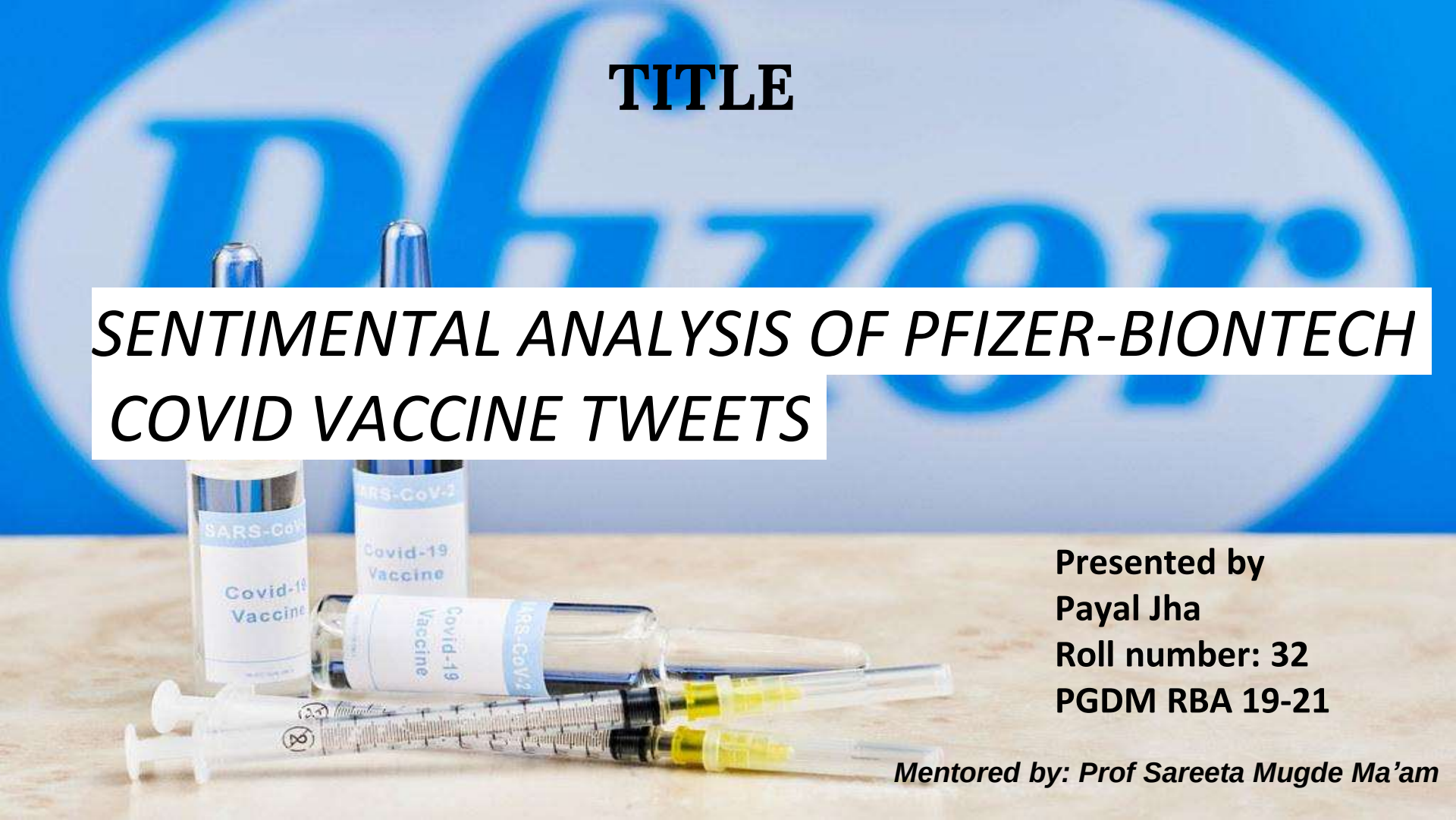
TITLE



SENTIMENTAL ANALYSIS OF PFIZER-BIONTECH COVID VACCINE TWEETS

Presented by
Payal Jha
Roll number: 32
PGDM RBA 19-21

Mentored by: Prof Sareeta Mugde Ma'am

OBJECTIVE OF STUDY

To study the sentiments of people and analyze the trends in the sentiments related to the Pfizer- BioNTech Covid Vaccine

REVIEW OF LITERATURE

Coronavirus Disease 2019 (COVID-19) which started in December 2019 in the city of Wuhan and ever since has been spreading. Ever since it's inception scientists and researchers have continuously been striving to get more information right from the origin to the cure. Although there has been a lot of research and work done on the studies related to analysis of the sentiments directed on social media, towards the Covid19 pandemic, there is not much of research done related to the behavioral development of the people towards the Coronavirus vaccine.

However, there have been some studies on text sentiment classification in general and also related to the Coronavirus, which eventually corresponds to the objective of this research project

After analysing and researching the existing work, we found that there have been numerous researches conducted in order to study the sentiments of people in business, or particularly how they were dealing with the whole covid scenario. To my best knowledge, as of now there haven't been any studies in order to determine and analyze the response of people towards Covid Pfizer Vaccine. Hence the project will help us gain insights on the sentiments of people directed towards the Pfizer Vaccine through analytical tools, and eventually bridge the gap.

