



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

## Diabetes &amp; Metabolic Syndrome: Clinical Research &amp; Reviews

journal homepage: [www.elsevier.com/locate/dsx](http://www.elsevier.com/locate/dsx)

## Original Article

## Indian citizen's perspective about side effects of COVID-19 vaccine – A machine learning study

Praveen Sv <sup>a,\*</sup>, Jyoti Tandon <sup>b</sup>, Vikas <sup>c</sup>, Hitesh Hinduja <sup>d</sup><sup>a</sup> National Institute of Technology, Trichy, India<sup>b</sup> Nielsen India Private Limited, Delhi, India<sup>c</sup> Department of Management Studies, Rajiv Gandhi Institute of Petroleum Technology, Uttar Pradesh, India<sup>d</sup> Artificial Intelligence, Ola Electric, Bangalore, India

## ARTICLE INFO

## Article history:

Received 12 May 2021

Accepted 6 June 2021

## Keywords:

Sentimental analysis

Topic modeling

COVID-19

Vaccines

Side effect

## ABSTRACT

**Background and aims:** Ever since the vaccination drive for COVID-19 has started in India, the citizens have been sharing their views on social media about it. The present study examines the attitude of Indian citizens towards the side effects of the COVID-19 vaccine.

**Methods:** Social media posts were used for this research. Using Python, we have collected social media posts of Indians focusing on side effects of COVID -19 vaccines. In study one, sentimental analysis was done to find overall attitude of Indian citizens towards the side effects of COVID-19 vaccine and in study two, topic modeling done to analyze the major side effects voiced out by the citizens after taking COVID-19 vaccine.

**Results:** The studies conducted have revealed that nearly 78.5% of tweets posted by Indian citizens about the side effects of the COVID-19 vaccine were either in neutral or positive sentiments. Our topic modeling studies have found that fear of efficiency in the workplace and the fear of death as the prime two issues that contributes Indian citizens to have negative sentiment about the side effects of the COVID-19 vaccine.

**Conclusion:** While it is important for the Indian government to actively encourage its citizens to have vaccine, it is also important to help the citizens understand the important of the vaccination program. The best way to educate citizens regarding the positive aspect of the vaccination program is by addressing the fears, Indian citizens have voiced in their social media post about the COVID-19 vaccines.

© 2021 Diabetes India. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

A pandemic is considered a phenomenon having devastating effects on people and economics with many causalities. Pandemics are said to have both health and economic calamities [1]. The first case of COVID-19 was registered in India on January 27, 2020. A nationwide lockdown was imposed in the country in the month of March 2020. The general public was being monitored for following the lockdown, social distancing, and wearing masks. As per the last year's records, it is established that by May 18, 2020, India registered 1 Lakh COVID-19 infected cases. However, within a span of less than 2 months the cases were increased by 8 times and India

had 8 Lakh infected cases [2]. As of May 4, 2021, the number of confirmed cases of COVID 19 rose to 2,02,82,833 in India, and the death toll reaching 2,22,408 [3]. The above statistics make India the second largest affected country in the world only after the United States of America. Currently, COVID-19 is spreading at a distressing rate in India. The spread of the deadly virus highlights the importance of vaccination at a national level. The vaccination is supposed to protect the nation from continued damage.

The Government has introduced the vaccination drive all over the country on January 16, 2021 [4]. The vaccination drive is being conducted in stages, focusing on health care workers and frontline staff in the first stage. They are at a supreme risk of getting exposed to the virus. The second stage focuses on elderly people above the age group of 45. They are at a high risk of getting affected owing to the virus. In the third stage, people in the age group of 18–45 will be targeted. Nationwide usage of two COVID-19 vaccines is approved, Covishield by Serum Institute of India and Covaxin

\* Corresponding author. National Institute of Technology, Trichy Tanjore Main road, National highway 67 Tiruchirappalli, 620015, Tamil Nadu, India.

