

An analytic study of people's changing sentiments towards the COVID-19 vaccine on Online Social Network with time

ECE 400 Final Project Report

Submitted by:

Ajit Singh (1801010)

Ankit Kumar (1801021)

Project supervisor: Dr. Surajit Panja

Indian Institute of Information Technology, Guwahati



Abstract

The COVID-19 pandemic has wreaked havoc since the start of 2020 and now in 2021, mass vaccination programs are in place in the majority of the nations throughout the world. At the time of writing this paper, countries like the USA, UK have double jabbed approximately 59.1% and 69% of their population, whereas India, with its population of over 1.4 billion people, has only been able to fully vaccinate 31.5% of its citizens^[1]. Majority of the COVID-19 vaccines are administered via two doses. However, now-a-days vaccine boosters have also been launched in many countries. The USA's food and drug administration (FDA) in September 2021 authorized booster doses of the Pfizer BioNTech COVID-19 vaccine for certain populations^[2].

Even though boosters are being launched for a certain sect of people who have already been vaccinated, there is still a large chunk of humanity who haven't been double jabbed yet. Vaccination in the majority of the places is encouraged but is not mandatory, hence there are many instances of people making it a point of debate. There is still skepticism regarding the acceptance of COVID-19 vaccines by some. Many people now-a-days voice out their opinions on social networks online. One of them is the popular microblogging platform, Twitter^[3] where users summarize their opinion (known as 'tweets') in 280 characters or less.

Hence, our study aims to analyze and summarize how people's opinion online on Twitter has shifted from the start of the general global vaccine rollout (December 2020) to now, when boosters are being launched for double vaccinated people (September 2021)^[5]. Additionally, we also take this study into the Indian context, by repeating the analyses for understanding evolving people's opinion on the COVISHIELD vaccine. Serum Institute of India (SII) and AstraZeneca produced COVISHIELD vaccine has become the backbone of India's fight against COVID-19. By the end of August, India had administered 650 million doses of the COVID-19 vaccine and approximately 600 million of those were COVISHIELD vaccines^[4].

For our study, we scraped over 2.4 million Tweets consisting of various keywords related to the COVID-19 vaccine discussion globally and India. In the first part of our study we aim to understand people's conversations and attitude towards the vaccine globally by using various tools and methods of Natural Language Processing (NLP) like sentiment analysis, phrase co-occurrence (Bigram) analysis and emotion analysis.

In the second part, we perform the same set of analyses for COVISHIELD data in order to take our study into the Indian context.

Our results clearly indicate that sentiments from the start of vaccine

rollout (December 2020) to mass acceptance of the vaccines now (September 2021) have shifted towards the negative side. Negative opinions and sentiments related to the vaccines have increased. However, the sentiments around the pre-rollout of vaccine boosters are exactly the same as that of pre-rollout of vaccines, indicating that people at the beginning of rollout of such a countermeasure are generally positive and accepting. Through our bigrams analysis, we identified that fear of long term side effects, allergic reactions and questions regarding natural immunity were major causes of concerns among the public regarding the COVID-19 vaccine. We also understood some more phrases and focal points of conversations used by pro and anti vaccination people through the study. Finally, we validated our results from sentiment analysis by performing emotion analysis on the tweets dataset.

Governments throughout the world can focus on these terms and topics to further spread education regarding them so that the people who have been misled (through rampant spread of misinformation online now-a-days) can understand better, increase awareness and be encouraged to take up the vaccines.

Contents

Sr.	Chapter	Pages
1)	Objectives	4
2)	Introduction	5-6
3)	Literature Survey and Related Work	7-9
4)	Introduction to our analyses	10-11
5)	Methodology	12-20
6)	Evaluation	21-52
7)	Conclusions	53-54
8)	References	55-56

Objectives :

As vaccines for COVID-19, which has already wrecked the deadliest pandemic related havoc for the world and India alike, have started to get rolled out, the society is expected to polarize into two factions-one wherein people have reservation about being vaccinated and voice this on online social media, and the one wherein people advocates being vaccinated. In order to better spread the message of getting vaccinated among the people, the governments of the world have to understand the key concerns of the skeptics.

Our objectives are to achieve the following:

1. Analyzing the people's reaction to covid vaccination over two succinctly different time periods (pre vaccine rollout and pre booster rollout) on the basis of twitter data.
2. Examining the Indian public sentiment related to the COVID-19 vaccine in India.
3. Analyzing the most frequent words people use in their conversations online (pro and anti Vaccination) to help understand certain communities and their concerns.
4. Analyzing the public emotions towards covid vaccination over time.

In this report, we have acquainted ourselves with the analysis of social network data. For our analyses here, we have used generalized, global Twitter data on COVID-19 vaccination as well as a topic in the context of India (Covishield vaccine) and performed sentiment, emotion and ngrams network analysis to get an overview of the information that spreads through social networks. We believe that understanding sentiments, emotions and knowing the kinds of phrases used by certain sections of people is the first step in analysing homophily in social networks.

Introduction :

Machine Learning is one of the most exciting subset of Artificial intelligence that brings out the power of data in new ways. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. There are several things that couldn't have been imagined even a few years back, but now it's possible with the help of machine learning^[6]. Sentiment and emotion analysis, speech recognition, virtual personal assistant, self driving cars, medical diagnosis etc. are some most trending real world applications of machine learning that are sweeping the entire world in this era.

The impact of social media over the last 20 years has been significant. It has gone from being an entertainment source to a fully integrated part of nearly every aspect of daily life for many. In this age, we all use social media like Twitter, Facebook, Reddit, Instagram etc. and share our opinion on trending topics through it. According to some sources^[7], Facebook is the largest social media platform globally with 2.85 billion monthly active users worldwide, while Twitter has around 397 million monthly active users. Twitter is encountering fast development social media. Many people generally use Twitter to share their opinions on trending topics and raise trending issues. Twitter is used highly for information spreading, political issues, etc.

Recently, The coronavirus disease 2019 (COVID-19) pandemic is one of the most trending topics on Twitter from the start of 2020. We know that Covid-19 has affected many lives all over the world. Almost every sector has suffered a lot due to this pandemic. The day when it all started is unforgettable. The first case of Covid-19 was reported in Wuhan, China on 31 Dec 2019. On 11 March 2020, Covid-19 was declared a global pandemic by World Health Organisation (WHO)^[8] and after that many countries were starting to impose lockdown in their country. In March 2020, According to WHO's situation report-71, globally, over 36,405 people had died, and more than 750,800 had been infected , while In India, total 29 people had died and 1071 cases had been reported^[9]. By 20 Nov 2021, there had been more than 5.12 million confirmed deaths and more than 258.47 million confirmed Covid-19 cases^[10], while In India, there had been more than 467,000 confirmed deaths and more than 33.6 million confirmed cases. Covid-19 vaccination was started in Dec 2020. After Covid-19 vaccine rollout, the situation has improved significantly. By 20 Nov 2021, approx 43% people had been fully vaccinated globally, while in India, approx 31% people had been fully vaccinated^[11]. Many countries are also planning to start booster dose (additional dose of 2 doses of vaccine) because of covid-19 third wave. The USA has approved booster doses in Sep 2021 for 65 and older. Widespread acceptance of a vaccine will be the next major step in fighting the coronavirus disease pandemic^[12]. People's trust on vaccination depends on the positive effect of vaccines. The information regarding vaccination which is spreading on social media affects people's thinking. Our aim is to analyze how people's sentiment and emotion on Covid-19 vaccination changes over time.

This paper aims to analyze public sentiment and emotion towards Covid vaccination and how it changes over time. We also try to find out which type of words people use frequently in their conversation about the vaccination. For analyzing, we scrape tweets based on different keywords for example 'covid vaccine', 'vaccine booster', 'covishield', etc globally to understand people's opinion on vaccination.. The tweets in this study are related to discussions about vaccination and booster doses during covid-19. We have performed sentiment, emotion and Bigram analysis to have an overview of people's opinion regarding Covid Vaccination.

Literature survey and Related work

1. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA^[12]:

In this paper, they conducted a randomized controlled trial in the UK and the USA to quantify how exposure to online misinformation around COVID-19 vaccines affects intent to vaccinate to protect oneself or others and find that some sociodemographic groups are differentially impacted by exposure to misinformation. They have also shown that in both countries—as of September 2020—fewer people would ‘definitely’ take a vaccine than is likely required for herd immunity, and that, relative to factual information.

They found that online social network structures, governed by social homophily, can lead to selective exposure and creation of homogeneous echo chambers and polarization of opinions, which may amplify (or dampen) the spread of misinformation among certain demographics. There is evidence of echo chambers on real social media platforms around information on vaccines, in general. If such information silos also exist for COVID-19 vaccines, then they may lead to self-selection of misinformation or factual information, inducing individuals to become progressively more or less inclined to vaccinate. Their study does not directly quantify such social network effects, it emphasizes the need to do so further. Furthermore, they found correlational evidence that misinformation identified by our participants after exposure as having the most impact on lowering their vaccination intent was made to have a scientific appeal, such as emphasizing on a direct link between a COVID-19 vaccine and adverse effects while using scientific imagery or links to strengthen their claims.

2. The echo chamber effect on social media^[13]:

In this paper, they explored the key differences between the leading social media platforms and how they are likely to influence the formation of echo chambers and information spreading. They performed a comparative analysis on more than 100 million pieces of content concerning controversial topics (e.g., gun control, vaccination, abortion) from Gab, Facebook, Reddit, and Twitter. This analysis focused on two main dimensions: 1) homophily in the interaction networks and 2) bias in the information diffusion toward like-minded peers.

Their results showed that the aggregation in homophilic clusters of users dominates online dynamics. However, a direct comparison of news consumption on Facebook and Reddit shows higher segregation on Facebook. They found out significant differences across platforms in terms of homophilic patterns in the network structure and biases in the information diffusion toward like-minded users. A clear cut distinction emerges between social media having a feed algorithm tweakable by the users (e.g., Reddit) and social media that don't provide such an option (e.g., Facebook and Twitter).

3. COVID-19 Echo Chambers: Examining the Impact of Conservative and Liberal News Sources on Risk Perception and Response^[14] :

In this paper, they explored their information-seeking tendencies, reliance on conservative and liberal online media, risk perceptions, and mitigation behaviors. The results of their study suggested that risk perceptions may vary across preferences for conservative or liberal bias; however, their results didn't support differences in the mitigation behavior across patterns of media use.

Further, their findings didn't support the notion of echo chambers, but rather suggest that people with lower information-seeking behavior may be more strongly influenced by politicized COVID-19 news. Risk estimates converge at higher levels of information seeking, suggesting that high information seekers consume news from sources across the political spectrum. These results are discussed in terms of their theoretical implications for the study of online echo chambers and their practical implications for public health officials and emergency managers.

4. A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis^[15] :

In this paper, they used NLTK, Text blob and VADER Sentiment analysis tools to classify the movie reviews which are downloaded from the website www.rottentomatoes.com that is provided by Cornell University, and made a comparison on these tools to find the efficient one for sentiment classification.

They did a comparison between these three analysis tools and found out that VADER performs exceptionally better than Text Blob and NLTK. The accuracy of VADER is 77% but for Text blob and NLTK is 74% and 62% respectively. VADER Sentiment Analysis works better for texts from social media and other Web sources as well, than Text blob because when it comes to analysing comments or reviews from social media, the sentiment of the sentence changes based on the emoticons. VADER also supports emoji sentiments. It also considers up to three exclamation marks that add the additional positive or negative intensity. VADER also follows grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. They concluded that VADER is a gold standard list of lexical features which is specially attuned to find semantics in micro blog text.

Introduction to our analyses

- Introduction to Sentiment analysis^[16] :

Sentiment analysis is the classification of subjective opinions within the text data using natural language processing. It basically gives an idea about any sentence. The main task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature level whether the expressed opinion in a document, a sentence or a feature is positive, negative, or neutral. There are mainly 3 approaches to extract the sentiment from given reviews and classify the result as positive or negative.

- a. Lexicon Based Approach
- b. Machine Learning Approach
- c. Hybrid Approach

The machine learning approach applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon, a collection of known sentiment terms. It is divided into dictionary-based approaches and corpus-based approaches which use statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches. The accuracy of a sentiment analysis is based on how well it agrees with human judgements. We used 2 lexicon based approach methods Text Blob and Vader for our analysis.

- Introduction to Emotion analysis^[21] :

Emotion is a strong feeling about a human's situation or relation with others. Human beings can easily identify the emotions from text and experience it but machines can't. Emotion analysis is an evolved form of sentiment analysis. It detects and analyzes the well known emotions such as Happy, Angry, Surprise, Sad and Fear from any textual data. We tried to find the emotion behind the user's tweets based on vaccination.

- Introduction to Bigram analysis^[22] :

An N-gram is an N-character slice of a longer string. As the name suggests, Bigram is a 2 character slice of any string. Bigram analysis is used to find out which words often show up together in any large datasets. For example, The word "SEARCH" would be composed in Bigram like SE, EA, AR, RC and CH. Using Bigram analysis, we can know which type of words people use most frequently in their conversation.

- Introduction to networkX^[17] :

NetworkX is a python library for studying graphs and networks. With NetworkX we can load and store networks in standard and nonstandard

data formats, generate many types of random and classic networks, analyze network structure etc.

- Introduction to Text Blob and VADER:

TextBlob and VADER both are lexicon based sentiment analysis methods. We used both methods for our analysis.

TextBlob^[18] :

TextBlob provides us with an API which allows us to make use of text processing operations such as sentiment analysis, part of speech (pos) tagging, noun phrase extraction, etc. for our data analysis. TextBlob is itself based on the NLTK (Natural Language Toolkit) platform, and provides a very good general overview of the data; Hence we found it suitable for our analysis purposes. The approach that the TextBlob package applies to sentiment analysis differs in that it's rule-based and therefore requires a pre-defined set of categorized words. Thus, it can be classified as a lexicon-based approach to sentiment analysis. In TextBlob, the range of polarity is between -1 to 1. Negative values near -1 indicate negative sentiment, positive values near 1 indicate positive sentiments, and 0 indicates neutral sentiment.

VADER^[19]:

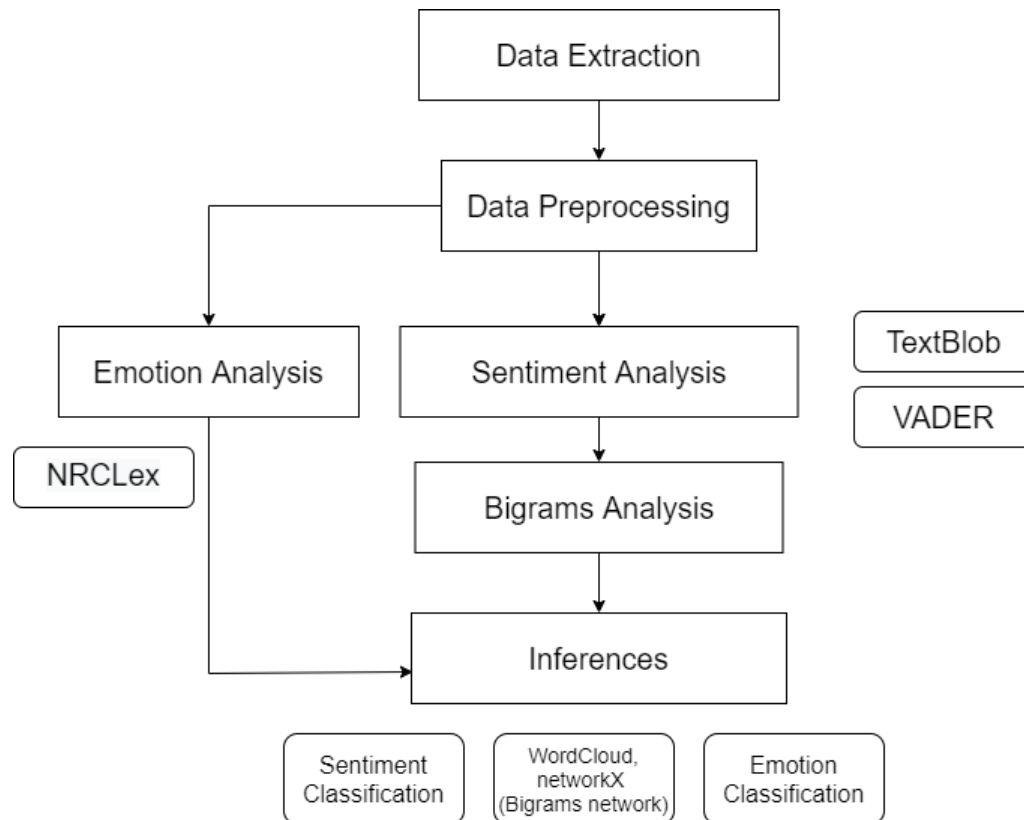
VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule based sentiment analysis method. VADER has been quite successful when dealing with social media texts, product reviews etc. VADER not only tells about the positivity and negativity score but also tells about how positive or negative a sentiment is. VADER analyses a piece of text to see if any of the words from the text are present and returns the metric values of the negative, neutral, positive, and compound for a given sentence.

- Introduction to NRCLex^[20]:

NRCLex is a Lexicon based tool that measures the emotion behind the speech or text. It basically finds all eight well known basic emotion categories such as Anger, Fear, Anticipation, Trust, Surprise, Sadness, Joy, and Disgust and two sentiments like Positive and Negative. It actually analyzes the textual data and recognizes the emotion embedded in it.

Methodology

In this section we describe our methods of data collection. We also describe the various methods used in our data analyses.



Overall workflow for our analysis

Data used in our analysis:

For our analysis we extracted data from Twitter, a popular microblogging platform where people can share their opinions (known as tweets) within 280 characters. Twitter is vastly used by over 300 million people to voice their opinions online, and allows data to be extracted for academic and research purposes. In order to scrape data from twitter, we made use of an open source Python based software package (library) called snsrape.

