iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XII **Month of publication:** December 2021

DOI: https://doi.org/10.22214/ijraset.2021.39391

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

Chennai Floods 2021: Sentiment Analysis of Twitter Data using Tweepy and TextBlob

Raghav Tinnalur Swaminathan[1], Vikram Balaji[2], Shankar Subramanian[3]
[1, 2, 3]*Student, Department of Information Science and Technology, College of Engineering, Anna University, Guindy, Chennai, India*

Abstract: ***The rise in the usage of Twitter for the exclamation of the problems worldwide and also as a 'review system,' where the customers can directly hold an entity responsible in front of the public by tweeting and tagging them, gives them immense power and counts towards being an advantage for researchers to analyze such data that can be scraped and used through APIs for a variety of purposes. Through this research, our motive is to analyze the 2021 Chennai floods with data sourced from twitter to understand the public sentiment during the 14-day span. The same is achieved with the help of Tweepy to authenticate data extraction from Twitter and TextBlob, for the classification of sentiment tags - positive, negative, and neutral. The result of this study focuses on the visualization of our findings, with various charts and metrics indicating the sentiment of the tweets we have scraped and analyzed.***

Keywords: Sentiment Analysis, WordCloud, Subjectivity, Polarity, Chennai Floods

I. INTRODUCTION

Understanding public sentiment through various social media platforms has become a necessity in today's world, considering social media's impact on our day-to-day lives. A recent trend in sentiment analysis indicates that data from social networking sites produce the most natural results. Twitter happens to be among the most noticed public platforms for candid public sentiment analysis on various issues.

This study aims to analyze public sentiment via Tweets from Twitter during the Chennai Floods in November 2021, with the help of Tweepy and TextBlob. Through Tweepy, we are initially required to use a Twitter Developer account, which later generates consumer keys and access tokens used for account authentication. Post authentication, we must search for tweets from a specific account or a trending hashtag for required extraction. The tweets extracted can be stored into a CSV or manipulated using Pandas. Upon creating the required DataFrame using Pandas, we perform data cleansing to remove unwanted characters in tweets to ensure accurate results. We then calculate the Subjectivity and Polarity of the tweets using TextBlob. Based on the Polarity score, we further indicate the positivity, negativity, and neutrality of a tweet and visualize the results to get a clearer idea of the sentiment that prevails.

Throughout the paper, we shall focus on elaborating upon a stepwise procedure, from data extraction to visualizing our results and determining public sentiment during the Chennai floods in November 2021.

II. PROBLEM STATEMENT

During stressful times and natural calamities, it is vital to note public sentiment for many reasons, including, but not limited to emergency response, critical public announcements, and detecting fake news that spreads chaos. Social networking websites tend to be the right place to start for achieving the same.

They act as a medium of communication for any widespread news among the public. This research study aims to analyze social media data, emphasizing on Twitter, to compute sentiment scores and visualize the same, eventually elucidating people's sentiment on social media during the 2021 Chennai Floods.

III.RESEARCH OBJECTIVES

The purpose of this study is to gain a detailed understanding of how tweets can be extracted through Tweepy and TextBlob, for further use in calculating a sentiment score with the help of Subjectivity and Polarity. The application chosen to implement the same is the recent Chennai floods. The aim is to extract tweets made by Twitter users during the 14-day span, analyze the public sentiment and action over Twitter, and visualize the results. This study also aims to understand the distribution of positive, negative, and neutral tweets during a calamity.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

IV. LITERATURE SURVEY

A. *Sentiment Analysis On Twitter By Using Textblob For Natural Language Processing*

Social Media is essential in terms of public data sourcing. It comprises people posting about what they think and feel about a particular topic that tags can segregate. It can also be used for opinion mining, where we can do a comparative analysis on unstructured text corpora [5]. Through this paper, we understand and implement TextBlob for computing the polarity and subjectivity of the tweets that have been extracted with the help of Tweepy.

B. *Sentiment Analysis Based On Expanded Aspect And Polarity-Ambiguous Word Lexicon*

Polar ambiguous words can hurt the polarity detection during sentiment analysis. Words such as "large, small" can change the polarity of the sentence. This paper gives us a clearer idea on estimating the words' impact on the sentence, as we bootstrap it with the core sentiment of the sentence [7]. We further use this in analyzing how the Polarity score for certain tweets were computed.

