Full name and semester: Lakku Sai Jagan Reddy and Fall 2022

Title: Sentiment Analysis on COVID-19 vaccination in Europe.

Word count: "Sentiment Analysis on COVID-19 vaccination in Europe(1350 words)

Introduction

COVID-19 pandemic is the most unexpected medical emergency in the world which has started in 2019. Recent research on the COVID-19 pandemic found an increase in anxiety, stress, and depression among people of all ages. The World Health Organization (WHO) officially named the 2019 new coronavirus outbreak "COVID-19," an abbreviation for "Corona Virus Disease," on February 11, 2020. Today, after a year of fighting this pandemic, WHO reported that over two million people have died and over one hundred million infection cases have been reported. Given the disease's unprecedented spread, there have been global efforts to combat the pandemic. The development of a vaccine is one of the promising interventions in dealing with the COVID-19 pandemic. Concerning COVID-19, vaccines such as Sputnik V have shown an efficacy level of 91.6% after several trials involving over 20,000 participants as of February 2021 [1]. Several countries have also reported that their citizens have been vaccinated. BioNTech Pfizer, on the other hand, demonstrated 95% accuracy after the second dose and 100% accuracy in children aged 12 to 15 [2]. Following the development of vaccines, another challenge was gaining public acceptance. Maximum vaccine delivery around the world could result in covid control. Most people are unwilling to accept it due to concerns about side effects and other rumours[3]. Vaccination has greatly aided in the control of the COVID-19 pandemic in many countries, and vaccinations are still being administered to combat the pandemic. I would like to analyse how the people of Europe reacted on the Vaccination, since it was mostly opposed by many people in the Europe.

Research Questions

How did the people of Europe react to COVID Vaccination?
What are the most prominent topics discussed based on Vaccination drive?

Method

❖ **Data**

The data required for analysis was collected from twitter using snscrape package.I have collected around 1700 tweets and did sentiment analysis on the tweets. Tweets were collected based on hastags #Covid , #Covid19, #Vaccination, #Pfizer ,#Moderna, #BioNTech,#Europe and using the geo location of the Europe to make sure the tweets are from Europe. And the hashtags are case sensitive too. Also these tweets were collected in the time frame December 2020 to December 2021. Tweet id, location , date, tweet, num_of_likes, num_of_retweet were collected from the tweets.

❖ **Analysis**

Collected data is been first converted to a data frame and stored in a csv file.

▪ **Data Pre-processing**

Data must be pre-processed to remove the unnecessary contents in the data. I have initially removed the punctuations, emojis, and then the English stop words using nltk.corpus. nltk.corpus is a package which contains default 40 stop words like "a","an","the","of" etc. Since these words does not make any significance in sentiment analysis, we have removed them. I have used wordcloud module to analyse the most used words in the tweets.

▪ **Sentiment Analysis in TextBlob**

The text is contained solely in the content column of the extracted columns. So I pre-processed the text in the content column by converting it to lowercase, then removing special characters, punctuation, and numbers. The textblob library is used to determine the polarity of each tweet. I've added a new column called processed tweets, which contains the processed text and is polarized. I also used the Python module Valence Aware Dictionary and Sentiment Reasoner (VADER) for sentiment reasoning, which states that sentiment is negative if the polarity is less than -0.05, positive if the polarity is greater than -0.05, and neutral otherwise. We could see from the basic analysis the distribution of sentiment for the tweets. Most of the tweets were neutral, followed by being positive and then negative. I then analysed the daily tweet count monthly basis and observed a peak in the tweets during the initial phase of the COVID vaccine and the tweets gradually decreased as the time passed.
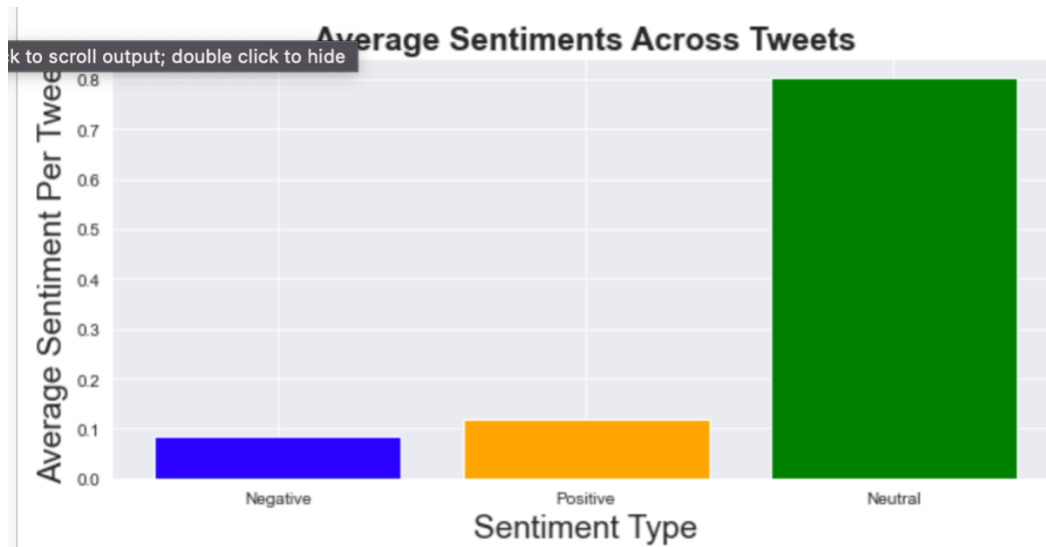
▪ **Machine Learning Models:**

