“Public sentiment towards climate change on Reddit”: A sentiment analysis study of posts and subreddits related to climate change (word count: 2682)

Full name: Srikeerthana Reddy Bandi
Semester: Spring 2023

1. Introduction:

Climate change is one of the most pressing issues facing the world today. Public perception and attitudes towards this issue can significantly impact policies and actions taken to mitigate the effects of climate change. Social media platforms like Reddit provide a valuable means of understanding public sentiment towards climate change. In this study, I explored public sentiment on climate change through Reddit posts.

Previous studies have used sentiment analysis to investigate public attitudes towards climate change on social media platforms such as Twitter, Facebook, and Instagram. For instance, Fownes, J. R., Yu, C., & Margolin, D. (2018) analyzed tweets related to the Paris climate agreement and found that the overall sentiment was mixed, with both positive and negative attitudes expressed. Similarly, Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Rosenthal, S. (2020) analyzed public opinion polls and social media data to understand climate change beliefs and found that there were significant disparities in belief depending on political affiliation. Zeng (2022) used data mining and text analysis to investigate the Chinese public's perception of climate change on social media and found that the sentiment towards climate change was generally negative.

However, there is limited research on sentiment analysis of Reddit posts related to climate change. As one of the largest online communities with millions of users, Reddit provides a unique platform to study public opinions and attitudes towards climate change. In this study, sentiment analysis was conducted on Reddit posts related to climate change, building on previous research in the field. Specifically, I aim to investigate the public sentiment towards climate change. By analyzing the sentiment of Reddit posts related to climate change, I hope to provide insights into public attitudes towards this important issue and contribute to the growing body of research on climate change sentiment analysis.

2. Research Question: How do public attitudes towards climate change vary across different subreddits on Reddit that are relevant to climate change?

3. Method:

3.1 Data:

The dataset for this research was collected using PRAW, a Python package designed for accessing the Reddit API. The data was collected by filtering subreddits related to climate change or global warming, which were manually selected based on their relevance to the research question. I focused on seven subreddits related to climate change and global warming, namely "climatechange", "globalwarming", "climate", "environment", "climate_science", "climateskeptics", and "ClimateOffensive". These subreddits were selected based on their relevance to the research question and their high activity levels. To collect the data from subreddits,

I set up the API credentials and user agent, and then used the subreddit search function to find posts that contained the keywords "climate change" or "global warming". The search was case-insensitive, meaning that it would capture variations in capitalization.

Title	Username	Date	Upvotes	Subreddit
the impact of climate change on biodiversity and the struggle for survival	JackONeill23	12/14/22 13:58	6	climatechange
true terror: ten worst things climate change will soon deliver to our world	quarry	5/10/22 10:35	17	globalwarming
is climate change triggering extreme cold? the debate is super hot. | research suggests that global warming is altering the jet stream, pushing arctic air down to southern climes more frequently. but the scientific jury is still out.	silence7	12/23/22 12:42	144	climate
europe experienced its hottest summer on record in 2021, while being ravaged by floods, heatwaves and fires, according to a report published friday by the european union's copernicus climate change service, showing that global warming is sharply on the rise.	Wagamaga	4/24/22 13:25	54	environment
climate change-driven droughts are getting hotter, study finds: temperature increases during dry periods outpace average climate warming	gmb92	8/7/18 10:56	17	climate_science
hurricane florence isn't the result of climate change, global warming	barttali	9/14/18 11:22	24	climateskeptics
the scientific consensus on the most effective way to mitigate climate change	ILikeNeurons	4/15/20 11:20	14	ClimateOffensive

Figure 1: Illustrates the collection of posts through Python libraries from various subreddits.

I have also specified a time frame for our dataset, covering the period from January 1, 2018, to March 1, 2023. This time frame was chosen to provide a broad and comprehensive view of public sentiment towards climate change over the past five years. The python script creates an empty list to store the post data and loops through each subreddit in the subreddits list. For each subreddit, it uses the reddit.subreddit() function to get the subreddit object, and the subreddit.search() function to search for posts that match the keywords within the date range. The posts are then saved into a list and written to a CSV file named "posts.csv" with the columns "title", "username", "date", "upvotes", and "subreddit". The date is converted from a Unix timestamp to a human-readable format using the "time" module.