E-mail address: [praveennitt04@gmail.com](mailto:praveennitt04@gmail.com) (P. Sv).

manufactured by Bharat Biotech. As per the official data, as of May 4, 2021, a total of 158932921 doses overall including both the first dose and second dose of the vaccine has been provided in India [3].

Although researchers are investigating and analyzing a lot about COVID-19 and some have focused on vaccines. However, no studies in the recent past have been conducted to analyze the perception of people of India for side effects of COVID-19 vaccines. The focus of the present research lies in analyzing the perceptions of the general public of India towards the side effects of COVID-19 vaccines. In the present study we aim at answering the succeeding two research problems:

RQ 1: How is the general public of India perceiving the side effects/after-effects of the COVID-19 vaccine?

RQ 2: What are the side effects/after-effects of the COVID-19 vaccines highlighted by citizens of India in social media?

## 2. Methodology

### 2.1. Data collection and data pre-processing

The analysis aims at taking into consideration the social media posts of citizens of India on Twitter to analyze the perception about the COVID-19 vaccine and its side effects. Using Python Library Twint, tweets having the words 'COVID Vaccine' and 'Side effects' were scrapped.

For the purpose of data collection, Twitter has been used. Post the outbreak of COVID-19, more and more people are using Twitter as a social media platform to endorse their views in the form of "Tweets" [5] Twitter is also represented as a powerful public health tool apart from traditional sources like Radio, Newspaper, and Television as the leaders directly communicate the information on COVID-19 to citizens.

It is well established from the previous research that social media acts as a most credible source to access and record masses' behavior during unusual periods like the current one [6–9].

For the present study, tweets with the words "COVID vaccine" and 'Side effects' were considered. By using the geographical filtering option on Python library Twint only tweets belonging to India were studied. Only the tweets posted in English were examined for the study and tweets from other languages were eradicating for analysis. After eliminating the tweets of other languages, the tweets in the English language were taken into consideration for this study.

To curtail down the sampling errors caused by unbalanced samples, an equal number of tweets for each week was used in the analysis. Post selecting the tweets, the process of data cleaning takes place which aims at removing the punctuation, emoticons, images, hyperlinks, numbers, and stop words. Only "Text" shall be considered for analysis. Stop words needs to be filtered out as they have no meaning of their own and removing them from the sentence leaves the meaning of the sentence unaltered. They are not required for analysis.

After eliminating the punctuation, hyperlinks, numbers, and stop words from the corpus, stemming & lemmatizing the data was performed. Stemming is a method by which prefixes and suffixes are removed to find their common base or root. And, lemmatization is a practice of combining dissimilar words to narrow down the dimensionality. Lemmatization and stemming are very closely related and are an important step in Natural Language Processing. Python Libraries Regular expressions and Gensim were used for the data cleaning process.

### 2.2. Research methodology

#### 2.2.1. Sentimental analysis

The motive of research conducted in study one is to comprehend the attitude of residents of India for the side effects of the COVID-19 vaccine. To analyze and understand the same, we used the process of sentiment analysis. Sentimental analysis is a method of categorizing the sample texts into positive, negative, or neutral brackets. The sentiment score is ascertained from the sentimental analysis. Every word in the sentiment corpus whether it is positive, negative, or neutral contains a sentiment score and using the score the model will determine whether the particular tweet in the corpus is having positive, negative or neutral sentiment.

Sentiment analysis also known as opinion analysis can be defined as "An automatic technique to select and analyze the subjective verdicts on various aspects of an item" [10]. It is a machine learning technique involving the use of Natural Language Processing. The basic aim of sentiment analysis is to find the opinion, attitude, or emotions of the writer for a particular text [11]. The aims also include identifying the degree of polarity of the tone of the text message which can be expressed as positive, negative, or neutral. The positive score in the analysis denotes satisfaction, happiness, and contentment on the part of the author as against negative indicating disappointment, sadness, and sorrow. Sentiment analysis is a technique that identifies and classifies the opinions of people computationally.

