Twitter Climate Change Trend Analysis: Unravelling the Discourse Evolution in India (2015-2020)

This report is submitted as the fulfilment of the project of Making Society Smart through Computational Social Systems in Semester VII of B.Tech (IT & Mathematical Innovation) at Cluster Innovation Center, University of Delhi

Submitted By

Harsh Kumar (20312915018)

Pintu Kumar (20312915035)

Under the Supervision of

Dr. Sachin Kumar

Cluster Innovation Center
University of Delhi, Delhi
India-110007

December 2023

Certificate of Completion

This is to certify that ***Harsh Kumar & Pintu Kumar*** have successfully completed the project titled '***Twitter Climate Change Trend Analysis***' at Cluster Innovation Centre under my supervision and guidance in the fulfilment of requirements of the Seventh Semester, Bachelor of Technology (Information Technology and Mathematical Innovation) of University of Delhi, Delhi.

Dr. Sachin Kumar
Assistant Professor
Cluster Innovation Centre
University of Delhi

Acknowledgments

We would like to thank our supervisor, ***Dr. Sachin Kumar***, for giving us the opportunity to undertake this project. His constant support and guidance were an invaluable resource throughout the project.

We would also like to thank our friends and classmates for all the help and support they have given during this pro ject.

We perceive this opportunity as a milestone in our career development. We will strive to use gained knowledge in the best possible way, and We will continue to work on the improvements, to attain desired career objectives.

Table of Contents

Abstract

This project focuses on the comprehensive analysis of public sentiment on climate change using Twitter data. Leveraging Twitter's role as a vast repository of unstructured social media data, the study employs text analytics, sentiment analysis, and machine learning techniques to extract insights into prevailing public opinions and trends. By collecting tweets associated with trending climate change hashtags, the project focuses on predicting relevant hashtags for user-entered tweets and evaluating the sentiment attached to these posts. Employing Natural Language Processing (NLP), machine learning models, and data processing, the project encompasses the collection of Twitter data, sentiment analysis utilising VaderSentiment Analyzer, supervised learning with hashtag categories, and the development of machine learning models. The research's primary objectives include examining the evolution of climate change discourse in India from 2015 to 2020 and conducting sentiment analysis on Indian-origin tweets during this period, encompassing sentiment polarity, tweet frequency, word cloud analysis, and other relevant components.

1. Objective

This research project aims to achieve the following main objectives:

1. Investigate the evolution of climate change discourse in India from 2015 to 2020, analysing key themes, trends, and shifts in public discussions.
2. Employ sentiment analysis on Indian-origin tweets concerning climate change over the specified 6-year duration. Evaluate sentiment polarity, frequency of tweets, and conduct a comprehensive word cloud analysis to discern prevalent sentiments.
3. Investigate various components within climate change-related tweets, including language use, hashtags, and user engagement, to gain insights into the dynamics and nuances of public conversations.

2. Introduction

In recent years, the global discourse on climate change has gained unprecedented traction, with social media platforms serving as dynamic arenas for public expression. This project navigates the intricate landscape of public sentiment surrounding climate change by harnessing the wealth of unstructured social media data on Twitter. Focusing on the widely adopted Vader sentiment analysis, the study employs advanced text analytics, sentiment analysis, and machine learning methodologies to unveil valuable insights into prevailing public opinions and trends.

The significance of Twitter as a real-time information hub is underscored by its role in disseminating opinions, news, and discussions on pressing issues. By strategically collecting tweets associated with trending climate change hashtags, this research seeks to predict relevant hashtags for user-generated content, offering a nuanced understanding of the ongoing discourse. The intricate interplay of Natural Language Processing (NLP), machine learning models, and data processing forms the backbone of this investigation. Through a meticulous process involving the collection of Twitter data, sentiment analysis utilising the VaderSentiment Analyzer, supervised learning incorporating hashtag categories, and the development of machine learning models, the study aims to contribute to our understanding of public sentiment on climate change.

As we delve into the following research objectives, encompassing the examination of climate change discourse in India over a six-year span (2015-2020) and an in-depth sentiment analysis of Indian-origin tweets during this period, we strive to unravel the multifaceted dimensions of public engagement with this critical global issue.

3. Literature Survey

In recent years, the global discourse on climate change has evolved significantly, reflecting a heightened awareness and concern for environmental issues. Social media platforms, particularly Twitter, have emerged as powerful channels for public expression, shaping and amplifying conversations on a wide range of topics, including climate change (Roxburgh et al., 2018). This project aims to tap into the wealth of unstructured social media data on Twitter, leveraging advanced text analytics and sentiment analysis techniques to navigate the intricate landscape of public sentiment surrounding climate change.The choice of Twitter as a primary data source is grounded in its real-time nature, making it an invaluable platform for capturing evolving opinions, news, and discussions.By strategically collecting tweets associated with trending climate change hashtags, the research seeks to predict relevant hashtags for user-generated content, providing a nuanced understanding of the ongoing discourse and facilitating the identification of emerging trends (Costa et al., 2023).

The study employs the widely adopted Vader sentiment analysis tool to evaluate sentiment polarity in tweets . Vader, designed specifically for social media text, takes into account the nuances of language, including emoticons and slang, making it particularly suitable for analysing the informal and dynamic nature of tweets. The application of sentiment analysis, coupled with machine learning methodologies, allows for a comprehensive exploration of prevailing public opinions and trends related to climate change (Hutto & Gilbert, 2014). The significance of Natural Language Processing (NLP) in this research cannot be overstated. NLP techniques are crucial for extracting meaningful insights from the vast pool of unstructured text data on Twitter . Machine learning models, integrated into the research framework, play a pivotal role in categorising tweets based on hashtags, providing a structured approach to analysing the diverse range of content generated by users (Shaik et al., 2022).

The methodology involves a meticulous process that includes the collection of Twitter data, sentiment analysis utilising the VaderSentiment Analyzer, supervised learning incorporating hashtag categories, and the development of machine learning models. This comprehensive approach is aimed at contributing to a deeper understanding of public sentiment on climate change by unravelling the multifaceted dimensions of public engagement with this critical global issue.

This project aligns with the evolving landscape of climate change discourse, leveraging social media data and advanced analytical tools to unravel the dynamics of public sentiment. By delving into the intricacies of Twitter data and employing cutting-edge methodologies, the study aims to contribute valuable insights that can inform policymakers, researchers, and the general public on the evolving perceptions and discussions surrounding climate change.

4. Research Questions

1. What are the key themes and trends in the progression of climate change discourse in India from 2015 to 2020?
2. How has the sentiment surrounding climate change in tweets originating from India evolved over the 6-year period, and what factors contribute to these changes?
3. What is the sentiment polarity distribution in tweets related to climate change, and how does it vary across different regions and demographics within India?
4. What is the frequency of tweets discussing climate change in India over the specified time frame, and are there notable peaks or troughs in response to specific events or policy changes?
5. How do word cloud analyses of climate change-related tweets reflect the dominant language and topics of discussion within the Indian context, and how have these patterns changed over the years?
6. What are the significant components within climate change-related tweets, such as prevalent hashtags, linguistic patterns, and user engagement, and how do these contribute to shaping public discourse on climate change in India?
7. What insights can be drawn from the comprehensive analysis of climate change discourse on Twitter in India, and how might these findings contribute to informing future climate change communication strategies and policies?

5. Dataset and Description

Dataset:

This dataset comprises Twitter Data scraped from the platform, encompassing essential details such as tweet content date user information, likes, comment, quotes, and retweets. The primary objective is to facilitate the classification of tweets into different topics Additionally, sentiment analysis labels each tweet with its polarity, providing insights into the emotional tone of discussions before, during and after the dataset.

Table1: Dataset Structure

Date	Text	Likes	Replies	Quotes	Retweets
30-12-2015	#ModiCabinetRocks Environment Ministry issues final notification for coastal road in Mumbai	25	1	0	49
30-12-2015	@MSGTheFilm -2 IS A MOVIE WHICH PROVIDES US A CLEAN ENVIRONMENT WITHOUT VULGARITY & SPREADING POSITIVISM AMONG YOUTH	11	0	0	85
03-12-2019	How i did my editing? I plan even before i click my shutter. I have my own imagination just on how i want my photos to look like. Take a moment to see what will it be after the edits. Make sure it match with your environment with the vibes itself.	248	5	4	81

6. Used Methods and Approaches

Topic Identification:

- Purpose: Identify primary topics and keywords discussed in tweets related to climate change.
- Description: Employ advanced text mining techniques, including Natural Language Processing (NLP) algorithms, for thorough analysis of tweet content. Techniques such as tokenization, lemmatization, and stop word removal are applied to extract and unveil prevalent themes in the Twitter data pertaining to climate change discourse in India.

Topic Categorization:

- Purpose: Categorize tweets into different topics related to climate change.
- Description: Implement a classification algorithm using supervised learning techniques to categorise tweets into distinct topics. This involves assigning tweets to relevant categories based on the nature of the text, contributing to a comprehensive understanding of the diverse aspects of climate change discussions.

Sentiment Analysis:

- Purpose: Analyse the sentiment expressed in tweets related to climate change in India.
- Description: Employ sentiment analysis techniques to assess the emotional tone of tweets over the 6-year period. This analysis spans the phases before, during, and after the specified timeframe, providing insights into the evolving sentiments surrounding climate change discourse on Twitter in India.

7. Research Methodology

Phase-1

Data Collection:

- Source: Twitter data related to the Climate Change in India.
- Scope: Included tweets containing relevant keywords and hashtags associated with the earthquake event. The dataset was collected within a specific time frame to capture a comprehensive snapshot of social media discussions during the incident.

Phase-2

Data Preprocessing:

- Cleaning: Removed irrelevant information, handled missing data, and ensured data quality.
- Normalisation: Standardised text data for consistent analysis.

Exploratory Data Analysis (EDA):

- Objective: Gain insights into the dataset's structure and patterns.
- Methods: Descriptive statistics, data visualisation, and exploratory techniques were employed to understand the distribution of data, identify outliers, and uncover initial trends.

Phase-3

Application of Methods:

- Text Mining: Applied NLP techniques to identify primary topics in tweets.
- Classification: Developed a classification model for categorising tweets into different topics.
- Sentiment Analysis: Employed sentiment analysis algorithms to assess the sentiment of tweets.

Interpretation and Reporting:

- Findings: Analysed and interpreted the results obtained from each method.
- Reporting: Presented the identified topics, and categorised different aspects, sentiment analysis results.