C. *An Introduction To Twitter Data Analysis In Python*

Twitter's popularity as a micro-blogging site provides much information from its users on various events be it a new product release or response to a new law [10]. This paper talks about pre-processing methods on tweets and how to apply analytics to extract information from Twitter. We further refer to the methodologies elucidated by this paper in drawing conclusions regarding the sentiments of various tweets during the Chennai Floods.

V. IMPLEMENTATION

To perform sentiment analysis, we will be extracting data with the help of Tweepy. We may define Tweepy as a python library used to access the Twitter API - offered by Twitter to extract and access data on a variety of topics. To access data through Twitter API, a developer account needs to be created and approved by Twitter through an application process. Once we obtain approval for a developer account, we generate consumer tokens and access tokens and declare the same as variables. We further use *OAuthHandler* and *set_access_token* to authenticate the account, by verifying the access tokens. The same is depicted in figure 2.

```
In [2]: #Keys
        consumer_key = #Enter Consumer Key
        consumer_secret = #Enter Consumer Secret Key
        access_token = #Enter Access Token
        access_token_secret = #Enter Access Token Secret

In [3]: #Authentication Handler and Setting Access Token
        auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_token_secret)

In [4]: #Authentication
        api = tweepy.API(auth,wait_on_rate_limit=True)
```

Figure 2: Twitter API authentication using Tweepy

We then scrape tweets using the *tweets_search* function and further search for tweets with the help of a specific hashtag, language, and time. We are also required to specify the count of tweets for extraction. Upon successful extraction of tweets, they are placed in a DataFrame and are labeled accordingly.

Once data is scraped and placed in a DataFrame, we go ahead with cleansing the tweets. Ideally, manipulating any form of textual data needs to be handled with more attention, as transforming the wrong information can cause biased analysis, eventually leading to inaccurate results. Further on, we perform data cleansing to remove various symbols and usernames, to ensure more accurate computation for sentiment analysis concerning the tweets.

Upon data cleansing, we use the *sentiment* function under TextBlob to compute Subjectivity and Polarity scores, which are of fundamental use towards the classification of the tweets extracted.

Polarity: It can be defined as a float value between the range [-1,1] that classifies whether a given text is positive or negative. In a nutshell, the Polarity score helps analyze the emotion or sentiment behind a given text.

Subjectivity: It can be defined as a float value between the range [0,1] and classifies a given text based on whether it is more subjective or objective. A subjective sentence is the kind of textual data where the tone is more inclined towards an opinionated expression. An objective sentence is the kind of textual data whose tone is more inclined towards factual expressions.

The Polarity score is further used to classify tweets into three tags, namely positive (>=0), neutral (=0), and negative (<=0). We visualize our final results to analyze patterns from our textual data by classifying tweets into the given tags.

Out[14]:

	Tweets	Subjectivity	Polarity	Analysis
0	: Conservancy worker clearing the thrash at DT...	0.540000	0.160000	Positive
1	: Spells of Rains to continue over Chennai &am...	0.000000	0.000000	Neutral
2	: Conservancy worker clearing the thrash at DT...	0.540000	0.160000	Positive
3	Fresh set of rain bands nearing chennai\nRains...	0.433333	0.360417	Positive
4	_Metro_Water _NEHRU Kindly bring machine, to c...	0.390000	0.206667	Positive
...	...	...	...	...
95	: Conservancy worker clearing the thrash at DT...	0.540000	0.160000	Positive
96	_: Floki Frunkpuppy $FLOKI to the moon , join ...	0.850000	0.250000	Positive
97	: Conservancy worker clearing the thrash at DT...	0.540000	0.160000	Positive
98	: Conservancy worker clearing the thrash at DT...	0.540000	0.160000	Positive
99	: Conservancy worker clearing the thrash at DT...	0.540000	0.160000	Positive

100 rows × 4 columns

Figure 3: Computed and classified values of Subjectivity and Polarity

Figure 3 represents the computed values of Subjectivity and Polarity, eventually aiding in the classification of tweets into sentiment tags as a separate column.

VI. RESULTS

Based upon the sentiment computations that have been calculated, various trends and findings are visualized with the help of various charts imported from Matplotlib. The texts extracted are further tokenized, and words with the highest frequency in terms of usage are visualized in the form of a word cloud.