Selection of Machine Learning Model plays a key role in predicting the accurate results. I have split the data into training and test samples and applied some machine learning algorithms to see the accuracy of the results and decide on the model. I have trained the data on models: Random Forest, Naïve Bayes, logistic regression, Decision Tree and Gradient Boosting and observed Random Forest Classifier has the highest accuracy of 0.708 , so I would like to use the Random Forest model , to predict analyse the sentiments.

Results

I have used the wordcloud to get the most used or significant words in the tweets and the results show that most used words are vaccine, covid19,Swedish. So the people of Sweden have talked most about covid vaccine.

Below are the results which show the distribution of sentiment among the tweets which shows the density of positive and negative results are close to 0 , whereas the density oof neutral results are close to 1.

```
0.08478217270194985
0.1169610027855153
0.8012618384401113
```


Below results show how the average sentiments across tweets , which says that the average of neutral is very high compared to the others , which is obvious from the above figure too.

Average Sentiments Across Tweets

Below figure shows the amount of daily tweets based on monthly wise. We Could see from the below figure that number of tweets were very in Jan 2021 , which was the initial stage of COVID vaccine, while later people have reduced reacting on the vaccination which shows the decrease in number of tweets by Jan 2022.

Daily Tweets
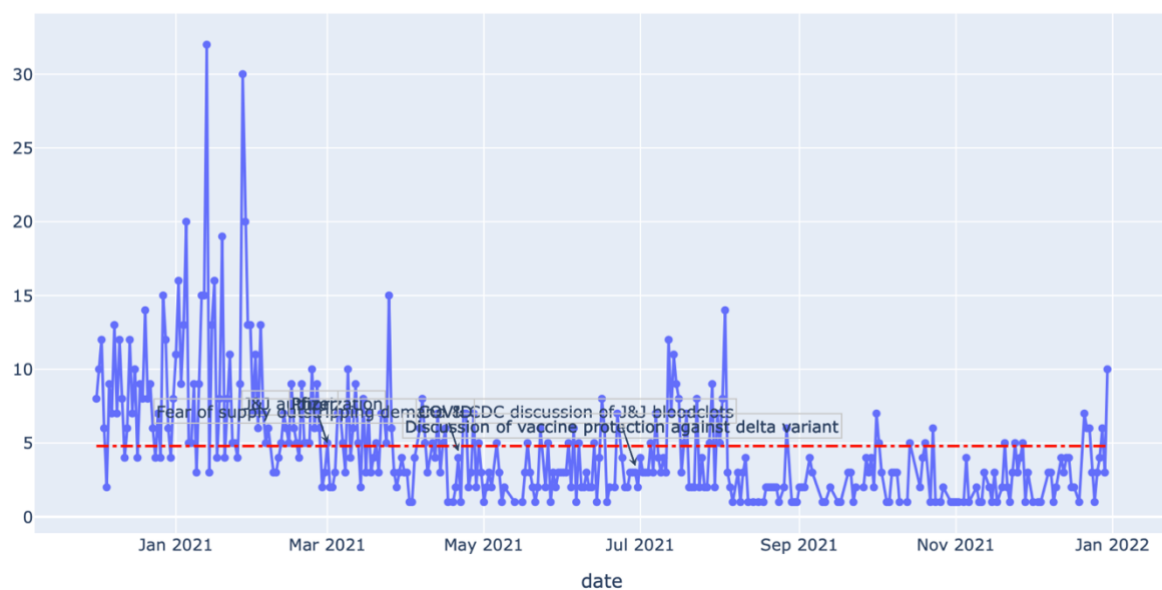
Below figure shows the distribution of sentiment among the tweets , which shows that the overall positive sentiments are most , followed by Neutral and then Negative.