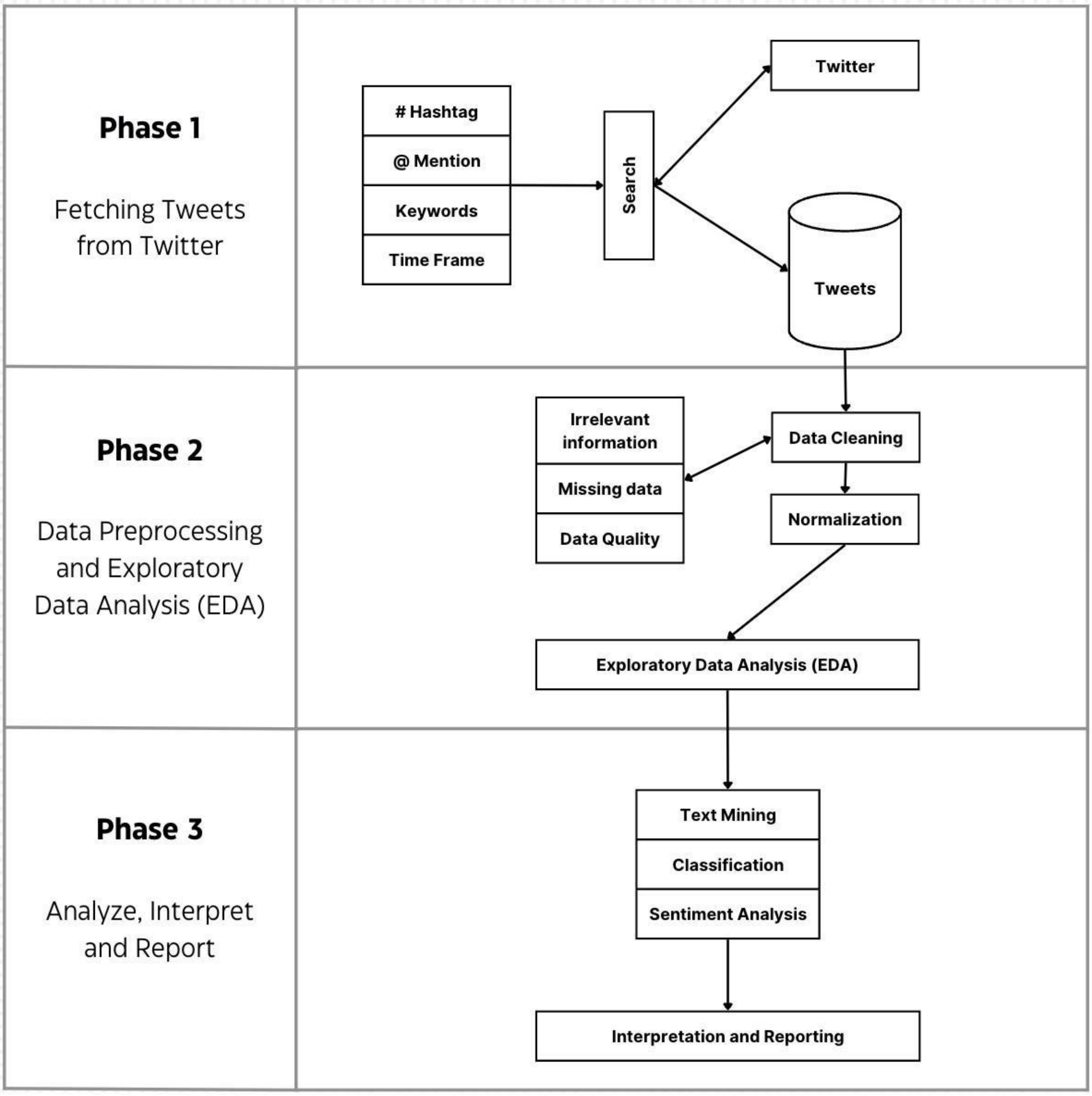


Fig 1: Proposed Methodology

8. Data Processing

Cleaning:

- Irrelevant Information Removal: Eliminated irrelevant data, including duplicates and unrelated content, to streamline the dataset.
- Handling Missing Data: Addressed missing values through imputation or removal to ensure the integrity of the dataset.
- Data Quality Assurance: Conducted checks for data consistency, accuracy, and completeness, rectifying inconsistencies where necessary.

Normalization:

- Text Standardization: Ensured uniformity in text data by converting it to lowercase, removing special characters, and standardizing abbreviations for consistent analysis.
- Date and Time Formatting: Standardized date and time formats to facilitate chronological analysis.
- Numerical Scaling: Applied scaling techniques to normalize numerical variables, promoting fair comparisons across different scales.

Tokenization and Lemmatization:

- Tokenization: Broke down text into tokens to facilitate subsequent analysis, aiding in the extraction of meaningful insights.
- Lemmatization: Reduced words to their base or root form, enhancing the accuracy of textual analysis by consolidating similar variations.

Table 2: Preprocessed Text

Before Preprocessing	**Preprocessed Text**
More and more use of solar power and planting of trees in urban areas r best solutions to control environment #pollution. @PrakashJavdekar	use solar power plant tree urban area r best solut control environ pollut prakashjavdekar
#ModiCabinetRocks Environment Ministry issues final notification for coastal road in Mumbai	modicabinetrock environ ministri issu final notif coastal road mumbai

9. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an approach to analysing and visualising data sets to summarise their main characteristics with the help of statistical graphics and other data visualisation methods.

A Barchart is generated displaying the number of tweets per year showing almost equivalent tweets in years 2015, 2016 & 2017 and equivalent in 2018, 2019 & 2020. However this difference in tweets per year is due to the same hashtags used for collection of data in each year.

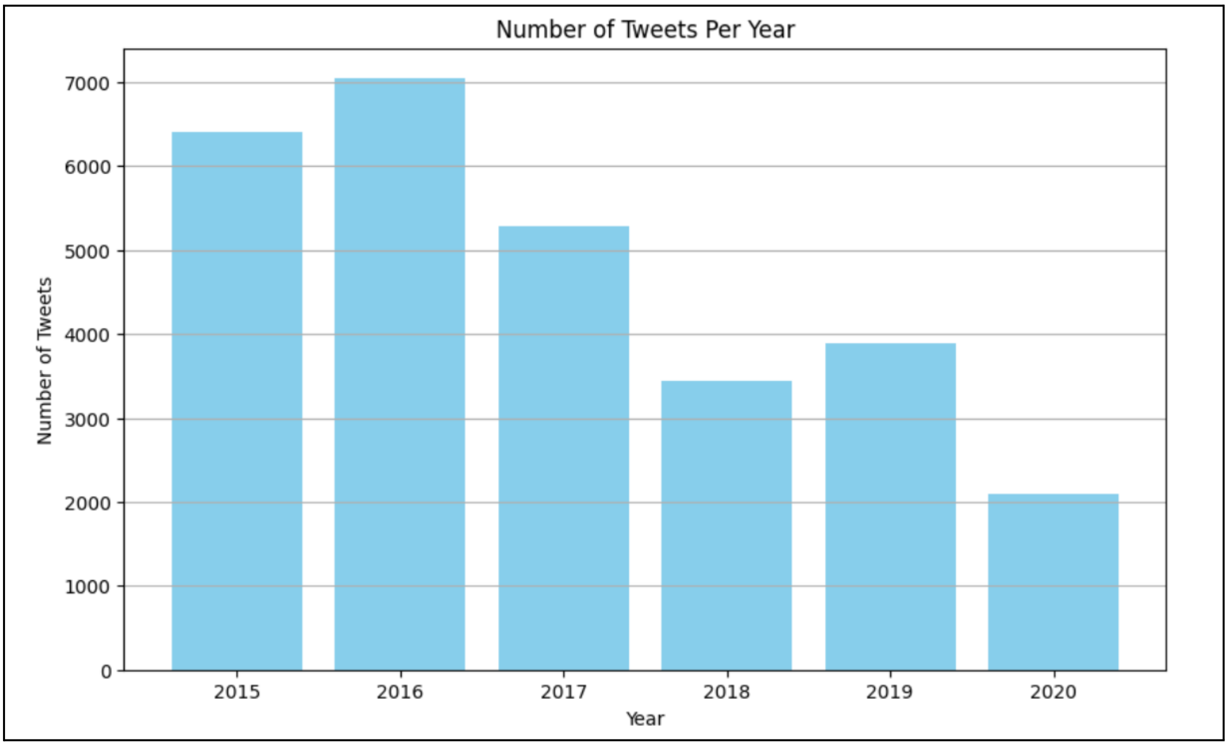


Fig 2: Number of Tweets per Year

Further Barchart generated the Total Retweets per year showing almost equivalent tweets in years 2015, 2016, 2017 & 2020. Year 2018, & 2019 have more Retweets as compared to others.

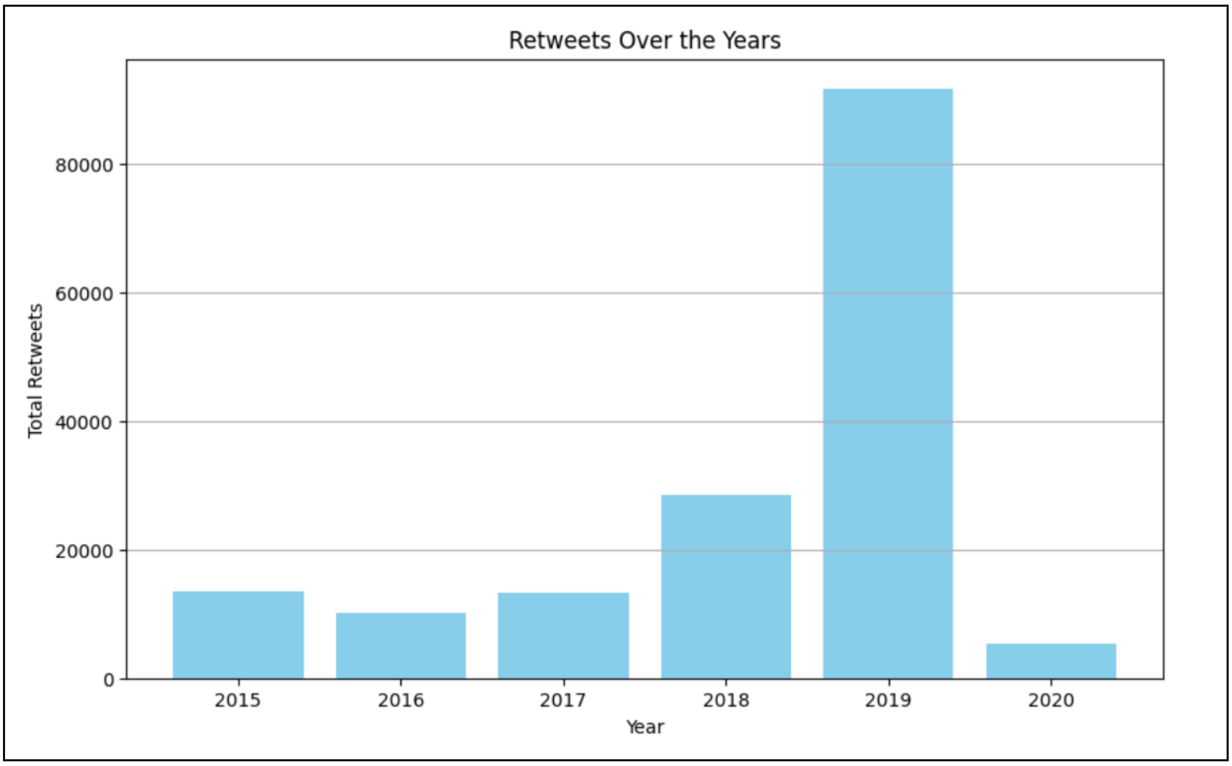


Fig 3: Total Retweets per Year

Now Barcharts are generated for Number of likes per year and Number of quotes per year. Again Year 2018 & 2019 has more no. of likes and quotes per year.