Figure 4 depicts a cross-comparative study between the Subjectivity and Polarity scores of the 1000 tweets extracted in the form of a scatter plot. As stated, the Subjectivity score is a value between the range [0,1] that determines whether a tweet is more opinionated or factual in nature, and the Polarity score, what generates a value between [-1,1] classifies a tag as positive, negative or zero.

What can be interpreted from the graph is that with the tweets having a diverse Subjectivity and Polarity score from random set of tweets depicted, they are equally spread out across all quadrants of the scatter plot. This helps indicate a diverse sentiment among people of the Greater Chennai Area during a 14-day span through a graphical depiction.

However, when we consider the average Subjectivity and Polarity scores of all tweets extracted, we get values of [-0.453] and [0.435] respectively, which indicates more inclination towards negative set of tweets among the people, with a combination of objective and subjective tweets.

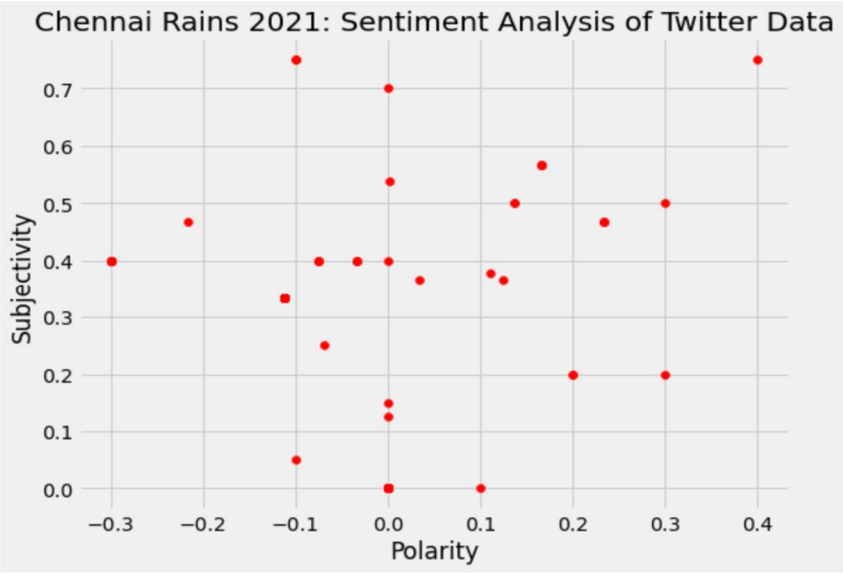


Figure 4: Cross-comparison between Subjectivity and Polarity

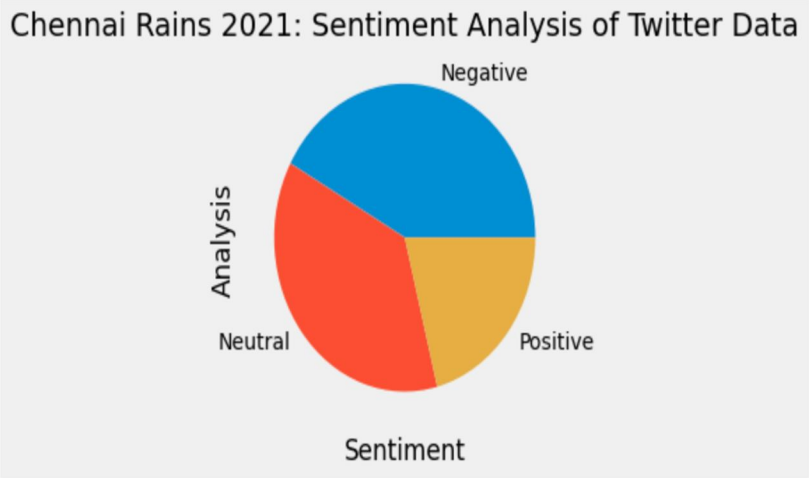


Figure 5: Visualizing classification of tags

The diverse results between the scatter plot and the average of Subjectivity and Polarity scores can be explained by figure 5, wherein the count of tweets, based on sentiment tags assigned through Polarity scores visualizes the distribution of various tweets based on the count. The pie-chart used for visualization shows a nearly equal distribution between neutral and negative tweets, with a considerable amount of positive tweets. Through this, we can draw a major conclusion, such that though there seemed to be a lot of negative tweets during the Chennai Floods, there were equally stronger positive tweets which helped the entire city get through these tough times.