Sentiment percieved
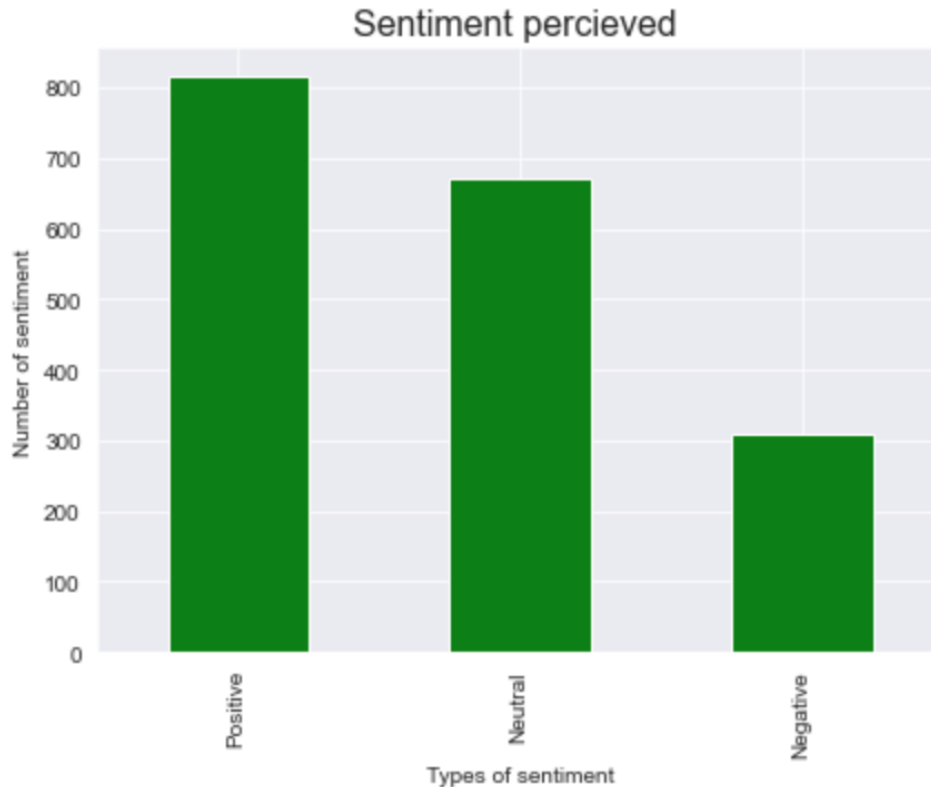
We can see that the accuracy of Random Forest Model is better than others, so I have decide to use the Random Forest Classifier for the sentiment analysis.

Accuracy of Machine Learning Models

```
Random Forest Classifier : {'criterion': 'entropy'}
Accuracy: 0.708        Precision: 0.765        Recall: 0.626        F1: 0.634

Naive Bayes : {'alpha': 0.5, 'fit_prior': True}
Accuracy: 0.632        Precision: 0.640        Recall: 0.560        F1: 0.569

logistic regression : {'max_iter': 2000}
Accuracy: 0.666        Precision: 0.710        Recall: 0.578        F1: 0.580

Decision Tree : {'min_samples_split': 5}
Accuracy: 0.604        Precision: 0.559        Recall: 0.542        F1: 0.545

Gradient Boosting : {'learning_rate': 0.1}
Accuracy: 0.694        Precision: 0.733        Recall: 0.620        F1: 0.635

CPU times: user 1min 7s, sys: 883 ms, total: 1min 8s
Wall time: 1min 6s
```

After performing the sentiment analysis by Random Forest Classifier , we could see below that the accuracy is 69% and precision of corresponding Negative, Neutral and Positive Sentiments.