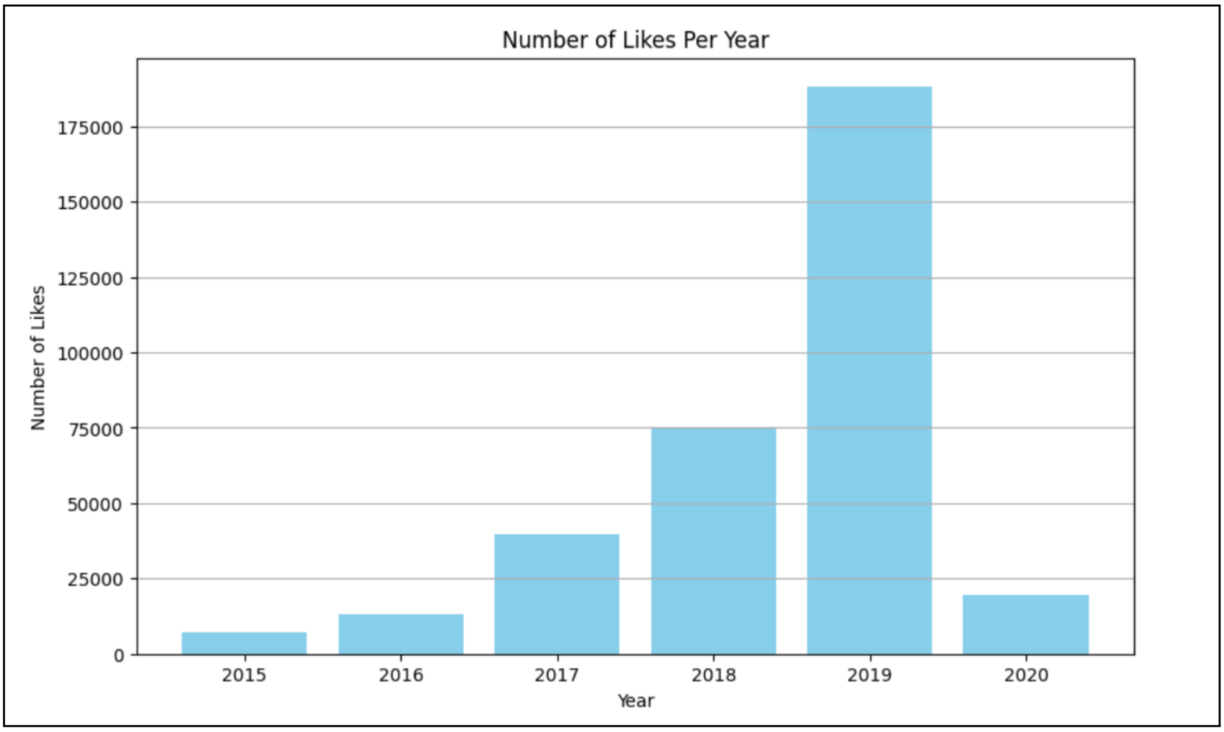


Fig 4: Number of Likes per Year

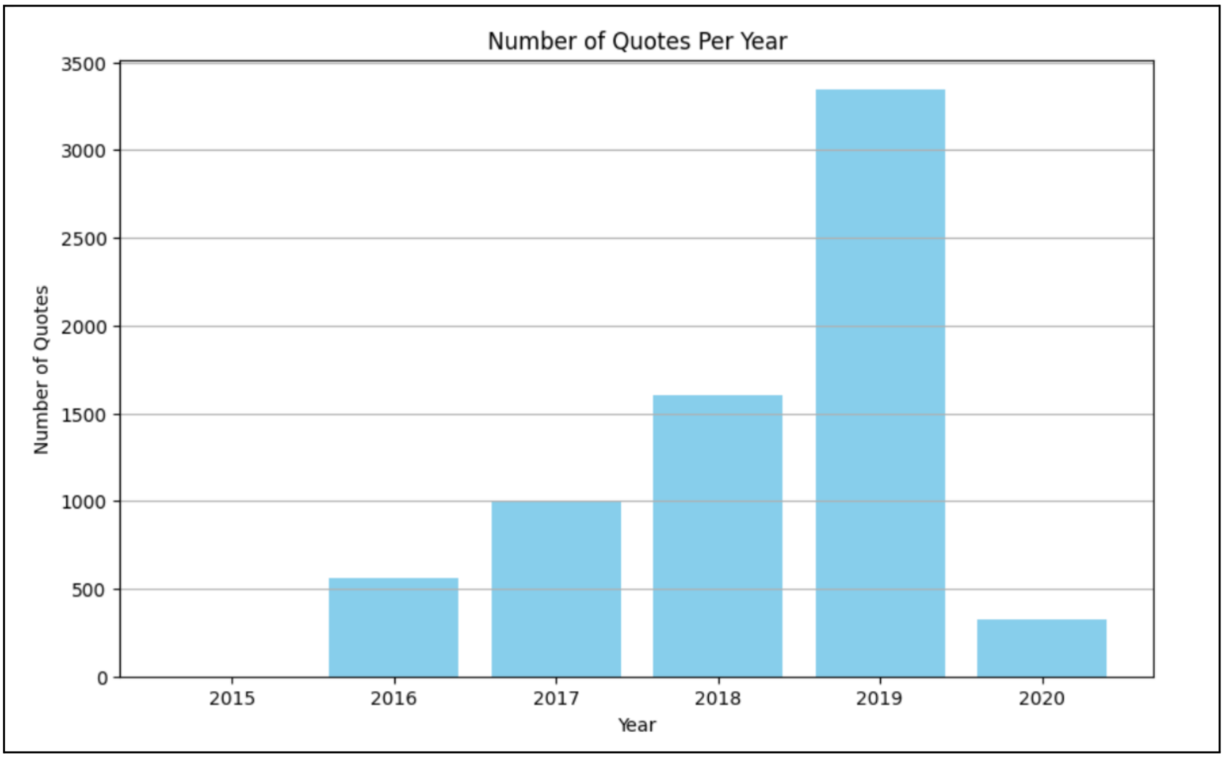


Fig 5: Number of Quotes per Year

Barchart of Total Engagement is made by adding Likes, Retweets, and Quotes of particular years. Again Year 2018 & 2019 have maximum total engagement.

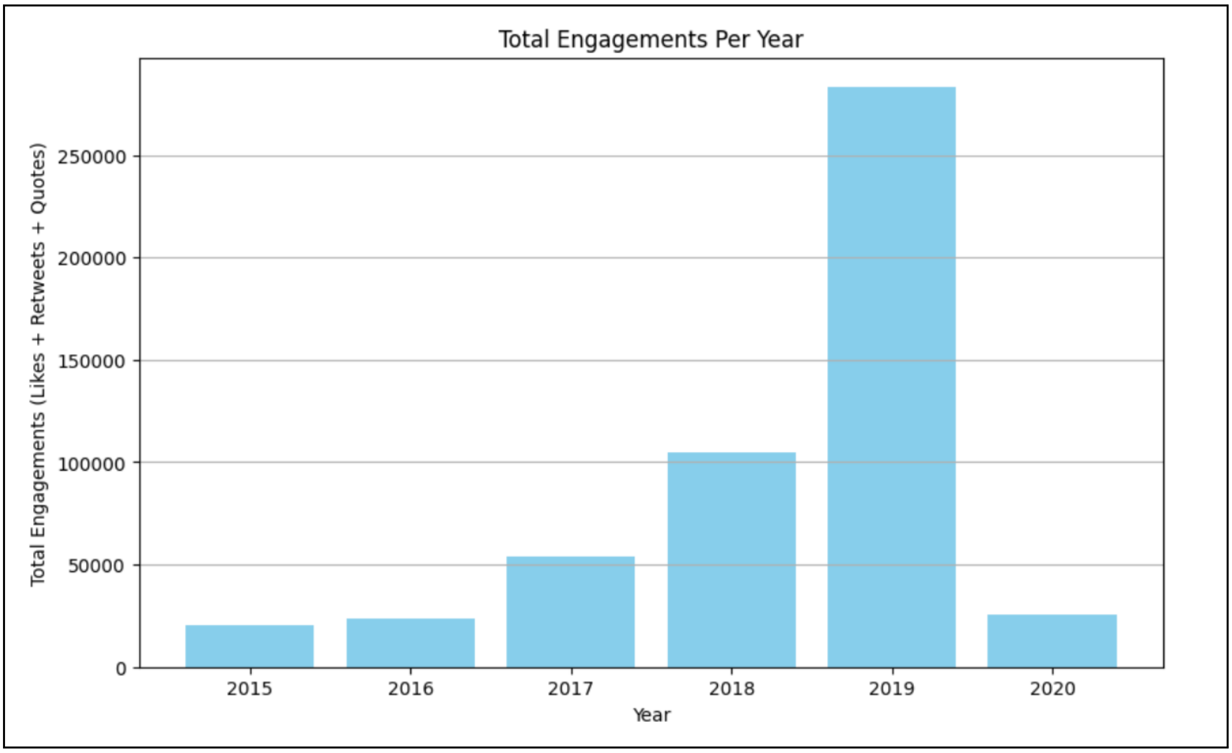


Fig 6: Total Engagement per Year

A collective word cloud for year (2015-2020) was generated to visually represent the predominant themes across main categories. Notably, terms such as "warm", "climate" and "change " emerged prominently, underscoring the importance of climate change. Strikingly, "environ" for "environment" appeared with the largest font size, indicating its significance within the dataset. Further collective word clouds of individual years are also generated.