Another important aspect of this study is to analyze the correlation of events associated with tweets which has not been explored extensively

RESEARCH METHODOLOGY

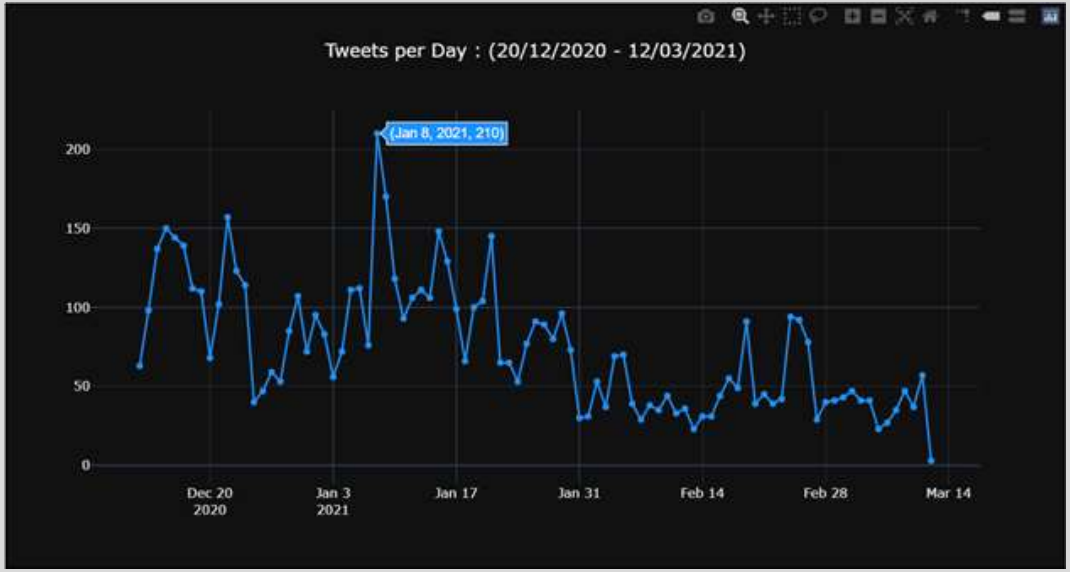
The idea here is to extract relevant data related to the Pfizer-BioNtech Covid vaccine tweets using Tweepy. The audience studied here will be the whole world from where the tweets are being generated. The time period for which the Twitter tweets will be studied is from 12th December 2020 to 12th March 2021. The basic ideology here is to extract the tweets, using natural language processing over it to get insights of user behavior related to the vaccine

DATA COLLECTION

The Data will be extracted directly from twitter using the Tweepy package in Python. The steps required for the same are listed below

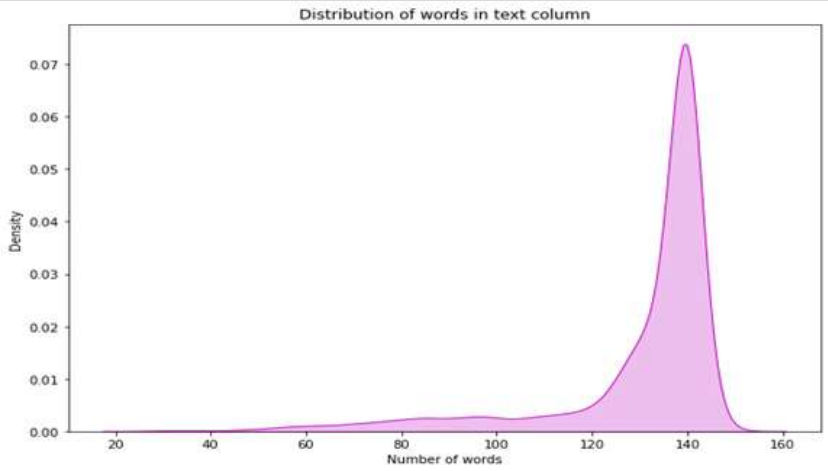
1. Signup for a twitter developer account
2. Make sure to make a note of your OAuth settings, which include – Consumer Key, Consumer Secret, OAuth Access Token, OAuth Access Token Secret.
3. Load the Tweepy package in Python & authorize the tokens mentioned above
4. Extracts the relevant tweets using the required hashtag
5. Pull the metadata into a CSV file.

ANALYSIS & FINDINGS

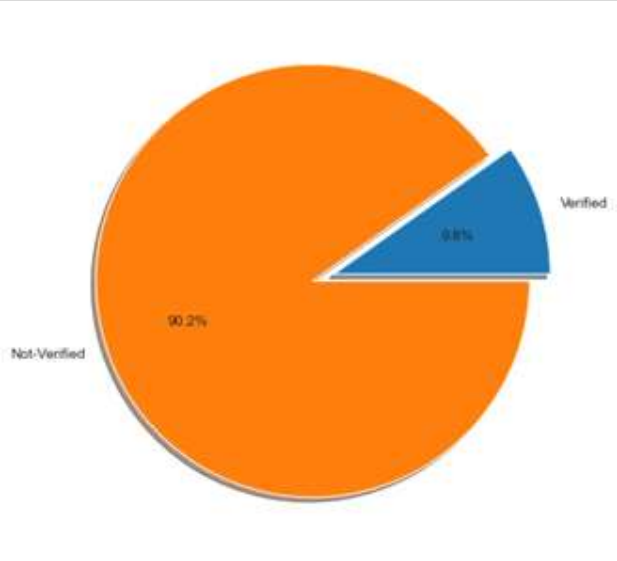


The image above shows the trend in the number of tweets with the highest number of tweets being recorded on 8th January 2021

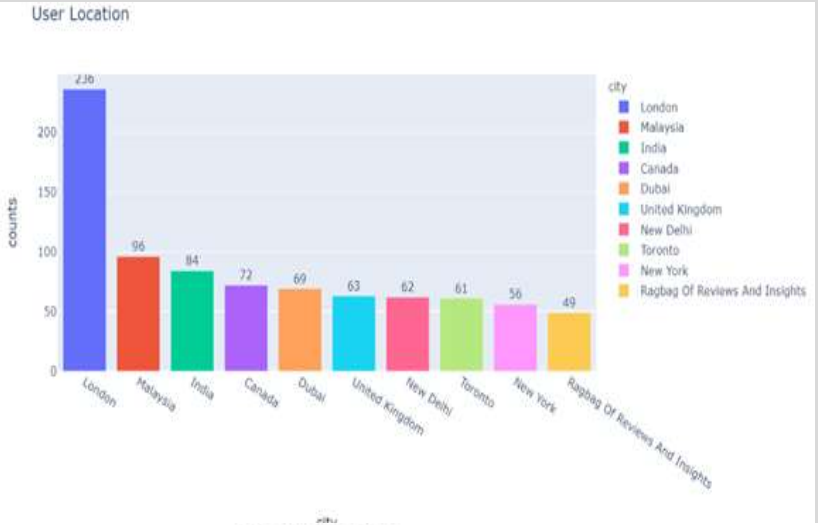
The reason behind this can be attributed to the fact that a study conducted showed that the Pfizer BioNtech vaccine also protected against the new variant of Coronavirus