Sentiment Accuracy

```
[[ 17  22  25]
 [  1 110  27]
 [  1  35 121]]
              precision    recall  f1-score   support

    Negative       0.89      0.27      0.41        64
     Neutral       0.66      0.80      0.72       138
    Positive       0.70      0.77      0.73       157

    accuracy                           0.69       359
   macro avg       0.75      0.61      0.62       359
weighted avg       0.72      0.69      0.67       359

0.6908077994428969
```
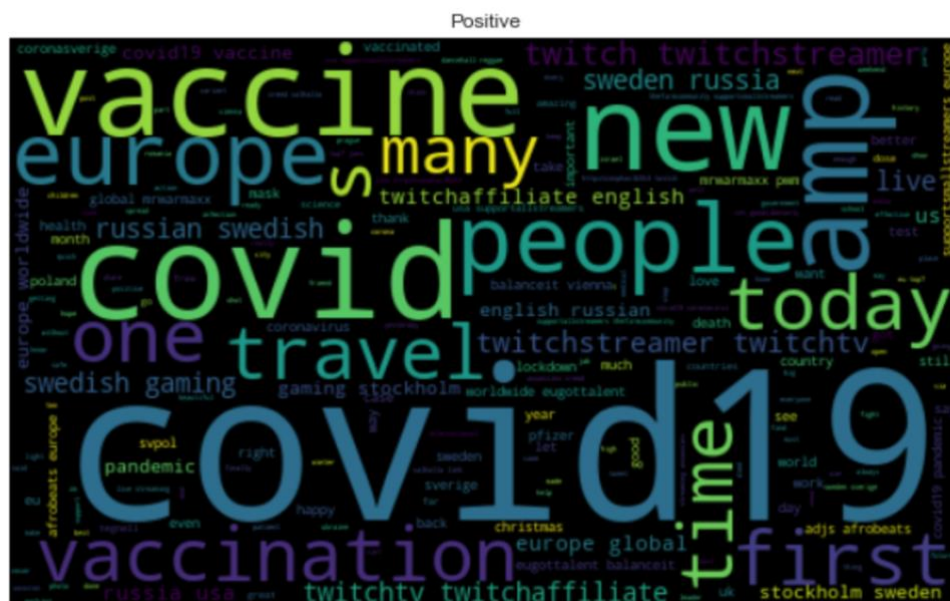
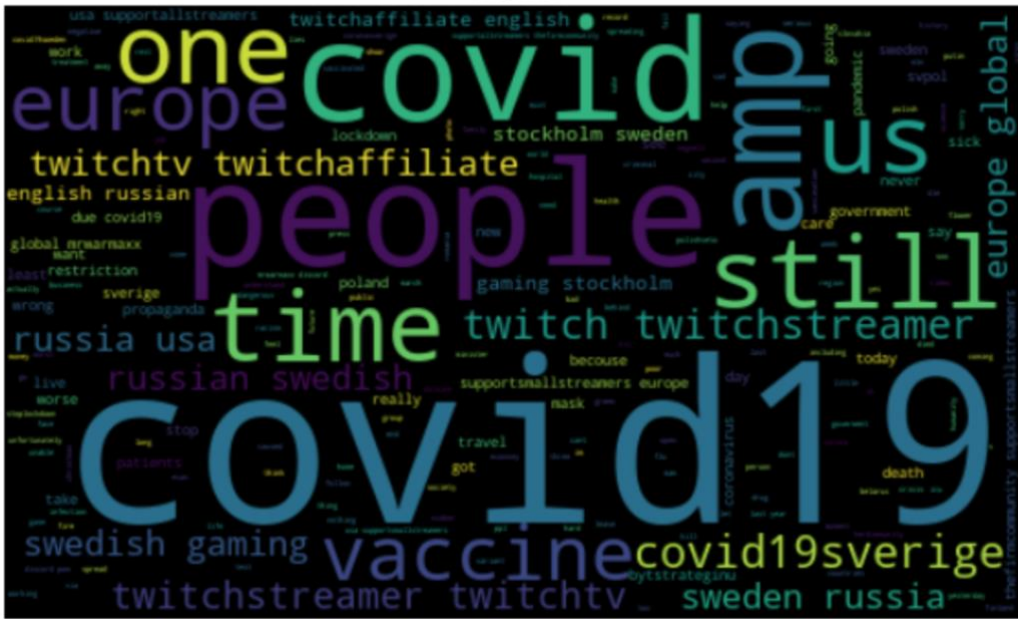| Sentiment	Count	Percentage
Negative	64	17.8%
Neutral	138	38.4%
Positive	157	43.7%
	359	

The above represents the percentage of sentiments distributed after applying Random Forest Model.We can say that most of the tweets(43.7%) were positive sentiment, followed by neutral and Negative.