We collected about **2.5** million tweets overall for various different parts of our analyses. First, we collected all tweets with the keywords '**covid vaccine**' for the entire months of **December 2020** and **September 2021**. We then collected all the tweets containing the keywords '**vaccine booster**' for the month of September 2021. Lastly, we collected all tweets containing the keyword '**covishield**' for the entire time period from **December 2020** to **September 2021**.

We thus separate our analyses for **two** different datasets, the first one containing tweets regarding **global topics** (covid vaccine, vaccine boosters) and the second one containing tweets on the **Indian** vaccine, Covishield.

Here is the summary of collected tweets in the dataset:

Twitter Dataset 1: Vaccination tweets used in our experimental evaluation

Keyword/s used to scrape tweets	Time period over which tweets were scraped	Total number of collected tweets
Covid vaccine	December 2020	1,048,499
Covid vaccine	September 2021	921,240
Vaccine booster	September 2021	82,117

Twitter Dataset 2:

Keyword/s used to scrape tweets	Time period over which tweets were scraped	Total number of collected tweets
Covishield	December 2020 - September 2021	371,580

Total number of tweets collected: **2,423,436**

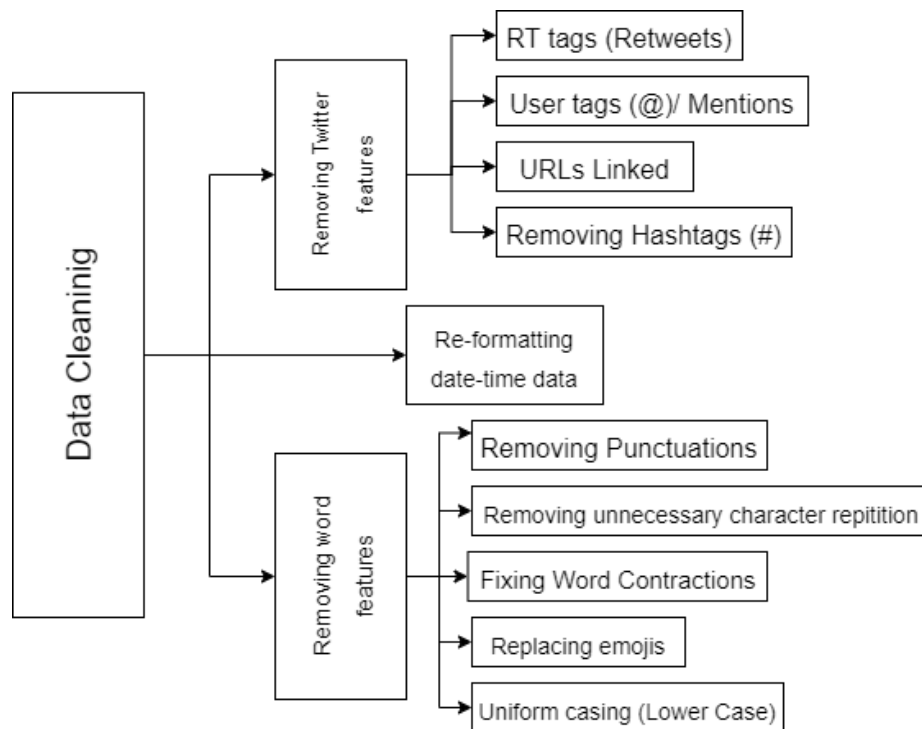
Twitter Dataset 1:

We chose tweets from different time periods so as to gain an insight into how public conversations and subsequently, sentiments change and evolve with time. In December 2020, mass vaccination campaigns across countries had not yet taken place and vaccine rollouts were going on across continents. Whereas in September 2021 a sizable chunk of humankind has been vaccinated with COVID-19 vaccines and boosters have been launched. Hence, we found it interesting to cover people's sentiments for two different time periods with different underlying circumstances. For our booster analysis, we wanted to understand what similarities are present in people's opinions online pre-vaccine rollout and pre-booster rollout.

Twitter Dataset 2:

To understand the shifting sentiments around the Indian vaccine COVISHIELD, we collected a continuous set of data containing all tweets from December 2020-September 2021.

Data preprocessing:



Workflow for data preprocessing

Text preprocessing is the most integral task for any NLP process. Since we used TextBlob, VADER for our sentiment analysis, we didn't need to perform pre-processing operations like tokenization, stemming or lemmatization on our tweets. Rather, we used some 3rd party Python based modules as well as made use of regular expression evaluations to clean and preprocess our tweets. We followed the following set of rules to clean and process our text data to be fed into the models:

a. Processing the tweet features:

Each tweet made by users on twitter has some fundamental properties, such as retweet mentions (RT), user tags (@) for tagged users, twitter hashtags (#). Also, many users often link URLs to external web pages in their tweets. So, the first step in data cleaning was to process (clean) these features according to our needs.

i. Cleaning Retweet mentions:

Twitter has a feature called 'retweet', which allows a user to pass on someone's opinion (tweet) on his own page. However, the credit of the original tweet still remains with the original poster. Thus, the user who forwarded the original tweet is said to have 'retweeted' the tweet. The Retweet feature basically allows users to share others' tweets. If someone retweets a tweet, twitter automatically adds the string 'RT' in front of the new retweeted tweet. For our text cleaning purposes, we removed the 'RT' (Retweet tag) from all our tweets since it does not have any

significance over the sentiment of the tweet.

ii. Removing URLs:

Twitter allows users to share URLs (Internet links) in tweets. However, we removed all the URLs from the collected tweets as it doesn't add any value to the sentiment of the tweet.

iii. Removing tagged users in tweets:

If a user tags another user in their tweets, twitter adds their username preceding with an '@' character to the tweet. For our analysis, we removed all the mentions of usernames in the tweet since usernames add no significant meaning to the sentences.

iv. Removing the '#' (Hash) character:

Twitter hashtags allow conversations to be grouped under specific topics (known as hashtags). These so called grouped conversation topic names (Hashtags) are preceded with the '#' symbol. We removed all the hashtags from our collected set of tweets.

b. Processing the word features:

People often use a very informal style of language on social media. There are no strict formal rules followed. However, these patterns of informal language are still perceivable by the human brain since we can think of our own. This is simply not true for machines. For our program, feeding it with a 'Hello' and 'Helloooo' is quite different. Even different word casings for the same word are classified as two different words. Hence, we processed some word features in the tweets present in our dataset by following the following set of rules:

i. Replacing emojis with their text description:

We replaced all the emojis present in our tweets with their text meanings so that they could add meaning to the tweets in terms of text. Emojis do contribute towards expressing sentiment in typed text and hence we didn't choose to omit them from our corpus of tweets.

ii. Fixing word contractions:

The words such as 'I've' or 'Don't' are said to be forms of word contractions. We can make interpretations of these contractions very well as human beings. (I have or Do not in

this case). However, making the meaning of word contractions is not possible in the case of machines and hence we replaced all such contractions by their actual full forms.

iii. Replacing word and punctuation repetition:

The language used by users on social media is extremely informal and is not exactly governed by the formal, grammatical rules. Hence, instead of saying 'Hello!' one can write 'Helloooo!!!' and can get away with it since we humans can perceive such changes in the use of language. However, in terms of computer interpretation, both these words are different. Hence, we replaced all the repetition of characters and punctuations with just a single occurrence of these said characters.

iv. Converting all text into lower case:

In order to maintain uniformity in word casing, we converted all the tweets into lower case.

c. Cleaning the date-time data:

One of the tweet components we fetched during our data extraction process was the local time of posting of the tweet. However, we only required the date data required for our needs (plotting tweet engagements for time series data). Hence, we processed the date-time data collected during extraction to represent only the date of the tweet.

Shown below is the output of text preprocessing:

```
tweet = "RT @AnkitKumar Don't listen to them 😡 Veganism is the way to go forwarddddd!!!!!"  
new_tweet = process_tweet(tweet)  
new_tweet
```

```
' do not listen to them :angry_face: veganism is the way to go forwardd!'
```

The original tweet contained a retweet tag, user tag, word contraction, emoji, multiple punctuations and had mixed word casings. However, our processing function returned the clean version of the tweet as per the rules we defined above.

Sentiment analysis: Tools used and Sentiment Categorization

In the first part of our analysis on social network data, we are performing

sentiment analysis on our tweets dataset. For this Natural Language Processing (NLP) task, we are making use of TextBlob and VADER (Valence Aware Dictionary of sEntiment Reasoning), which as described earlier are lexicon based models for sentiment classification. Generally, VADER is better for predicting sentiments for social network data, but we used both models (VADER as well as TextBlob) in our analysis in order to confirm and affirm whether the results and trends which we were getting were consistent or not. TextBlob gives us a polarity score, which helps us to classify sentiments. Similarly, VADER also gives a compound score, which is used to set classification labels.

We set the TextBlob polarity score as well as the VADER compound score's neutral threshold to be 0.05. Hence, we used the following rule to classify sentiments:

Score Range	Sentiment Category
$-0.05 < \text{score} < 0.05$	Neutral
$\text{score} < -0.05$	Negative
$\text{score} > 0.05$	Positive

Table describing our rules to categorize sentiments based on Polarity Score (TextBlob) and Compound Score (VADER)

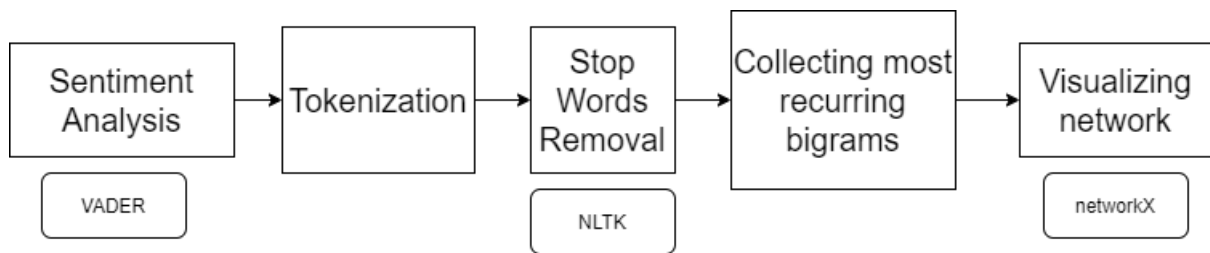
In order to infer results from our sentiment analysis, we make use of data visualizations such as time-series plot (to capture trends over a period of time), bar chart (to chart percentages) and a rather unique form of data visualization technique used for text data: Word clouds.

Word Clouds:

Plotting Word clouds^[26] is a technique of visualising text data. It conveys important information about the text corpus such as the most recurring words and also their intensity of recurrence. A word cloud is a collection, or cluster, of words depicted in different sizes. The bigger and bolder the word appears, the more often it's mentioned within a given text and the more important it is pertaining to the text corpus.

Thus, we will plot out word clouds for positive and negative tweets to better understand what kinds of words people are using in tweets which are positive/negative. We will use sentiment classification results from TextBlob as well as VADER and the wordcloud module to chart our word clouds:

Bigrams Analysis:



Workflow for bigrams analysis

For our bigrams analysis, we first followed the same text (tweet corpus) preprocessing steps as we did for the sentiment analysis described above. Then, we tokenized our tweet corpus.

Tokenization

Natural Language Processing (NLP)^[25] and other forms of computerized processing are analyses which are based on words and not sentences. Thus, we often need to split our data into small chunks, called 'tokens'.

Tokenization essentially means splitting a phrase, sentence, paragraph, or an entire text document into smaller units, such as individual words or terms. Each of these smaller units are called tokens. Thus, we can say that the process of breaking a sentence that has been cleaned into different words is called tokenization.

```
['This is a list']
```

A list containing a full sentence

```
['This', 'is', 'a', 'list', 'of', 'words']
```

A list containing a set of tokens

Next, we remove the stopwords from our tokenized tweet corpus.

Removing Stop Words

Stop words are the words that do not add any meaning to the sentence in terms of Natural Language Processing (NLP). Some examples of stop words in the English language are:

"i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "your", "yours", "yourself", "yourselves", "he", "him", "his", "himself", "she", "her", "hers", "herself", "it", "its", "itself", "they", "them", "their", "theirs", "themselves", "what", "which", "who", "whom", "this", "that", "these", "those", "am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had", "having", "do", "does", "did", "doing", "a", "an", "the", "and", "but", "if", "or", "because", "as", "until", "while"

To remove the collection of stop words, we make use of the Natural Language

Toolkit (NLTK)^[23]. NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning and wrappers for industrial-strength NLP libraries^[25]. NLTK provides a list of stopwords in English. Hence, we identify all these words from our tokenized set of words and remove them from our tweet corpus.

Capturing and counting Bigrams

After removing stopwords, we capture all the bigrams for every tweet. For this, again we made use of the Natural Language Processing Toolkit (NLTK)^[23]. After capturing all bigrams for a tweet and for every tweet, we counted the recurrence of bigrams for tweets with negative sentiments as well as positive sentiments.

Graphing network visualizations

To truly understand and visualize the kind of recurring phrases used by people in different types of tweets (phrases used by people in Positive/Negative/Neutral tweets), we charted a network plot using networkX^[17], the leading software package in Python for studying graphs and networks. The network was plotted using networkX's spring layout functionality, which makes use of the Fruchterman-Reingold force-directed algorithm^[24] for positioning nodes.

The edges are assigned between words. (A bigram is represented as 2 nodes (each word represents a node) which are connected by an edge (representing that the words are often used in succession; i.e. the inter-connected pair are bigrams) The properties of the edges themselves and the positioning of the nodes do not matter in the network; It is the inter-connectivity between words (leading to formation of phrases) that is of interest to us. Thus, by plotting network plots of bigrams, we can see most frequently used words and how some of these words are interconnected with each other. Thus, it helps us to extract different kinds of phrases used recurrently/ identify recurrent themes in different categories of tweets: Positive, Neutral or Negative.

Emotion Analysis:

For our task of emotion classification of tweets, we are making use of the tool NRCLex. NRCLex^[20] measures the 'emotional affect' from a body of text. The affect dictionary contains approximately 27,000 words, and is based on the National Research Council Canada (NRC) affect lexicon and the NLTK library's

WordNet synonym sets. Basically, for each tweet it returns us the possible emotions the tweet passes on, which may be a category from the following: fear, anger, trust, surprise, anticipation, joy, sadness, positive, negative. Note that positive and negative refers to the emotion that is portrayed by a tweet and not its underlying sentiment.

Now that we have described our work methodology, we will showcase the results of our analyses. First we have performed all our analyses on twitter data pertaining to global topics (covid vaccines, vaccine boosters are talking points globally). Then we repeat our analysis for the Indian vaccine COVISHIELD. Thus, we can gain insight on conversations online about both people worldwide talking about vaccines as well as those commenting about the Indian vaccine COVISHIELD.

Evaluation

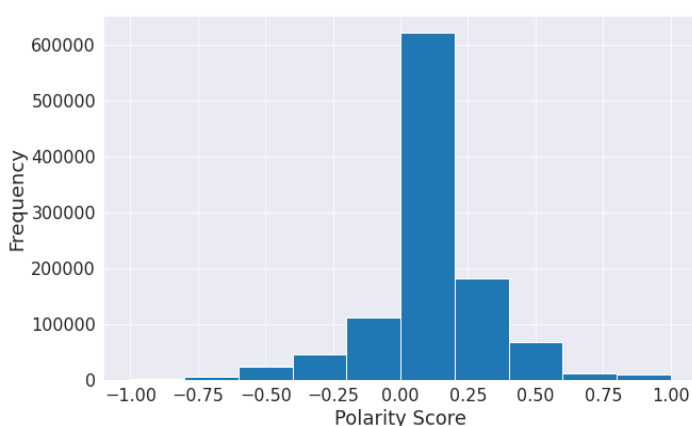
Analysis on the first dataset:

As described earlier, the first dataset consists of tweets pertaining to global topics such as covid vaccine and vaccine booster. First, we perform analyses on this dataset.

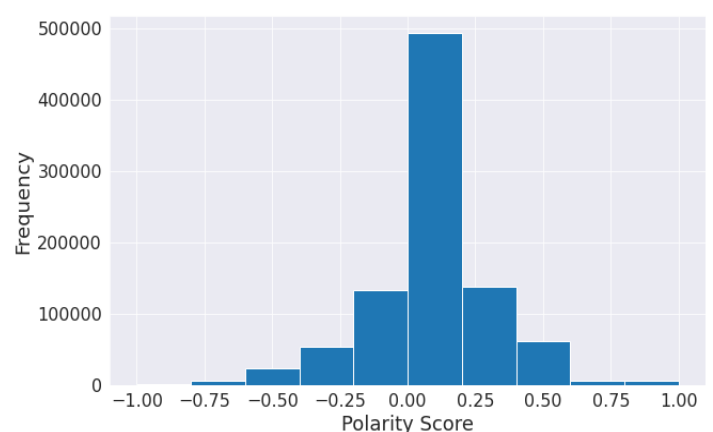
Sentiment analysis on covid vaccine data

As described earlier, we use TextBlob and VADER to evaluate our sentiment analysis. After feeding in data which has been preprocessed, we obtain a polarity score in case of TextBlob and a compound score in the case of VADER. And based on these scores, we classify the tweet as having positive, negative or neutral sentiment. First, we perform sentiment analysis on the December 2020 'covid vaccine' tweets data followed by the same on September 2021 data. We observed that many of the tweets which were classified as either slightly positive or neutral by TextBlob were classified as negative by VADER. We find that TextBlob is more likely to classify tweets as "neutral" than VADER, which is more likely to classify sentences as being either of the two extremes. In general, VADER performed better classification as it was able to judge intense/extreme sentiments better and the polarity scores' distributions for VADER were relatively more evenly and well spread across varying polarity scores as compared to TextBlob.