Overall, our dataset consists of the titles, username, dates, upvotes, and subreddits of posts that contained the specified keywords and fell within the specified time frame. I have gathered a dataset consisting of 779 posts from seven different subreddits using PRAW.

3.2 Preprocessing:

In order to prepare the data for sentiment analysis, I have performed preprocessing on the text data. The preprocessing steps included removing duplicates, converting all text to lower case, removing punctuations, and eliminating URLs.

To perform preprocessing, I have run a python script. The script uses the csv module to read and write CSV files. It opens the 'posts.csv' file in read mode, creates a CSV reader object, and skips the header row. Then, it creates a set to store the unique post titles and a list to store the preprocessed posts data. For each row in the CSV file, the script extracts the post title, preprocesses it by converting it to lowercase and removing URLs using regular expressions. It then checks if the preprocessed title contains any of the keywords "climate change" or "global warming". If it does, the script checks if the title is already in the set of unique titles. If not, it adds the title to the set and the preprocessed post data to the list.

These preprocessing steps were crucial to clean the data and make it more suitable for analysis. By removing duplicates, I ensure that each post is unique, and by converting the text to lower case and removing punctuations, I standardize the text and make it easier to analyze. Furthermore, by removing URLs, I eliminate any irrelevant information that might interfere with the sentiment analysis. This preprocessing step was necessary to prepare the data for further analysis and to gain

valuable insights into the sentiments expressed in the posts. Post, preprocessing, I have a total of 573 posts to analyze.

3.3 Analysis:

To examine the general public's sentiment towards climate change, I conducted a sentiment analysis using two distinct tools: VADER and TEXTBLOB. These two tools were chosen due to their ease of use, accuracy, and simplicity. VADER is a rule-based sentiment analysis model that incorporates a combination of lexicon-based and rule-based approaches to determine text sentiment. TEXTBLOB, on the other hand, is a straightforward and user-friendly sentiment analysis model that employs machine learning techniques to classify text sentiment. This model is built on top of the Natural Language Toolkit (NLTK) and leverages a pre-trained Naive Bayes classifier to classify text into positive, negative, and neutral categories.

I utilized the pandas python library to read our preprocessed file, which contained 573 data points (post titles) for sentiment analysis. First, using VADER, I calculated the positive, negative, neutral, and compound scores for each post title. VADER calculates sentiment scores based on a set of rules that consider the presence of positive and negative words, punctuation, capitalization, and other linguistic features that convey sentiment. It also considers the context of the text, such as negation, amplification, and contrast, to determine the overall sentiment of the text. Based on the compound score which lies between -1 to +1, I added one more column to our data frame to classify the sentiments as positive, negative, or neutral. To accomplish this, I used the following categorization rule:

Greater than 0	Positive
Less than 0	Negative
Equals to 0	Neutral

These are some sample post titles along with their corresponding sentiment, analyzed using Vader.

S.No	*Title*	*Compound Score*	*Sentiment*
1	*global warming overshoots increase risks of climate tipping cascades in a network model - nature*	*0.2023*	*Positive*
2	*we were told there'd be droughts due to climate change. that's not happening. indeed it turns out lakes are growing in size and that's bad for... climate change*	*-0.4215*	*Negative*
3	*climate change awareness music video*	*0*	*Neutral*

Figure 2: Illustrates the Compound scores and sentiment of the posts using Vader.

Similarly, using TEXTBLOB, I computed the sentiment score for each post title and added one more column to classify the sentiments as positive, negative, or neutral. TextBlob uses a pre-trained Naive Bayes classifier to classify words and phrases as positive, negative, and neutral. It uses a labeled dataset to train the classifier, where each word or phrase is labeled positive, negative, and neutral.
These are some sample post titles along with their corresponding sentiment, analyzed using Textblob.

S.No	Title	Sentiment Score	Sentiment
1	*biden opens new federal office for climate change, health and equity | the office will be the first government effort to focus specifically on the public health dangers of global warming.*	0.096590909	Positive
2	*the amount of baked-in global warming, from carbon pollution already in the air, is enough to blow past international agreed upon goals to limit climate change, a new study finds.*	*-0.022727273*	*Negative*
3	*i got lost while trying to understand climate change*	*0*	*Neutral*

Figure 3: Illustrates the Sentiment scores and sentiment of the posts using Textblob

To gain a better understanding of the data, I have analyzed the sentiment counts for each year by sorting the dates in chronological order and computing the total count of sentiments for each year. Moreover, I have also calculated the count and percentage of positive, negative, and neutral sentiments individually to facilitate in-depth analysis. The count is calculated by simply counting the number of instances that fall into each category of interest. To calculate the percentage, I divided the count of instances in each category by the total number of instances in the dataset and multiplied by 100.

Year	Total Sentiment Count	Positive Sentiment Count	Neutral Sentiment Count	Negative Sentiment Count	% Positive	% Neutral	% Negative
2018	68	40	4	21	61.54	6.15	32.31
2019	130	68	15	47	52.31	11.54	52.31
2020	98	58	8	32	59.18	8.16	32.65
2021	92	47	8	37	51.09	8.7	40.22
2022	190	99	15	74	52.66	7.98	39.36

Table 1: Vader Sentiment Analysis

Based on the table provided, we can see the sentiment analysis results for the years 2018-2022. The sentiment analysis was conducted using VADER (Valence Aware Dictionary and Sentiment Reasoner), a popular lexicon and rule-based sentiment analysis tool.

First, we can observe that the sentiment count varies significantly between years, ranging from 68 in 2018 to 190 in 2022. Based on the sentiment analysis over the years from 2018 to 2022, it appears that the communication related to the data source has been generally positive, with only minor variations in the percentage breakdown of positive, neutral, and negative sentiment. The percentage of positive sentiment has been consistently higher than negative sentiment, except for 2019 where negative sentiment was slightly higher. The percentage of neutral sentiment has remained relatively consistent over the years. The year 2020 had the highest percentage of positive sentiment and the lowest percentage of negative sentiment, while 2019 and 2021 had higher levels of negative sentiment. The significant increase in positive sentiment in 2022 suggests that the communication related to the data source has become more positive.

Interestingly, the percentage of neutral sentiment has remained consistently low, ranging from 6.15% in 2018 to 7.98% in 2022. This suggests that the majority of the sentiment being expressed is either positive or negative, with very little falling into the neutral category.

Year	Total Sentiment Count	Positive Sentiment Count	Neutral Sentiment Count	Negative Sentiment Count	% Positive	% Neutral	% Negative
2018	68	25	25	15	38.46	38.46	23.08
2019	130	48	52	30	36.92	40	23.08
2020	98	41	41	16	41.84	41.84	16.33
2021	92	34	39	19	36.96	42.39	20.65
2022	190	71	77	40	37.77	40.96	21.28

Table 2: TextBlob Sentiment Analysis

The percentage of positive posts has remained relatively consistent, ranging from 38.46% in 2018 to 37.77% in 2020. This suggests that overall positivity in the communication related to the data source has been steady over the years. The percentage of neutral posts has also remained relatively stable, ranging from 38.46% in 2018 to 40.96 % in 2022. This indicates that a significant portion of the communication related to the data source has been neither strongly positive nor strongly negative. The percentage of negative posts has ranged from 16.33% in 2020 to 23.08% in 2019. This suggests that there has been some negative sentiment present in the communication related to the data source, although it has fluctuated from year to year.

The ratio of positive to negative posts has generally been around 2:1, indicating that positive posts have been more prevalent than negative posts. However, in 2019, the ratio was closer to 1.5:1, suggesting that negative posts were more prominent that year.

Overall, the two tables provide complementary information about the sentiment trends in the communication related to the data source. The VADER sentiment analysis seems to be more sensitive to positive sentiment, while TextBlob is slightly more sensitive to neutral sentiment. However, the general trends and conclusions drawn from both tables are consistent, indicating a largely positive and neutral tone in the communication related to the data source.

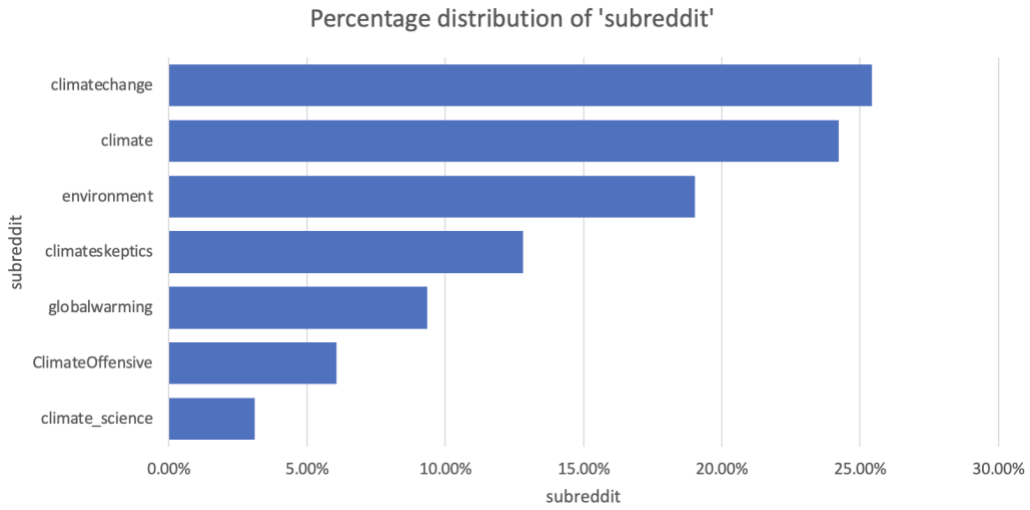
Statistics	Compound Score
Lowest	-0.9081
Average	0.018
Median	0.1027
Highest	0.8987

Table 3: Statistics of the compound score

The highest compound score is the largest sentiment score in the dataset, representing the most positive sentiment which is 0.8987. The lowest compound score is the smallest score in the dataset, representing the most negative sentiment which is -0.9081. The median compound score is the middle value in the dataset, separating it into two equal halves. The median compound score of 0.1027 suggests that the sentiment scores are evenly distributed around the middle point, further indicating the mixed nature of public sentiment on climate change. The average compound score is the sum of all scores divided by the number of scores in the dataset and represents the overall sentiment score of the dataset, the average compound score of 0.018 suggests that, overall, the sentiment is slightly positive, but with a relatively low magnitude.

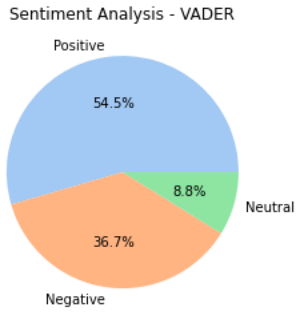
4. Results:

The subreddits climatechange, climate, and environment are the most popular ones among Reddit users to discuss climate change and global warming. On the other hand, the relatively low count of the ClimateOffensive subreddit, which focuses on discussing potential solutions to climate change, indicates that finding solutions to climate change may not be as much of a priority for Reddit users as discussing the problem itself. Additionally, the low count of the climate_science subreddit suggests that there may be a lack of interest or understanding among Reddit users regarding the scientific aspects of climate change.



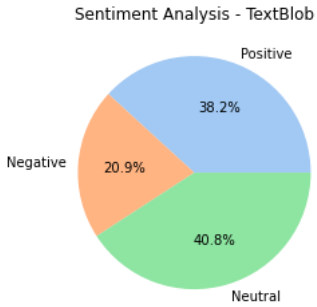
Graph 1: Percentage distribution of subreddit

A total of 573 data points were analyzed. Following are the graphs which visualize the percentage distribution of polarity scores.



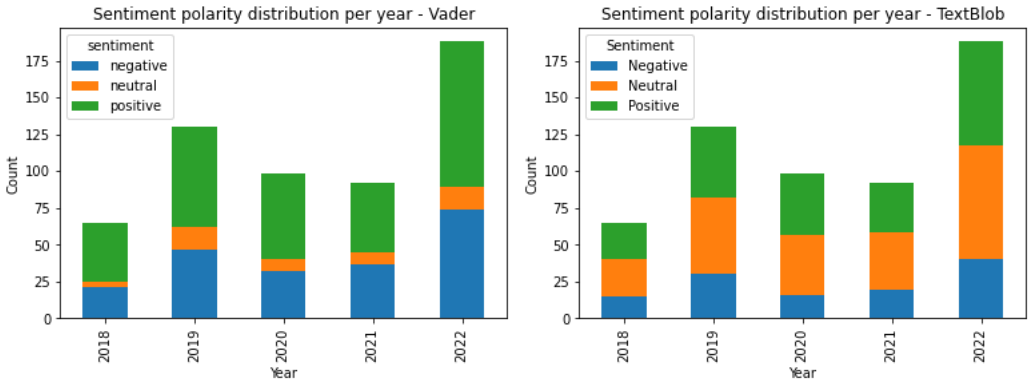
Graph 2: PIE chart representation of polarity score distribution using Vader

The graph shows the results of sentiment analysis on a set of text data using the VADER. Of these, 54.5% are classified as positive sentiment, 36.7% as negative sentiment, and 8.8% as neutral sentiment for Vader.



Graph 3: PIE chart representation of polarity score distribution using Textblob

The graph shows the results of sentiment analysis on a set of text data using Textblob. Of these, 38.2% are classified as positive sentiment, 20.9% as negative sentiment, and 40.8% as neutral sentiment.



Graph 4 and 5 : Sentiment polarity distribution per each year

Based on the sentiment analysis of the five-year posts, it can be observed that the sentiment polarity of the posts varied across different years. In 2018, the sentiment was predominantly positive, with only a few negative and neutral posts. In 2019 and 2020, the sentiment polarity was more balanced, with a significant number of positive and negative posts and some neutral posts. In 2021, there were more positive than negative or neutral posts, while in 2022, the sentiment polarity was mostly negative, with a slightly higher number of positive posts. But in general, there have been slightly more positive than negative posts across the years whereas the neutral posts remain consistent.

The highest number of sentiment values were observed in the year 2022, with 77 neutral, 71 positive, and 40 negative sentiment values. The year 2019 had 52 neutral, 48 positive, and 30 negative sentiment values. The year 2020 had 41 positive, 41 neutral, and 16 negative sentiment values. The year 2021 had 39 neutral, 34 positive, and 19 negative sentiment values. Finally, the year 2018 had 25 neutral, 25 positive, and 15 negative sentiment values. Overall, the data shows that the sentiment values were evenly distributed across the different years, with no significant outliers. The data suggests that Neutral sentiment is the most common sentiment across all years, followed by Positive sentiment, and Negative sentiment is the least common.

5. Conclusion:

In conclusion, the sentiment analysis conducted on a dataset of 573 post titles related to climate change using VADER and TEXTBLOB has provided insights into the public's sentiment towards climate change. The analysis shows that there has been a generally positive and neutral tone in the communication related to the data source over the years 2018-2022, with only minor variations in the percentage breakdown of positive, neutral, and negative sentiment. The sentiment count varies significantly between years, and the majority of the sentiment being expressed is either positive or negative, with very little falling into the neutral category. The results suggest that finding solutions to climate change may not be as much of a priority for Reddit users as discussing the problem itself, and there may be a lack of interest or understanding among Reddit users regarding the scientific aspects of climate change. The sentiment analysis results can be useful for policymakers, climate change activists, and organizations working towards raising awareness about climate change.

6. Limitations:

Ambiguity of sentiment: Sentiment analysis tools are not perfect and may not always accurately capture the intended sentiment of a text. This is especially true for sarcastic or ironic statements, which can be difficult for sentiment analysis tools to classify.

Lack of context: The sentiment analysis was conducted solely based on the titles of posts. This means that the analysis may not capture the full context or meaning of a post, which can impact the accuracy of the sentiment classification.

Limited timeframe: The sentiment analysis was conducted only over a 5-year period. This means that the findings may not reflect long-term sentiment trends.

7. References:

Zeng, L. (2022). Chinese Public Perception of Climate Change on Social Media: An Investigation Based on Data Mining and Text Analysis. Journal of Environmental Public Health, 2022, 6294436. **doi: 10.1155/2022/6294436.**

Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Rosenthal, S. (2020). Climate change in the American mind: April 2020. Yale University and George Mason University. **https://climatecommunication.yale.edu/publications/climate-change-in-the-american-mind-april-2020/**

Fownes, J. R., Yu, C., & Margolin, D. (2018). Twitter and climate change: an analysis of tweets relating to the Paris climate agreement. Sociology Compass, 12(3), e12587. **doi: 10.1111/soc4.12587**