Below analysis shows the most significant words based on the sentiments:
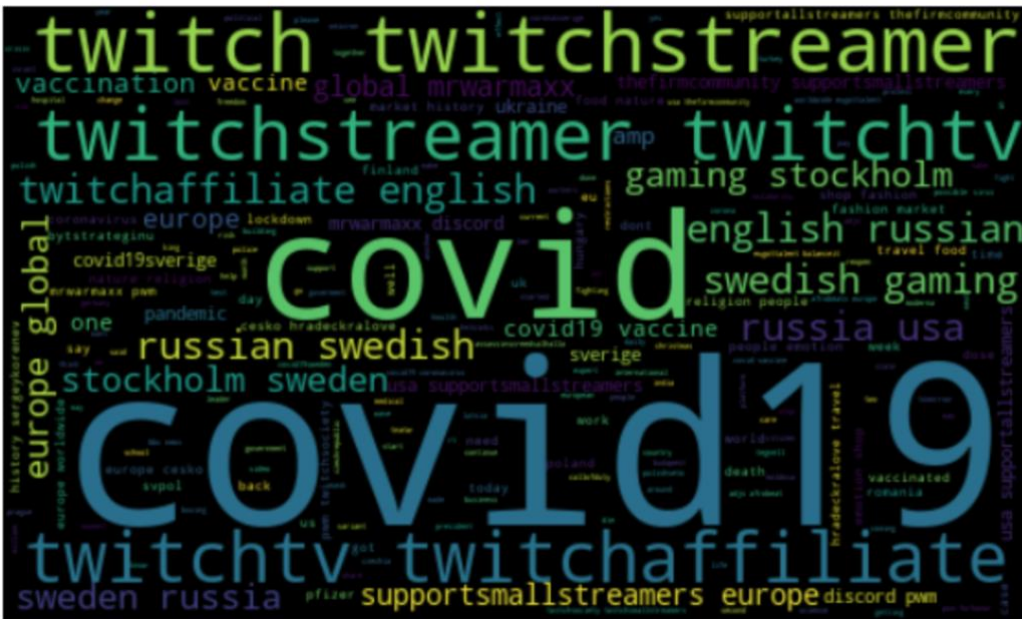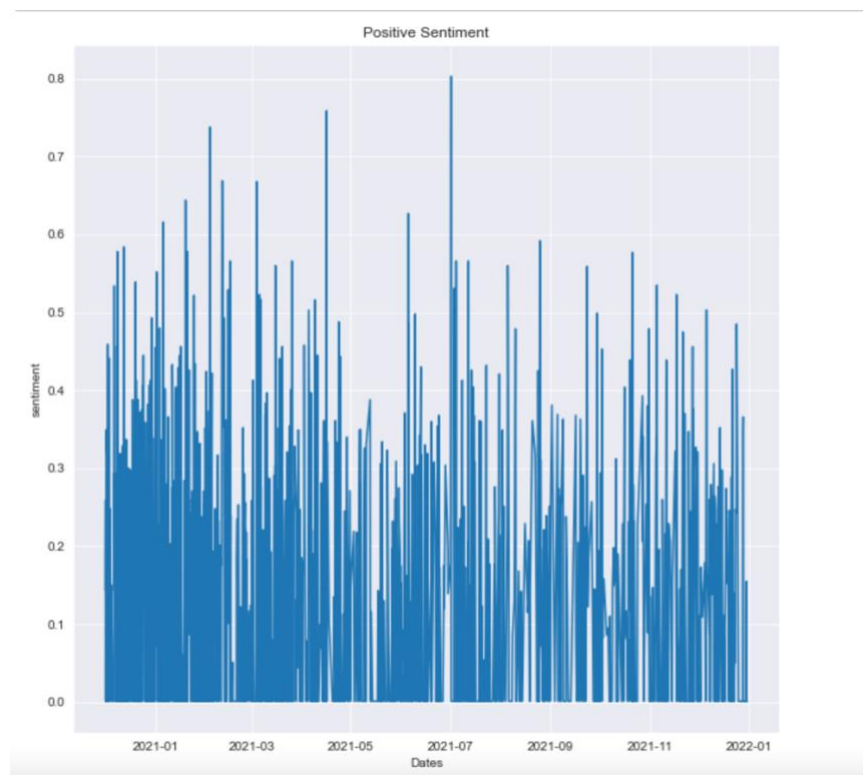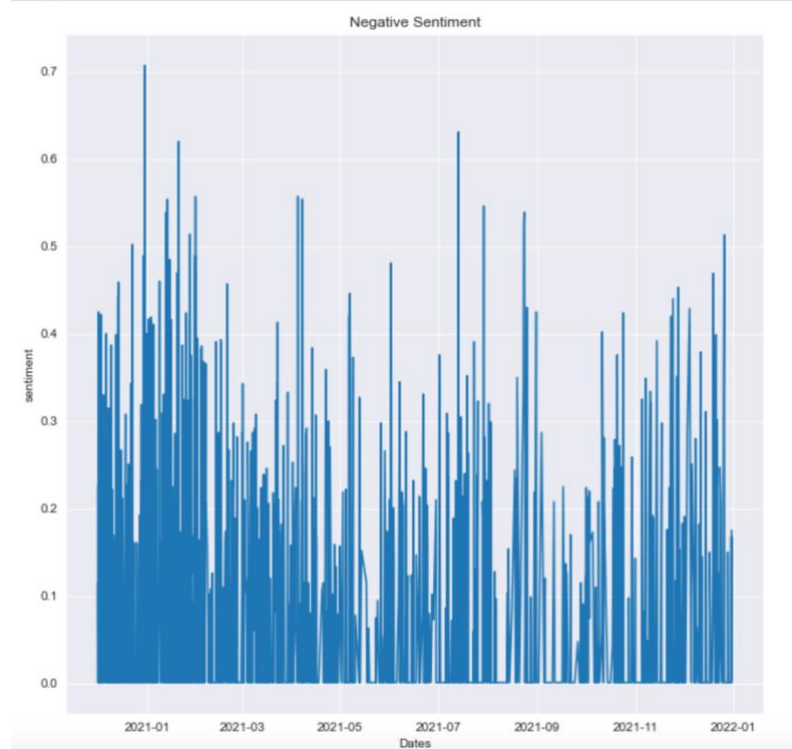