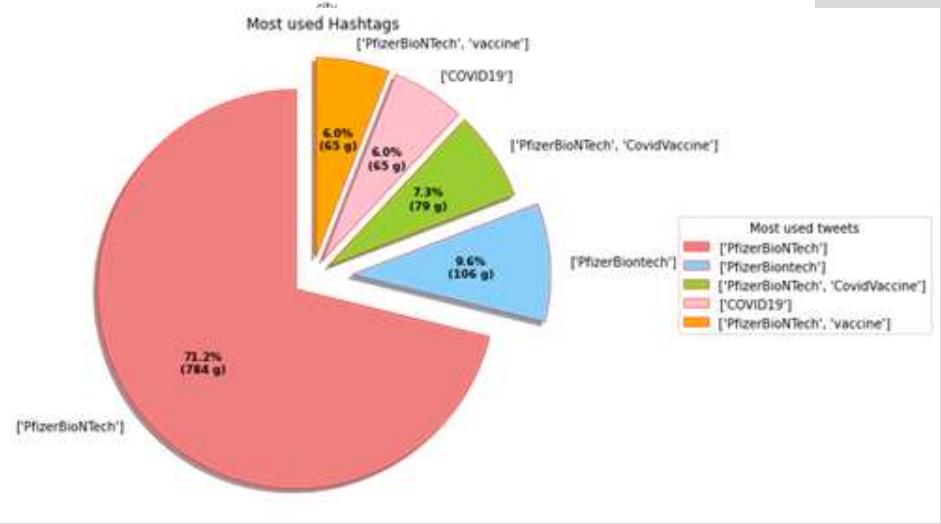
This Figure shows the variations in the length of the tweets. Here we can see that the Maximum number of tweets are between the range of 130-140 in length



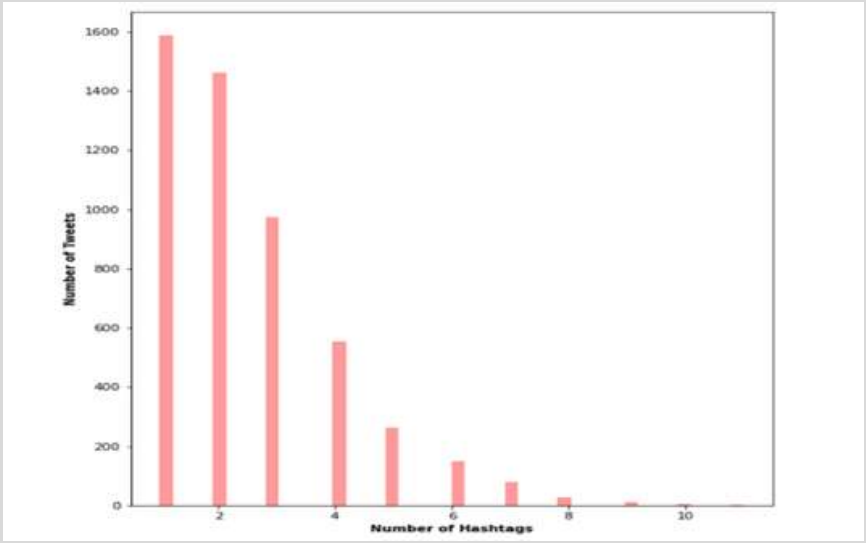
The figure shows that 90.2% of the users in study are not verified whereas 9.8% of the users are verified



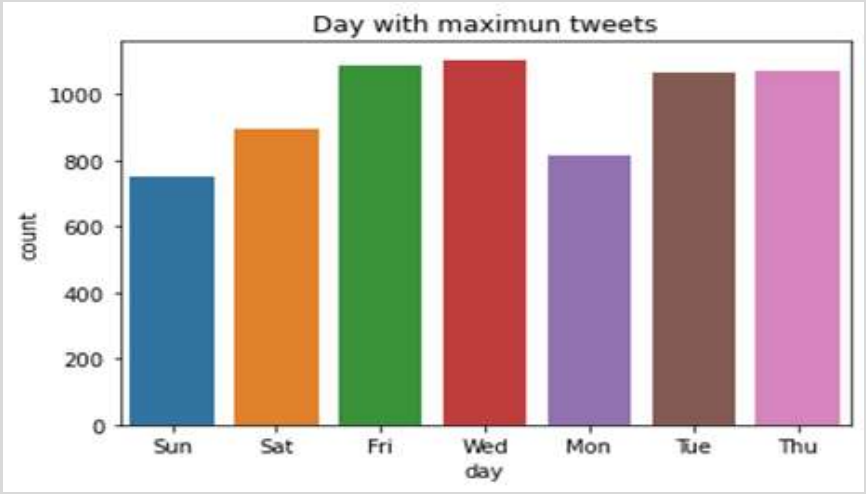
In this figure we can see the geographic distribution of the users. Maximum tweets during the study period were from London