Initially, when sentimental analysis started, it was done at a document level [12], sentence-level was the second level [13] followed by the phrase level [14,15].

For this study, Python (a computer programming language) has been used to collect the tweets. For processing the textual data, the Text Blob-Python library will be used. As per the text blob, every English word will have a sentimental score. Text blob while applying the principles of advanced machine learning as well as Natural Language Processing (NLP) aims at studying each word collected in the corpus and will classify the opinions as being neutral, positive, or negative [16].

#### 2.2.2. Latent Dirichlet Allocation (LDA)

Study one helped in understanding the perception of citizens of India towards the side effects of the COVID-19 vaccine if the attitude was positive or negative. We conducted a sentimental analysis for the same, it can be inferred that sentimental analysis can assist in recording the general attitude of people and cannot highlight the exact side effects of the vaccine.

To understand the major side effects/after-effects of the COVID-19 vaccine on Indian's study two will be conducted.

Latent Dirichlet Allocation (LDA) topic modeling was undertaken in study 2 to identify the major side effects/after-effects which are visible to Indian citizens after the COVID-19 vaccine. Blei, Ng, and Jordan get the credit of familiarizing the world with LDA for the first time in 2003.

Topic modeling uses a set of algorithms to recapitulate extensive texts by determining and finding the unseen subjects and themes in a corpus [17]. Traditionally, before the origination of Latent Dirichlet Allocation, Probabilistic Latent Semantic Indexing was used to derive the issues. The concept behind PLSI is that each word in a document is modeled using an algorithm as a sample from the mixture model. The mixture elements in the mixture models are the multinomial random variables that can be considered as topics.

A major drawback of PLSI which led to its loss of popularity and

increased the usage of LDA is the algorithms in Probabilistic Latent Semantic Indexing makes the probabilistic model unavailable for the whole document [18]. LDA works on the principle of the “Bag of words” assumption and follows Bayesian probability theory [19]. The fundamental characteristic of LDA topic modeling is based on the usage of algorithms to derive a similar set of topics in every document which is most talked about or opined about. Latent Dirichlet Allocation very well takes into account the assumption that it is possible for some set of words to develop a linkage with some particular topics always. While using the LDA technique, it is a possibility to discover latent topics from the group of vast and huge unstructured data in the corpus. Library LDAvis is used to better analyze, understand, and later summarise the identified side effects.

### 3. Results

In study one, we have performed sentiment analysis for the data we have collected across the months of March and April 2021. We have tracked out the weekly sentiments of the Indian population about COVID-19 vaccines across these two months. After removing all the tweets from different languages, we have finally chosen 189,888 tweets for our study. Our study shows that 44.9% (n = 85,407 tweets) of our sample were in a neutral tone. Tweets with positive sentiments about the side effects of the vaccine were 33.6% (n = 63,848 tweets). Finally tweets with negative sentiments recorded for 21.3% (n = 40,633 tweets). It is an encouraging sign that, even while posting about the side effects of the COVID-19 vaccine, nearly 78.5% of the tweets were with either neutral or positive sentiments. From our results, it can also be concluded that the positive sentiments towards the side effects of the COVID-19 vaccine increased to a greater extent from the 2<sup>nd</sup> week of April (When the total COVID-19 cases began to see a drastic increase).

**Note:** Fig. 1 depicts the number of neutral, negative and positive tweets for each week of months (March and April 2021) and their variation. For example, the 1<sup>st</sup> week of March recorded 10,696 neutral tweets.