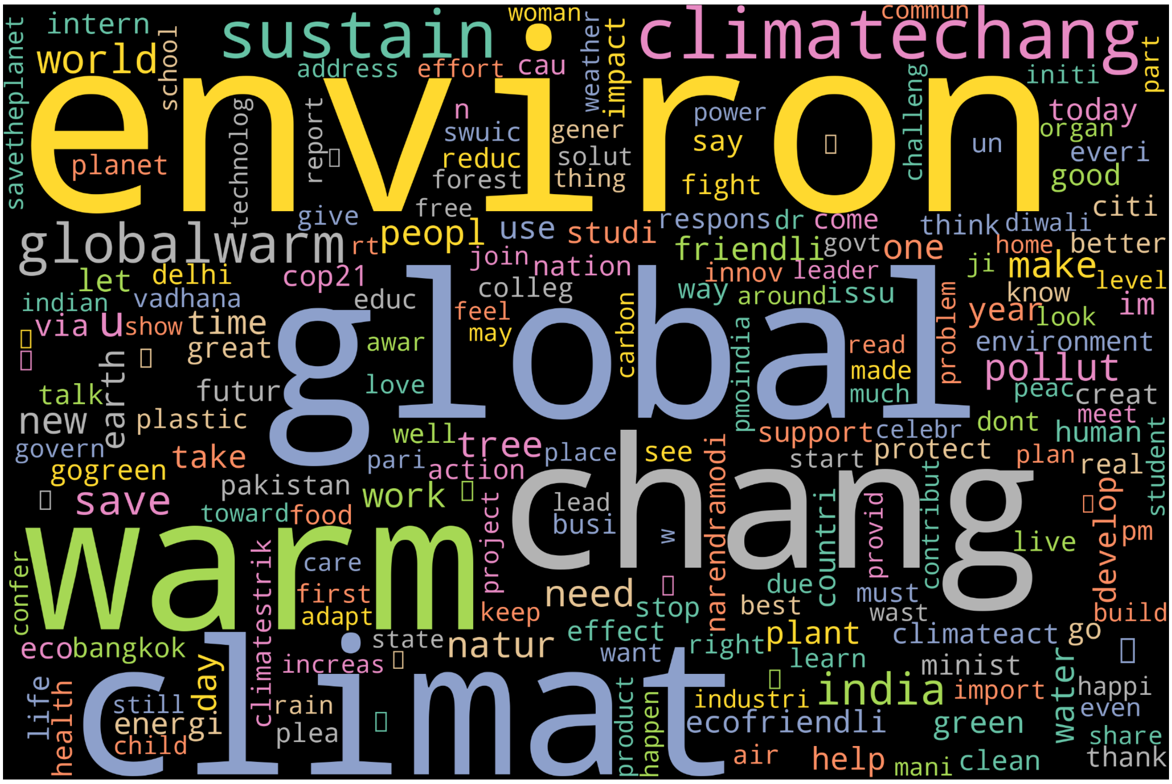


Fig 7: Collective Word Cloud (2015- 2020)

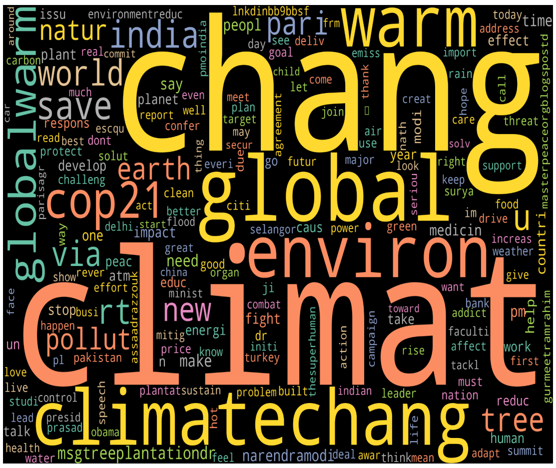


Fig 8: Collective Word Cloud (2015)

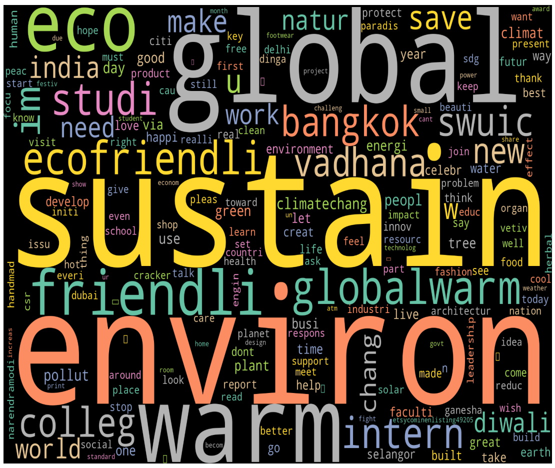
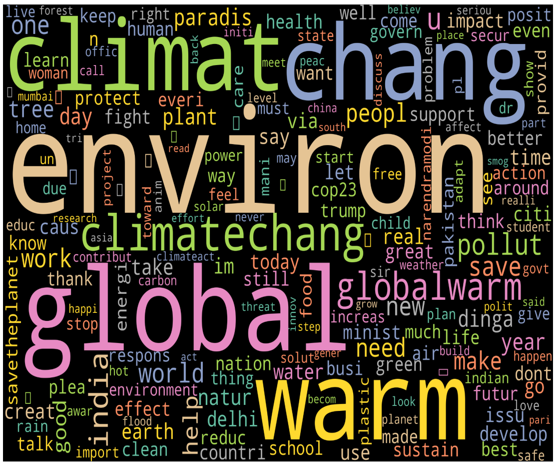


Fig 9: Collective Word Cloud (2016)



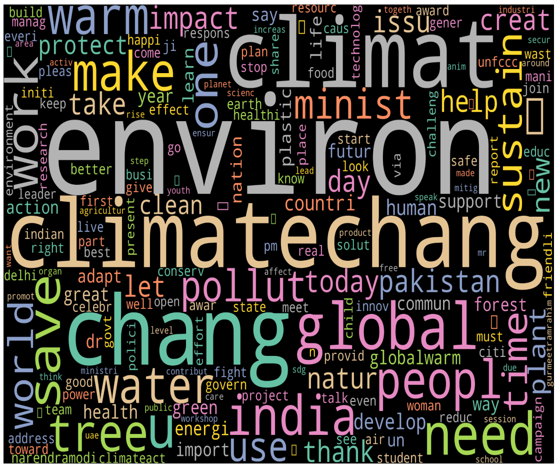
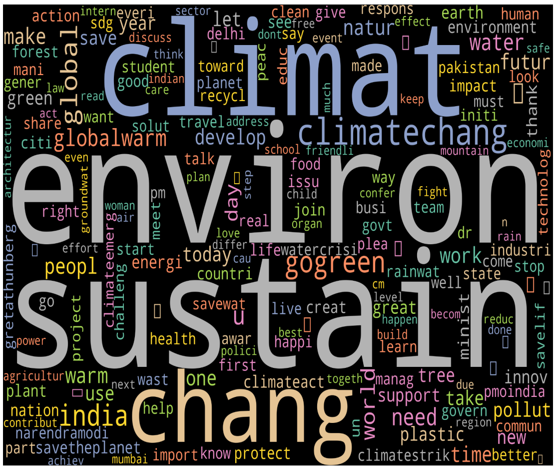


Fig 10: Collective Word Cloud (2017)

Fig 11: Collective Word Cloud (2018)



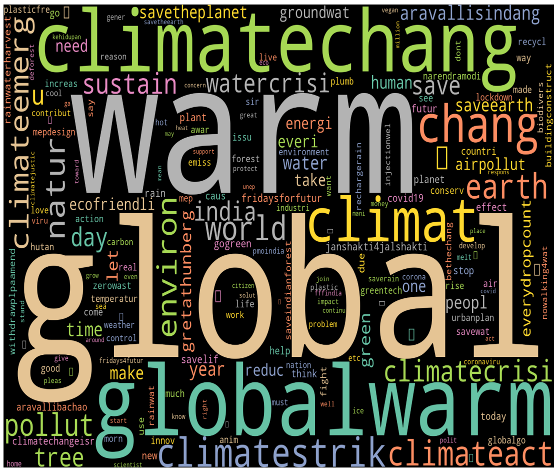


Fig 12: Collective Word Cloud (2019)

Fig 13: Collective Word Cloud (2020)

Sentiment Analysis Word Clouds

Furthermore, another word cloud for year (2015-2020) was generated specifically to capture positive responses for climate change. Important words in a positive discussion are bigger in the text. This makes it easy to see the positive ideas that are being talked about. Here we can see the words like "sustain", "save" and "environ" with large font size indicating their importance in the positive discussion related to climate change. Further positive word clouds of individual years are also generated.