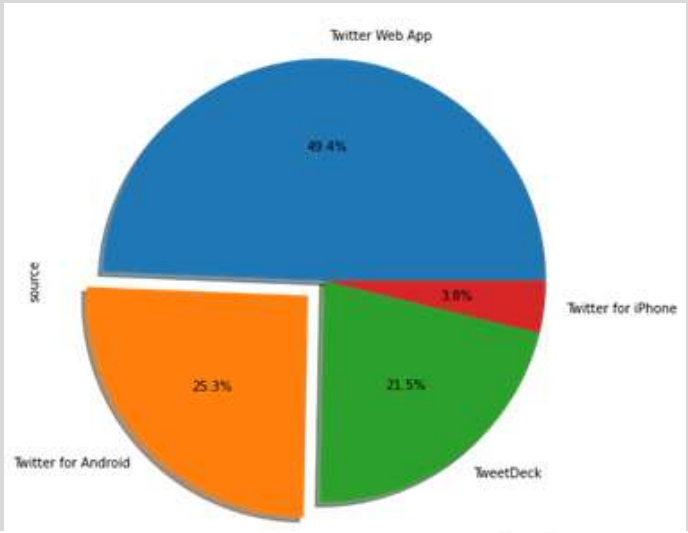
The pie chart shows the most common hashtags from the tweets. Significantly 71% of the tweets consisted of only one hashtag which was #PfizerBioNTech



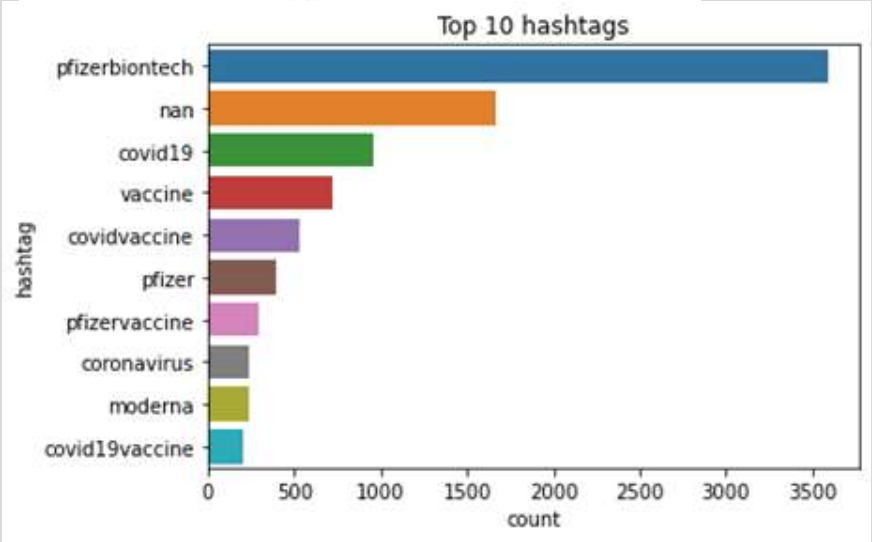
The bar chart shows the variations in the number of hashtags and the number of tweets. Here we can see that more than 1500 tweets consist of only one single hashtag



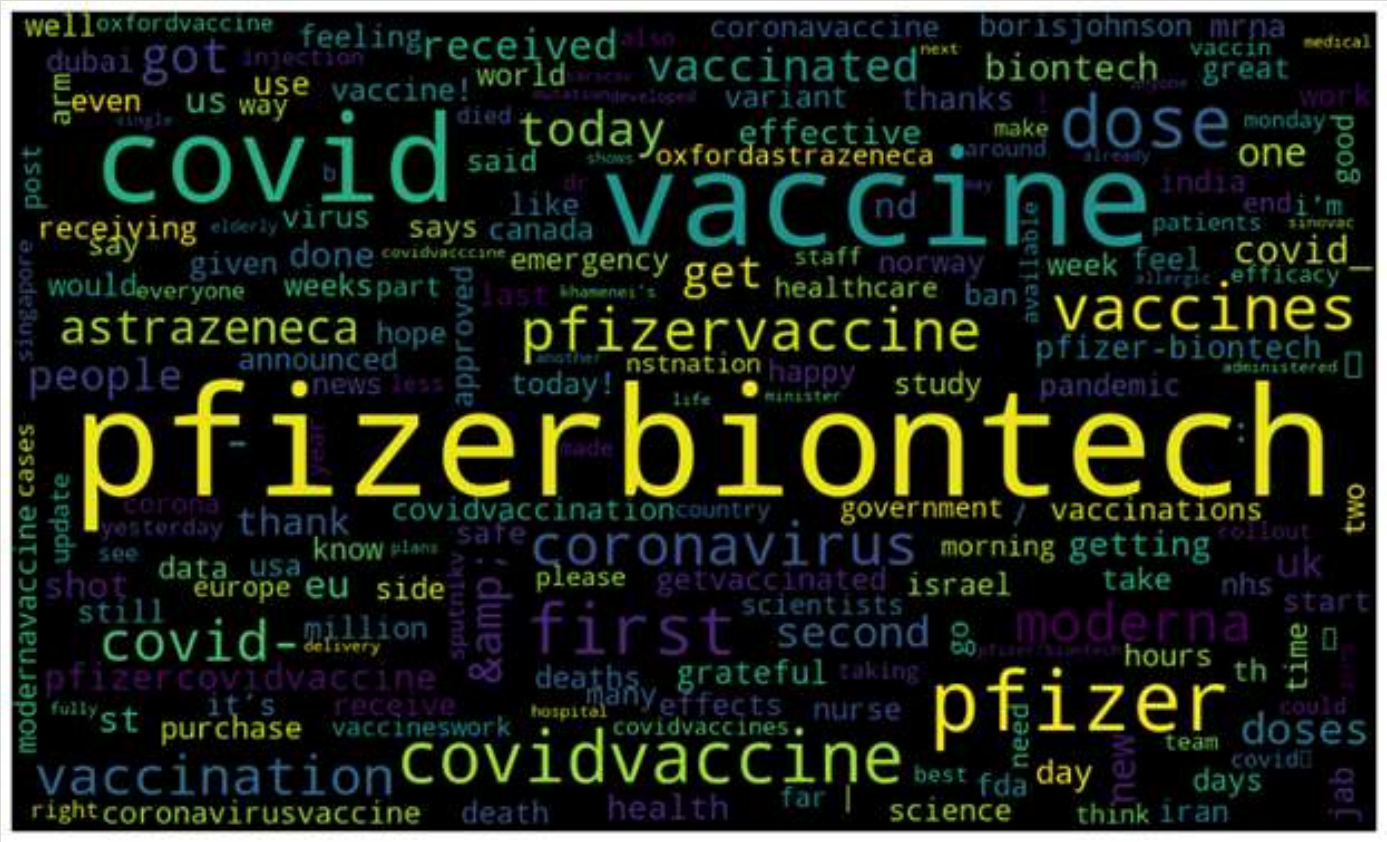
The count plot shows that the day of the week when tweets were maximum was, Wednesday. But a significant number of tweets were also posted on Fridays, Tuesdays and Thursdays