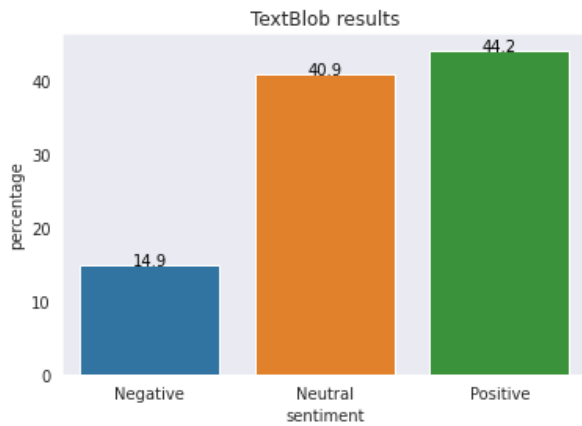
Results from Sentiment Analysis:



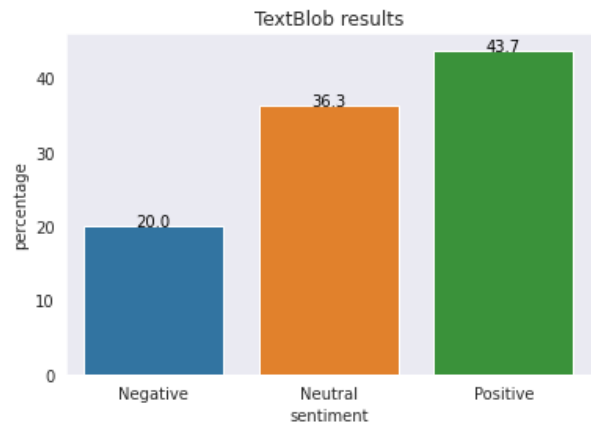
Polarity scores' distribution - December Data (TextBlob)



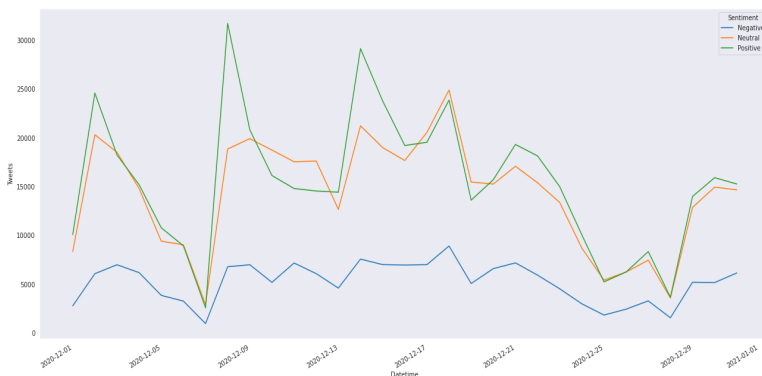
Polarity scores' distribution - September Data (TextBlob)



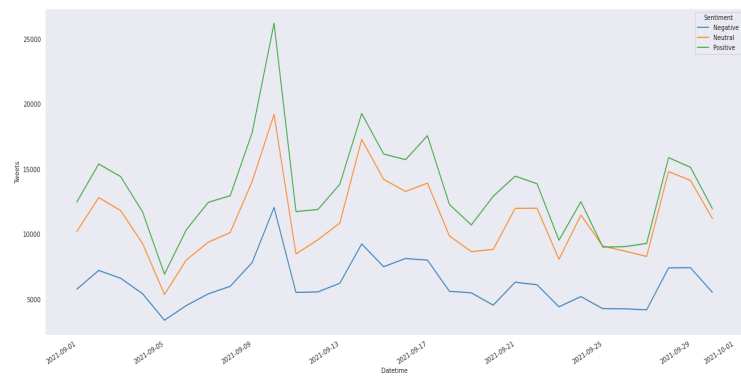
Percentage of Tweets across sentiment categories - December Data (TextBlob)



Percentage of Tweets across sentiment categories - September Data (TextBlob)



Time Series plot of tweets across the sentiment categories over the month of December (TextBlob)

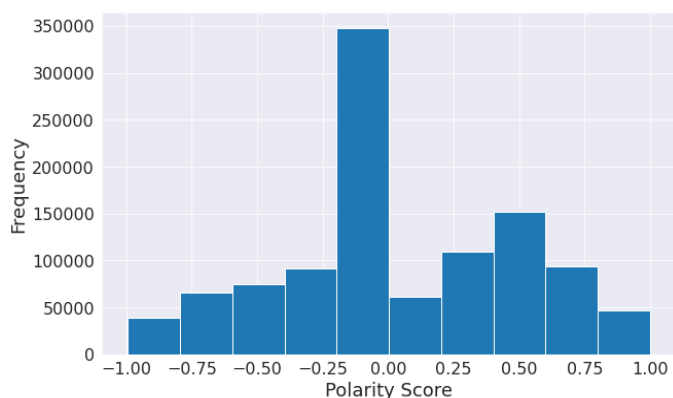


Time Series plot of tweets across the sentiment categories over the month of September (TextBlob)

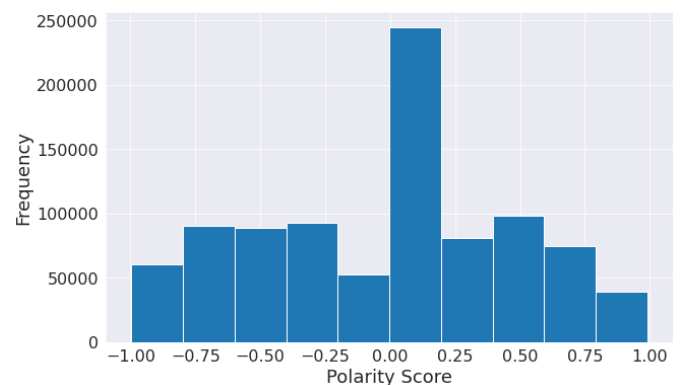
We observe that the public sentiments expressed on Twitter via tweets regarding covid vaccines were mostly neutral or slightly positive at the beginning of the mass vaccine rollout throughout the world. Majority of the factual (fact based) Tweets were classified as “neutral” by TextBlob. Also, upon seeing the time-series plot we can say that the percentage of positive, negative or neutral tweets at any given time is more or less the same as all 3 trendlines for the respective sentiments seem to be following the same pattern. It’s the frequency of tweets though, which changes. For eg: For some days, the number of tweets with keyword “covid vaccine” was very large in number (as can be seen by the major spikes in the time-series plots). These spikes were mostly the result of some major public event or a decision taken by some authority related to covid vaccines. We were able to isolate some key major events using these spikes. For December, one of the spikes corresponds

to the date when the USA president and vice president took the vaccine^[27]. It obviously sparked many conversations online as these people are prominent public figures. In September, one of the spikes near September 17 corresponds to the time when boosters were officially recommended for senior citizens by the USA FDA panel^[2]. However, as vaccine rollouts were scaled up and as a large section of the population started to get fully vaccinated by the end of September, opinions seemed to have shifted towards the negative side. We can clearly see that negativity has risen (almost 5% increment in TextBlob score). Even though the conversations around the COVID-19 vaccines are still positive in nature, the general negativity around the vaccines has increased. Textblob is known to have a positive bias (Unable to classify negative opinions relatively well) and hence we have less percentage of negative tweets. However, we want to confirm the trend which this model is trying to put into picture here.. Has negativity surrounding the vaccines risen with time as more people have been vaccinated? In order to validate our analysis, next we repeat the process using VADER.

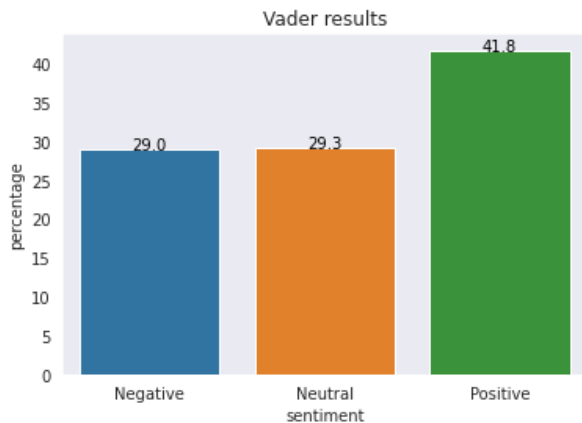
VADER confirms our trend that negativity around the vaccines has risen in September. It is interesting that VADER classifies most of the tweets as being negative rather than positive/neutral. However, the change in percentage of positive tweets classified by VADER is only around 5%, around the same change in negative tweets from December to September as shown by TextBlob, just in the opposite direction. However, now a majority of the neutral tweets have been classified by VADER to be negative. It can be accorded to the intense frequency of usage of words like 'death', 'sick', 'ill', along with "side effects" etc. in tweets surrounding the conversations around COVID-19 vaccine nowadays rather than majority of the tweets being fact-based or simply in anticipation of the vaccine pre-rollout which were classified to be neutral.



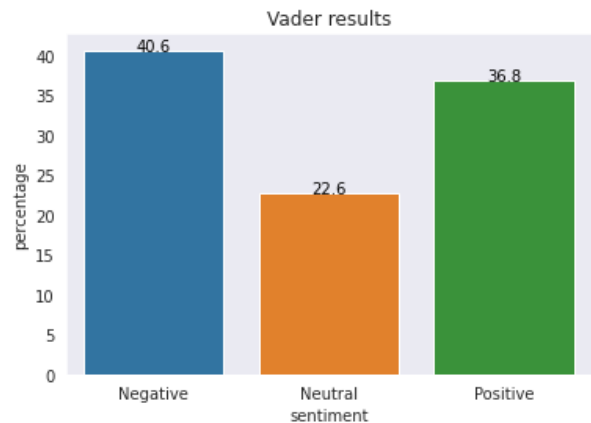
Polarity scores' distribution - December Data (VADER)



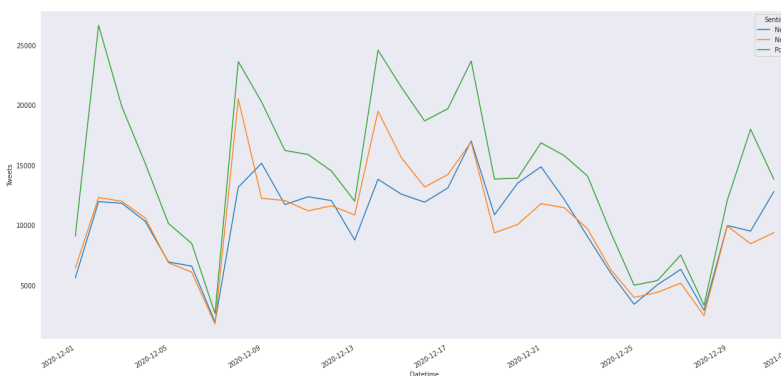
Polarity scores' distribution - September Data (VADER)



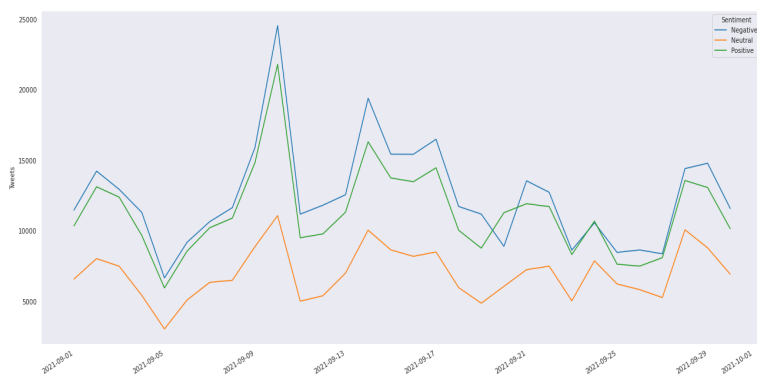
Percentage of Tweets across sentiment categories - December Data (VADER)



Percentage of Tweets across sentiment categories - September Data (VADER)



Time Series plot of tweets across the sentiment categories over the month of December (VADER)



Time Series plot of tweets across the sentiment categories over the month of September (VADER)

Even though the individual numbers across sentiment categories differ for TextBlob and VADER, the trend of increasing negativity around covid vaccines online on Twitter can be well established by observing the results of the sentiment classifiers. Now, the next task is to assess the reasons behind these results and answer questions like:

1. What may be the possible reasoning behind this increased negativity?
2. What kinds of words or topics are being used/talked about in negative tweets surrounding covid vaccines?

To better understand the kinds of words frequently used by the people in their tweets, we plot out word clouds for tweets having positive/negative sentiments classified by each model.



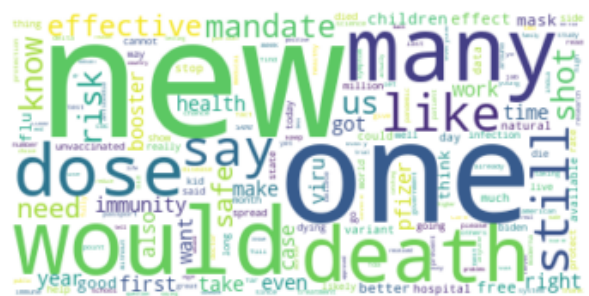
Word cloud for positive sentiments
(December Data, TextBlob)



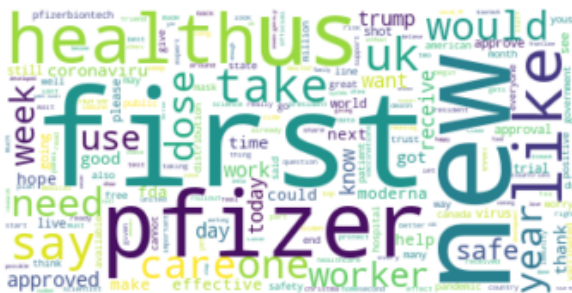
Word cloud for negative sentiments
(December Data, TextBlob)



Word cloud for positive sentiments
(September Data, TextBlob)



Word cloud for negative sentiments
(September Data, TextBlob)



Word cloud for positive sentiments
(December Data, VADER)



Word cloud for negative sentiments
(December Data, VADER)



Word cloud for positive sentiments



Word cloud for negative sentiments

The word clouds for the same time periods may look pretty similar since some of the commonly used English words (which aren't stopwords) such as 'would', 'like', 'still', etc. are expected to be there. However, closer inspection (the slightly smaller words) tells us a lot about these conversations. We can make some inferences from these word clouds which back the sentiment analysis results:

1. Positive word cloud contains generally considered to be positive words such as 'immunity', 'better', 'effective', 'many', 'health', etc. whereas the negative word cloud contains words such as 'anti', 'sick', 'died', 'bad', 'virus', 'stupid', etc. which are considered to be in negative light.
2. In the word clouds for tweets pre-vaccine rollout (December 2020), vaccine producing brands Pfizer, Moderna, etc. have much stronger presence (bigger, bolder) in the positive word clouds as compared to negative word clouds. Also, vaccine producing brands are much more present in word clouds for tweets from December 2020 as compared to September 2021, which is expected. We can conclude that conversations regarding vaccine producing companies are generally in a positive light. This may be due to positive public anticipation of the vaccine after dark periods of lockdown and failure of states to control the outbreak of the virus. Also, words like 'clinical trials', 'passed', 'approved' are present in the positive word clouds marking that people were mostly positive and buzzing about the impending vaccine rollout. Also, we can find words such as 'safe', 'good', 'effective' in the positive word clouds. These are attributed to people positively expressing their sentiments towards the vaccine in response to the reports and studies publishing results of the vaccines' clinical trials.
3. Pre-vaccine rollout negative sentiment word clouds have political terms and words in them. Upon analysing these tweets we came to the conclusion that mostly the political tweets surrounding vaccines were in a negative light. Divisive tweets often included words such as 'trump', 'biden', 'administration', 'government'.
4. Many of the negative tweets surrounding vaccines compared covid-19 to the common flu and cold. This can be observed in the negative word clouds prior to the vaccine rollout. Negative word clouds also have words like 'allergic', 'reaction' which are used by vaccine sceptics in fear that there has been a lack of proper clinical trials before release of vaccines. The negative word cloud also includes words such as 'stop', 'refuse', etc.
5. Positive word clouds for pre vaccine rollout include 'thanks', 'scientists', 'workers'. Most of these tweets were people showing their gratitude

towards front line healthcare workers and scientists. On the other hand, the same set of words were also present in the negative word clouds although they were used in a different context. People were sceptical about the work-life balance and working conditions of the front line workers and thus, we can say that most conversations surrounding front line workers were in a positive light. However, these conversations cannot be found in tweets from September 2021. This just means that when Covid-19 was at its peak and countries were in lockdown, healthcare workers and scientists were very much present in online conversations but as 'normality' resumes, these people fade away from the public spotlight once again.

6. Words such as 'distribution' and 'availability' are present in both positive and negative word clouds pre vaccine rollout. People were a bit sceptical about the manner of vaccine rollout and it was a big point of public debate.
7. Positive word clouds for post mass vaccine rollout (September 2021) contains words in support of the Covid-19 vaccine. 'Got', 'better', 'effective', 'vaccinated', etc. are examples of these words. 'Safe', 'Immune', 'Free vaccine', 'Approved' are more such words.
8. Word clouds for September 2021 include words such as 'variant', 'rate'. There were conversations around the discovery of new variants of the virus (delta, omicron, etc.) and the rate of growth and spread of these new variants. For a while, there was negativity and fear surrounding these topics.
9. Words such as 'booster', 'vaccine passport', 'mandate' are present in word clouds from 2021. Launch of vaccine passports and mandates passing usage of vaccine booster were controversial topics.
10. Masks were a topic of discussion in December 2020 as well as September 2021 and were present in most of the word clouds.

Next, we observe the top 10 most positive and negative tweets to understand the language used in tweets which have been assigned the most extreme polarity scores by our sentiment classifiers.

First, we plot out the top 10 most positive tweets classified by TextBlob and VADER for December 2020:

- **Top 10 Positive Tweets from December 2020 containing the keyword 'covid vaccine'**

TextBlob classification results

	Text	VComp	Username
0	@Yodaonexile @LizardK59688109 But, you've chosen to suggest her death doesn't matter cos she's elderly. If she died of flu, it would've been sad to die of external factor, to die of Covid-19 that has no vaccine makes her death more sad. Her heart didn't pack up, brain hadn't failed, so death is premature.	-0.9935	vospie
1	@ZubyMusic Weak men are generally scared men. Scared of immigrants, scared of guns being taken, scared of election fraud, scared of covid hoax, scared of the spectre of socialism, scared of vaccines, scared of higher education, scared of FEMA death camps, scared of death panels, etc, etc.	-0.9921	I_ArgYOU
2	@L_A_D_J But still flu has a vaccine every year and kills thousands 2018 it killed 58,000 so covid will still kill thousands yearly and nothing will be said bit odd to me it's been here over a year and killed 61,000 allegedly very suspect maybe knock the died with covid off it's quarter.	-0.9919	tm_hatton
3	Merica. Covid is fake. 5g is poison. Fluoride is poison. Gmo is poison. City water is poison. Vaccines are poison. Msm is poison. Chemtrails is poison. Gov is mafia. "Elite" are evil. Joseph Magele famous "dr Angel of death. Wake up. This means war. Avenged sevenfold. https://t.co/5XVOBo5nnc	-0.9911	Ronbert27734234
4	#COVIDIOT #COVID19 #CovidVaccine #COVIDIOTS #COVID19Vaccine. Fuck Trump. Fuck Pence. Fuck GOP. Fuck Cruz. Fuck Rubio. Fuck Paul. Fuck Mitch. Fuck Jordan. Fuck Gaetz. Fuck McConnell. Fuck Eric. Who did I forget, please add and RT https://t.co/yD670lvXw	-0.9899	brett_zen
5	covid fucked my birthday up and i'm mad. so fuck covid, fuck that vaccine, fuck my boyfriend, fuck them kids, fuck leos & fuck keyshia cole!	-0.9895	artixaj
6	@joncoopertweets That the Sociopath wants credit for the vaccine but will take no responsibility for the 10s of thousands who have suffered & died, often alone, because of his utter & deliberate failure to address the COVID-19 pandemic he knew was a catastrophic public health crisis! #Cruelty	-0.9886	UpSpireMe
7	I'll dead ass leave the country and be poor as fuck before I'm forced to take a covid vaccine I DO NOT CARE THAT IS THE MOST FUCKED HORROR MOVIE SHIT EVER CALL ME EVERY NAME IN THE BOOK I DONT CARE	-0.9883	pudge3000
8	TRUMP'S INCOMPETENT, LYING FRAUD ABOUT THE COVID VIRUS HAS COST MORE LIVES THAN MEN KILLED IN BATTLE IN WW2. LET THAT SINK IN. TRUMP MURDER. Coronavirus updates: US death toll surpasses World War II combat fatalities; Pfizer vaccine clear... https://t.co/hQixjmgxR via @YahooNews	-0.9878	dougalmac54
9	@mysageadvice @BelievePHQ @RachelintheOC @SueInge @RonThornton @Journeyingdave @Carolyn1948 @CROWDREAM @BreakingKind @AndyDRC @RodneyClaeys @Sam_ivzir @Rickbluewave140 @WholePlateWay @JerryAlberson To my knowledge, there are not significant risks of neurological complications with the #coronavirus vaccines. But many people who get #COVID suffer terrible neurological consequences. So I worry ...	-0.9877	LEAD_Coalition

VADER classification results

We observed that most of the positive tweets surrounding covid vaccines were in support of the vaccines. Note that it is not obvious. Positive tweets need not reflect public support towards the vaccines and negative sentiments doesn't necessarily mean that those tweets are anti-vaccine. A few negative tweets were anti-vaccine. But many negative tweets were also bashing those who were against vaccines. Also, pro-vaccine tweets tended to receive higher engagements (likes, retweets, etc.) than anti-vaccine tweets.

Then, we repeat the process for September 2021 :

- **Top 10 Positive Tweets from September 2021 containing the keyword 'covid vaccine'**

	Text	TBScore	Username
0	Vaccines are truly our best shot to beat COVID-19. Join @Network18 #Sanjeevani - A Shot Of Life, a @FederalBankLtd CSR initiative to spread vaccine awareness across the country, and get vaccinated soon. Vaccine #LagayaKya @Apollo24x7 @UWMumbai https://t.co/MyfonuPFu8	1.0	News18Kannada
1	Vaccines are truly our best shot to beat COVID-19. Join @Network18 #Sanjeevani - A Shot Of Life, a @FederalBankLtd CSR initiative to spread vaccine awareness across the country, and get vaccinated soon. Vaccine #LagayaKya @Apollo24x7 @UWMumbai https://t.co/wN4GWjW6X8	1.0	HindiNews18
2	Blockchain would have been an excellent technology choice for issuing COVID vaccine certificate.	1.0	anandmallaya
3	Covid doesn't exist and vaccines are poison is the best answer ever. https://t.co/teG07LPJjp	1.0	DanEvans1986
4	'Impressive' start for B.C. vaccine card, COVID hospitalizations increase https://t.co/yVbKMzSIMZ https://t.co/6wEYgPhVhw	1.0	VernonMattersca
5	@wmarshcc8 Bill, did you listen to my speech? Nowhere did I say don't get the vaccine, I encouraged everyone to talk to their physicians and make the decision best for them. I won't support a mandate, when I got the vaccine I was lead to believe it would keep me from getting covid again.	1.0	kimberlyarice1
6	@TomiLahren Did you stomp your foot when you tweeted that? Covid vaccines have been given to over 2 billion people. It's safety profile is excellent.	1.0	Nimrod247
7	World's Largest Vaccination Drive: In a magnificent feat, India administers 85 Crore COVID-19 Vaccines. In @mansukhmandviya @MoHFW_INDIA https://t.co/nWOqJ7sqv1	1.0	airworldservice
8	CDC says Moderna vaccine staying especially effective. COVID updates https://t.co/lmaLATgpSA	1.0	Humming_Zone
9	Oregonians who have received two doses of the Pfizer or Moderna vaccines, or one dose of the Johnson & Johnson vaccine, continue to have excellent protection from COVID-19. #CovidVaccine https://t.co/lxcuFwsInL	1.0	OHSUNews

TextBlob classification results

Similar trends can be observed for September 2021 covid vaccine tweets. Positive tweets were mostly in support of the vaccines and tweets in support of the vaccines received generally higher engagements as compared to anti-vaccine tweets. Some negative sentiment tweets spread false anti-vaccine propaganda and conspiracies such as covid vaccines are bioweapons. These conspiracy cults are unfortunately one of the largest proponents of anti-vaccine content online on social media. Many negative tweets were also political in nature.

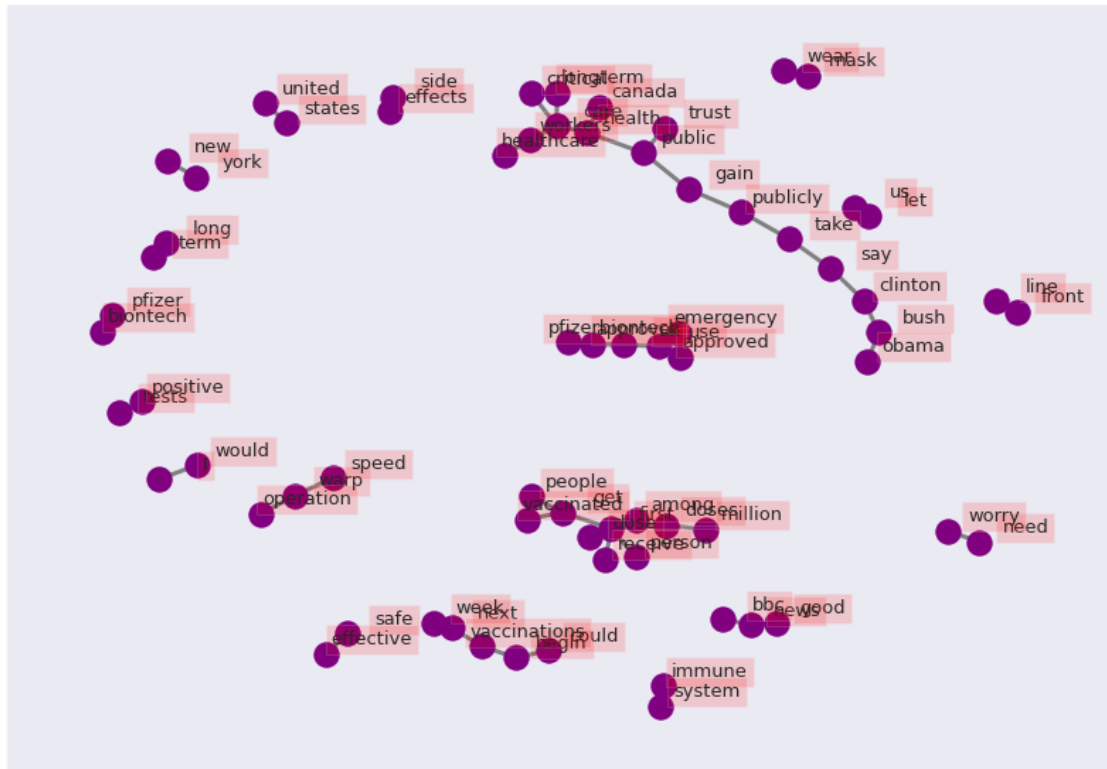
To further understand the kinds of words/phrases recurrently used by people in negative tweets surrounding covid vaccine, we perform bigrams analysis.

Bigrams Analysis on covid vaccine data

VADER classifies sentiments well for social network data (informal language). Hence, we use sentiments classified from VADER for our bigrams analysis. As described in our methodology, we perform the Bigrams analysis on the data.

0	(emergency, use)	7756
1	(side, effects)	6972
2	(allergic, reactions)	5082
3	(allergic, reaction)	3578
4	(new, strain)	3241
5	(healthcare, workers)	2949
6	(long, term)	2796
7	(use, authorization)	2794
8	(flu, shot)	2390
9	(wear, mask)	2335
10	(let, us)	2306
11	(front, line)	2251
12	(severe, allergic)	1932
13	(people, get)	1931

Most recurring Bigrams - Tweets with Negative Sentiments (December 2020)



Network plot showing interconnectivity between recurring bigrams

Bigrams Analysis for tweets pre-vaccine rollout reveals interesting set of observations about tweets with either of the extreme sentiments:

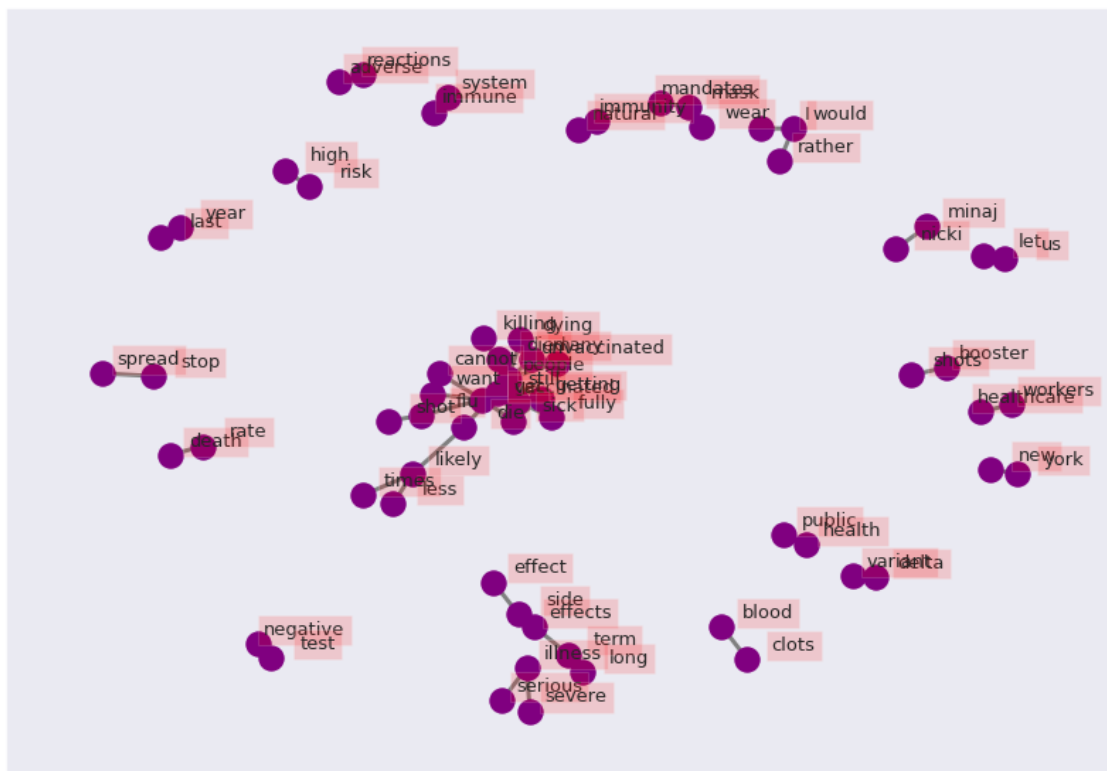
1. The interconnection between multiple pairs of bigrams enables us to get to know the recurring phrases/short sentences or topics used/talked about in our tweets. For eg: A long chain of words in the negative bigrams network plot which goes like 'lives-countless-costing-countries-income-lower-price-cut-refuses-expensive-billions-rake' refers to a piece of news from December which was widely circulated and talked about on Twitter, alleging that vaccine manufacturing company Pfizer could rake in billions from its expensive covid-19 vaccine and that it was planning to not cut price for lower income countries. Hence people were sceptical that this move could cost countless lives; Hence there were many negative tweets surrounding this set of words with people expressing their concern and thus, negative sentiment tweets.
2. We found that most of the people against vaccines/sceptics and in the extreme, anti-vaxxers and conspiracy theorists generally use **side effects, long term** side effects and **allergic reactions** to vaccines as the major reasons why they oppose it. Although many tweets expressed concerns regarding vaccines' unproven long term side effects, many tweets which were anti vaccine mainly harped on about potential vaccine induced severe allergic reactions to children and adults, etc. in

their tweets which received high negativity score by VADER. This can be seen in the network plot as well.

3. Conversations around 'new strain' of the virus were also marked as negative by VADER. Many of these tweets were also anti-vaccine and expressed concerns whether the virus could mutate before we could reach herd immunity through mass vaccinations.
4. We also observe many political topics in the negative sentiment bigrams network plot as well. Bigrams such as 'trump administration' and few other topics pertaining to American and European politics (From which majority of the Twitter users belong) can be seen.
5. Conversations surrounding healthcare workers were present in both positive and negative sentiment tweets. Some people were expressing their gratitude while some were expressing concerns surrounding the safety of healthcare workers. Vaccine sceptics expressed concerns surrounding clinical trials of the vaccine; Whether they were adequately tested before being administered to the workers or not.
6. 'Wear masks' appears in both positive as well as negative network plots. It is a huge topic of debate among pro-vaccine as well as anti-vaccine people. Sceptics often question whether wearing masks are required after taking the vaccines or whether they are effective in curbing the spread of the virus in the first place.
7. The most used bigrams in negative sentiment tweets were surrounding possible side effects of vaccines, long term effects and allergic reactions. On the other hand, most tweets with positive sentiments were conversations surrounding healthcare workers and those operating on the front line.
8. Operation warp speed is one of the recurring ngrams. It refers to an initiative by the federal government of the United States to accelerate the manufacturing and distribution of the vaccine. Mostly it was met with positive tweets. However, some sceptics tweeted negatively expressing concerns surrounding the budget and funds allocated to the program.
9. N grams such as 'I would', 'public trust', 'good news', 'need not worry' etc. appear in the positive network diagram. These n grams are mostly used in tweets supporting the vaccines.
10. Tweets regarding approval of vaccines and certain vaccines like pfizer-biontech by governments and related news also dominate the positive sentiment tweets. News surrounding countries approving vaccine usage and rolling out mass vaccination programs were met in positive light by the public.

0	(side, effects)	9049
1	(get, vaccinated)	6343
2	(people, died)	5297
3	(long, term)	5205
4	(still, get)	4513
5	(flu, shot)	4362
6	(fully, vaccinated)	4219
7	(vaccinated, people)	4168
8	(many, people)	3734
9	(people, get)	3681
10	(delta, variant)	3209
11	(let, us)	3033
12	(natural, immunity)	3010
13	(get, sick)	2704

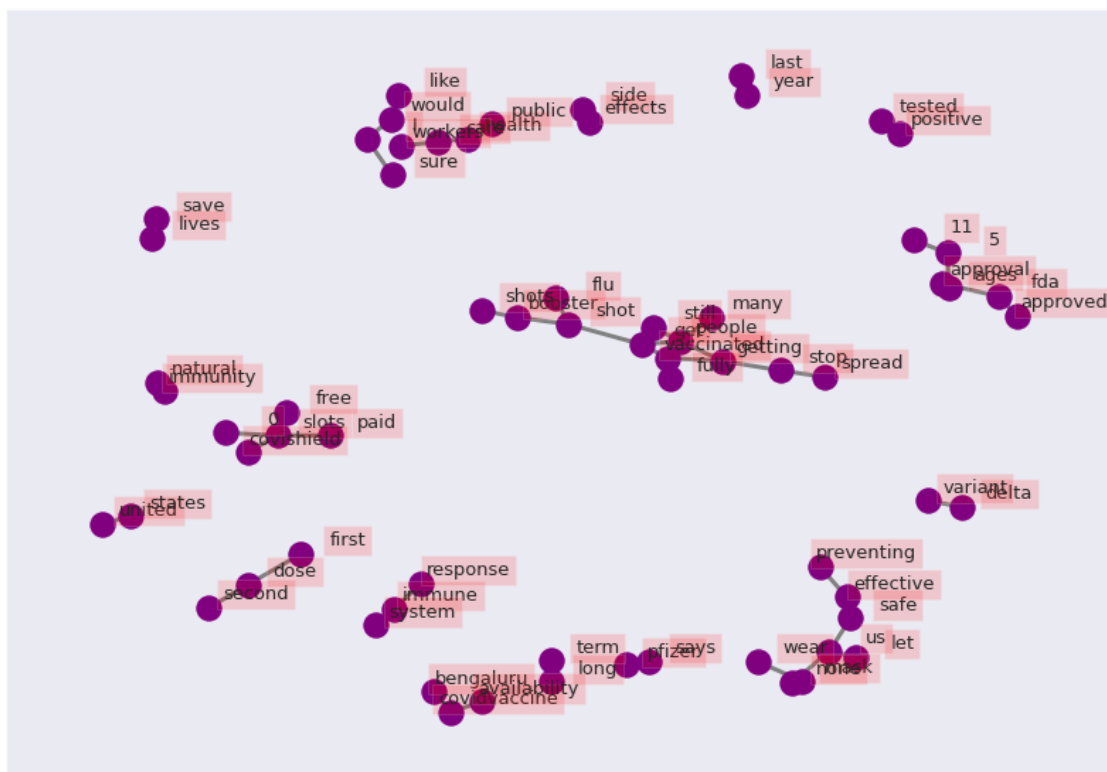
Most recurring Bigrams - Tweets with Negative Sentiments (September 2021)



Network plot showing interconnectivity between recurring bigrams

0	(natural, immunity)	11248
1	(get, vaccinated)	6972
2	(side, effects)	5053
3	(safe, effective)	4605
4	(fully, vaccinated)	3877
5	(immune, system)	3838
6	(let, us)	3440
7	(long, term)	3393
8	(still, get)	3235
9	(health, care)	3226
10	(vaccinated, people)	2979
11	(people, get)	2961
12	(delta, variant)	2734
13	(booster, shots)	2610

Most recurring Bigrams - Tweets with Positive Sentiments (September 2021)



Most recurring Bigrams - Tweets with Positive Sentiments (September 2021)

Bigrams Analysis for tweets post-vaccine rollout reveals how conversations (positive as well as negative) surrounding the vaccines have evolved since the rollout:

1. Once again, majority of the anti-vaccine / vaccine sceptical tweets expressed concerns surrounding long term side effects, allergic reactions and compared the covid-19 virus to the common flu. Key bigrams include '**side effects**', '**adverse reactions**', '**long term**', '**blood clots**', '**get sick**'.
2. Pre-rollout, majority of the sceptics expressed doubts regarding the potential side effects. However, post mass vaccine rollout it was reported in the media concerns surrounding a rare blood clotting condition which develops after taking the vaccine. This bigram was majorly found in anti vaccine tweets and highly negative tweets.
3. Anti-vaccine people talked about vaccinated people still contracting the virus and harped on about the vaccine killing people rather than the disease. Also, wearing masks appears again as a point of debate.
4. Delta Variant, which appeared in the summer of 2021 was a major talking point. Many negative tweets stressed that vaccines were ineffective against these variants and raised questions surrounding the mutations of the virus? Will the vaccine be able to catch up with the mutations or not?
5. Natural Immunity is a new bigram which is also a hotly contested topic of debate online surrounding vaccines. Majority of the vaccine sceptics and tweets against vaccines claim that natural immunity is alone enough to combat the virus. Whereas pro-vaccine people tend to counter this claim. Thus, this bigram was present in tweets supporting both sides of the debate. However, it was mostly used in anti-vaccine tweets.
6. Positive tweets and pro-vaccine tweets once again contained terms surrounding approval and spread of the mass-rollout programs. Key bigrams such as 'let us', 'save lives', 'stop spread', 'effective in preventing', 'fda approval', etc. were used in many pro-vaccination campaigns and tweets on twitter.
7. 'Last year' is one of the bigrams used in many anti-vaccine tweets. Upon further inspection, we observed that it was people comparing death rates and their own experiences in combating covid without being vaccinated.

Alleged 'Covid-19' deaths more than 3,000% higher than this time **last year** and 80% of those dying had the fake vaccine according to the latest Public Health data (the fake vaccine is killing them)
<https://t.co/yykWaRljAy>

@EliseRCook COVID vaccines in the **last year** have killed more people than all other vaccines combined since 1990, and that's just the deaths that are getting reported. <https://t.co/OUyRtLoNxS>

@HeyNurseKat @FourWinns298 Says Kat who has a vaccine in her body for **last years** covid. Keep that same energy around your 8th or 9th shot. Can you still catch covid once vaccinated?? Hmmm.
 **** your science.

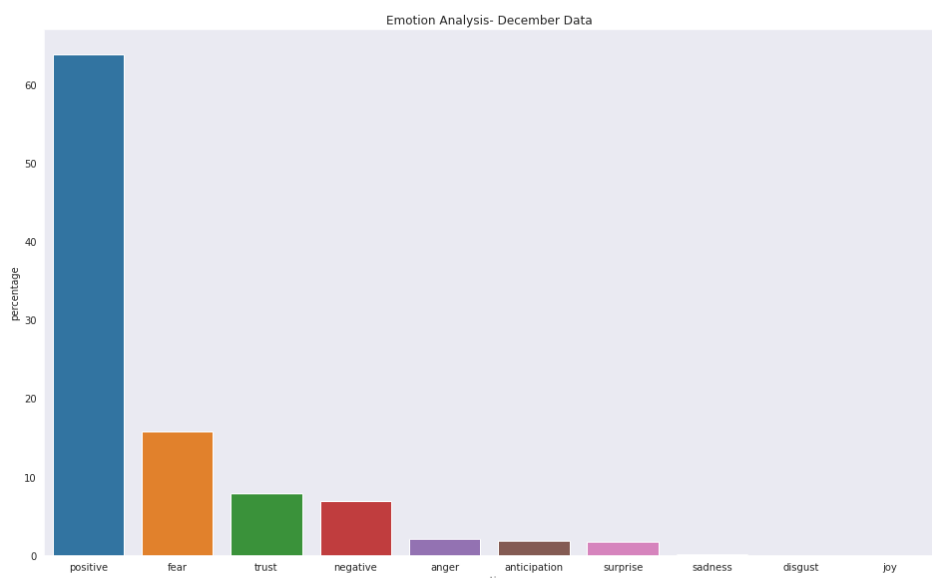
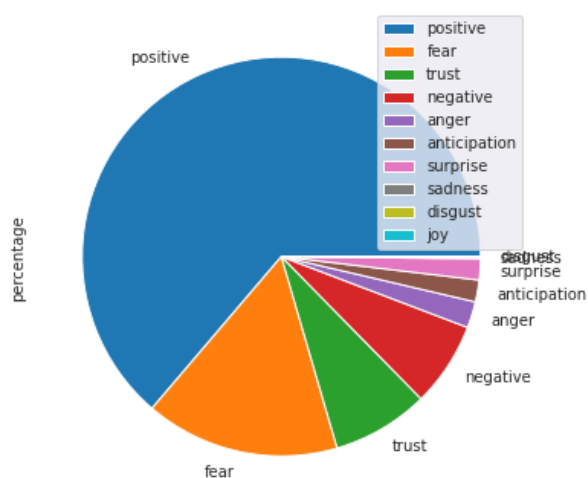
An example of a set of tweets containing these bigrams described above -
This bigram was mostly used in anti-vaccine tweets.

Emotion Analysis on covid vaccine data

We use NRCLex for this part of our analysis. Following are the results of emotion category classification of tweets using NRCLex:

	emotion	counts	percentage
0	positive	689037	63.76
1	fear	170433	15.77
2	trust	84801	7.85
3	negative	73994	6.85
4	anger	23063	2.13
5	anticipation	19137	1.77
6	surprise	18454	1.71
7	sadness	1468	0.14
8	disgust	345	0.03
9	joy	21	0.00

Table depicting results from emotion analysis performed using NRCLex on covid vaccine (December 2020) data

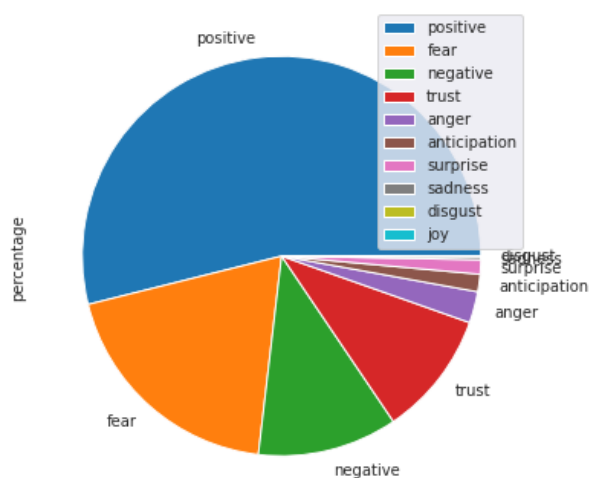


Pie chart depicting percentage number of tweets classified into one of 10 emotion categories

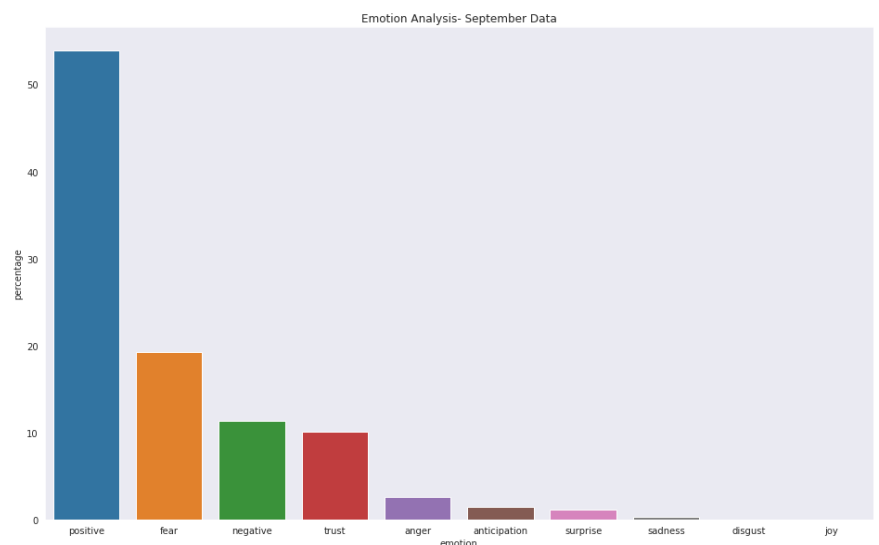
Bar chart depicting percentage number of tweets classified into one of 10 emotion categories

	emotion	counts	percentage
0	positive	496427	53.89
1	fear	177300	19.25
2	negative	104455	11.34
3	trust	93271	10.12
4	anger	23436	2.54
5	anticipation	13003	1.41
6	surprise	10656	1.16
7	sadness	2413	0.26
8	disgust	248	0.03
9	joy	32	0.00

Table depicting results from emotion analysis performed using NRCLEX on covid vaccine (September 2021) data



Pie chart depicting percentage number of tweets classified into one of 10 emotion categories



Bar chart depicting percentage number of tweets classified into one of 10 emotion categories

In sentiment analysis, we categorized tweets in 3 categories but here we observe all well known emotions such as fear, positive, trust, surprise, negative etc. Like sentiment analysis, In this analysis also we can clearly see the positivity towards vaccination has decreased and negativity/fear has increased over time. The positivity towards vaccination decreased approx 10% and fear or negativity increased approx 8% in this time period (pre-vaccine rollout to post-vaccine rollout). But the overall scenario is in the favor of positivity. People are mostly afraid about covid vaccination because of fake things spreading on social platforms and different mutations of covid-19. Major events directly affect People's opinion. Over this time period we observe people's trust on vaccination has increased but anger as well although there is no significant change in anticipation, surprise, sadness and joy for both time periods.

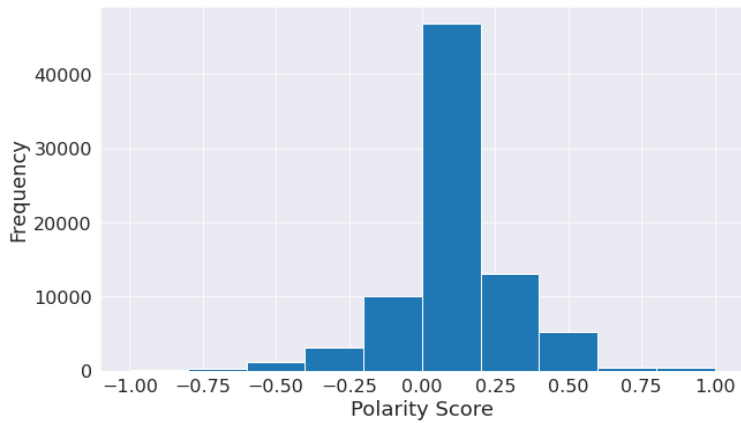
People's opinion changed over this time period (Dec 2020 to Sep 2021):

Emotions	Changed over this time period
Positivity	-9.87 %
Fear	+3.48 %
Negativity	+4.49 %
Trust	+2.27 %
Anger	+0.41 %
Anticipation/Joy/Surprise/Sadness	Not Significant

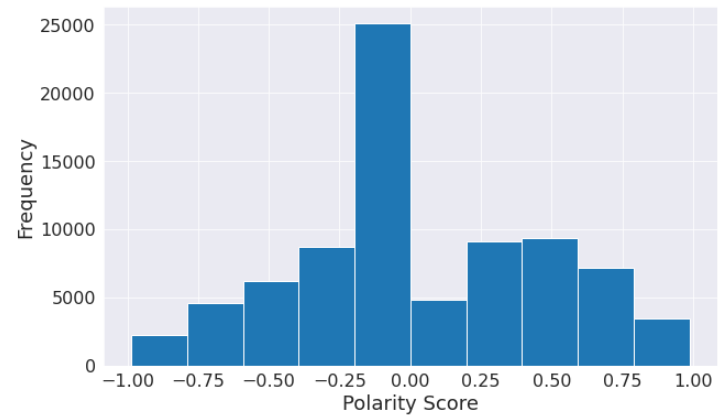
As we know, Major events have the ability to change people's opinion or emotion on any topic. We observe, people generally change their opinion on vaccination after some kind of major event like decisions taken by gov, any anti or pro vaccine news etc.. After the vaccine rollout, so many rumors were spreading against vaccination and we clearly can see its effect on positivity / negativity towards vaccination.

Sentiment analysis on vaccine booster data

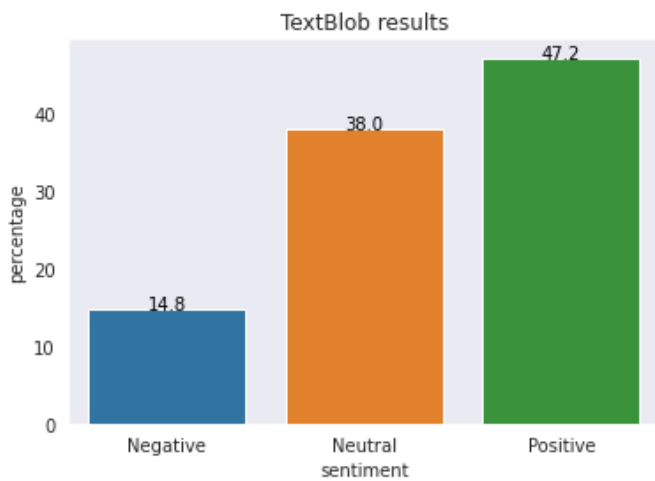
We repeat the analyses performed above on the covid vaccine data and summarize our results.



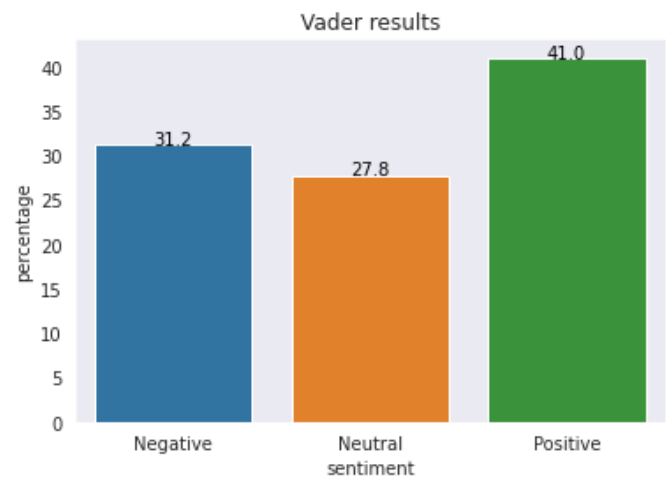
Polarity scores' distribution - Vaccine booster Data (TextBlob)



Polarity scores' distribution - Vaccine booster Data (VADER)



Percentage of Tweets across sentiment categories - Vaccine booster Data (TextBlob)



Percentage of Tweets across sentiment categories - Vaccine booster Data (VADER)

- **Top 10 Most Positive Tweets**

	Text	TBScore	Username
0	NIH director says Israeli Covid booster data is 'impressive,' shows tenfold reduction in infection: The National Institutes of Health director called Israel's data on Covid vaccine booster shots "impressive," saying it shows a tenfold reduction in... https://t.co/555vH1weHO https://t.co/PkrwluGqBo	1.0	stockleaksnews
1	Quite impressive indeed... \n\nProtection of BNT162b2 Vaccine Booster against Covid-19 in Israel NEJM https://t.co/B8RwhoX03z	1.0	DanielAletaha
2	This quote on booster vaccines perfectly encapsulates the maddening pit HK is in with regards to Covid: "If we know when the border will reopen, perhaps we will need to start doing work two months before the reopening" https://t.co/kzF41U2TwZ	1.0	Birdyword
3	@jreddlovette @TeyaSophia I entered, but it was stressful so I just did the best I could!	1.0	FKGaillard
4	@pogokero @RealMattCouch If it's the best why a booster? Why are people who have the vaccine still dying from covid?	1.0	burkett_chase
5	@MichealMartinTD Yet you still won't remove restrictions come 22nd October 2 weeks before this date, you'll announce that the Wu & Lambda variants are a cause of concern and might not work against the vaccine, best to air on the side of caution and wait until the booster is rolled out to everyone	1.0	james99570263
6	Tucker said it best about the vaccine & booster shots https://t.co/z2L2f7wQHH	1.0	PinniesH
7	#VaccineBooster shot may be our best hope for removing #CreepyJoe https://t.co/JORESBKUb0	1.0	Torpor24918710
8	@MartyMakary Hello I wanted to share my data:\n 1 At 15 days after the Janssen vaccination\n 2 to 30 days after the Janssen vaccine\n 3 at 45 days after the Janssen vaccine\n 4 2 months after the Janssen vaccine I got a booster from Pfizer. \n\nbest regards from Costa Rica, Pura Vida! https://t.co/TIZLUT3rtz	1.0	cristianjbcr_
9	Concerned about Delta? Best practices below. \n\nVaccine + n95 mask. Booster when recommended. https://t.co/Hr7HhDwJ5	1.0	mbrockenbrough

Top 10 most Positive tweets containing keyword Vaccine Booster - TextBlob

	Text	VComp	Username
0	Learning more\nTeaching more\nCaring more\nHelping more\nLoving more\n\nGreater benevolence & kindness\nMore understanding & patient calmness\nKindness & extroverted social engagement\nWearing a filtering mask for public safety\nGetting a booster vaccine shot #3 total\nSharing ideas more!	0.9865	Cosmicinglewood
1	@POTUS I GOT BOTH THE PFIZER VACCINES & I FEEL GREAT I'M JUST WAITING FOR MY BOOSTER SHOT THANKS FOR MAKING OUR WORLD A BETTER SAFER PLACE & THANKS FOR HELPING THE PEOPLE WHO ARE STRUGGLING LOVE YOU PRESIDENT BIDEN LOVE YOU TO THE MOON & BACK ❤️ STAY SAFE & GOOD LUCK 🙌 🙌	0.9851	LamontT91982184
2	@TMDTH1 @jodi_geczi @ToddShadley Fantastic read, it literally confirms the study that I shared and confirms that the vaccine works but that it just wares of and the third booster is having great success in folks over 60. The study I shared has multiple parts, I only shared two of them. Thanks for sharing.	0.9825	sharpcutco
3	@VanessMar You will be fine. U need the vaccine, not only to protect yourself but also to protect ur loved ones \nI got mine & I'm fine. I'm ready for the booster soon.\nThanks for not believing the lies. U are helping the world get back to normal. Encourage others to do the same. God bless.	0.9822	WheelinGreg
4	Got my Pfizer booster today. So thankful to have survived the pandemic. So thankful the vaccine exists & is accessible to me. Wish it were the same for everyone. I am blessed, grateful and humbled. #Pfizer #BoosterShots #CovidVaccine #blessed	0.9804	SnapdragonGo
5	@GrahamAllen_1 Hi Graham, I chose to get the vaccine and will prob chose the booster as well but respect your right to choose what whatever you and your doctor think best. Interesting studies on natural immunity coming out of Israel. According to Joe, weekly testing will be option for job Best	0.9779	emmak_ane
6	@MABlumenfeld It's pretty clear the vaccine effectiveness wears down over time. Pfizer may be a self interested party, but it says the booster is helpful and has a benefit. Would love to see the data behind that, but my gut reaction is to trust Pfizer.	0.9765	SteveMatthews12
7	@michaeljknowles Isn't it the other way around? The vaccines are so effective that most people don't need a booster, but for those who *might* benefit from one (older/weaker immune system), it is safe enough that we can be comfortable giving them one. (Safety concerns seem to be mostly ages <25)	0.9761	TwiceFriedEthos
8	Thank you Namrata for your kind words! We're delighted you got your booster and it's great to highlight the hard work each team across all health and care services do to help people like yourself get their vaccine, so thank you for the reminder! 🙌 🙌 https://t.co/b49DRqyXy8	0.9756	FrimleyHC
9	Pfizer seeks FDA approval for vaccine booster dose – Mind Body Love\n https://t.co/dlovzLQUrgn#modeling #model #photography #fashion #photoshoot #photooftheday #instagram #instagood #love #style #portrait #follow #photo #photographer #like #beauty #picooftheday #beautiful #art #like	0.9756	Jlinfo4us

Top 10 most Positive tweets containing keyword Vaccine Booster - Vader

• Top 10 Most Negative Tweets

	Text	TBScore	Username
0	'We Sent a Terrible Message': Scientists Say Biden Jumped the Gun With Vaccine Booster Plan #NewsBreak https://t.co/WM98ZauCv4	-1.0	Evertbeetho
1	#COVID19 \n\nUS President Joe Biden, who received booster shots of the vaccine, says unvaccinated citizens are causing an 'awful lot of damage for the rest of the country' \n\n https://t.co/RCbrwKnQKc	-1.0	scroll_in
2	Between vaccine passports, appting, booster shots, police with masks and helmets and censoring of information I'm getting real sick of it all. Borderline pissed off! \n\nThe people will rise together and those anonymous cops will cry for forgiveness. And we should forgive! https://t.co/2wAZcq4GMh	-1.0	DAOFUQ
3	@cvsparmacy your phone agents are giving out incorrect information about COVID vaccine booster doses. I tried to speak with a supervisor and 3 separate agents cut the call. Horrible customer service	-1.0	anmfis
4	I got whooping cough 9 yrs ago. Had had childhood vax but no booster.\nIt was terrifying at times.\nI have permanent damage to the muscles around my vocal cords \nGet your vaccines. All of them. Boosters included. 🙌 https://t.co/eastDyrwHx	-1.0	whickity01
5	https://t.co/M5aJbclRHR Now they are requiring a COVID vaccine booster every 5 months! Are the pharmaceutical companies, Fauci, FDA & the Biden administration hoping fearmongering & mandates will booster their already outrageous income levels?	-1.0	ZechiniVicki
6	WHO calls for halt to booster jabs amid 'shocking vaccine inequity' \n\n#Covid_19 \n\n https://t.co/GP0D0Xh01f	-1.0	PressTV
7	@POTUS So you can inject us, booster every year till we die from these poison vaccines! You are an evil man!	-1.0	Douglas00498697
8	Hell no vaccine..no booster shot..no masks 🙌\nAfter a year or two goes by, what terrible side effects are going to show up in humans that took the vaccine? Heart problems, birth defects, kidney problems, lung problems. Kids with lung problems from wearing masks 12 hrs. a day 🙌	-1.0	tom70006983
9	@ZubyMusic They would call that a disingenuous assumption. Because with a booster shot and the vaccine, you would protect the people from going insane. So they're calling it for Covid. At this point we need to cure mass hysteria.	-1.0	Stephanospi

Top 10 most Negative tweets containing keyword Vaccine Booster - TextBlob

	Text	VComp	Username
0	@SBaysden They say three booster shoots per year, but really what it means is "as we all knew, there is no ONE vaccine for life for coronavirus (ex: flu), and we're so bad we need 3 shots for each yearly vaccine"\nSo, basically, it's a very bad flu vaccine (more than 1 shot is bad dev)	-0.9853	Max_from_France
1	It will never be enough.\n\nThis is how Control Through Fear works.\n\nFear the Communists for 50 years.\n\nRussia Russia Russia! Every year.\n\nFear the jihadists for 30 years.\n\nWar on terror! War on terror! Every year.\n\nFear the invisible virus.\n\nVaccine! Booster! Flu season! Every year. https://t.co/ltk1kce5am	-0.9834	valithedgod
2	"but the vaccine has really bad side effects" yah no shit all vaccines do I had muscle aches and hot and cold chills when I had to get my meningitis and tdap booster at 18. sorry if you're a little bitch for being scared of a sore arm, a headache, and drowsiness for 24 hours	-0.9829	goldennoyn
3	@Sandowler @LozzaFox But my 'shit' is backed with evidence. The vaccine killed my uncle and left my mum with heart failure. There is no way you will get my kids to go anywhere near it. My mum has been told by her heart specialist that if she gets the booster she will be dead. Plain and simple.	-0.9803	YouDontKnow131
4	@orp2150 @VABVOX Just had it done after work yesterday. My arm is sore, but no other reactions to the booster or the flu shot. My first two covid shots had no reaction. My arm was never sore. I believe my arm is sore now due to the flu shot. It always is sore whenever I get the flu vaccine.	-0.9783	pvalt42
5	It's just my opinion but... It's fuckin' nuts that people are getting the Covid vaccine BOOSTER and some dumb fuck Americans refuse to even get the first dose. They try some of the most ignorant fucking shit in a search of alternative remedies.	-0.9771	runsamuck2
6	And let's unleash a bio-weapon called COVID-19 and various mutations of it to kill off those who manage to survive our two vaccines and endless booster shots, along with sterilization, abortion, suicide and murder, war, terrorism, etc....	-0.9761	KirkCoon9698
7	I am afraid of needles, the last thing I'd ever do is volunteer to get a shot. But I put my fear away, and got the shot, the vaccine, and it was absolutely nothing to fear. It was done in no time, no pain, no ill effects. If required I'll happily get the booster. Worry not. https://t.co/VcKpHhDTdY	-0.9753	JamesHu18332770
8	covid sufferers are being denied booster shots. This is all wrong. Why are we being punished because of some selfish, stupid #antivaxxers? Why are we long covid sufferers who fought with our lives when #vaccines were not yet available,	-0.9738	black_visions
9	@ZarinTaylor @SenRonJohnson 15,000 Americans dead from the vaccine. I'd call that murder. 660,000 seriously injured. I'd call that a failed vaccine. How many boosters you going to take? They are planning at least 5. Every booster weakens your immunity more and more till your dead. Go ahead follow Biden.	-0.9729	dianesu44580009

Top 10 most Negative tweets containing keyword Vaccine Booster - Vader

For the time-period during which our data was collected (September 2021), the vaccine booster was in a pre-rollout phase, just like the pre-vaccine rollout phase in December 2020 . Hence, we expected similarity in results. However, looking at the fact that negativity has risen in tweets surrounding the covid-19 vaccine, we expected the results to have higher negativity scores.

Upon performing the sentiment analysis, we make some key observations:

1. The Sentiment Analysis results for Booster data **exactly (or almost fully) matches** the results for December 2020 pre-vaccine rollout data **across both** the sentiment classifiers:

Sentiment	Classifier	December 2020	Vaccine Booster
Negative	TextBlob	14.9	14.8
Positive	TextBlob	44.2	47.2
Neutral	TextBlob	40.9	38.0

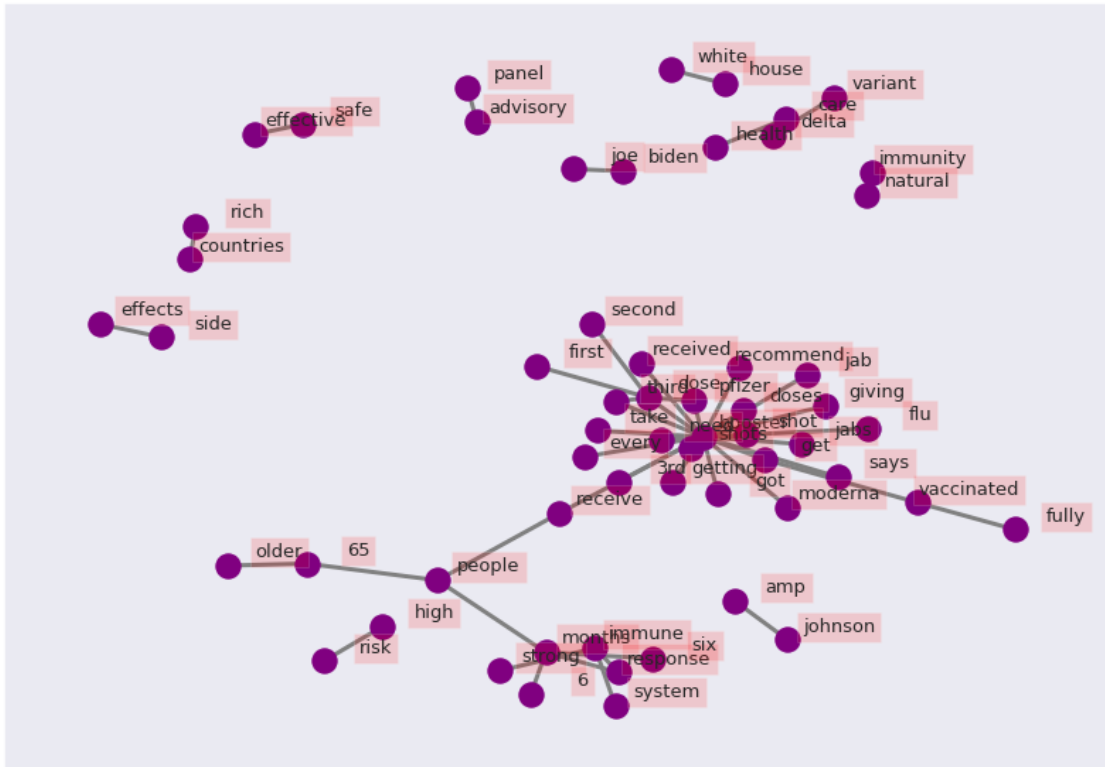
Sentiment	Classifier	December 2020	Vaccine Booster
Negative	VADER	29.0	31.2
Positive	VADER	41.8	41.0
Neutral	VADER	29.3	27.8

Tables comparing sentiment classification results for pre-vaccine rollout and pre-booster rollout.

It is interesting to observe that people's opinions for both the pre-rollouts were similar in sentiments.

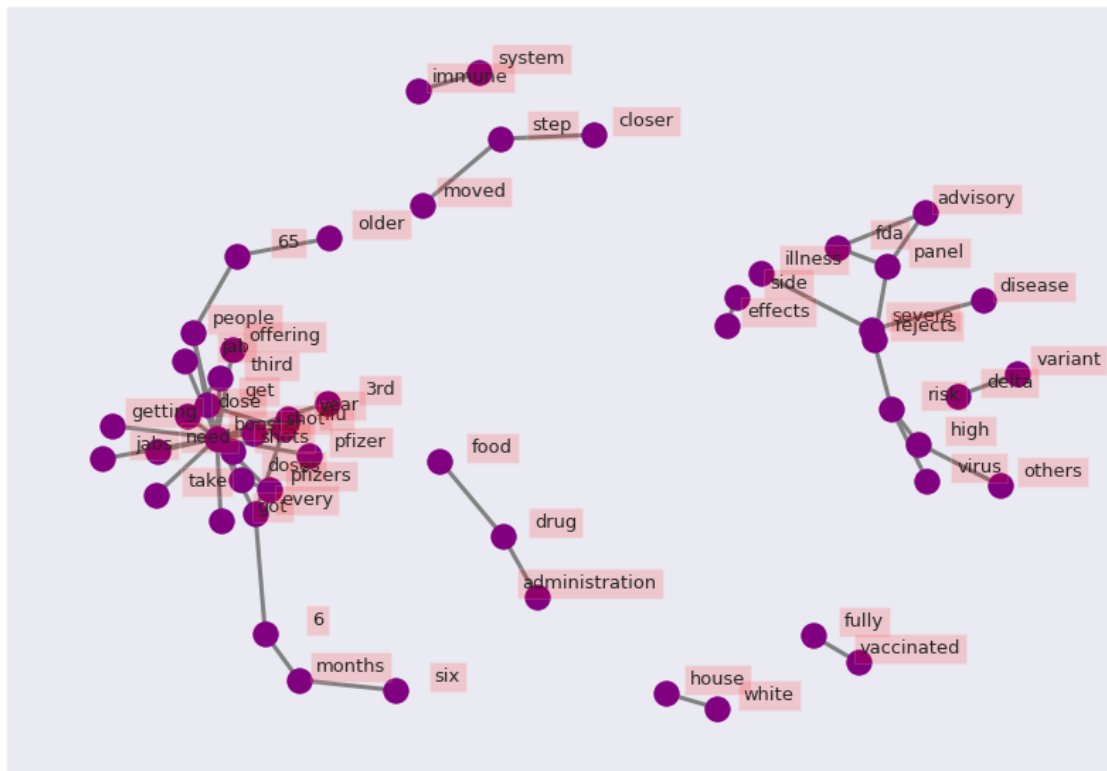
Bigrams Analysis on vaccine booster data

- **Positive Sentiment Bigrams Network:**



Most recurring Bigrams - Tweets with Negative Sentiments (Booster Data)

- **Negative Sentiment Bigrams Network:**



Most recurring Bigrams - Tweets with Positive Sentiments (Booster Data)

Some key observations:

1. The pro and anti-vaccine agendas were highly similar to those as compared to the covid vaccine data. However, one major thing that stood out was that the majority of the anti-vaccine people/sceptics use vaccine booster in their agendas against the covid-19 vaccines. We observed that the N grams such as 'Natural immunity', 'Side effects', etc. were used by many in the context of boosters so as to promote their anti-vaccine agendas. An example is the usage of natural immunity against vaccines in the context of boosters:

natural immunity wins 700 over vaccines already this data is from the worlds largest live lab test Israel. they have already had a 3rd booster and they are still getting sick eat crow bro the vaxx and boosters are are failing you will get covid anyways

natural immunity lasts significantly longer than the vaccine which is already needing a booster

we already know natural immunity lasts longer then the vaccine by far you would not need booste shots if the vaccine lasted longer natural immunity can last years you can the lord for that

you need to read and listen on lbc today they were discussing the fact that natural immunity lasts longer are they not reporting now the vaccine needs a booster after 6 months

face_with_tears_of_joy face_with_tears_of_joy

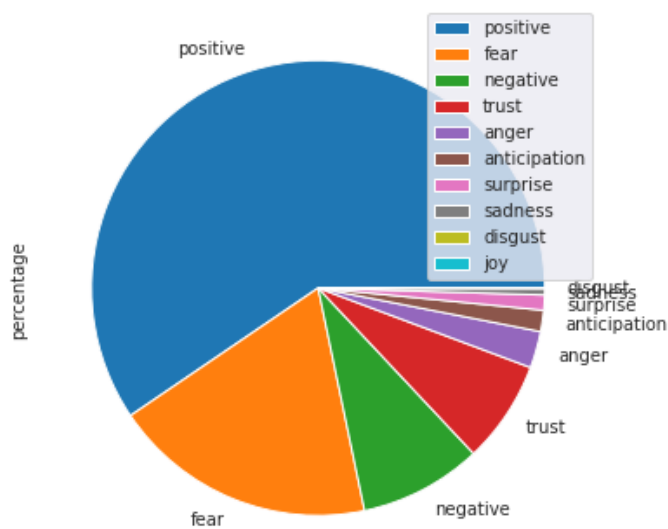
Set of Anti-Vaccine Tweets using the bigram 'natural immunity'

2. Many anti-vaccine tweets also criticised the need for a booster dose after being fully vaccinated. Bigrams such as 'every year' were used in many such tweets.
3. Majority of the positive tweets were, just like in the case of pre-vaccine rollout tweets, in anticipation of the vaccine booster. Many in favour of vaccine-tweets also promoted that boosters were necessary to control spread of variants. Key bigrams include 'get the booster', 'need booster', 'take booster', 'high risk', etc.

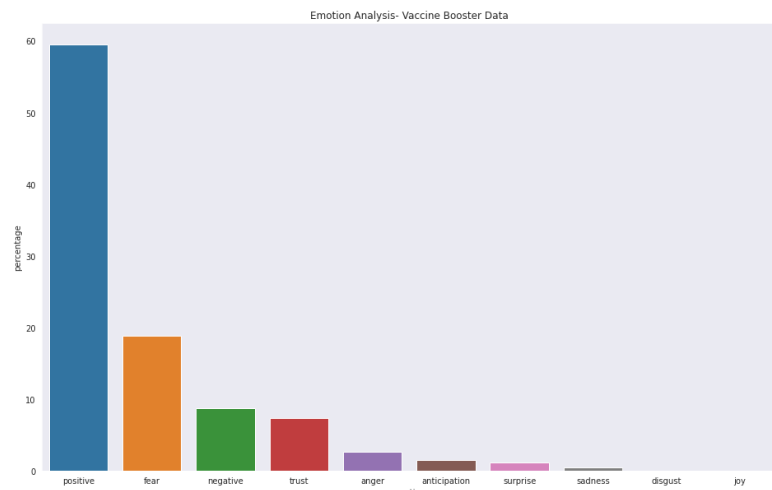
Emotions Analysis on vaccine booster data

	emotion	counts	percentage
0	positive	48031	59.47
1	fear	15194	18.81
2	negative	7055	8.74
3	trust	5888	7.29
4	anger	2105	2.61
5	anticipation	1220	1.51
6	surprise	867	1.07
7	sadness	386	0.48
8	disgust	11	0.01
9	joy	2	0.00

Table depicting results from emotion analysis performed using NRClex on vaccine booster data



Pie chart depicting percentage number of tweets classified into one of 10 emotion categories



Bar chart depicting percentage number of tweets classified into one of 10 emotion categories

Through emotion analysis on Booster data, we again find similar results like sentiment analysis. In Sentiment analysis, we observe people's opinion on booster dose is the same as people's opinion on covid vaccine in pre vaccine rollout phase. Like sentiment analysis on booster, emotion analysis shows the same thing. For example, negativity was approx 14% according to sentiment analysis and in this analysis we can see that negativity + anger is approx 12%. Here, we observe, the positivity towards booster dominate the chart followed by fear. According to this analysis, fear is approx 19% and we know the reason behind it is because people are scared of different mutations of covid-19 and they have to take it even after fully vaccinating. People's trust/anger on booster dose is the same as in the pre vaccine rollout phase on covid vaccine but the positivity decreased by approx 4.5% and fear+negativity+anger increased by approx 5.5% than pre-vaccine rollout phase.

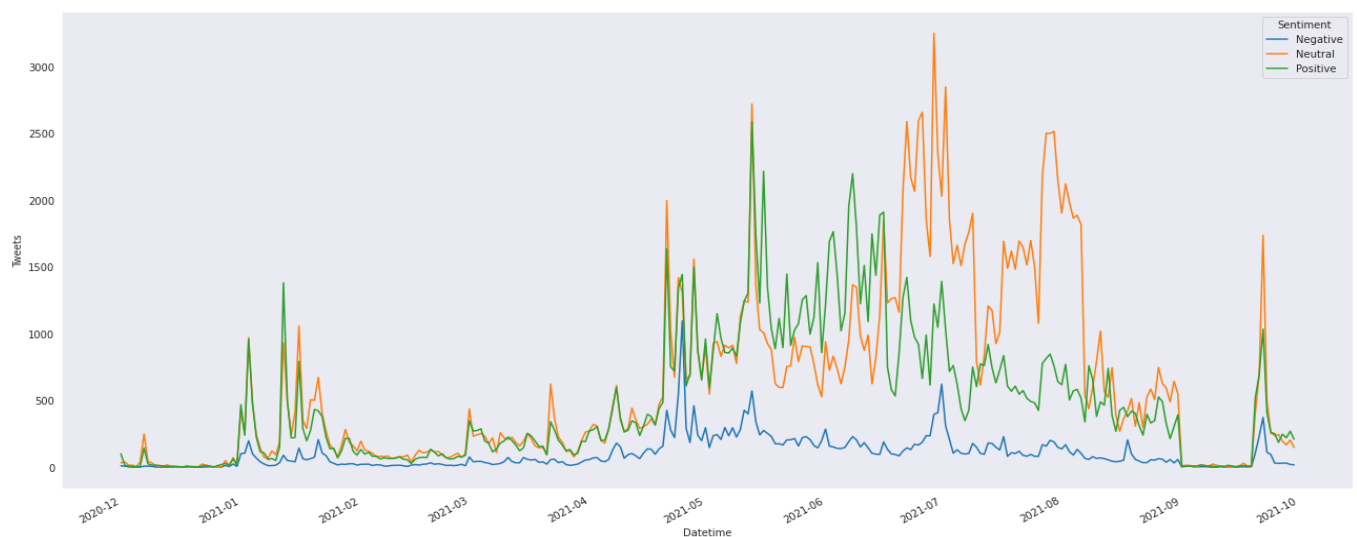
Comparison between Booster emotion analysis with Pre-Vaccine rollout phase (Dec 2020) and Post-vaccine rollout phase (Sep 2021) :

Emotions	Comparison with Pre-vaccine rollout	Comparison with Post-vaccine rollout
Positivity	-4.29 %	+5.58 %
Fear	+3.04 %	-0.44 %

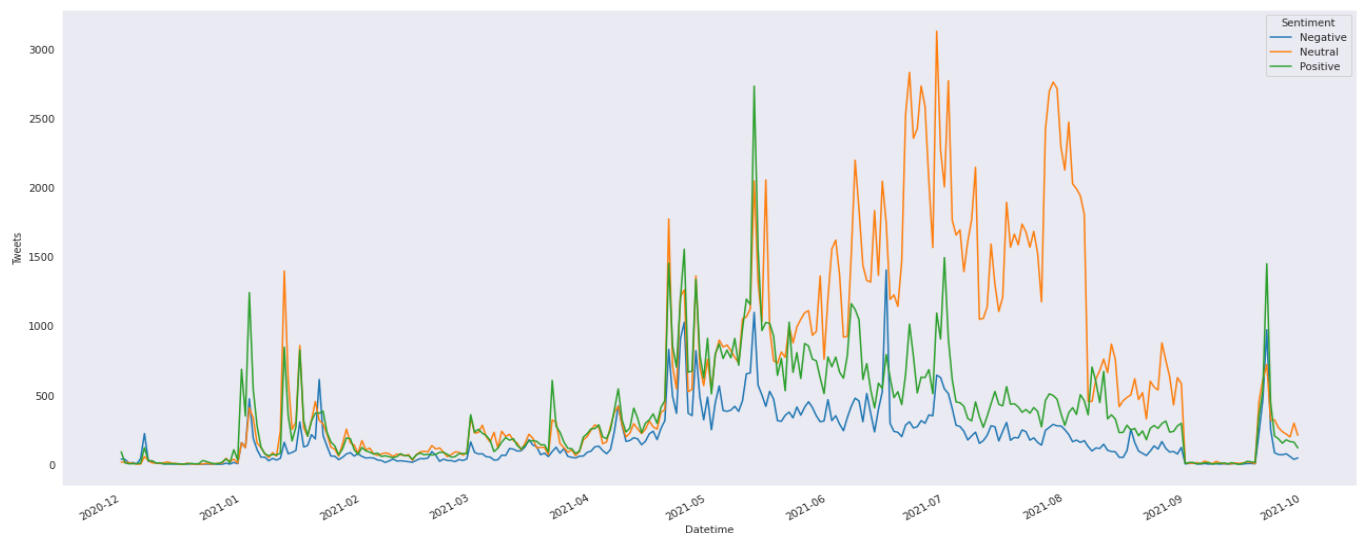
Negativity	+1.89 %	-2.61 %
Trust	-0.56 %	-2.83 %
Anger	-0.48 %	-0.07 %
Anticipation/Joy/Surprise/Sadness	Not Significant	Not Significant

Analysis on the second dataset (Covishield Data):

As described earlier, the second dataset consists of tweets about the Indian vaccine COVISHIELD. For this keyword, we have a set of tweets for the continuous time period of December 2020 - September 2021. Hence, we are able to better visualize the trendlines and make conclusions. We analysed the sentiments for the entirety of the time period of December 2020 to September 2021 and hence the overall metrics (Sentiment Category percentages, Bigrams, Emotion Analysis percentages, Word Clouds, etc.) do not matter. We are interested in observing the evolution of sentiments through time. Hence, we plotted Time-Series Plots using Sentiments classified using TextBlob and VADER.



Evolving public sentiments for the tweets containing the keyword 'covishield'
- Classified using TextBlob



Evolving public sentiments for the tweets containing the keyword 'covishield'
- Classified using VADER

We observe that the majority of the public conversations surrounding Covishield vaccines were positive in nature. The spikes in the time-series plot correspond to major news or announcements regarding Covishield. Some key observations:

1. The earliest spikes point out to Mid-January. On 16th of January 2021, the Indian Prime Minister Narendra Modi had officially launched a mass vaccination drive. Hence, there was a lot of positive anticipation among the people and the majority of the conversations were positive.
2. We can observe that the negativity (Represented by the blue line) peaks around May. It was the deadliest period of Covid outbreak in India and the time period corresponded to the peak of the second wave of the outbreak in India. Hence, there was a lot of negativity and fear among the people. The tweets reflected that.
3. The peak second wave of the Covid-19 outbreak in India lasted for approximately a month till the end of May. Tweets surrounding the Covishield vaccine went up starting with the second wave. A possible reason can be that people started to take vaccination more seriously due to the second wave.
4. The amount of neutral tweets (Orange spikes) rose considerably after May. Majority of the neutral tweets are factual based tweets and most of them were regarding news about availability of vaccines. Beginning at the end of June (Starting June 21)^[28], the Indian government accelerated its mass vaccination program and thus, the severe increase in the amount of neutral tweets reflects that.

Next, we observe the top 10 most positive, negative and neutral tweets to understand the language used in tweets which have been assigned the highest / most extreme polarity scores by our sentiment classifiers.

• Positive Tweets

	Text	TBScore	Username
0	1st dose of covishield vaccine done. Excellent arrangements at Delhi Govt. School of Excellence, MPK. No queue, just a 20 mins observation time after taking the jab.\nDo take the vaccine as soon as you get the chance! @msisodia @SatyendarJain @AamAadmiParty https://t.co/TmO2SdnEBz	1.0	midnight__muse
1	@AnchorAnandN Which vaccine is best? \n\nCovaxin Or covishield	1.0	Trueliberal9
2	I got vaccinated! With Covishield! Let's hope for the best now.	1.0	dasguptanandini
3	@dude_basu @sukhpreethind @vikramchandra @poornibv Covishield . Astra zeneca is a proven vaccine . It is not gaumutra . We in India can't get the best because our PM slept and didn't care . It is not the best for all ages . We have no choice get the vaccine .	1.0	KishanThar
4	@MadhuriDixit You are the greatest actress in 1995 to 2000 year. I have vaccinated with covishield. Thank q.	1.0	NRajakumar20
5	Which is best #Covaxin or #Covishield ?? \n\nPlease respond !!!	1.0	thota_J
6	Best vaccines in India....??? \n\n#COVISHIELD , #COVAXIN or #SPUTNIKV	1.0	hrpatel29
7	@chetan_bhagat Can you compare what is best between Covaccine , covishield vs Pfizer Moderna	1.0	Nituveer
8	Excellent thread on a raging and pertinent topic thru d lens of #BioEthics #VaccineRollout #covaxin #Covishield https://t.co/ai7TP3aHvV	1.0	prodeeptoc
9	@chetan_bhagat What's the definition of the Best in this context compared to Covishield and Covaxin	1.0	ganeshram_gg

Top 10 most Positive tweets containing keyword Covishield - TextBlob

	Text	VComp	Username
0	#HappyNewYear2021 #Master #aespa #success #covishield #AUSvIND #BiggBoss14 #Welcome2021 #KrackTrailer #Bye2020 God is source of peace perfection preserver of safety exalted in might supreme glory to God; creator, evolver. Whatever in Heavens on Earth declares His praise n glory	0.9870	KBegum9
1	Better efficacy than Oxford/covishield's 62%. Best in the inactivated type vaccine. Works against uk variant. Works on children. Huge export potential now! Will boost confidence in takers. Win win situation for all. Hope acceptance increases! #Covaxin https://t.co/sG7akQDFiK	0.9816	beingabhinav07
2	Anna nijamgane buzy #prabhas anna max I wanna try to participate in dis trend most of the time but I couldn't 🙄🙄🙄\n3hours back took my first shot of covishield dose ... always a proud fan of you long live prabhas🙄🙄🙄\nLuv u luv u luv u always anna \n#PrabhasBdayFestin100Days https://t.co/bvlf5BvfvJ	0.9811	haribabutaurus
3	@JawanSnow Covaxin ka aaya nahi hai final result but phase 1 and 2 are promising thus far, for safety and mutations like UK strain. Covishield hai tested and approved as effective. Covaxin if approved will be a blessing because its cost would be 500INR for two doses in open market.	0.9800	dylanhead94
4	@drthomasisaac Dear Beloved Thomas sir. Hope you are safe and sound , may I kindly seek your help in availing second dose of covishield vaccine in Kannur area , if you may help sir. Please - My sincere Gratitude!	0.9790	krishna12264793
5	Got VACCINATED 1st DOSE of COVISHIELD at Khallikote clg ground.\nFaced some technical glitches but it was solved by the help of kind volunteers. Thanks TEAM GANJAM for such a dedicated service and smart procedure like Drive-in Vaccination. Jai Hind \n @BrahmapurCorp @Ganjam_Admin https://t.co/WHo67C8U50	0.9787	FrancisReagon
6	@AmbSibiGeorge Good to know that you are safe and recovering fast. Take good care. Can you also please help thousands of Indians stuck in India and waiting for MOH approval on Covishield vaccine....we need your sincere help support and action on this please. Namaste 🙏	0.9786	lamSubikash
7	Wow! the greatest new year gift for Indians , expert committee today recommends approval of Astra Zeneca Oxford Covishield Vaccine manufactured by Serum Inst of India after being satisfied about its safety and efficacy . 2021 truly starts off as the year of hope for humankind!	0.9783	_Heartmatters
8	Wow ! the greatest new year gift for Indians , expert committee today recommends approval of Astra Zeneca Oxford Covishield Vaccine manufactured by Serum Inst of India after being satisfied about its safety and efficacy . 2021 truly starts off as the year of hope for humankind !	0.9783	Dr_AshokSeth
9	@neerajmital @UnusualMonk @ShashankSArora Sir, truth is its easy to complain. Single does of covishield is also effective but we just want to hear the term fully vaccinated.\nCan we do better? of course yes.\nBut at the moment I cannot see any other leader who could do that.\nWd be happy to vote to any other capable leader.	0.9782	lamShreyas_K

Top 10 most Positive tweets containing keyword Covishield - VADER

• Negative Tweets

	Text	TBScore	Username
0	#EU_stopColonial With or w/o UK makes no sense, colonial mindset.\nEurope for 100s year ruthlessly ruled n ruined world not yet change mindset ruler Vs slaves.\nAstra Zeneca accepted why not Covishield? EU passengers for days quarantine in Indian conditions? https://t.co/UwYsh1WqTV	-1.0	KatratL
1	The price of the Covishield vaccine:\n\n1) India: ₹ 600\n2) Saudi Arabia: ₹ 395\n3) South Africa: ₹ 395\n4) US: ₹ 300\n5) Bangladesh: ₹ 300\n6) Brazil: ₹ 237\n7) UK: ₹ 226\n\nIndian GOV is worst to handle covid 19 https://t.co/412lwJkGst	-1.0	husenmulani7
2	@nuts2406 @sibinmohan He took Pfizer I bet. At worst, Covishield. My bet is on Pfizer though .	-1.0	FreedomSutra
3	@MISSXBUTTERFLYX Already happening.\nGo here, scroll to bottom and click Search Database.\nType in Covishield and go to ADR'S. Terrifying!\nhttps://t.co/eZvvLYC3MG	-1.0	SooziScoones
4	@MajorPoonia Sir sirf VIP and Politicians logon ko hi Covaxin mil raha hai...aam janta ko sirf Covishield hi diya ja raha hai....I lost my Maa inspite of she being vaccinated with Covishield. It's disgusting what's happening everywhere this differential treatment is happening	-1.0	shubha_1982
5	📺 1. Mumbai airport worst affected by Covid, lost 81% flyers in 10 months \n📺 2. Interval between doses of Covishield extended to 4-8 weeks \n📺 3. Covid-19 vaccine: Virchow Biotech to make 200 m\n#TodayPharmanews\n#pharmarelatednews\n#bestpharmacynews\n#latestpharmacynews https://t.co/37uzbue0vg	-1.0	aiims_pharmacy
6	Terrible movement\nFire breaks out in serum institute of Pune which is the world's largest #COVID19 vaccine producer....\n\n#fireworks\n#Covishield\n#COVID19Vaccination \n#SerumInstituteofIndia https://t.co/MxYzO3UPJ4	-1.0	NilamPa57454782
7	This is not the queue for getting their 1st dose of Covishield or Covaxin.. \n\nProviding Wealth to TN minister and CM.. by getting Alcohol... \n\nPathetic situation in TN \n\n@mktalin @PMOIndia https://t.co/e6Qq7JsFD9	-1.0	haritcn
8	ವ್ಯಾಕ್ಸಿನ್ ವಾರ್: Tv9 Reveals Shocking Details Of Covaxin Vs Covishield Efficacy (Part-4)\n\nVideo Link ► https://t.co/0HdLisr9LY\n\n#TV9Kannada #ವ್ಯಾಕ್ಸಿನ್ವಾರ್ #VaccineWar #India #CovidVaccine #VaccineDistribution #CentralGovernment #KannadaNews https://t.co/t17u0WWvhZV	-1.0	tv9kannada
9	ವ್ಯಾಕ್ಸಿನ್ ವಾರ್: Tv9 Reveals Shocking Details Of Covaxin Vs Covishield Efficacy (Part-3)\n\nVideo Link ► https://t.co/gNEZLPtWWhf\n\n#TV9Kannada #ವ್ಯಾಕ್ಸಿನ್ವಾರ್ #VaccineWar #India #CovidVaccine #VaccineDistribution #CentralGovernment #KannadaNews https://t.co/ejBkottu87M	-1.0	tv9kannada

Top 10 most Negative tweets containing keyword Covishield - TextBlob

	Text	VComp	Username
0	@SaffronTommy Mask scam. HCQ scam. SII commissions scam. BB bribery scam. Covishield adulteration scam. Covaxin theft scam. Gandhi family out of turn vaccination scam. Rahul Gandhi PA arrested with ten kilos of cocaine.\n\nIndia defaults on sovereign debt. Moody and Fitch downgrade India credit.	-0.9854	ksub
1	0.61 per million Covishield doses reported clotting, bleeding: Health ministry Hindustan Times '\n\nBut still brain dead sheeple will queue up to take this poisonous death shot! 🐏👤🔪 I weep at the plight of humanity! 🥲🇮🇳 https://t.co/4e5fECRQej	-0.9786	absolutaesthete
2	4 Death in my family due to #COVID19 with in 1month all after taking #Covishield vaccination.\nI am not sure but this vaccine have something wrong\nAll 4death confirmed negative report but symptoms for death is same\n\n\nI know 5 more people died after vaccination. Something Wong	-0.9771	hipertexter
3	@HappymonJacob Don't know about Covishield but the EU's institutions do foster a deadly combination of old timey imperial racism and new timey ethno-nationalist racism. Much less discussed than racism in America where there is at least a tradition of documenting and challenging racism.	-0.9769	SRoyChowdhury01
4	@Swamy39 Nonsense Information. Myself and Relatives-Friends Taken Covishield at Kolkata from Different Hospitals but No Side and Bad effect, No Bad News in Kolkata. Mr.Swamy should Reconsider the Statistics,Facts and Propaganda about Vaccination and Death Reports as MP	-0.9749	AnjanBa92533398
5	@htTweets Why only COVAXIN??? '\n\nWhy not COVISHIELD or SPUTNIK?? '\n\nHe failed to control covid spread, failed in administration, failed to provide basic needs to poor and needy. '\n\nCommunally crying on every one. '\n\nMost sickest, stinking communally biased incapable CM https://t.co/ABAGWMrZct	-0.9744	SAYNOTOYSRCP
6	@timesofindia Why putting pressure on vaccination, to stop death , Government must ban vaccination and close killer covid helpline until investigation is not done. Serum institute is not accepting request to conduct examination of death due to adverse effects of covishield vaccine	-0.9744	MehraDileep
7	@amabirdman Side effects of #Covishield is such an hellish experience\nMine started with Fever Exactly after 8 hours of the injection , followed by terrible head ache and body ache\nHead ache is an understatement , and I literally could not get up as far as body pain went\nPrick region pain too	-0.9723	masala_dosai
8	@shooky0t7 I took covishield.\nBut damn my left arm hurts.. and my body is painin.. two different pain... I don even know what I should do...\nMy bro as no side effects whatsoever..he s snoring away to glory..not that he should 🤔 but damn this is just 🤔🤔🤔	-0.9721	HopeandSunshin3
9	@kanimozhi My uncle got infected in Hyderabad and died even after both doses of covishield. The vaccine offers some protection. But in case you are very old, sick (diabetic, overweight) you could still fall very ill and in some cases die.	-0.9720	bvenkateshwaran

Top 10 most Negative tweets containing keyword Covishield - VADER

● Neutral Tweets

	Text	Sentiment	Username
0	"Such lack of transparency is neither conducive for clinical research nor for public trust in the vaccine," says @AnantBhan regarding the alleged adverse event in the #Covishield trial. @SaakhiChadha\nhttps://t.co/u4Jl71NHg1	Neutral	QuintFit
1	Govt's perspective on the adverse event in @SerumInstIndia vaccine #covishield '\n\nAN ADVERSE EVENT SEEN IN CLINICAL TRIAL OF SERUM INSTITUTE VACCINE WILL NOT AFFECT THE TIMELINE OF THE #VACCINE PRODUCTION: Rajesh Bhushan, health secretary @MoHFW_INDIA	Neutral	Milan_reports
2	After meeting with Prime Minister @narendramodi on November 28, #SII CEO @adarpoonawalla revealed that the company is in the process of applying for an emergency use license for #Covishield in the next two weeks.	Neutral	BiIndia
3	@HLN_BE This is just the beginning. Coronavirus 'Covishield' vaccine volunteer sues Serum Institute of India, Oxford Group over 'adverse reaction' - The Hindu https://t.co/Tzp1tnt1sw	Neutral	patriot_zeo
4	Health Ministry reacts to claims against Covishield: 'Won't affect vaccine timeline' https://t.co/dx8PKpC831	Neutral	republic
5	The #Covishield vaccine being manufactured by #SII is among the leading candidates for a #COVID19Vaccine around the world, locked in a race against the likes of #Pfizer, #Moderna and #SputnikV	Neutral	BiIndia
6	Dr Shahid Jameel, Senior Virologist & Director at @AshokaUniv: With the current manpower & resources, it'll take the Govt 27 yrs to vaccinate all Indians. @ShereenBhan #Serum #Covid19 #CoronavirusVaccine #Covishield #Vaccine https://t.co/ryox93cDFz	Neutral	CNBCTV18News
7	Dr @cspramesh, Director at @TataMemorial to @ShereenBhan: Have seen unprecedented level of collaboration & co-operation between countries cutting across borders. #Serum #Covid19 #CoronavirusVaccine #Covishield #Vaccine https://t.co/0BTTrQlJTq	Neutral	CNBCTV18News
8	Dubbed #covishield , the vaccine being developed by UK's #OxfordUniversity and the US pharma giant #AstraZeneca recently came under scrutiny after a 40-year old volunteer from Chennai claimed to suffer from neurological and psychological side effects as a result of the trials.	Neutral	BiIndia
9	Covishield has no side effects: Serum Institute of India https://t.co/QoLNknELiN #covishield #nosideeffects #seruminstituteofindia #covidvaccine	Neutral	newstrack_eng

Top 10 most Neutral tweets containing keyword Covishield - TextBlob

	Text	Sentiment	Username
0	Serum Institute denies side effects claim, says Covishield 'safe and immunogenic https://t.co/NOcYw0qhsg	Neutral	jimiless
1	Covishield 'Safe And Immunogenic, Serum Institute Denies All Side Effects @SerumInstIndia '\n\n#covid19 #covid19vaccine #Covishield #seruminstituteofindia #Coronavirusupdates #vaccineupdates '\n\nhttps://t.co/tSEGM0O9oi	Neutral	HealthwireMedia
2	BIO-BREAKING: Addressing concerns related to the serious adverse event reported by a volunteer in Chennai, @SerumInstIndia has called the 'Covishield' vaccine safe & immunogenic. Meanwhile, #DCGI is investigating the incident.\n\nMore: https://t.co/YzyFMuhmm via @BioVoiceNews	Neutral	BiovoiceNews
3	Health Ministry reacts to claims against Covishield: 'Won't affect vaccine timeline' https://t.co/dx8PKpC831	Neutral	republic
4	Dr Shahid Jameel, Senior Virologist & Director at @AshokaUniv: With the current manpower & resources, it'll take the Govt 27 yrs to vaccinate all Indians. @ShereenBhan #Serum #Covid19 #CoronavirusVaccine #Covishield #Vaccine https://t.co/ryox93cDFz	Neutral	CNBCTV18News
5	SII is developing Covid-19 vaccine - Covishield in partnership with UK's Oxford University and US pharma giant AstraZeneca.\n\nhttps://t.co/EbbP9uBrBo\n\n@SerumInstIndia @adarpoonawalla @UniofOxford '\n\n#CovidVaccine #COVID19 #astrazeneca	Neutral	TrueScoopNews
6	Allegations Against 'Covishield' Vaccine Yet Human Trials Continue. Why? https://t.co/GrMjC8uiKC https://t.co/7FH8lyXIdl	Neutral	GulistanNewsTV
7	Pune-based #SerumInstituteofIndia could make the #CovidVaccine by #AstraZeneca-Oxford University available in India's market after March-April next year.\n\n#covishield #CoronaVaccine #COVID19\n\nRead more: https://t.co/pPRvY4Vhos https://t.co/T4hvnAVs0X	Neutral	NewIndianXpress
8	Initial causal assessment of findings related to the Chennai clinical trial participant did not necessitate stoppage of Covishield vaccine trials, says ICMR DGN\n\n(by @viswanath_pilla) #Covishield #Vaccine #COVID19 '\n\nhttps://t.co/Fq5ISYH2bE	Neutral	moneycontrolcom
9	Serum Institute denies side effects claim, says Covishield safe and immunogenic\n\nClick on the link to Know More 📌 https://t.co/qzN2sDWVhe https://t.co/YbGuA9l7jj	Neutral	vohglobal

Top 10 most Neutral tweets containing keyword Covishield - VADER

Key observations:

1. Majority of the positive tweets were in support of vaccines.

2. Many negative tweets were political in nature. Also, like covid vaccine and booster tweets, many negative tweets containing keyword 'covishield' also highlighted potential side effects of the vaccine.
3. We established earlier that the majority of the neutral tweets containing the keyword 'covishield' were factual in nature or were surrounding the availability of the vaccines. This can be proven even further by observing the accounts posting the neutral tweets. We observe that the majority of the accounts posting neutral tweets were belonging to prominent Indian news/media organisations.

Conclusions

This is a study of COVID-19 vaccinations in the aftermath of the 2020 COVID-19 outbreak with quantitative analysis. We analysed public opinions surrounding the COVID-19 vaccines using various methods after collecting data (Approximately 2.5 million Tweets) from the popular social networking and microblogging website, Twitter. We separated the analyses into 2 different datasets based on the characteristics of data collected: Global Data and Covishield Data. Here are some key inferences from our analyses:

1. We found out that the majority of the tweets surrounding covid vaccines are **positive** in nature (Of positive sentiment).
2. The **negativity** in online conversations surrounding the Covid-19 vaccine has **risen** since the beginning of the mass vaccine rollout by various governments across the world. This was first shown by TextBlob and validated using VADER. This was also depicted by the results of the Bigrams Analysis and the NRCLEX-Emotion Analysis.
3. Majority of the tweets with **positive** sentiments were **in favour** of vaccines. **Pro vaccine** tweets generally received **higher engagements** (likes, retweets) as compared to anti-vaccine tweets. However, tweets with negative sentiment need not necessarily be anti-vaccine.
4. Many positive tweets earlier (In December 2020) were in anticipation of the vaccine rollout. In September 2021, many positive tweets were praising the effectiveness of the vaccine in combating COVID.
5. Majority of the **neutral tweets** related to the vaccines were **factual** in nature.
6. **Political tweets** surrounding the vaccines are likely to be of **negative** sentiment.
7. Using Text Visualisations, Bigrams Analysis and Observations we separated some of the commonly used phrases / words and analysed the anti-vaccine tweets in depth.
8. Many vaccine sceptics are likely to use phrases / n grams such as '**side effects**', '**long term**', '**adverse reactions**', '**allergic reactions**' in their anti-vaccine tweets.
9. As a large number of people get vaccinated and reports of side effects are released, negativity surrounding vaccinations increase. '**Blood clots**' and '**Allergic Reactions**' were found to be present in a lot of Anti-Vaccine tweets. Some expressed legit concerns whereas some were straight promoting an agenda against vaccines.
10. Conspiracy theorists and Anti-Vaxxers with agendas push for scientifically unproven theories such as Vaccines being **bioweapons**,

Natural Immunity being better than vaccines to combat the virus, Covid being a type of **Flu**, etc.

11. Thus, we understand that in order to encourage more people to take vaccines, **awareness** and **education** must be spread surrounding topics such as the potential **side effects** of the vaccines, its **long term effects**, its effectiveness against mutating variants of the virus, etc.
12. When many countries were under strict lockdowns, many conversations online revolved around the safety of **healthcare workers** and **frontline workers**. Many positive tweets were in appraisal of these said workers and negative tweets were concerns surrounding their safety, etc. However, as time passes by and normality resumes, healthcare workers and scientists are no longer being talked about much.
13. Analyses of tweets in anticipation of the vaccine and in anticipation of the vaccine boosters showed **highly similar** results.
14. Vaccine booster has attracted **negativity** from people who use it to promote their agenda that vaccines in itself are insufficient to combat the virus.
15. Through emotion analysis on covid vaccine for pre-vaccine rollout phase and post-vaccine rollout phase, we found out that the positive emotions (As assessed by NRCLex) towards vaccination decreased approximately **10%** and fear, negativity and anger increased approx **8.5%** over this time period. But the overall scenario is in the favor of positivity.
16. People's trust/anger on booster dose is the same as in the pre vaccine rollout phase on covid vaccine but the positivity decreased by approximately **4.5%** and fear, negativity and anger increased by approximately **5.5%** than pre-vaccine rollout phase.
17. Using **time-series plots**, we mapped out the **evolving** sentiments surrounding the Indian vaccine, COVISHIELD. We observed that the peaks in the graph corresponded to major events and news regarding the COVISHIELD vaccine. Negativity peaked during the peak of the second outbreak (May 2021). Since the acceleration of mass vaccination drive (June 2021), the amount of neutral (fact based) tweets have risen.
18. Majority of the **neutral tweets** with the keyword COVISHIELD were found to be made by key Indian **news/media** organisations.

References

1. <https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/>
2. <https://www.fda.gov/news-events/press-announcements/fda-authorizes-booster-dose-pfizer-biontech-covid-19-vaccine-certain-populations>
3. <https://twitter.com/>
4. <https://economictimes.indiatimes.com/topic/Indias-Covid-19-Vaccine/2>
5. <https://www.reuters.com/world/us/top-us-health-officials-plan-narrower-covid-19-booster-launch-source-2021-09-03/>
6. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160 (2021).
<https://doi.org/10.1007/s42979-021-00592-x>
7. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
8. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7569573/>
9. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200331-sitrep-71-covid-19.pdf?sfvrsn=4360e92b_8
10. <https://www.worldometers.info/coronavirus/>
11. <https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html>
12. Loomba, S., de Figueiredo, A., Piatek, S.J. et al. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav* 5, 337–348 (2021).
<https://doi.org/10.1038/s41562-021-01056-1>.
13. Cinelli, Matteo & Morales, Gianmarco & Galeazzi, Alessandro & Quattrociocchi, Walter & Starnini, Michele. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*. 118. e2023301118. 10.1073/pnas.2023301118.
14. COVID-19 Echo Chambers: Examining the Impact of Conservative and Liberal News Sources on Risk Perception and Response. Kenneth A. Lachlan, Emily Hutter, and Christine Gilbert. *HealthSecurity*. Feb2021.21-30. <http://doi.org/10.1089/hs.2020.0176>.
15. Bonta, Venkateswarlu & Kumares, Nandhini & Janardhan, N.. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology*. 8. 1-6. 10.51983/ajcst-2019.8.S2.2037.
16. <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>
17. <https://pypi.org/project/networkx/>
18. <https://pypi.org/project/textblob/0.9.0/>
19. <https://pypi.org/project/vaderSentiment/>
20. <https://pypi.org/project/NRCLex/>
21. <https://devblogs.microsoft.com/cse/2015/11/29/emotion-detection-and-recognition-from-text-using-deep-learning/>
22. <https://towardsdatascience.com/introduction-to-language-models-n-gram-e323081503d9>

23. <https://www.nltk.org/>
24. <https://www.sciencedirect.com/topics/computer-science/reingold-layout>
25. <https://realpython.com/nltk-nlp-python/>
26. <https://www.datacamp.com/community/tutorials/wordcloud-python>
27. <https://economictimes.indiatimes.com/news/international/world-news/us-vice-president-elect-kamala-harris-gets-moderna-covid-19-vaccine/video-show/80019138.cms?from=mdr>
28. <https://www.thehindu.com/news/national/coronavirus-live-june-22-2021-updates/article34898077.ece>