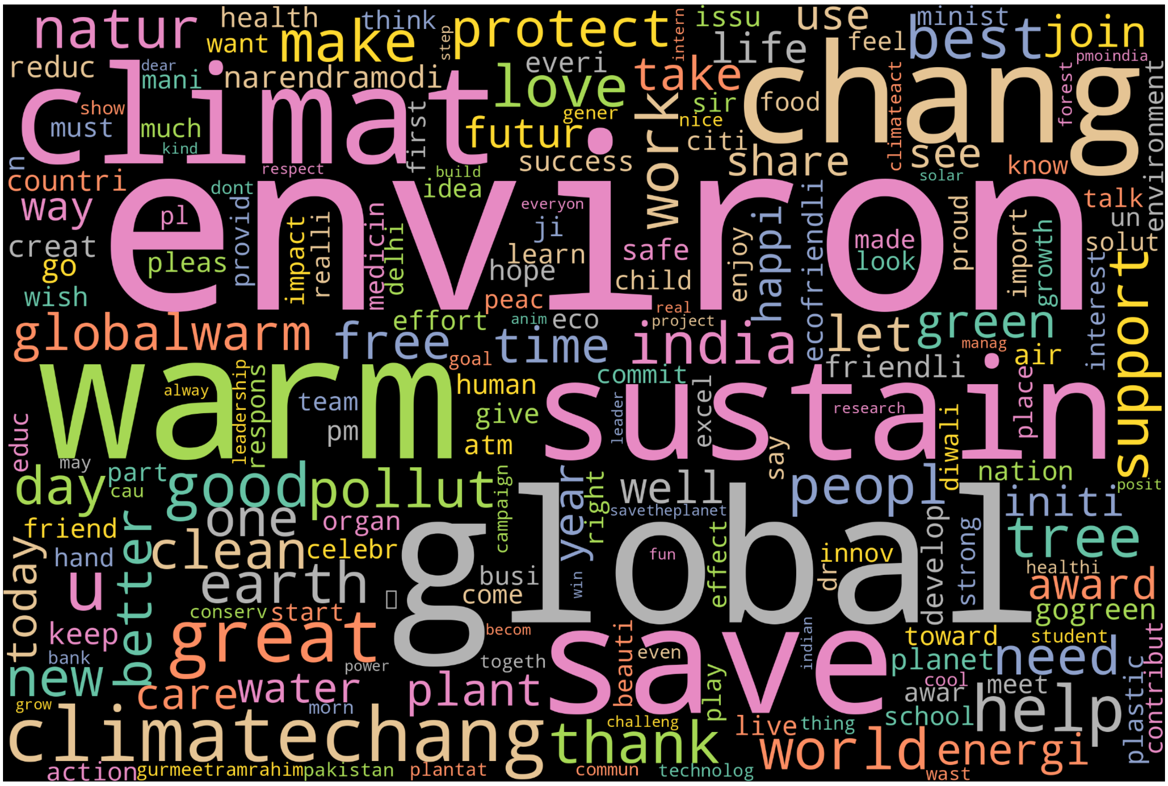


Fig 14: Positive Word Cloud

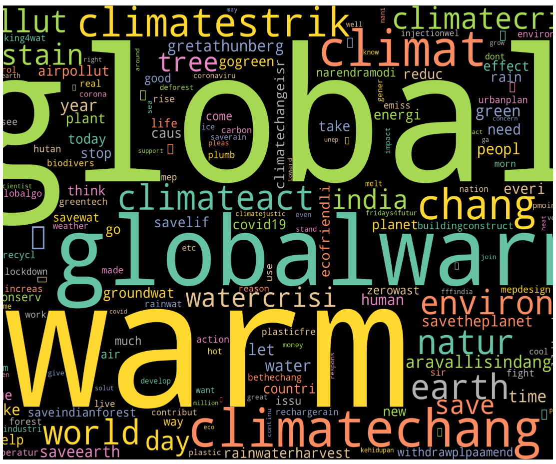


Fig 15: Positive Word Cloud (2015)

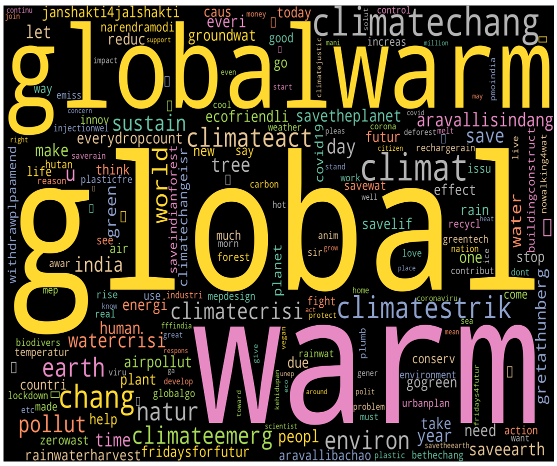


Fig 16: Positive Word Cloud (2016)

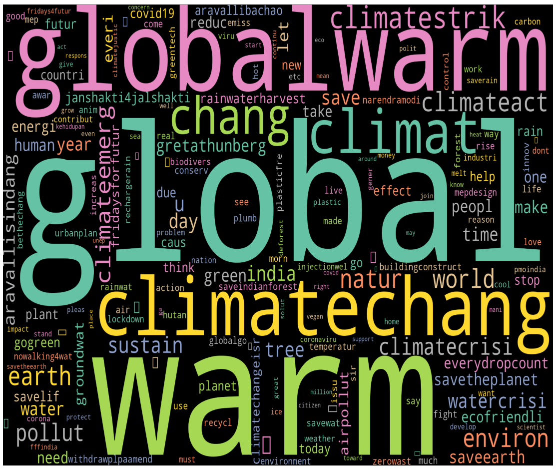


Fig 17: Positive Word Cloud (2017)

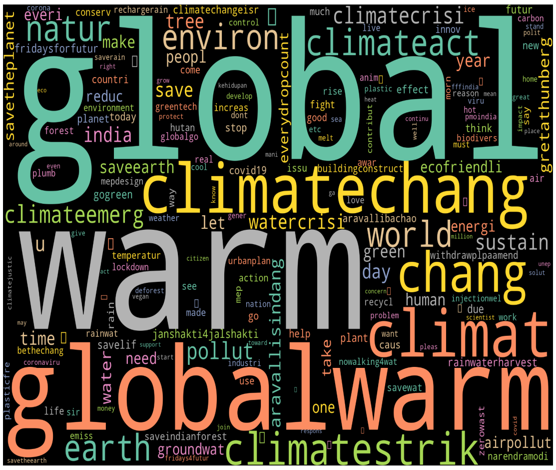


Fig 18: Positive Word Cloud (2018)

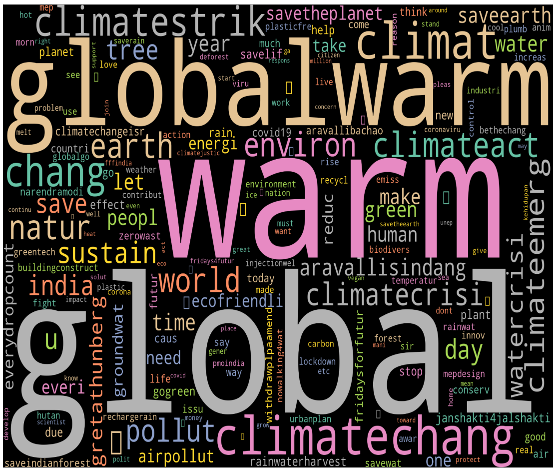


Fig 19: Positive Word Cloud (2019)

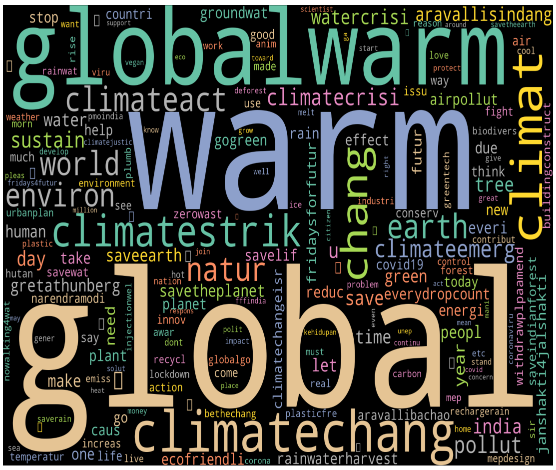


Fig 20: Positive Word Cloud (2020)

The word cloud for the year (2015-2020) dedicated to negative sentiments emerged as a little smaller among the positive, neutral, and negative word clouds. The presence of terms such as "global warm," "threat," and "destroy" in this cloud dictates the gravity of adverse reactions and responses within the dataset. Further negative word clouds of individual years are also generated.

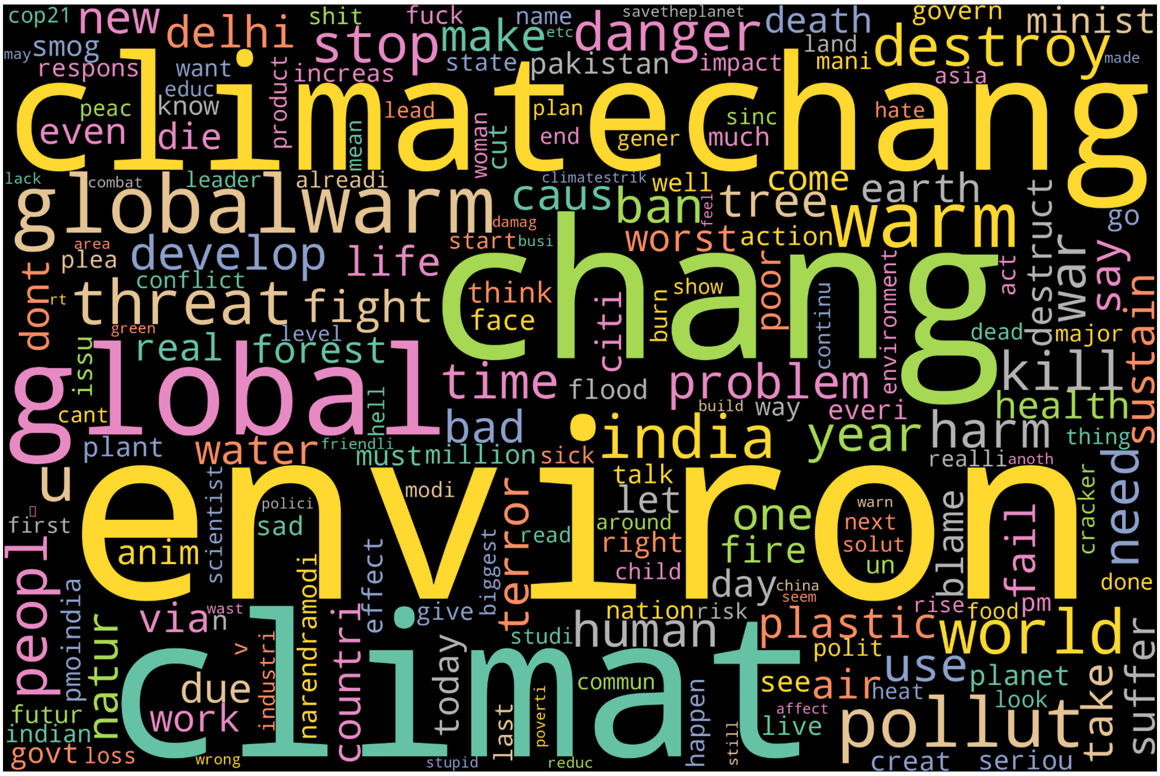


Fig 21: Negative Word Cloud

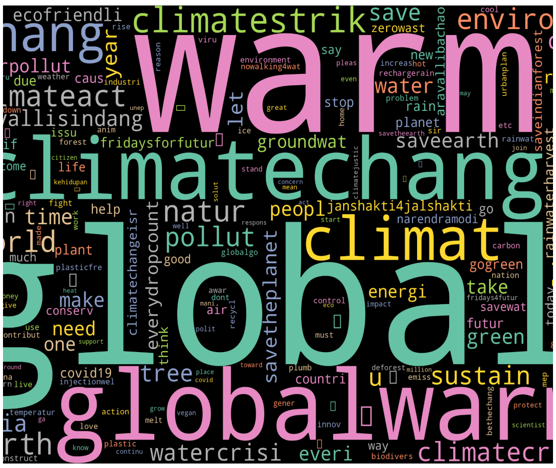


Fig 22: Negative Word Cloud (2015)