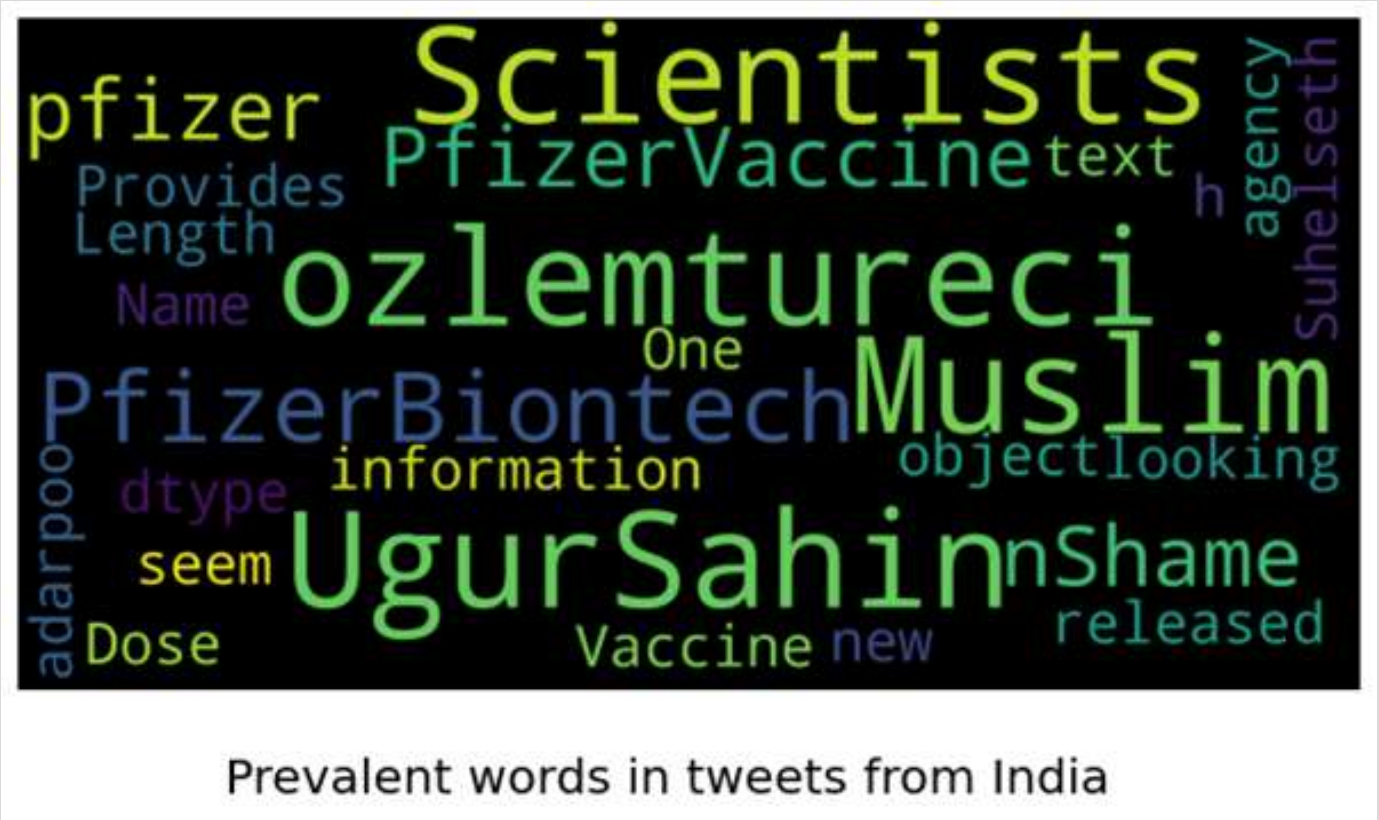
The Pie chart gives us a clear picture of the top 5 sources which were used to post the tweets. It shows that people generally prefer using the web app for their tweets followed by android users



The figure show the top 10 hashtags in the tweets



The Wordcloud gives us a clear representation of the most frequently used words in the tweets.