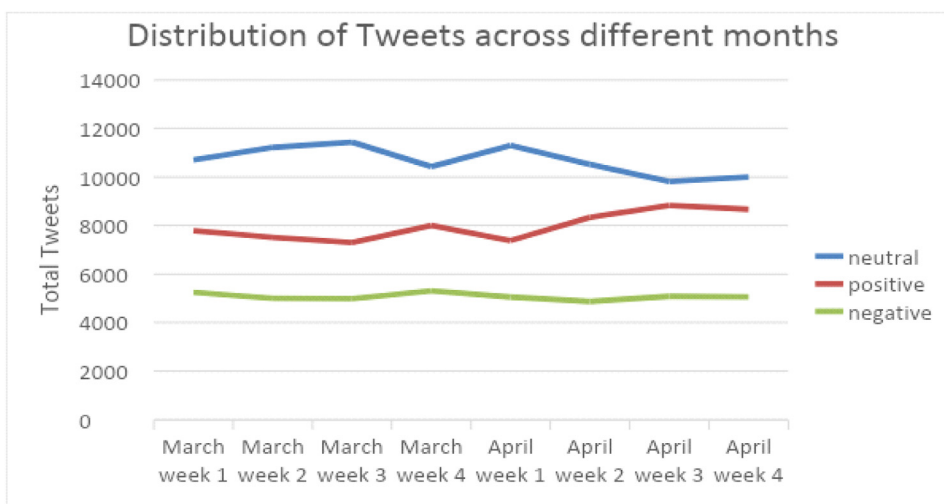
**Note:** Fig. 2 illustrates the percentage of neutral, positive and negative tweets for each week and their variation. For instance, the first week of March records 12.5% (n = 10,696 tweets) of total neutral tweets recorded (n = 85,407).

It is to be highlighted that the first study helped us understand the perception of general Indian population about the side effects of the COVID-19 vaccine, the sentimental analysis study hasn't helped us understand the various variants that determine the sentiment of Indian population about the side effects of COVID-19 vaccine. To identify the variants that determine the negative sentiment of Indian population regarding the side effects of the COVID-19 vaccine, we have performed a study two. In study two we have conducted Latent Dirichlet Allocation topic modeling for the tweets with negative sentiments about the side effects of the COVID-19 vaccine. This study will help us understand the prime factors that led the Indian general population to have negative opinion on side effects of the COVID-19 vaccine. The outcomes of the investigation are specified in Table 2.

### 4. Discussions and conclusion

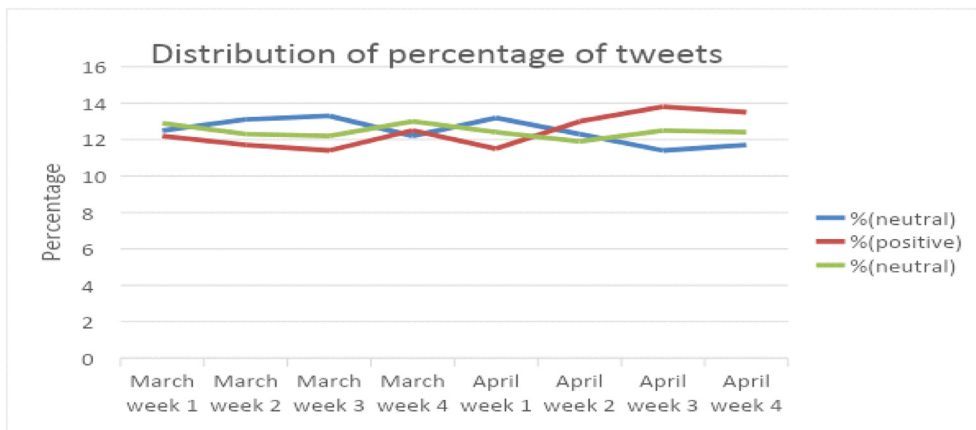
Our study has revealed an interesting fact that despite the issue being discussed is about the side effects of the vaccine, only 21.3% of the tweets about the side effects of the vaccine were of negative sentiments. However, it is still important for the Indian government and the NGOs to help educate the general public on the fears they share about the side effects of the COVID-19 vaccines on social media. Our analysis has shown that the fear of whether taking the COVID-19 vaccine will impact the efficiency in the workplace, possible death as a side effect of COVID-19 vaccines, the risk feeling in taking vaccines, the banning of Covishield in countries like Norway and whether it is good to take, fear of long term effects of the COVID-19 vaccine, whether the pre-existing conditions like Hypertension and Diabetes will elevate the side effects of the vaccine, fear of blood clot, the fear created by media, safety measures in the vaccination place and the efficiency of the vaccine.

