp.2024.100097.

[13] Moosburger R, Wagner J, Heldt K, Richter A, Manz K, Mensink GBM, Loss J. Perspectives of individuals on reducing meat consumption to mitigate climate change: protocol for a scoping review. BMJ Open. 2023 Apr 18;13(4):e071122. doi: 10.1136/bmjopen-2022-071122. PMID: 37072363; PMCID: PMC10124295.