Prevalent words in tweets from India

The following word cloud is specifically directed towards the tweets coming from India. Here we can see words like Shame, Muslim, Ugur Sahin (A German Oncologist who helped in the development of the Vaccine) & his wife Ozlem Tureci

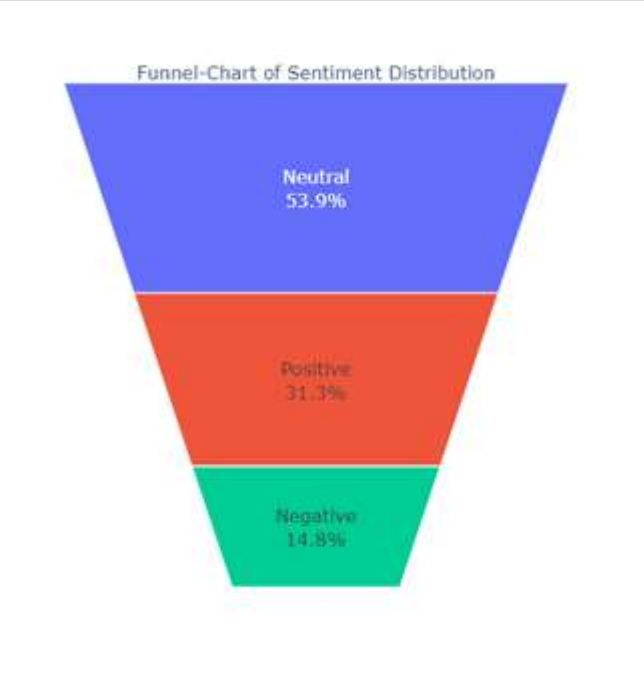
TEXT BLOB OR VADER SENTIMENT ANALYSIS?

Two of the most popular packages which are used for sentimental analysis are TextBlob & Vader.

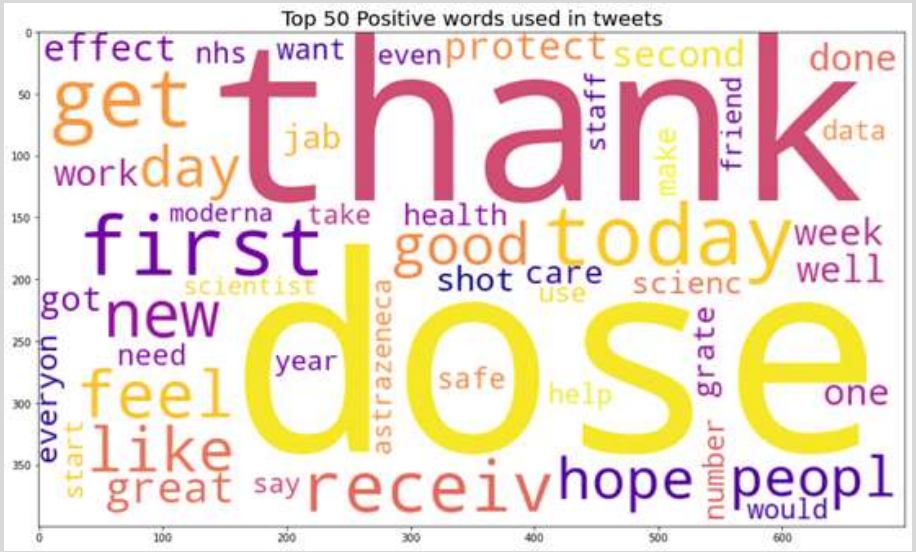
- Textblob sentiment analyzer generally results into two prospects for a given sentence: Polarity is a number that lies between [-1,1], -1 which indicates negative sentiment and +1 which indicates positive sentiment.
- Subjectivity is also a number that lies in the range of [0,1]. Subjectivity generally tends to point out towards the opinions in the statements and it also lies in the range of 0 to 1 inclusive

- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon-based, openly available sentiment analyzer which has been made into a pre-built library and is readily available for anyone to use.
- It generally uses pre-defined lexicon-based words which are later used to classify statements positive or negative as per the words they use.
- Vader sentimental analysis is generally preferred more over the TextBlob for social media analysis.

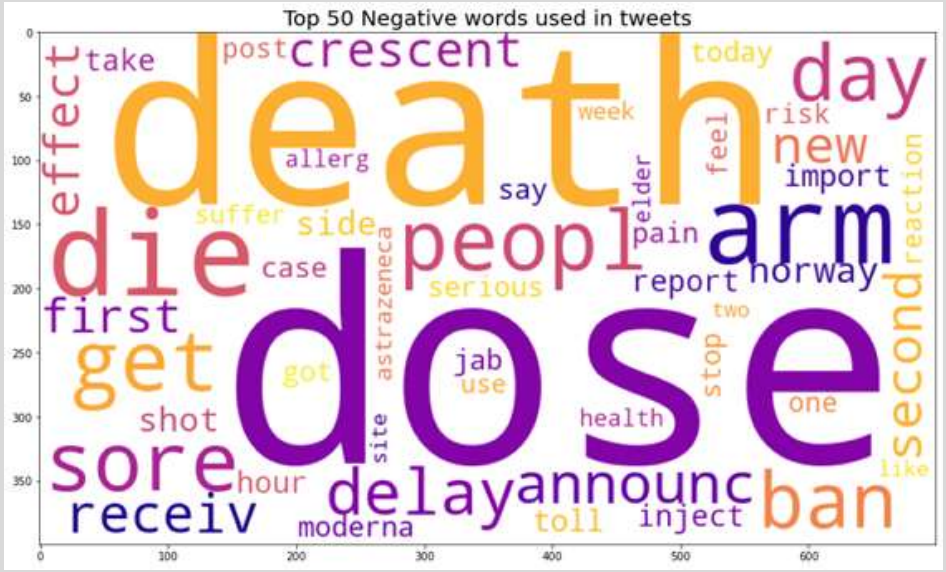
	Sentiment	text
1	Neutral	3655
2	Positive	2125
0	Negative	1007



After using the Vader Sentimental analysis, it helped us to classify the tweets into Positive, Negative & Neutral. Out of the 6787 tweets , we see that almost 54% of the tweets are neutral whereas 31% are positive & the rest are negative