Ever since the emergence of the COVID-19 crises, various studies



**Note:** Figure 1 depicts the number of neutral, negative and positive tweets for each week of months (March and April 2021) and their variation. For example, the 1<sup>st</sup> week of March recorded 10,696 neutral tweets.

Fig. 1. Graphical representation for Table 1 (a).



**Note:** Figure 2 illustrates the percentage of neutral, positive and negative tweets for each week and their variation. For instance, the first week of March records 12.5% (n=10,696 tweets) of total neutral tweets recorded (n=85,407).

Fig. 2. Graphical representation for Table 1 (b).

**Table 1**  
Sentiment analysis.

Month	Total Tweets	Neutral	%	Positive	%	Negative	%
March (1st week)	23736	10,696	12.5	7792	12.2	5248	12.9
March (2nd week)	23736	11,216	13.1	7512	11.7	5008	12.3
March (3rd week)	23736	11,440	13.3	7304	11.4	4992	12.2
March (4th week)	23736	10,424	12.2	8008	12.5	5304	13.0
April (1st week)	23736	11,296	13.2	7384	11.5	5056	12.4
April (2nd week)	23736	10,519	12.3	8344	13.0	4873	11.9
April (3rd week)	23736	9816	11.4	8832	13.8	5088	12.5
April (4th week)	23736	10,000	11.7	8672	13.5	5064	12.4
	189,888	85,407		63,848		40,633	

**Table 2**  
Latent Dirichlet allocation topic modeling.

Topic label	Top words
Lack of efficiency in workplace	Covid, vaccine, work, kind, effect, nature, perform
Fear of death	Effect, death, called, fuck, immune, is
Feeling of risk	Feel, risk, total, don't, vaccine, Covid
Banning of Covishield	Norway, Nordic, banned, Covishield, fake, should
Fear of long-term effects	Pandemic, vaccine, side, after, long, effect
Existing conditions	Diabetic, Covid, are, personal, BP, shit
Fear of Blood clot	vaccine, Intake, clot, blood, old, not
Fear created by media	Death, in, Vivek, act, focused, news,
Safety measures	Sanitize, booth, with, vaccine, centre, fear
Efficiency of the vaccine.	Skeptic, medicine, Covid, doubt, efficient, risk

(Note: Topic label column is labelled manually, and the 'top words' are generated using the LDA model).

were conducted in analyzing the best ways to understand the general public perception of psychological effects caused by COVID-19 and the effective ways to deal with the crisis [19–23]. With the decline of the first wave and the sudden emergence of the second wave, government officials and policymakers need to understand how the perception of the public changes regarding various aspects of the crisis at different stages of the crisis. It is important for governments all around the world to actively fasten the vaccination program to prevent the third and even possible fourth wave of the COVID-19 crisis.

4.1. Limitations of the study

We have used Tweets of the Indian population in general to determine the attitude of Indian citizens concerning the side effects of the COVID-19 vaccines. Future researchers can focus on understanding how the aspect of subculture plays a role in determining an individual's perception of COVID-19 vaccines.

Data availability statement

The datasets generated and studied during the current study are

available on reasonable request from the corresponding author.

#### Ethical statement

Not applicable.

#### Research involving human participants and/or animals

Not applicable.

#### Informed consent

Not applicable.

#### Financial disclosure

None.

#### Declaration of competing interest

Not applicable.

#### Acknowledgements

None.