Figure 6 further represents the high-frequency words that have been used, based on the tweets extracted during the 14-day span.

```
In [80]: #Cleate a word cloud
         wc = ' '.join([tweets for tweets in df['Tweets']])
         wordcloud = WordCloud(width = 1000, height = 500, random_state = 30, max_font_size = 120).generate(wc)
         plt.imshow(wordcloud, interpolation = "bilinear")
         plt.axis('off')
         plt.show()
```

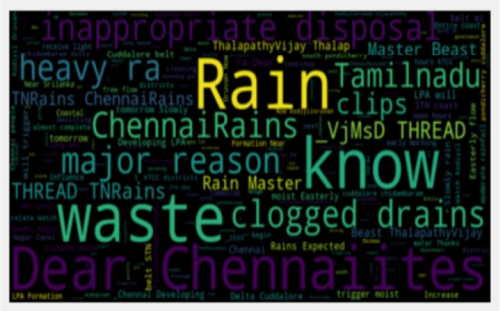


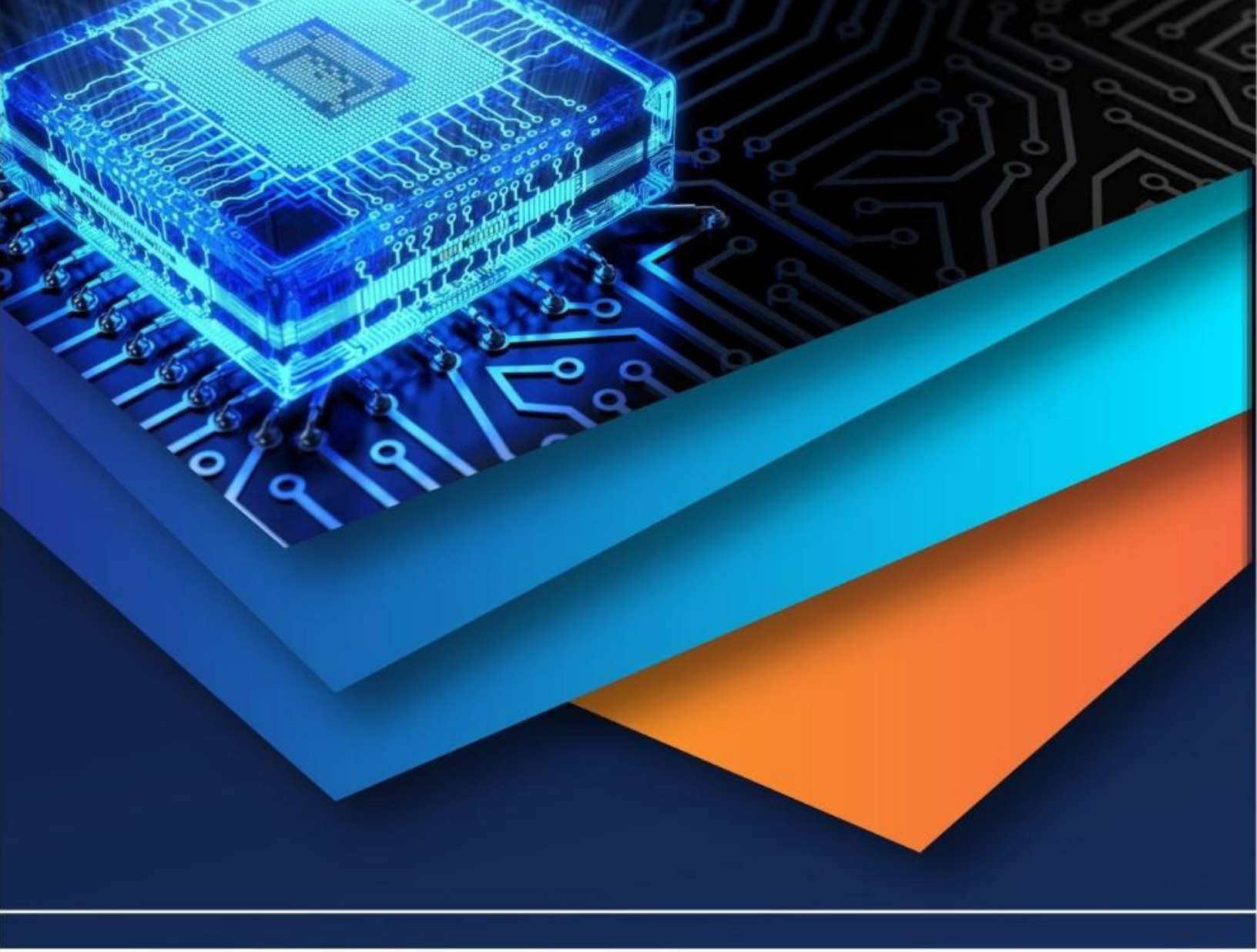
Figure 6: Word cloud representing high frequency words