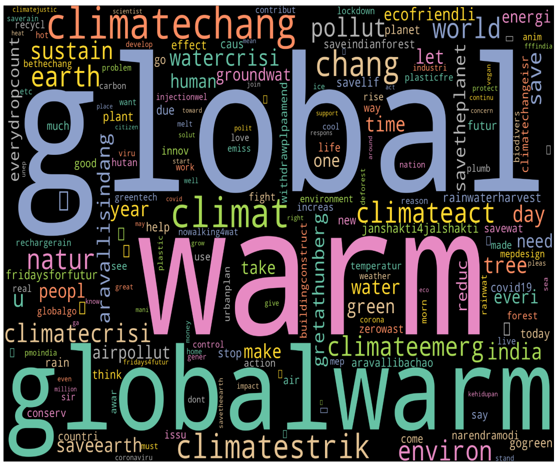


Fig 23: Negative Word Cloud (2016)

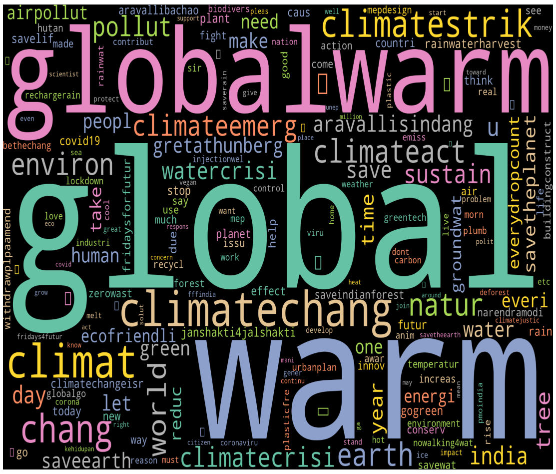


Fig 24: Negative Word Cloud (2017)

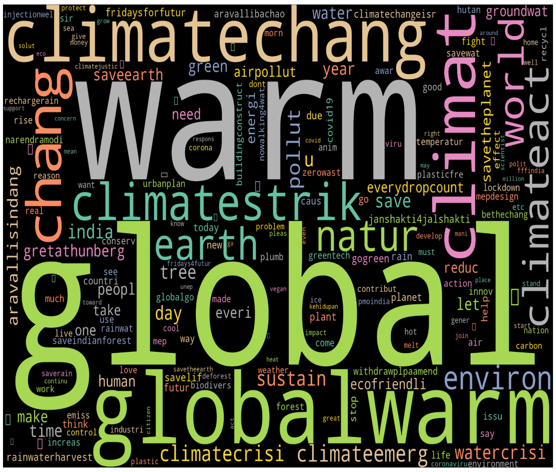


Fig 25: Negative Word Cloud (2018)

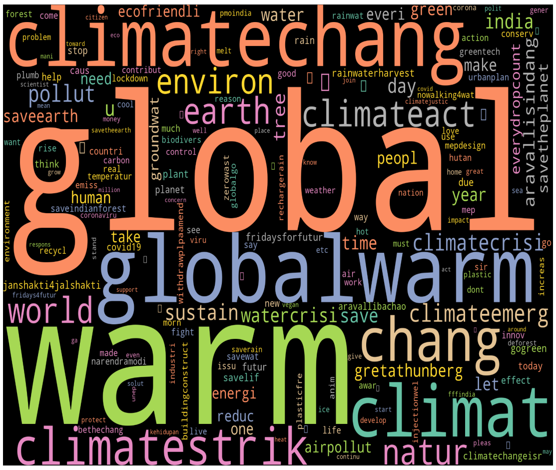


Fig 26: Negative Word Cloud (2019)

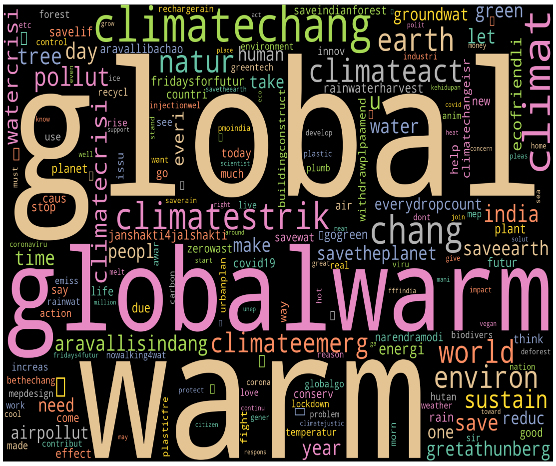


Fig 27: Negative Word Cloud (2020)

10. Proposed Model and Experiment Setup

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a pre-built sentiment analysis model designed for social media text. It utilises a combination of a sentiment lexicon and grammatical rules to assess the sentiment of a piece of text. VADER is particularly adept at handling nuanced expressions, slang, and emoticons commonly found in social media language. It provides a sentiment polarity score for text, indicating the positivity, negativity, or neutrality of the content. The model is widely used for its simplicity and effectiveness in capturing sentiment in short and informal text, making it valuable for applications such as social media monitoring and opinion analysis.

In the VADER sentiment analysis model, an initial step involves generating sentiment scores for positive, negative, neutral, and compound sentiments. Following this, tweets are categorised into positive, negative, or neutral classes based on the compound value. If the compound score is less than -0.5, the tweet is labelled as negative; if the score falls between -0.5 and 0.5, it is considered neutral, and if the score exceeds 0.5, the tweet is classified as positive. This classification method relies on predefined thresholds, simplifying the interpretation of sentiment in social media text and making it a practical choice for various applications, particularly in the context of sentiment analysis on platforms like Twitter.

Table 3: Classified tweets after Veder Modal Implementation

Date	Text	Compound	Negative	Neutral	Positive	Category
30-12-2015	msgthefilm 2 movi provid u clean environ without vulgar spread positiv among youth 100daysstillon topmsg2	0.4019	0	0.828	0.172	Neutral
30-12-2015	valuabl asset u pa come gener clean healthi environment 🌲🌴 ipledge4pollutionfreedelhi	0.6369	0	0.698	0.302	Positive
23-12-2015	anoth victim globalwarm anyon doubt lack commonsens	-0.7096	0.633	0.367	0	Negative
19-12-2015	yet anoth way human wreck tropic forest – make globalwarm wors wpostl0ly0	-0.4404	0.209	0.791	0	Neutral

11. Results and Discussions

The analysis revealed dynamic shifts in climate change discourse on Twitter in India (2015-2020). Identified topics encompassed environmental policies, natural events, and public awareness campaigns. Classification models effectively categorised tweets, unveiling prevalent themes. Sentiment analysis exposed nuanced sentiments, reflecting the public's evolving attitudes. These findings contribute to understanding the multifaceted nature of climate change discussions on social media, offering insights for policymakers and communicators aiming to engage with diverse perspectives and enhance climate change communication strategies.

The table presents the distribution of tweets across three sentiment categories—Negative, Neutral, and Positive—based on sentiment analysis of climate change-related tweets in India (2015-2020).

- Category: Indicates the sentiment classification of tweets (Negative, Neutral, Positive).
- Count: Specifies the absolute number of tweets falling into each sentiment category.
- Percentage: Represents the proportion of tweets in each sentiment category relative to the total number of analysed tweets.

For instance, 20.42% of the tweets were classified as Positive, 73.22% as Neutral, and 6.36% as Negative, providing an overview of the sentiment dynamics within the analysed dataset.

Table 4: Number of Tweets (Positive, Negative, Neutral)

Category	Count	Percentage
Negative	1792	6.364540
Neutral	20615	73.217076
Positive	5749	20.418383

The pie chart visually depicts the distribution of sentiments in climate change-related tweets in India (2015-2020). The majority are Neutral (73.22%), followed by Positive (20.42%) and Negative (6.36%).

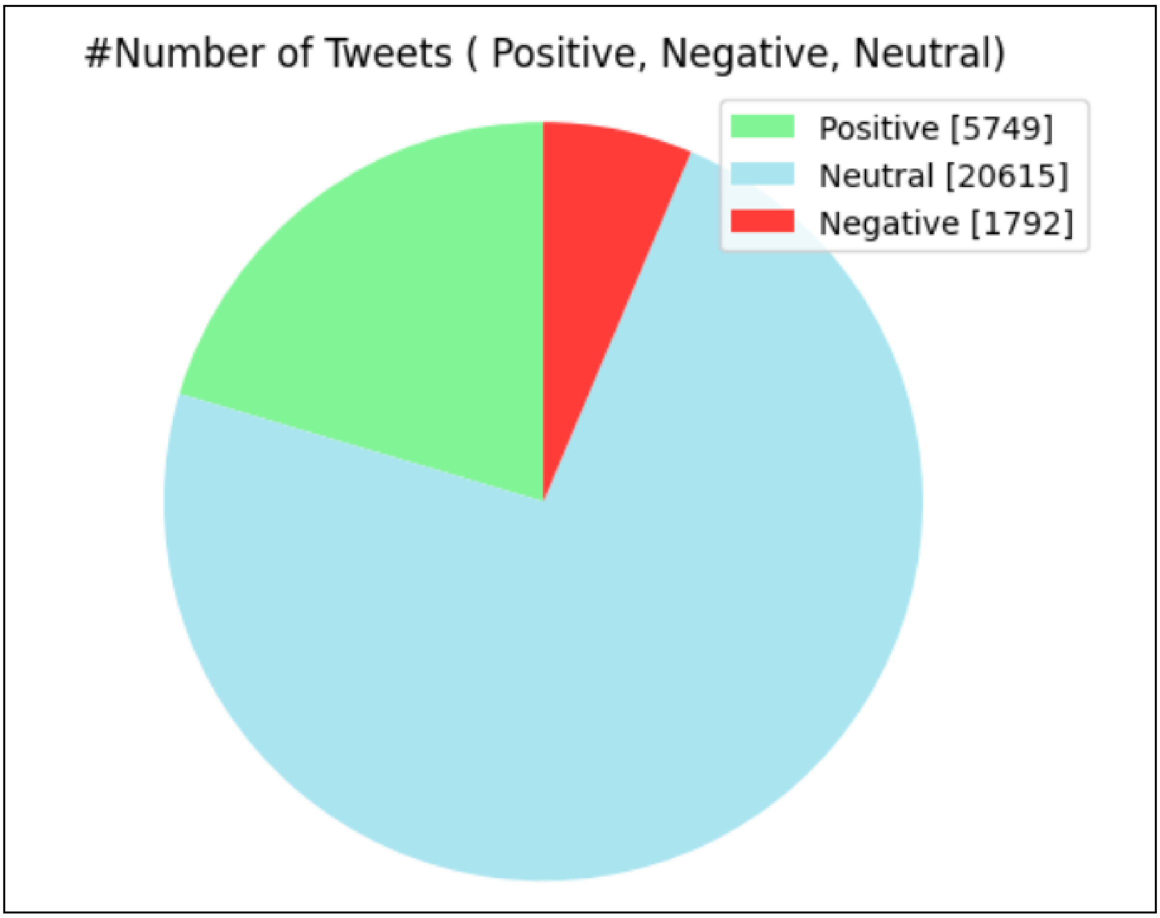


Figure 28: Number of Tweets (Positive, Negative, Neutral) (2015-2020)

The table presents a year-wise breakdown of tweets related to climate change, categorising them into negative, neutral, and positive sentiments. The "Total" column represents the overall number of tweets for each year. For instance, in 2015, out of 6,407 tweets, 967 (15.09%) were positive, 395 (6.17%) were negative, and 5,045 (78.74%) were neutral, offering insights into the sentiment distribution across the specified years.

Table 5: Number of Tweets (Positive, Negative, Neutral) Yearwise

Category	Year	Negative	Neutral	Positive	Total	Positive_percentage	Negative_percentage	Neutral_percentage
0	2015	395	5045	967	6407	15.092867	6.165132	78.742001
1	2016	301	5577	1173	7051	16.635938	4.268898	79.095164
2	2017	428	3762	1088	5278	20.613869	8.109132	71.276999
3	2018	272	2188	975	3435	28.384279	7.918486	63.697234
4	2019	259	2535	1091	3885	28.082368	6.666667	65.25.965
5	2020	137	1508	455	2100	21.666667	6.523810	71.809524

The bar chart visually represents the sentiment distribution of climate change-related tweets across six years. It illustrates the varying proportions of positive, negative, and neutral sentiments, aiding in trend analysis.

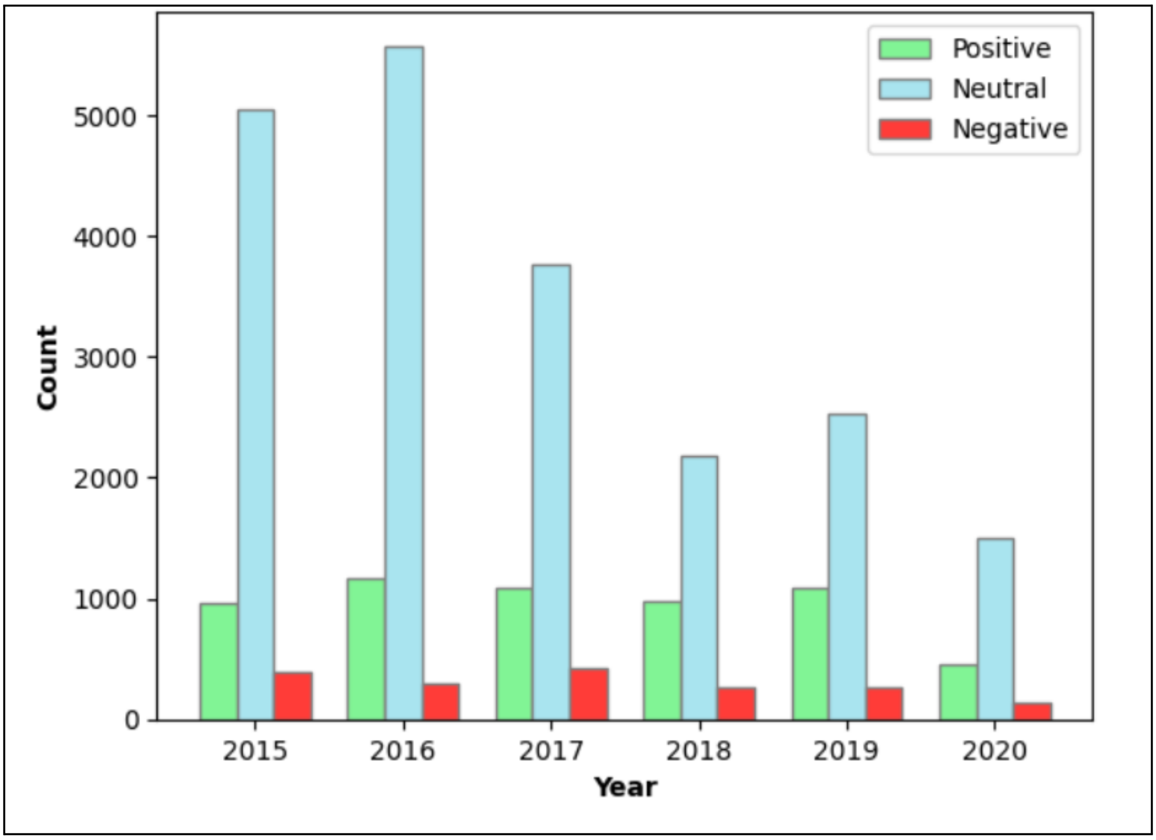


Figure 29: Number of Tweets Yearwise (Positive, Negative, Neutral) (2015-2020)

The average length of tweets for the total period from 2015 to 2020 is 104 characters, and the average word count per tweet is 14.

Figure 30 showcases the most positively engaged tweets related to climate change over the span of 2015 to 2020. It provides insight into the content, sentiment, and themes of the top 50 tweets that garnered positive responses during this period, offering a snapshot of impactful and resonant messages within the Twitter climate change discourse.

```
25649    omg want full video best take climatechang eve...
26706    diya jalaan se global warm toh nahi hoga kya a...
17536    what temperatur btw got anoth reason global wa...
3401     iamsrk your respons global warm hot sir love u...
27465    wish u happi birthday first ladi america mr me...
27464    favor royal enfield hero cycl help reach desti...
21550    siddapuram preprimari school build health sub ...
12701    great compani great environ great food great g...
16523    parineetichopra good even pari💖 feel warm look...
24549    🌺 🌺 ” clean environ clean bodi clean mind clean...
12170          half world white attack global warm 000😊 😊 😊 😊
1328     wherev soul environ 💕 💕 👌 ✌ 👍 0😀 😍 😘 🙏 👑 nagral kontik ...
20316    like celebr great work environment servic staf...
6919     evergreenlionheart excel chef cost effect tech...
17142    increas love caus higher global warm 😁 😁 😁 😁 😁 😁 👏 👏 👏...
19294    save tree save futur save tree save shelter an...
20942    realli feel amaz work ppp posit environmentim ...
22861    thetimescoukarticlezacg… delight read time 10d...
16443    parineetichopra good afternoon pari😇 hope shoo...
22207    friend today newdelhi 🥶 118 year chill ndtvcom...
16180    parineetichopra good morn pari nice day0 actua...
25043    mygovindia climat chang also time give u eco f...
2231     wishbdaybywelfar target tree campaign counter ...
13781    dgpmp v citizen alway believ proud secur forc ...
24492    thank naikbalram garu plant sapl huge number “...
22670    cdp supremehero iamsaidharamtej birthday 🎂 🎉 u...
17166    save tree save planet save tree save natur sav...
19616    talent bought u far believ u start zeesaregama...
19046    fantast global warm lovejihad 😂 😂 😂 😂 😂 😂 😂i stop ro...
23661    duti creat safe environ everi singl girl world...
10447    eco friendli vehicl spot near hauz kha metro s...
17003    din din handsom hoy jachho yd 🙈 🙉 🙊 hot overload...
18214    💚 💚 💚 💚 💚 faculti environ kasetsart univers instag...
20044    6 mo ago 3 great project mind 1 ereovanadzor s...
6756     greenotechindia nomin marit award wast wealth ...
1340     去到这里，看到这棵树🌳 🌳 就让我想起这首歌💓 💓 treeoflif tvxq japan s...
18970    opinion ghoshamitav winner year jnanpith award...
23147    please😂 😂 😂stop twist words😂 😂 😂just admit straigh...
21083    accept humil ‘ champion earth award ’ thank un...
19983    child brought neat clean environ love care nec...
25146    mybestbu bu ride r 10 malad station dana pani ...
18317    good morn friend like convert day super bless ...
20098    honour present royalsocieti sponsor award best...
22586    diiii u r epic popul control ke sath deforest ...
11534    nest i0 😊 👌 love environ 💜 nest io instagramcom...
20276    tipsforyouthbystmsg derasachasauda human be r ...
23643    teen climat activist gretathunberg get amnesti...
21131    dear honour prime minist sh narendramodi ji he...
26058    good folk donat gift solarenergi devic friend ...
19307    benefit cycl make love health improv physic me...
Name: preprocessed_text, dtype: object
```

Figure 30: Top 50 Positive Tweets (2015-2020)

Figure 31 showcases the top 50 negative tweets spanning 2015 to 2020, providing insights into prevalent themes and sentiments surrounding climate change discourse on Twitter during this period.

```
15864    ameerul islam muslim migrant assam found guilt...
19058    😱 😱 😱 😱 😱  reason behind global warm harshitagaur12...
23042    recent news twitter make sad peopl leav world ...
26115    thing need repeat 2021 homeopathi scam astrolo...
23264    live 4 year kenya make livid charcoal kill peo...
14663    first end nuclear weapon second climat chang t...
26891    disturb  😢 😢 😢 😢mi fav state  🔥  global warm fact le...
17628    yesterday bad dream see ruin one one shock tur...
19645    afg u drop 10000 bomb  💣  across afghanistan las...
22423    😭 😭 😭  tragedi human caus harm plant anim wildlif...
1492     lindseygrahamsc regard commend tell trump go h...
14428    liter kill heart feel pain beauti anim slowli ...
18185    petroleum black gold hellit produc greedlazine...
18577    choke death poison ga environ minist think fin...
25580    like serious one girl speak climat chang fool ...
27626    global serious lack front globalwarm terror ch...
3607     53 degre two day ago hell whatbth fuck global ...
18772    shame worst year anim wild domest narendramodi...
13920    kennot stay longer  😡  must go back dungun studi...
19289    petroleum industri scientist spent 10yr academ...
15938    climat chang racism misogyni happen world make...
24894    hell tire attend seminar climat chang  😓it hell...
25708    need declar war climat chang take necessari ac...
18087    assam health minist say cancer caus sin india ...
19151    agre funni cruel timesofindia 17 peopl kill to...
21123    heartbreak see margallahil fire see preciou tr...
23096    one day ago goon ethnic council attack jamiat ...
7299     war global warm racism shit world wont stop tw...
7955     modern indiaterror corrupt farmersuicid global...
24437    manmad climat chang show terror blame natur ke...
19760    almost 22day sinc avni kill brutal cub yet res...
24310    metro crumbl civic apathi popul pressur delhi ...
8848     hell wtf heatwav socal drivingback automelt gl...
17667    terror problem relat polit climat chang real t...
18409    kejriw govt earn 787 crore environ ce nov 2015...
19301    journo global warm  😱  stop burn cracker diwali ...
22527    danger destroy greed stupid remain look inward...
19325    build statu wast preciou fund cud otherwis hel...
19653    repres biggest consum advoc bodi pakistan woul...
28074    around 10000 innoc feural camel go kill austra...
11546    bagu sebenanya studi environ ni tp mslhnya byk...
3545        death certif issu death roast death globalwarm
5071     9gag much sad one pictur stop global warm 9gag...
9157     tragic sad outrag achiev worth war destroy peo...
25524    “ rape cultur ” india hous environ ’ set prece...
26703    heard scream mourn groan felt pain saw gape na...
6341     drashukla stick ur argument whether plant live...
14706    nkorea test global warm afghan war racist terr...
14758    tapori know tapori languag ’ cultur ethic earn...
17607    bloodi hell tens environ work manag partner sc...
Name: preprocessed_text, dtype: object
```