#### References

- [1] Clark R. Business continuity and the pandemic threat. *IT Governance Ltd*; 2016.
- [2] Ghosh A, Nundy S, Mallick TK. How India is dealing with COVID-19 pandemic. *Sensors International* 2020;1:100021.
- [3] state-wise Vaccination. Ministry of Health and Family Welfare. Retrieved May 4, 2021. . (The data on this site changes daily).
- [4] World's largest vaccination programme begins in India on January 16. 2021. *The Hindu*. 15 January 2021. Retrieved 4 May.
- [5] Wong Q. "Twitter's user growth soars amid coronavirus, but uncertainty remains. *CNET*; 2020. [www.cnet.com/news/twitters-user-growth-soars-amid-coronavirus-but-uncertainty-remains/](http://www.cnet.com/news/twitters-user-growth-soars-amid-coronavirus-but-uncertainty-remains/).
- [6] Wang B, Zhuang J. Crisis information distribution on Twitter: a content analysis of tweets during Hurricane Sandy. *Nat Hazards* 2017;89(1):161–81.
- [7] Chatfield A, Brajawidagda U. Twitter tsunami early warning network: a social network analysis of Twitter information flows. 2012.
- [8] Earle PS, Bowden DC, Guy M. Twitter earthquake detection: earthquake monitoring in a social world. *Ann Geophys* 2012;54(6).
- [9] Buntain C, Golbeck J, Liu B, LaFree G. Evaluating public response to the Boston Marathon bombing and other acts of terrorism through Twitter. In: *Proceedings of the international AAAI conference on web and social media*. vol. 10; 2016, March. No. 1.
- [10] Soleymani M, Garcia D, Jou B, Schuller B, Chang SF, Pantic M. A survey of multimodal sentiment analysis. *Image Vis Comput* 2017;65:3–14.
- [11] Li J, Hovy E. Reflections on sentiment/opinion analysis. In: *A practical guide to sentiment analysis*. Cham: Springer; 2017. p. 41–59.
- [12] Pang B, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. 2004. arXiv preprint [cs/0409058](https://arxiv.org/abs/cs/0409058).
- [13] Hu M, Liu B. Mining and summarizing customer reviews. August. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*; 2004. p. 168–77.
- [14] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of human language technology conference and conference on empirical methods in natural language processing*; 2005, October. p. 347–54.
- [15] Agarwal A, Biadys F, Mckeown K. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In: *Proceedings of the 12th conference of the European chapter of the ACL*; 2009. p. 24–32. *EACL* 2009.
- [16] Tutorial. Quickstart. Available at: accessed, <https://textblob.readthedocs.io/en/dev/quickstart.html>. [Accessed 1 December 2019].
- [17] Blei DM, Lafferty JD. Topic models. In: *Text mining*. Chapman and Hall/CRC; 2009. p. 101–24.
- [18] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003;3:993–1022.
- [19] Press Trust of India. Recovered COVID-19 patients last immunity for 8 months raise hopes for vaccine: Study. *India Today*; 2020. <https://www.indiatoday.in/coronavirusoutbreak/story/COVID-19-antibody-immunity-lasts-8-months-study-1752290-2020-12-23>.
- [20] Praveen SV, Ittamalla R, Deepak G. Analyzing Indian general public's perspective on anxiety, stress and trauma during COVID-19—a machine learning study of 840,000 tweets. *Diabetes & Metabolic Syndrome: Clin Res Rev* 2021;15(3):667–71.
- [21] Praveen SV, Ittamalla R. General public's attitude toward governments implementing digital contact tracing to curb COVID-19—a study based on natural language processing. *Int J Pervasive Comput Commun* 2020. <https://doi.org/10.1108/ijpcc-09-2020-0121>.
- [22] Praveen SV, Ittamalla R. An analysis of attitude of general public toward COVID-19 crises—sentimental analysis and a topic modeling study. *Information Discovery and Delivery*. 2021.
- [23] Praveen SV, Ittamalla R. Psychological issues COVID-19 survivors face—a text analysis study. *J Loss Trauma* 2020:1–3.