VII. CONCLUSIONS

The sentiments during the Chennai Floods on Twitter were extracted with the help of Tweepy and analyzed to calculate a Subjectivity and Polarity score using TextBlob, eventually classifying various tweets into three sentiment tags. A cross-comparative depiction between Subjectivity and Polarity of the tweets indicates a spread-out sentiment among Twitter users with the help of a scatter plot. Additional findings in the study were also represented with the help of a pie-chart, bar graph, and a word cloud - which helped gain an in-depth understanding of public sentiments. To conclude, we can clearly state that Social Media plays a crucial role, especially during such challenging times, and is a vital source for crucial data analysis. This methodology will undoubtedly be valuable to researchers who intend to efficiently study sentiments over large amounts of textual data.

REFERENCES

[1] Saravanou, Antonia & Valkanas, George & Gunopulos, Dimitrios & Andrienko, Gennady. (2015). Twitter Floods when it Rains: A Case Study of the UK Floods in early 2014. 1233-1238. 10.1145/2740908.2741730.
[2] Ingule, Devki & Chhajed, Gyankamal. (2021). Survey of Public Sentiment Interpretation on Twitter.
[3] Chikersal, Prerna & Poria, Soujanya & Cambria, Erik & Gelbukh, Alexander & Chng, Eng. (2015). Modelling Public Sentiment in Twitter: Using Linguistic Patterns to Enhance Supervised Learning. 10.1007/978-3-319-18117-2_4.
[4] Kaur, Chhinder & Sharma, Anand. (2020). Twitter Sentiment Analysis on Coronavirus using Textblob.
[5] Deb, Shuvam & Hazarika, Ditiman & Konwar, Gopal & Bora, Dibya. (2021). Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing. 10.15439/2020KM20
[6] Farooqui, Nafees & Mehra, Ritika. (2019). Sentiment Analysis of Twitter Accounts using Natural Language Processing. 8. 473-479.
[7] Cao, Yanfang & Zhang, Pu & Xiong, Anping. (2015). Sentiment Analysis Based on Expanded Aspect and Polarity-Ambiguous Word Lexicon. International Journal of Advanced Computer Science and Applications. 6. 10.14569/IJACSA.2015.060215.
[8] Agarwal, Apoorv & Xie, Boyi & Vovsha, Ilia & Rambow, Owen & Passonneau, Rebecca. (2011). Sentiment Analysis of Twitter Data. Proceedings of the Workshop on Languages in Social Media.
[9] Mohey El-Din, Doaa. (2016). Analyzing Scientific Papers Based on Sentiment Analysis (First Draft). 10.13140/RG.2.1.2222.6328.
[10] Wisdom, Vivek. (2016). An introduction to Twitter Data Analysis in Python. 10.13140/RG.2.2.12803.30243.
[11] Di Capua, Michele & Di Nardo, Emanuel & Petrosino, Alfredo. (2015). An Architecture for Sentiment Analysis in Twitter. 10.13140/RG.2.1.2384.6482.
[12] Twitter Data Mining: Analyzing Big Data Using Python - (https://www.toptal.com/python/twitter-data-mining-using-python)
[13] Tapping Twitter Sentiments: A Complete Case-Study on 2015 Chennai Floods - (https://www.analyticsvidhya.com/blog/2016/07/capstone-project/)
[14] Sentiment Analysis of Twitter Posts on Chennai Floods using Python - (https://www.analyticsvidhya.com/blog/2017/01/sentiment-analysis-of-twitter-posts-on-chennai-floods-using-python/)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY