



A statistical analysis of tweets on covid-19 vaccine hesitancy utilizing opinion mining: an Indian perspective

Ravi Verma¹ · Amit Chhabra¹ · Ankit Gupta¹

Received: 14 August 2022 / Revised: 2 December 2022 / Accepted: 5 December 2022 / Published online: 27 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

The world witnessed the emergence of a deadly virus in December 2019, later named COVID-19. The virus was found to be highly contagious, and so people across the world were highly prone to be affected by the virus. Being a virus-borne disease, developing a vaccine was one of the most promising remedies. Thus, research organizations across the globe started working on developing the vaccine. However, it was later found by many researchers that a large number of people were hesitant to receive the vaccine. This paper aims to study the acceptance and hesitancy levels of people in India and compares them with the acceptance and hesitancy levels of people from the UK, the USA, and the rest of the world by analyzing their tweets on Twitter. For this study, 2,98,452 tweets were fetched from January 2020 to March 2022 from Twitter, and 1,84,720 tweets from 1,22,960 unique users were selected based on their country of origin. Machine learning based Sentiment analysis is then used to evaluate and analyze the tweets. The paper also proposes an NLP-based algorithm to perform opinion mining on Twitter data. The study found the public sentiment of the Indian population to be 63% positive, 28% neutral, and 9% negative. While the worldwide sentiment distribution is 45% positive, 34% neutral, and 21% negative, the USA has 42% positive, 34% neutral, and 23% negative and the UK has 50% positive, 29% neutral, and 21% negative. Also, sentiment analysis for individual vaccines in Indian context resulted in “Covaxin” with the highest positive sentiment at 43% followed by “Covishield” at 36%. The outcome of this work yields an insight into the public perception of the COVID-19 vaccine and thus can be used to formulate policies for existing and future vaccine campaigns. This study becomes more relevant as it is the consolidated opinion of Indian people, which is versatile in nature.

Keywords Natural language processing (NLP) · Opinion mining · Sentiment analysis · COVID-19 vaccine · Coronavirus · Latent dirichlet allocation (LDA)

1 Introduction

COVID-19 is a communicable disease that infects the respiratory organs of human beings that originated in Wuhan city of China. The disease started spreading worldwide in early December 2019 Xu et al. (2021). COVID-19 brought catastrophe in the lives of ordinary people by forcing them to limit physical contact through a lockdown and home isolation. As a result, many businesses suffered, and people lost their livelihoods during the pandemic.

As of July 13, 2022, COVID-19 has claimed the lives of 6,351,801 people (World 2020) across the world. A tremendous amount of effort has been put into vaccination programs worldwide to avert the unfolding of coronavirus. During the pandemic, many pharmaceutical industries and research organizations started research toward developing

Ravi Verma, Amit Chhabra and Ankit Gupta have contributed equally to this work.

✉ Amit Chhabra
amitchh80@gmail.com

Ravi Verma
ravivermaj488@gmail.com

Ankit Gupta
theankit@yahoo.com

¹ Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology (Degree Wing), Sector 26, Chandigarh, Chandigarh 160019, India

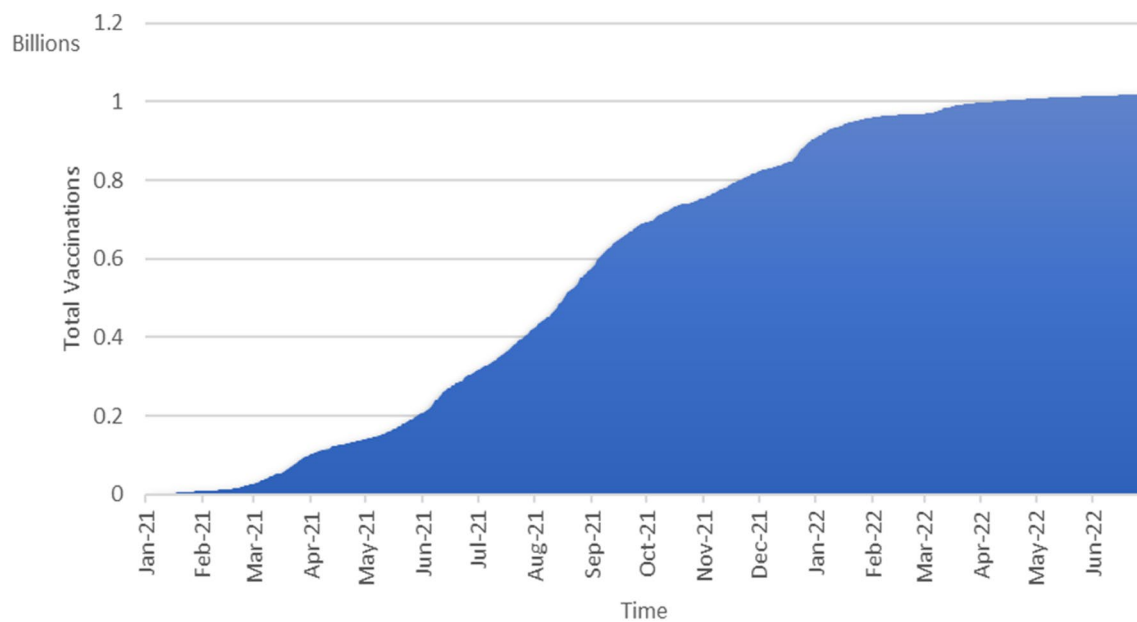


Fig. 1 COVID-19 vaccination in India *Source:* Mathieu et al. (2021)

an efficient vaccine. The COVID-19 vaccine strategies can be classified into major 4 types¹:

1. Inactivated vaccine
2. mRNA vaccine
3. Viral Vector- based vaccine
4. Nanoparticle-based peptide vaccine

Governments from various countries started vaccination programs to immunize their citizens. India's first COVID-19 infection was discovered in January 2020 Darapaneni et al. (2020). The Indian government began administering the COVID-19 vaccination drive nationwide on January 16, 2021 Jain et al. (2021).

Two vaccines were granted emergency use authorization by the Central Drugs Standard Control Organization (CDSCO) in India, Covishield (AstraZeneca's vaccine manufactured by Serum Institute of India and a non-replicating viral vector based vaccine) and Covaxin (manufactured by Bharat Biotech Limited, an inactivated virus-based vaccine). Sputnik - V (produced by Gamaleya Research Institute of Epidemiology and Microbiology and a non-replicating viral vector based vaccine) has been granted EUA in the month of April 2021 MoHFW (2022). Further, Government of India announced a CoWin (2022) portal for COVID-19 vaccine registration, using electronic devices such as Mobile, Tablets, Laptops, and Desktops (Kanabar and Bhatt 2022). The vaccination campaigns progress is depicted in Fig. 1.

¹ For further reading regarding types of COVID-19 vaccine and how they work, readers are encouraged to refer Khandker et al. (2021).

During vaccine development and usage, there was much debate revolving around the effectiveness of vaccines on social media, and public opinion was divided between pro-vaccination and anti-vaccination.

The public interest related to Covid-19 Vaccine for the past two years is shown in Figs. 2, 3. Social media plays an active part in disseminating information among the public. Twitter is a popular social media platform from where information is disseminated (Rosenberg et al. 2020). During the trials, release, and post-release of vaccinations, there were lots of rumors and myths being spread on Twitter relating to the vaccination drive. The rumors and anti-vaccine sentiments were fueled by various social, economic, educational, political, and religious factors. In the United States, there were large anti-vaccination and anti-mask protests (Khadafi

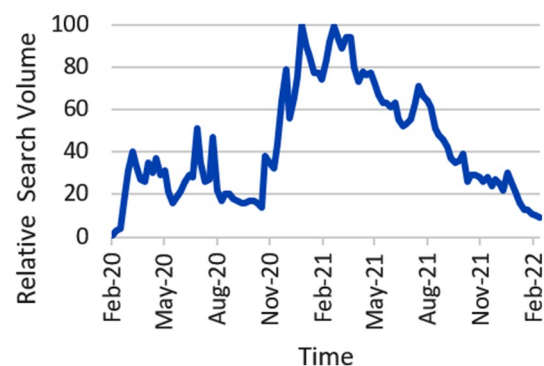


Fig. 2 Google search trend for "COVID-19 vaccine" *Source:* Google (2022)

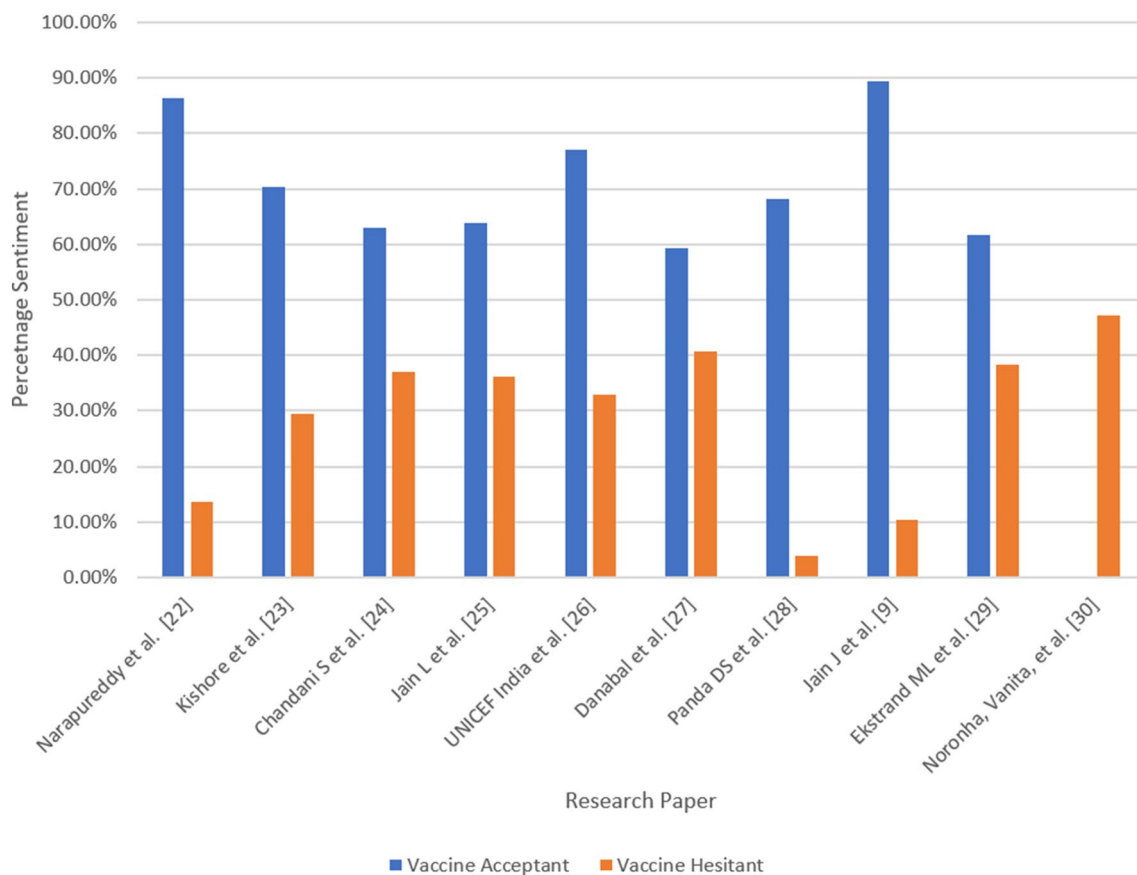


Fig. 3 COVID-19 public opinion surveys from India

et al. 2022; Pressman and Choi-Fitzpatrick 2021). Also, in lower-middle-income countries like India, people are hesitant to take vaccines due to religious reasons and a lack of education. These anti-vaccine sentiments are often kept alive for political reasons (Chatterjee et al. 2021).

The present study becomes more relevant as World Health Organization (WHO) has identified Vaccine hesitancy as one of the ten threats to global health (WHO 2019). WHO has defined vaccine hesitancy as: “the reluctance or refusal to vaccinate despite the availability of vaccines—threatens to reverse progress made in tackling vaccine-preventable diseases.” Further, WHO has categorized key reasons behind vaccine hesitancy as (i) complacency, (ii) inconvenience in accessing vaccines and (iii) lack of confidence. Yilmaz et al. (2022) studied these classifications of vaccine hesitancy behavior on Health care workers employed with a hospital in Turkey and found that 68.8 % respondent had fear/lack of confidence as a reason for vaccine hesitancy while 4.3 % and 9.1 % respondents had complacency and Inconvenience in accessing vaccines, respectively, as a reason for vaccine hesitancy.

The paper aims to study public opinion regarding vaccine hesitancy originated in India by fetching tweets from the Twitter platform. The tweets associated with the term “covid

vaccine” and their hashtag “#CovidVaccine” is used. The sentiment analysis is employed to understand the sentiments floating around the COVID-19 vaccines.

Based on the polarity score, which ranges from 1.0 to -1.0, the tweets were categorized into positive, neutral, and negative. Utilizing the polarity score, the sentiment distributions of countries, including India, the USA, and the UK, were plotted. The classified tweets helped assess people’s opinions on the COVID-19 vaccine. Section 2 of this paper is about motivation and related work. The material and methods are in Sect. 3, and it expounds on the dataset, data preparation steps, and the sentiment analysis process. Section 4 highlights the result and Sect. 5 presents the discussion. Ultimately, the limitations and conclusion are presented in Sects. 6 and 7 respectively.

2 Literature review

Ascertaining what the public thinks about a subject of interest is crucial for research purposes. For textual data, the use of NLP techniques like sentiment analysis is continued to be used. This section summarizes numerous earlier studies on sentiment

Table 1 Related work on COVID-19 vaccine survey from India

S.No	Source	Total participants	Vaccine acceptant (%)	Vaccine hesitant (%)
1	Narapureddy et al. (2021)	351	86.3	13.7
2	Kishore et al. (2021)	467	70.44	29.55
3	Chandani et al. (2021)	1638	63	37
4	Jain et al. (2021)	655	63.8	36.2
5	UNICEF (2021)	–	77	33
6	Danabal et al. (2021)	564	59.3	40.7
7	Panda et al. (2021)	359	68.2	3.9
8	Jain et al. (2021)	1068	89.4	10.5
9	Ekstrand et al. (2021)	438	61.6	38.4
10	Noronha et al. (2021)	435	–	47.1

analysis and public opinion surveys related to COVID-19 vaccines across India.

Guarino et al. (2021) carried out a study on COVID-19-related discussions on Facebook. This study retrieved 1.5 million posts for four months starting from January 2020, shared by 80,000 public pages and groups. The research work is limited to the conversation related to COVID-19 among Italian-speaking people on Facebook.

ÇILGIN et al. (2022) used Twitter posts with associated hashtags with COVID-19. They retrieved 60,242,040 tweets from Twitter from January 2020 to July 2020. They classified tweets into “Highly Positive,” “Positive,” “Neutral,” “Negative,” and “Highly Negative” using VADER (Hutto and Gilbert 2014) and used WordCloud (Heimerl et al. 2014) to visualize the contents of tweets. However, the tweets are collected only from the first six months of the COVID-19 outbreak in this research work.

Chen and Sokolova (2021) studied the post from the r/Depression community on the Reddit platform related to COVID-19. The community is popular among young users, and people share their stories in the community to seek support and opinions. They used VADER (Hutto and Gilbert 2014), Textblob (Heimerl et al. 2014), and SentiWordNet (Esuli and Sebastiani 2006) to analyze the data. However, the Reddit platform is used by a certain age group, and it is not as popular as other platforms like Twitter or Facebook. Also, Twitter is known for more political discussions.

The naïve Bayes algorithm, based on Bayesian Probability, is used in sentiment analysis. The sentiment analysis of 6000 tweets with the keywords “Vaccine COVID-19” from Indonesia reported 56% negative, 39% positive, and 1% neutral performed by Ritonga et al. (2021). Nevertheless, this research work only captures Indonesian tweets with only two weeks of January 2021.

Understanding the content of textual data helps gain insights that sentiment analysis might not reveal. The technique of topic

modeling LDA (Latent Dirichlet Allocation) was employed on tweets to obtain the topic of discussion. Melton et al. (2021) applied Latent Dirichlet Allocation and Sentiment Analysis on posts collected from December 1, 2020, to May 15, 2021, from the 13 communities on Reddit discussing the COVID-19 vaccine. Data were compiled, and LDA was applied each month to sense the changes in topics and sentiment. The results showed that the prime focus of discussion was the vaccine’s side effects. At the same time, the sentiment remained positive throughout. However, due to the unavailability of geographical data, this research cannot be compared with regional polls and opinion surveys.

During the COVID-19 pandemic, many surveys were organized in India. The Public opinion surveys on COVID-19 Vaccines in which people participated were conducted in the past years, shown in Table 1 from India.

3 Material and methods

The methodology section is organized into five sections describing each opinion mining workflow phase. The first section deals with data collection. Subsequently, data exploration, preprocessing, Latent Dirichlet Allocation, and sentiment analysis. The workflow of this section is described in Fig. 4.

3.1 Dataset

Tweets related to COVID-19 vaccine were web scrapped from www.twitter.com that were published on website during January 2020 and March 2022. Search term for the tweets were set as “#CovidVaccine”, #CoronavirusVaccine, #CoronaVaccine #Pandemic #Covid19, “vaccine,” “vaccination,” “covid.” Initially, a total of 2,98,452 tweets were retrieved. After preprocessing and data cleaning, 184720 tweets from 1,22,960 unique users were finally selected and segregated based on their demography, i.e., country of origin. Tweets were segregated into following categories:

1. Tweets originated from India
2. Tweets originated from the USA
3. Tweets originated from the UK
4. Tweets originated from rest of the world

for a comparative analysis of the sentiment analysis and to understand the Covid-19 vaccine sentiments from an Indian perspective.

3.2 Exploratory data analysis

The objective of the exploratory data analysis phase is to get acquainted with the columns of the dataset. The dataset contained the following attributes, as mentioned in Table 2. We are

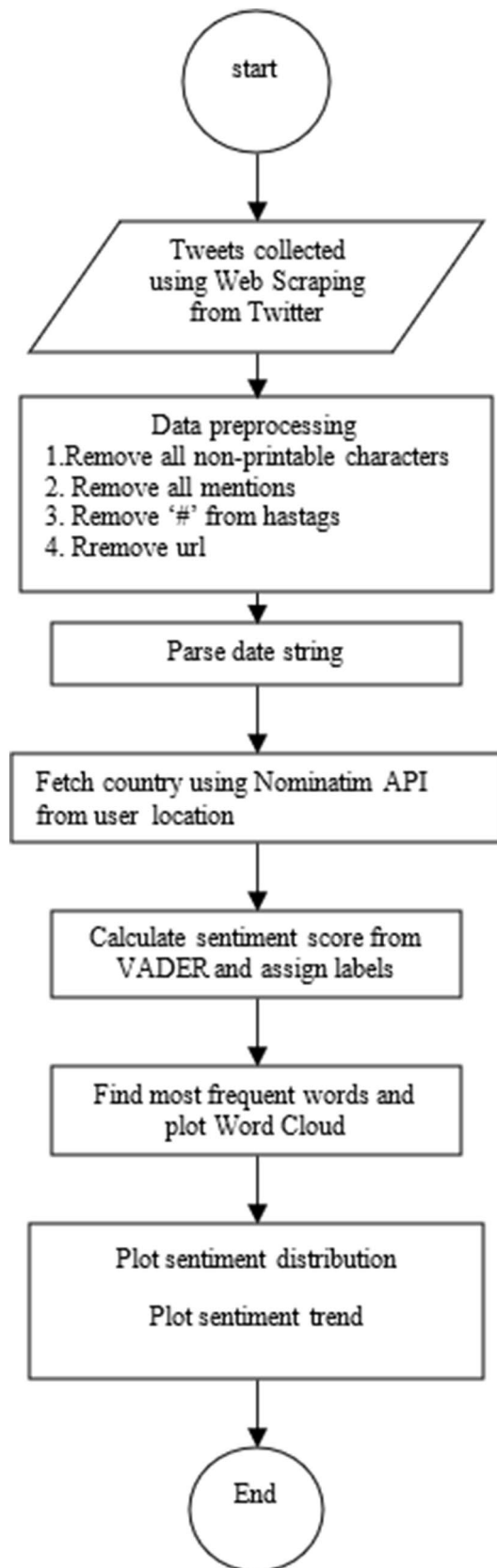


Fig. 4 Workflow of twitter opinion mining

concerned with the “text,” “user_location,” and “date” attributes. The text attribute contains the tweet content we used to conduct the sentiment analysis. Also, we required date attributes to plot the sentiment trend.

3.3 Data preprocessing

Much unwanted information is present in the raw textual data that is not required to conduct sentiment analysis, like URL links, user mentions, and unescaped HTML characters. This unwanted information does not affect the overall sentiment of the tweets, but its removal significantly decreases the time for computing sentiment scores. The “date” attribute contains the date-time in different formats. The different Date Time format was parsed using the Pandas library to this format “yyyy-mm-dd-hh:mm:ss.” Further, the “user_location” and “text” columns are selected from the dataset that would be useful in determining the sentiment of the tweet, sentiment trend, and geographical distribution. The “user_location” attribute helps determine the tweet’s country using the Nominatim open-source geocoding API.

Afterward, the raw tweets are cleaned by removing the following content from the text, like

1. All the non-printable characters from the tweets, like newline.
2. All the URLs and the symbol ‘@’ (mentions).
3. ‘#’ from the hashtags.
4. Rows with null tweet.

3.4 LDA

The topic-modeling technique LDA (Latent Dirichlet allocation) technique is unsupervised and is used to discover undiscovered subjects in textual data. The LDA is a probabilistic model based on the Bayesian model. We have the total documents represented by “D” and a “K” number of total topics. Each θ_d is a random variable that we want to learn, drawn from some Dirichlet distribution. The Dirichlet distribution assumes that any random variable vector from this distribution behaves as a probability mass function. The LDA assumes that each word in every document belongs to a particular topic. Over the whole vocabulary of words, each topic k has a probability distribution, and β_k is the probability distribution over words for a topic k . So, the problem becomes learning optimal θ and β probability distribution vectors, where w_{nd} is the n th word in document d . In the LDA, for each word, we assume they are independent of each other. It implies we can separate the joint probability of all words written here as the product of the probabilities of each separate word. We introduce latent variable z , which we let represent the topic index assigned to each word. We can now say the probability of some word w given

Table 2 COVID-19 vaccine tweet dataset attributes

S.No	Attribute	Description	Data type
1	user_name	Twitter handle	String
2	user_location	The location of the user	String
3	user_description	Description of the user	String
4	user_created	Date of account creation	String
5	user_followers	Total followers of user	Integer
6	user_friends	Total people the user follows	Integer
7	user_favourites	A total number of tweets liked by the user	Integer
8	user_verified	Is the user officially verified by Twitter	Boolean
9	text	Tweet content	String
10	date	The date publishing the tweet	String
11	hashtags	The hashtags used in the tweet	List
12	is_retweet	Is the tweet retweeted	Boolean
13	Source	The device used for tweeting	String

Table 3 Tweet classification example

S.No	Tweet	Label
1	Australia to manufacture Covid-19 vaccine and give it to the citizens for free of cost: AFP quotes Prime Minister #CovidVaccine	Positive
2	Still skeptical about the vaccine...#COVID19 #CovidVaccine	Negative
3	Not acceptable #ForcedVaccinations #ForcedVaccination #CovidVaccine #COVID19 #mandatoryvaccination	Negative
4	U.S. advisory panel released a draft plan Tuesday for how to ration the first doses of #CovidVaccine	Neutral

this topic z can be found precisely using β_z by applying reverse marginalization on the probability of each word.

$$\begin{aligned} \hat{\beta}_{1:K} \hat{\theta}_{1:D} &= \underset{\beta_{1:K} \theta_{1:D}}{\operatorname{argmax}} P(\{w_{nd}\}_{n=1}^N \{d=1\}^D | \beta_{1:K} \theta_{1:D}) \\ &= \underset{\beta_{1:K} \theta_{1:D}}{\operatorname{argmax}} \prod_{n=1}^N \prod_{d=1}^D P(w_{nd} | \beta_{1:K} \theta_d) \\ &= \underset{\beta_{1:K} \theta_{1:D}}{\operatorname{argmax}} \prod_{n=1}^N \prod_{d=1}^D \sum_{z=1}^K P(w_{nd} | z) P(z | \theta_d) \end{aligned}$$

This technique is used to discover the topics of all the tweets, and these topics can help better understand the context of the Twitter discussion.

3.5 Sentiment analysis

In this phase, we conducted sentiment analysis on the tweets and produced the polarity score. Also, we propose an algorithm to perform sentiment analysis on the Twitter dataset. The detailed algorithm is presented in algorithm 1. After the sentiment analysis, we further assign sentiment labels to each tweet. We used a popular social media sentiment analysis library called VADER (Hutto and Gilbert 2014). It is sensitive to emojis, punctuation

marks, and capital words. Often social media posts have these elements. The preprocessed tweets are used, which are cleaned for URLs, Mentions, and Retweets. Now the polarity score is used for classifying the tweets. A positive label was assigned to the tweets with a polarity score of more than or equal to 0.10. While a negative label is assigned to the tweets with a polarity score less than or equal to -0.10, any score between 0.10 and -0.10 is considered neutral. Classified tweets are used to plot the evolution of sentiments with time. The graph depicting the monthly evolution of the positive and negative sentiments is produced to examine the opinion toward the COVID-19 vaccination in the tweets. The pattern and trends in this graph are beneficial in discerning how the various factors like policies, decisions, and milestone events on COVID-19 vaccinations have affected the public perception of the COVID-19 vaccination (Table 3).

The sentiment distribution is plotted, and the positive, negative, and neutral sentiment and their percentage distribution were studied country-wise to understand the public’s response in various geography. To understand the causes of each sentiment, we extracted the most frequent words from tweets for India’s positive, negative, and neutral sentiments. Also, vaccine sentiments for individual vaccines were found using tweets containing the vaccines’ names and compared.

Algorithm 1 Twitter opinion mining algorithm

Input: Set of tuples of tweet $T = \{(T_1, D_1, C_1), (T_2, D_2, C_2), (T_3, D_3, C_3), \dots, (T_N, D_N, C_N)\}$ Where T_i is the tweet, D_i is the date string, and C_i is the location string

Output: Set of processed tweets $T = \{(T_1, D_1, C_1, P_1, L_1), (T_2, D_2, C_2, P_2, L_2), (T_3, D_3, C_3, P_3, L_3), \dots, (T_N, D_N, C_N, P_N, L_N)\}$ where T_i is the tweet, D_i is the date string, C_i is the location string, P_i is the polarity score and L_i is the sentiment label

```

function CLEANTEXT(T)  $S = \text{tokenize } T \text{ to } T_1, T_2, T_3 \dots T_N$ 
  for  $i=1$  to  $N$  do
    remove non-printable characters using regex from tokens in  $T_i$ 
    remove URL using regex from  $T_i$ 
    remove mentions using regex from  $T_i$ 
    remove retweets using regex from  $T_i$ 
  end for
   $R = \text{concatenate } S$ 
  return  $R$ ;
end function
function GETCOUNTRY(T)
   $L = \text{get city using Nominatim API geocoding from } L$ 
  return  $L$ ;
end function
function PARSEDATE(Date)
   $D = \text{get day from date}$ 
   $M = \text{get month from date}$ 
   $Y = \text{get year from date}$ 
  return  $(D, M, Y)$ ;
end function
for  $i=1$  to  $N$  do
   $T_i = \text{CLEANTEXT}(T_i)$ 
   $D_i = \text{PARSEDATE}(D_i)$ 
   $C_i = \text{GETCOUNTRY}(C_i)$ 
   $P_i = \text{compute polarity using Vader}$ 
  if  $P_i \geq 0.10$  then
     $L_i = \text{"positive"}$ 
  else  $[P_i < 0.10]$ 
    if  $P_i \leq -0.10$  then
       $L_i = \text{"negative"}$ 
    else  $[P_i > -0.10 \text{ and } P_i < 0.10]$ 
       $L_i = \text{"neutral"}$ 
    end if
  end if
   $O = O \cup (T_i, D_i, C_i, P_i, L_i)$ 
end for
return  $O$ 

```


Table 5 Most frequent words in each sentiment in India

S.No	Positive	Negative	Neutral
1	Free	Shortage	First
2	Safe	No	Dose
3	Availability	Emergency	Vaccinated
4	Vaccinated	Get	Administered
5	Take	Vaccine	Covid

and 21% as negative worldwide. While the sentiment distribution in the USA was 42% positive, 34% neutral, and 23% negative, while in India, 63% is found to be positive, 28% neutral, and 9% negative, and the UK had 50% positive, 29% neutral, and 21% negative as represented in Fig. 5, from January 2020 to March 2022.

4.2 LDA and popular words in tweets

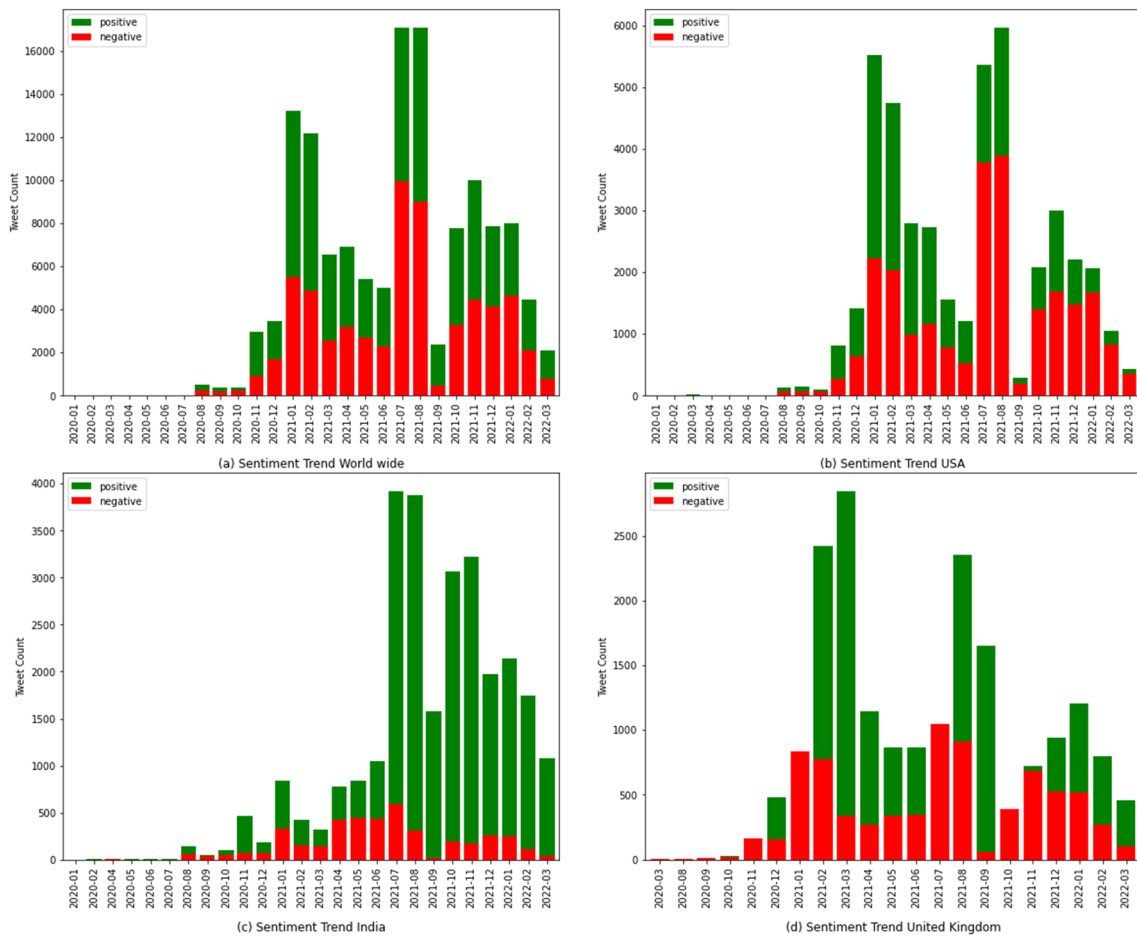


Fig. 7 COVID-19 tweet volume per month

Table 6 Individual vaccine sentiment distribution

S.No	Name	Positive (%)	Neutral (%)	Negative (%)
1	Covaxin	43	45	12
2	Covishield	36	51	13
3	Sputnik V	30	42	28

Table 4 shows the debate themes that emerged from the LDA on textual data of tweets. The first topic discusses dose accessibility, and the second topic is concerned with deaths following COVID-19 vaccine immunization. The third topic is vaccine eligibility. The final topic concerns the second dose.

Using the Word Cloud Library (Heimerl et al. 2014), the most frequently used phrases tweeted by users worldwide are compiled. The most repeated words are shown in Fig. 6, and according to it, the bulk of tweets of Indian origin used the terms “dose 1” and “dose 2” as the topics of conversation in India.

Also, studying the most frequent words for each sentiment for India is presented in Table 5, Which can help us better understand the cause of each sentiment. The top 5 words are selected for each sentiment.

4.3 Sentiment trend

The sentiment trend is the plot of the volume of negative and positive sentiment per month from January 2020 to March 2022. The sentiment trend worldwide, in the USA, India, and the United Kingdom, is depicted in Fig. 7. The sentiment trend suggests an increase in the volume of sentiments overall and a decline afterward. The Indian sentiment trend is different in contrast to the rest of the other sentiment trends. In comparison, the sentiment trend of the United Kingdom shows more acceptance than that of the USA.

4.4 Individual vaccine sentiments

The individual vaccine sentiment is found by searching all the tweets that contained the vaccine’s name in the text, using the classified labels of that tweet, and counting the positive, negative, and neutral labels for each COVID-19 vaccine tweet. Table 6 displays the sentiment distribution for each approved vaccine. The domestically made vaccine in India, “Covaxin” Thiagarajan (2021), exhibits the highest positive and least negative sentiments paired with “Covishield” Sapkal et al. (2021). While high levels of negative sentiments compared to Indian-origin vaccines were noted for non-Indian origin vaccines, including “Sputnik V” Cazzola et al. (2021). It supports the finding in Figs. 5 and 7.

5 Discussion

The paper’s primary objective is to understand the opinion of the Indian populace toward COVID-19 vaccines. Also, enquire about how the opinion evolved throughout the past two years. The timeline of sentiment trends can be divided into pre-vaccine and post-vaccine periods. Before First phase trial for the vaccine was done in July 2020 Pawar et al. (2021). The vaccine-related discussion was very low with high negative sentiment toward the vaccine. The volume of discussion starts to rise when the first wave of Covid-19 hit in mid of 2020. Then subsequently the volume of discussion keeps rising.

The pre-vaccine period is considered before the vaccination program starts in India, which is before January 2021 Narapureddy et al. (2021). The pre-vaccine phase is characterized by low vaccine-related discussion. The same is observed with the Google search trend for COVID-19 vaccines (Google 2022).

Followed by the pre-vaccine phase and the post-vaccine phase, when the vaccine was used, After the second wave hit India in May 2021 the discussion shows an exponential increase and negative sentiments also starts to decline, the discussion rises till July 2021 and then starts to decline. A substantial portion of the population being vaccinated every day contributed to the decline in interest in COVID-19 vaccines (Mathieu et al. 2021).

Then the volume of a discussion rises before the third wave hit India in January 2022. After this period the volume of discussion decreases also a large portion of the population in India is vaccinated with the first dose. With declining discussion also the negative sentiment declines.

One of the critical observations drawn is a contrast in the trends of India with the USA, UK, and worldwide shown in Fig. 7. An overall increase in positive sentiments characterizes the Indian sentiment trend, and during the same period, the sentiments of the USA, UK, and worldwide did not change much.

The high positive and falling negative sentiment in India suggests the high acceptability of vaccines in India which increased during the past two years. While the USA witnessed anti-mask and anti-vaccine protests (Gresele et al. 2021; Cazzola et al. 2021), contributing to a high vaccine hesitancy among the population of the USA. The same is observed for the rest of the world and the UK.

The LDA gave us a glimpse of the topics discussed in the tweets. The theme for each topic is “dose accessibility,” “post-covid vaccine deaths,” “COVID-19 Vaccine eligibility,” and “second dose.”

Table 5 helps us to understand the underlying cause of each sentiment in India. The following observation can be drawn from the most frequent words from positive tweets like “Free COVID-19 Vaccines in India,” “COVID-19 Vaccine availability,” and “COVID-19 vaccine is safe.” While the most frequent words in negative sentiment tweets give some of the possible causes like “COVID-19 vaccine shortage,” “Getting vaccinated in Emergency only,” and “No to getting vaccinated.”

Also, comparing the Indian home-grown vaccine and non-Indian vaccines. Table 6 suggests the highest positive and lowest negative sentiment for the Indian home-grown vaccine “Covaxin” (Thiagarajan 2021), which is corroborated by the previous observations. At the same time, non-Indian vaccines had high negative sentiments except the “Covishield” Sapkal et al. (2021) vaccine, which was used in India.

This information can assist critical stakeholders like the pharmaceutical industry in understanding public concerns. Also, it helps shape social networking campaigns and the management of the pandemic caused by COVID-19 by the government.

6 Limitation

The present study has some limitations such as people across the globe use various social media platforms to express their feelings while this study consider tweets retrieved from twitter and excludes data from other platforms like Facebook, Instagram, Snapchat and Quora, etc. Furthermore, this study also excludes relevant information that was available on Print Media. Also, the tweets are collected for the single window time period from January 2020 to March 2022 only. Retrieved tweets may not be comprehensive as some tweets related to the vaccine hesitancy might have been deleted by their users while some accounts might have been suspended or become private in this time frame. In this study, only English language tweets were considered. Also, the study uses Text Mining for analysis purpose while special characters, emojis, and pictures were not taken into consideration.

7 Conclusion

The sentiment analysis of social media platforms like Twitter can assist in studying public response or behavior. The stakeholders, like the government and pharmaceutical industries, are beneficiaries of such information. In India, opinions toward the COVID-19 vaccine have changed during the last two years. The overall change in opinion is positive toward the COVID-19 vaccines. In contrast, the opinion of the USA, the UK, and the rest of the world, had high negative sentiment toward the COVID-19 vaccine, which did not change much during the same period. The sentiment of the Indian population was found to be 63% positive, 28% neutral, and 9% negative. While the worldwide sentiment distribution was 45% of tweets positive and 34% neutral. 21% as negative. Also, the USA had 42% positive, 34% neutral, and 23% negative; the UK had 50% positive, 29% neutral, and 21% negative. The overall response of the Indian population is found to be vaccine-acceptant. The Indian home-grown vaccine showed high positive and low negative sentiments compared to the other non-Indian vaccines. This work can be further extended by incorporating other social media platforms and government sources. In future, we intend to add more varieties of datasets for our extended work.

Funding For this research no funding was provided.

Data Availability The dataset generated during and/or analyzed during the current study are not publicly available due Twitter Terms of Service but are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors have no financial or proprietary interests in any material discussed in this article.

Ethics approval For this research, no ethical approval was taken.

References

- Cazzola M, Rogliani P, Mazzeo F, Matera MG (2021) Controversy surrounding the sputnik v vaccine. *Respir Med* 187:106569. <https://doi.org/10.1016/j.rmed.2021.106569>
- Chandani S, Jani D, Sahu PK, Kataria U, Suryawanshi S, Khubchandani J, Sharma D (2021) Covid-19 vaccination hesitancy in India: state of the nation and priorities for research. *Brain Behav Immun-Health* 18:100375
- Chatterjee N, Mahmood Z, Marcussen E (2021) Politics of vaccine nationalism in India: global and domestic implications. *Forum Dev Stud* 48:357–369
- Chen Z, Sokolova M (2021) Sentiment analysis of the covid-related r/ depression posts. arXiv preprint <https://arxiv.org/abs/2108.06215>
- ÇILGIN C, Metin B, BİLGEHAN H, Ceyda Ü (2022) Twitter sentiment analysis during covid-19 outbreak with vader. *AJIT-e: Acad J Inf Technol* 13(49):72–89
- CoWin (2022). Cowin. Ministry of Health and Family Welfare. <https://www.cowin.gov.in/> Accessed 19 Oct 2022
- Danabal KGM, Magesh SS, Saravanan S, Gopichandran V (2021) Attitude towards covid 19 vaccines and vaccine hesitancy in urban and rural communities in Tamil Nadu, India-a community based survey. *BMC Health Serv Res* 21(1):1–10
- Darapaneni N, Reddy D, Paduri AR, Acharya P, Nithin HS (2020) Forecasting of covid-19 in india using arima model. In: 2020 11th IEEE annual ubiquitous computing, electronics & mobile communication conference (uemcon), pp 0894–0899
- Ekstrand ML, Heylen E, Gandhi M, Steward WT, Pereira M, Srinivasan K (2021) Covid-19 vaccine hesitancy among plwh in south India: implications for vaccination campaigns. *J Acquir Immune Defic Syndr* 88(5):421
- Esuli A, Sebastiani F (2006) Sentiwordnet: a publicly available lexical resource for opinion mining. In: Proceedings of the fifth international conference on language resources and evaluation (Irec'06)
- Google Trends (2022) Covid 19 vaccine - explore - google trends. Google. <https://trends.google.com/trends/explore?q=covid%2019%20vaccine&geo=IN> Accessed 13 Jul 2022
- Gresele P, Marietta M, Ageno W, Marcucci R, Contino L, Donadini MP et al (2021) Management of cerebral and splanchnic vein thrombosis associated with thrombocytopenia in subjects previously vaccinated with vaxzevria (astrazeneca): a position statement from the italian society for the study of haemostasis and thrombosis (siset). *Blood Transfus* 19(4):281
- Guarino S, Pierri F, Di Giovanni M, Celestini A (2021) Information disorders during the covid-19 infodemic: the case of italian facebook. *Online Soc Netw Media* 22:100124
- Heimerl F, Lohmann S, Lange S, Ertl T (2014) Word cloud explorer: text analytics based on word clouds. In: 2014 47th Hawaii international conference on system sciences, pp 1833–1842

- Hutto C, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media vol 8, pp 216–225
- Jain J, Saurabh S, Kumar P, Verma MK, Goel AD, Gupta MK, Raghav PR (2021) Covid-19 vaccine hesitancy among medical students in India. *Epidemiol Infect.* <https://doi.org/10.1017/s0950268821001205>
- Jain L, Vij J, Satapathy P, Chakrapani V, Patro B, Kar SS et al (2021) Factors in uencing covid-19 vaccination intentions among college students: a cross-sectional study in India. *Front Public Health.* <https://doi.org/10.3389/fpubh.2021.735902>
- Kanabar K, Bhatt N (2022) Co-win as communication platform for managing covid-19 in India. *Media Asia* 49(2):155–163. <https://doi.org/10.1080/01296612.2021.1990465>
- Khadafi R, Nurmandi A, Qodir Z, Misran (2022) Hashtag as a new weapon to resist the covid-19 vaccination policy: a qualitative study of the anti-vaccine movement in brazil, USA, and Indonesia. *Hum Vaccin Immunother* 18(1):2042135. <https://doi.org/10.1080/21645515.2022.2042135>
- Khandker SS, Godman B, Jawad MI, Meghla BA, Tisha TA, Khondoker MU et al (2021) A systematic review on covid-19 vaccine strategies, their effectiveness, and issues. *Vaccines* 9(12):1387
- Kishore J, Venkatesh U, Ghai G, Kumar P et al (2021) Perception and attitude towards covid-19 vaccination: a preliminary online survey from India. *J Fam Med Prim Care* 10(8):3116. <https://doi.org/10.4103/jfmpe.jfmpe253020>
- Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, Rodés-Guirao L (2021) A global database of covid-19 vaccinations. *Nat Hum Behav* 5(7):947–953. <https://doi.org/10.1038/s41562-021-01122-8>
- Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A (2021) Public sentiment analysis and topic modeling regarding covid-19 vaccines on the reddit social media platform: a call to action for strengthening vaccine confidence. *J Infect Public Health* 14(10):1505–1512
- MoHFW (2022). Cowin. Ministry of health and family welfare. <https://www.mohfw.gov.in/>. Accessed 19 Oct 2022
- Narapureddy BR, Muzammil K, Alshahrani MY, Alkhathami AG, Alsabaani A, AlShahrani AM, Alam MM (2021) Covid-19 vaccine acceptance: beliefs and barriers associated with vaccination among the residents of KSA. *J Multidiscip Healthc* 14:3243
- Noronha V, Abraham G, Bondili SK, Rajpurohit A, Menon RP, Gattani S et al (2021) Covid-19 vaccine uptake and vaccine hesitancy in Indian patients with cancer: a questionnaire-based survey. *Cancer Res, Stat, Treat* 4(2):211
- Panda DS, Giri RK, Nagarajappa AK, Basha S (2021) Covid-19 vaccine, acceptance, and concern of safety from public perspective in the state of Odisha, India. *Hum Vaccin Immunother* 17(10):3333–3337
- Pawar HA, Pawar AH, Pawar SA, Pawar PA (2021) Covid-19 vaccines approved for use in d under w development in India: an overview. *J Pharm Adv Res* 4:1277–1284
- Pressman J, Choi-Fitzpatrick A (2021) Covid19 and protest repertoires in the united states: an initial description of limited change. *Soc Mov Stud* 20(6):766–773
- Ritonga M, Al Ihsan MA, Anjar A, Rambe FH et al (2021) Sentiment analysis of covid-19 vaccine in indonesia using naïve bayes algorithm. *IOP Conf Ser: Mater Sci Eng* 1088:012045
- Rosenberg H, Syed S, Rezaie S (2020) The twitter pandemic: the critical role of twitter in the dissemination of medical information and misinformation during the covid-19 pandemic. *CJEM* 22(4):418–421. <https://doi.org/10.1017/cem.2020.361>
- Sapkal GN, Yadav PD, Sahay RR, Deshpande G, Gupta N, Nyayanit DA et al (2021) Neutralization of delta variant with sera of covishiel™ vaccinees and covid-19-recovered vaccinated individuals. *J Travel Med* 28(7):taab119. <https://doi.org/10.1093/jtm/taab119>
- Thiagarajan K (2021) What do we know about India’s covaxin vaccine? *BMJ: Br Med J* (online). <https://doi.org/10.1136/bmj.n997>
- UNICEF (2021, Nov). Building confidence in the covid-19 vaccine in India. <https://www.unicef.org/india/stories/building-confidence-covid-19-vaccine-India> Accessed 18 Jul 2022
- WHO (2019). “ten threats to global health in 2019”. World Health Organization. <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> Accessed 09 Dec 2020
- World Health Organization (2020). Who coronavirus (covid-19) dashboard. <https://covid19.who.int/>. World Health Organization. Accessed: 13 Jul 2022
- Xu H, Liu R, Luo Z, Xu M, Wang B (2021) Short epidemiological overview of the current situation on covid-19 pandemic in south-east european (see) countries. In: 2021 IEEE international conference on systems, man, and cybernetics (smc), pp. 3200–3205. <https://doi.org/10.1109/SMC52423.2021.9658764>
- Yilmaz S, Çolak FÜ, Yilmaz E, Ak R, Hökenek NM, Altıntaş MM (2022) Vaccine hesitancy of health-care workers: another challenge in the fight against covid-19 in istanbul. *Disaster Med Public Health Prep* 16(3):1134–1140

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.