Positive

Negative

Neutral

Below figures show the distribution of Positive Sentiment and Negative from 2021-01 to 2022-01. <mark>We can observe that the sentiments were not too positive during the initial days , but as the months pass the sentiment has been more positive.</mark>

Positive Sentiment

From below we can see the distribution of Negative Sentiment , which says that people were more negative during the initial days and was less negative during the last few months

Negative Sentiment

Conclusion :

Analysis here shows the sentiment analysis done using nltk based on polarites and categorising the tweets into Neutral, Positive and Negative Sentiments. I have Implemented the Random Forest Classifier for analysing the sentiments which shows 43.7% were positive,38.4% were Neutral and 17.8% were negative. To answer the research question on how people reacted to vaccination, I have done the analysis monthly wise , which shows that people were more negative towards the initial drive, and less negative as the time pass. Also Majority of the students were Neutral at first and more positive during 2021-07.The most prominent topics during this timeframe from the tweets were covid-19,vaccine,travel,today,new,people.We can conclude from the analysis that people were more positive towards the vaccination.In order to adopt covid-19 vaccination , we can spread good news on social media platforms and reduce the negative sentiments.

Limitations :

We have collected the data from twitter and performed machine learning classifiers by splitting data into training and testing set, but we have not considered fake tweets .Fake news was not handled in this paper. We can extent this work using deep learning models which might increase the accuracy.

References:

1. Jones I, Roy P. Sputnik V COVID-19 vaccine candidate appears safe and effective. Lancet. 2021 Feb 20;397(10275):642-643. doi: 10.1016/S0140-6736(21)00191-4. Epub 2021 Feb 2. PMID: 33545098; PMCID: PMC7906719.

2. Chagla Z. The BNT162b2 (BioNTech/Pfizer) vaccine had 95% efficacy against COVID-19 ≥7 days after the 2nd dose. Ann Intern Med. 2021 Feb;174(2):JC15. doi: 10.7326/ACPJ202102160-015. Epub 2021 Feb 2. PMID: 33524290.

3. Umair A, Masciari E. Sentimental and spatial analysis of COVID-19 vaccines tweets. Journal of Intelligent Information Systems. 2022 Apr:1-21. DOI: 10.1007/s10844-022-00699-4. PMID: 35462784; PMCID: PMC9012072.