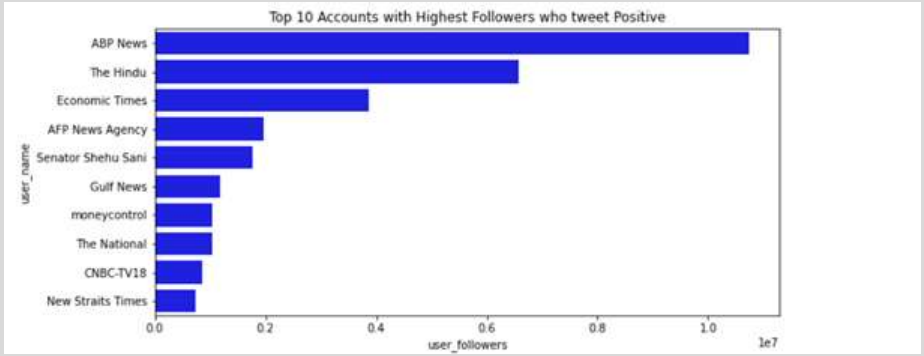


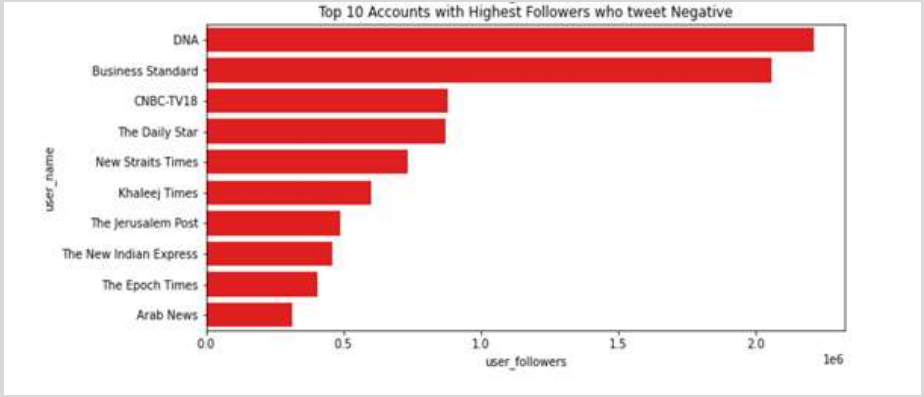
The wordcloud in the left has been generated keeping in mind the positive tweets. Here we can see words like Thank, First dose, Hope, great & good which gives us an idea that the tweets pertaining to the positive sentiments were filled with optimism towards the Vaccine

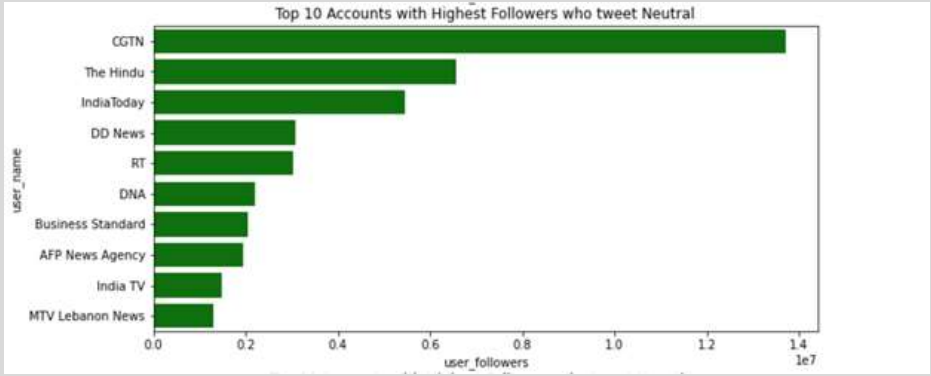


The word cloud in the right has been generated keeping in mind the negative tweets. Here we can see words like death, die, delay, ban, serious & reaction. This signifies two things, one that people were mostly worried about the delay in vaccinations and second about the side effects of the vaccine

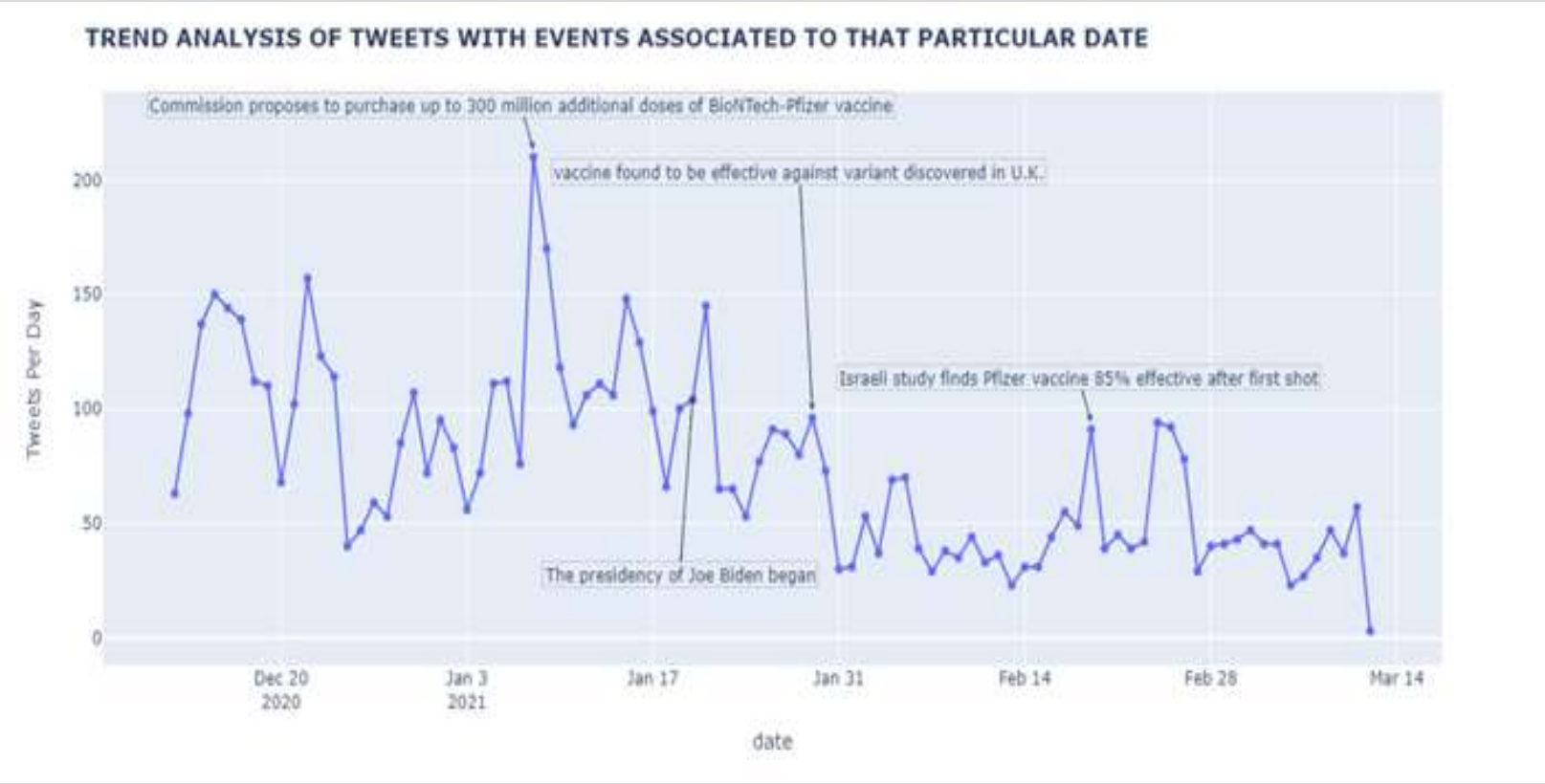
One interesting observation here is that there is a decline in the strength of average negative sentiment from the 20 of December to the 27 of December
The Reason for the same can be attributed to the news mentioned in the article above