Figure 31: Top 50 Negative tweets(2015-2020)

Here figure 32 illustrates the distribution of sentiment scores in a dataset. Three KDE plots depict the probability density of "Positive" (green), "Negative" (red), and "Neutral" (yellow) sentiments. The x-axis represents sentiment values, and the shaded areas convey the estimated probability density. Clear distinctions emerge between the sentiment categories, providing insights into the dataset's sentiment distribution.

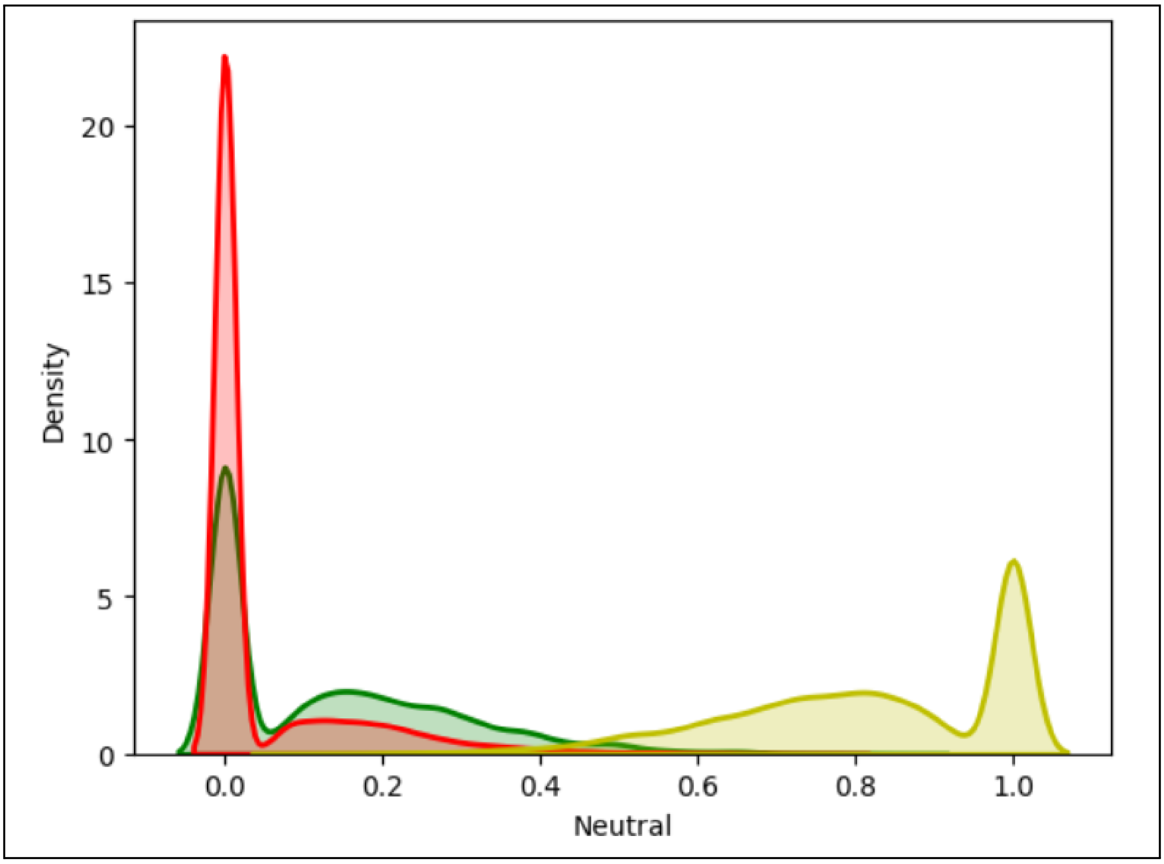


Figure 32: Visualising Sentiment Scores Of Positive, Neutral And Negative Tweets

The figure displays a visual representation of sentiment scores across the dataset. The visualization provides a comprehensive overview of the distribution and patterns in sentiment, aiding in the interpretation of sentiment dynamics in the context of climate change discourse.

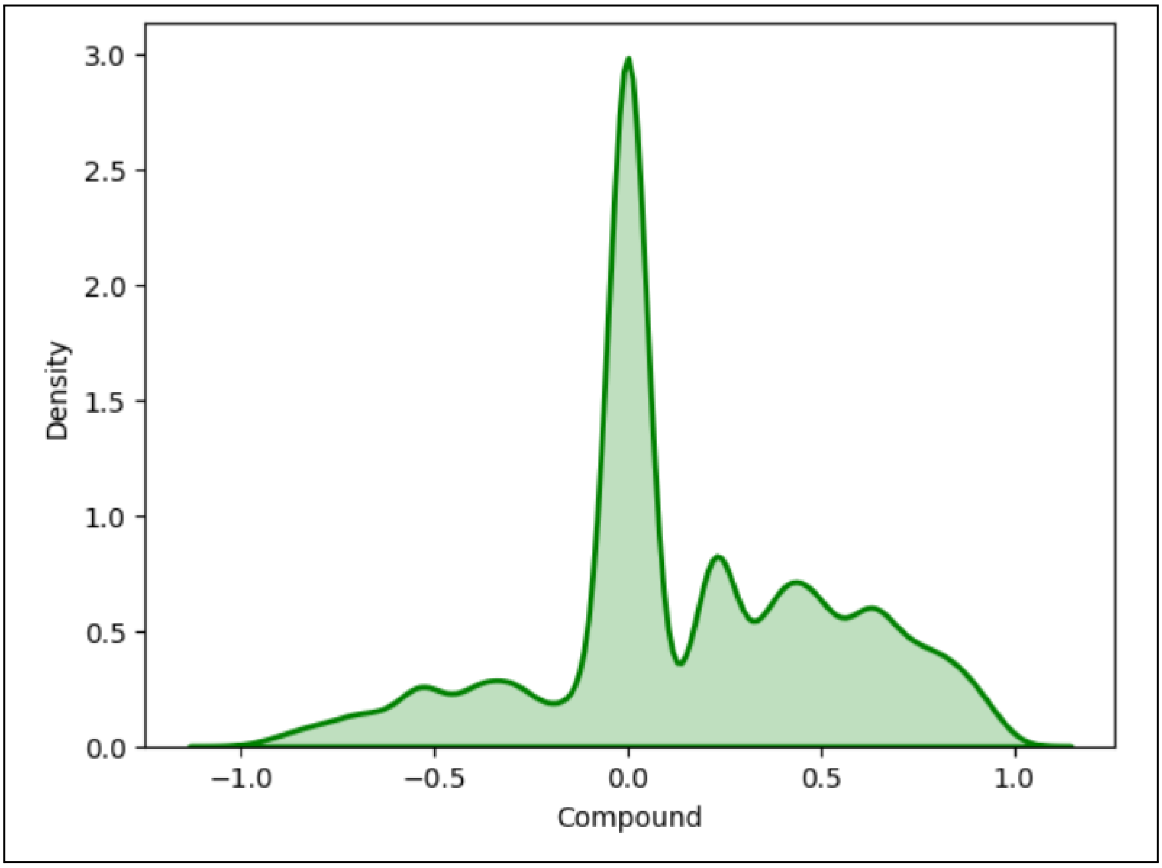


Figure 33: Visualisation of Sentiment Score

12. Conclusion

In conclusion, the study unravels noteworthy insights into the climate change discourse on Twitter in India from 2015 to 2020. Neutral tweets dominated, followed by positive and negative sentiments. The years 2018 and 2019 witnessed heightened engagement. The collective word cloud highlighted "environ" for "environment" with the largest font size. Positive sentiment was associated with words like "sustain," "save," and "environ," while negative sentiment featured terms like "global warm," "threat," and "destroy." Additionally, the average tweet length was 104 characters with an average word count of 14, indicating concise yet meaningful expressions throughout the analysed period.

13. Future Directions

1. **Global Comparative Analysis:** Extend the study to G7 and G20 countries, applying the same methodology to unveil comparative trends in climate change discourse. Investigate regional variations in sentiment, engagement, and key themes to contribute a global perspective.
2. **Continental Trends Exploration:** Explore climate change discourse trends across the seven continents, allowing for a nuanced understanding of regional perspectives. Analyse influential factors and sentiments to identify unique challenges and opportunities in diverse geographical contexts.
3. **Temporal Evolution Analysis:** Extend the temporal analysis beyond 2020 to track evolving patterns in climate change discussions. Evaluate the impact of recent events, policies, and global developments on public sentiment and discourse surrounding climate change.

14. References

[1] Roxburgh, N, Dabo, G, Shin, KJ et al. (4 more authors) (2019) Characterising climate change discourse on social media during extreme weather events. Global Environmental Change, 54. pp. 50-60. ISSN 0959-3780.

[2] Costa, A.J.B., Gomes, S.M.S., Kouloukoui, D. *et al.* Twitter conversations on sustainable development goals in Brazilian public universities using natural language processing. *Discov Sustain* 4, 51 (2023).

[3] Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.

[4] Shaik, Thanveer & Tao, Xiaohui & Li, Yan & Dann, Christopher & McDonald, Jacquie & Redmond, Petrea & Galligan, Linda. (2022). A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis. IEEE Access. 10. 1-1. 10.1109/ACCESS.2022.3177752.