The graphs give us an idea as to which users with maximum following out most positive and negative tweets. we can see that ABP news from India was the one posting most positive tweets whereas DNA news put out maximum negative reviews. CGTN (China Global television network) chose to keep a Neutral stand



The graph shows us a trend of the tweets from 12th December 2020 to 12th March 2021 with events causing the highs and lows. The highest peak in the number of tweets can be seen 8th January 2021 when additional 300 million doses of Pfizer vaccine was approved. And post a constant downfall two events which peaked the number of tweets was when the vaccine was found to be effective against the new strain and the presidency of Joe Biden.

CONCLUSIONS & RECOMMENDATIONS

- The majority of the Sentiments directed for the Pfizer Vaccine were Neutral.
- Overall, the tweets are more inclined towards being positive rather than being negative.
- A decreasing trend in the negative sentiments was observed from the 13th of December to the 27th, owing to the reason that the EU started vaccination drive during that time period.
- The positive sentiments' deviation is more stable than the negative sentiment, which is much less stable. In other words, the strength of the positivity in people's tweets tends to be more or less the same, unlike the negative strength.
- The peaks in the number of tweets when correlated with the news help us analyze the reason behind the peak in the tweets
- The highest peak in the number of tweets can be seen 8th January 2021 when additional lot of 300 million doses of Pfizer vaccine was approved.
- Although there are recurring words in the positive and negative word cloud, the other words help us give an idea about the type of thinking of the users and the primary reasons for being positive or negative.
- The sentiment strength of a tweet offers us a domain from which we can learn how the population is reacting to the vaccine; such insight can allow the different government to channel their advertisements towards more negative groups that usually refuse to believe in the integrity of the vaccine and observing the change over time similarly as we saw between the 13th and 27th of December.

RECOMMENDATIONS: The tweets from a different timeline post 12th March 2021 can be studied in order to deeply study the vaccination drives in the different countries and gauge the effectiveness through social media

REFERENCES

1. K. Sailunaz and R. J. J. o. C. S. Alhajj(2019), "Emotion and sentiment analysis from Twitter text," vol. 36, p. 101003.
2. P. Tyagi and R. J. A. a. S. Tripathi(2019), "A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data," .
3. A. Alsaeedi and M. Z. J. I. Khan(2019), "A Study on Sentiment Analysis Techniques of Twitter Data," vol. 10, no. 2, 2019.
4. C. Kaur and A. Sharma(2020), "Twitter Sentiment Analysis on Coronavirus using Textblob," EasyChair2516-2314, 2020.
5. D. Prabhakar Kaila, D. A. J. I. J. o. A. R. i. E. Prasad, and Technology(2020), "Informational Flow on Twitter–Corona Virus Outbreak–Topic Modelling Approach," vol. 11, no. 3, 2020.
6. R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. J. m. Lehmann(2020),"An" Infodemic": Leveraging High-Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak," 2020.
7. M. Alhajji, A. Al Khalifah, M. Aljubran, and M. Alkhalifah(2020), "Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19,".
8. C. K. J. A. a. S. Pastor(2020), "Sentiment Analysis of Filipinos and Effects of Extreme Community Quarantine Due to Coronavirus (COVID-19) Pandemic,"
9. N. K. Rajput, B. A. Grover, and V. K. J. a. p. a. Rathi(2020), "Word frequency and sentiment analysis of twitter messages during Coronavirus pandemic,".
10. M. Ra, B. Ab, and S. Kc(2020), "COVID-19 Outbreak: Tweet based Analysis and Visualization towards the Influence of Coronavirus in the World."
11. A. D. J. A. a. S. Dubey(2020), "Twitter Sentiment Analysis during COVID19 Outbreak,".
12. W. Medhat, A. Hassan, and H. J. A. S. e. j. Korashy(2014), "Sentiment analysis algorithms and applications: A survey," vol. 5, no. 4, pp. 1093-1113.
13. Chew C, Eysenbach G.(2015) Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One.;5(11):e14118.
14. Dredze M, Broniatowski D, Smith M, Hilyard K(2016). Understanding vaccine refusal: why we need social media now. Am J Prev Med.;50(4):550–552.

THANKYOU