Original Article

MODESTUM

OPEN ACCESS

Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content

Md Tarique Jamal Ansari [1]*, Naseem Ahmad Khan [2]

[1] Department of Information Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, INDIA
[2] University Institute of Engineering & Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, INDIA
*Corresponding Author: tjtjansari@gmail.com

ARTICLE INFO	ABSTRACT
	One year during the pandemic of COVID 19, numerous viable possibilities have been created in worldwide efforts to create and disseminate a viable vaccine. The rapid development of numerous vaccinations is remarkable; generally, the procedure takes 8 to 15 years. The vaccination of a critical proportion of the global population, which is vital for containing the pandemic, is now facing a new set of hurdles, including hazardous new strains of the virus, worldwide competition over a shortage of doses, as well as public suspicion about the vaccinations. A safe and efficacious vaccine COVID-19 is borne fruit globally. There are presently more than a dozen vaccinations worldwide authorized; many more continue to be developed. This paper used COVID-19 vaccine related tweets to present an overview of the public's reactions on current vaccination drives by using thematic sentiment and emotional analysis, and demographics interpretation to people. Further, experiments were carried out for sentiment analysis in order to uncover fresh information about the effect of location and gender. Overall Tweets were generally negative in tone and a huge vaccination trend can be seen in global health perspectives, as evidenced by the analysis of the role of comprehensive science and research in vaccination. **Keywords:** COVID-19, vaccination, SARS-CoV-2, vaccination refusal, NLP, Twitter

INTRODUCTION

We are seeing a significant surge in the content provided by users on the Web through sophisticated digitization, which gives views of individuals on different themes. Sentiment interpretation is a computer study that analyses the feelings and perspectives of individuals for an entity. In recent decades, the substantial study has been the focus of sentiment analysis. The research of sentiment or feeling is the calculation of opinions, feelings, and attitudes of individuals towards bodies such as brands, services, problems, events, themes, and their attributes [1].

The machine learning capability has demonstrated many amazing accomplishments in numerous areas. Natural language processing (NLP) is not the only area in which machine learning has shown not totally, but at least partially, that generic machine intelligence is capable to achieve great results for really complex tasks [2-5]. NLP is not a recent area and neither is machine learning now. However, the merger of the two areas is quite current and simply wants to improve. Analysis of sentiment is a key subject in the NLP area. Due to its usefulness and the range of financial issues it resolves and was able to address, it has quickly become one of the popular research topics in the area. Sentiment Analysis provides a broad array of application cases from automated systems to content regulation. Systems of AI that can understand emotion and sentiment in many industries have many uses.

Consequently, the building of emotionally artificial intelligence is gaining attention [6,7]. It assists in constructing the plans in marketing firms to facilitate public sentiments for items or brands, how individuals engage in promotions or releases of items and why individuals do not buy things. It is used in politics to monitor political views, identify consistency and contradiction between statements as well as government decisions. It can also be used to forecast the results of elections. Sentimental analysis can also be used to track and evaluate social occurrences, identify potentially problematic events, and determine the blogosphere's overall sentiment.

Sentiment analysis has now become a prominent and informative research approach for businesses in a variety of industries in recent years. This approach has been intensified in the electronic era, where customer views and attitudes are apparent all over the internet but frequently unorganised. Patients have strong feelings about the medical treatment they experience. Almost every meeting with a doctor or a hospital would elicit a response, whether favourable or bad. As a consequence, sentiment analysis in the healthcare field is extremely useful. Healthcare professionals can use the information obtained from analysing patient sentiments to overcome the communication barrier between hospitals and patients. This gives them the ability to improve the patient satisfaction and organisational outcomes on a broader scale. Sentiment analysis is based on information gathered from the web and numerous social media platforms. Massive volumes of data are taken from numerous sources, including mobile
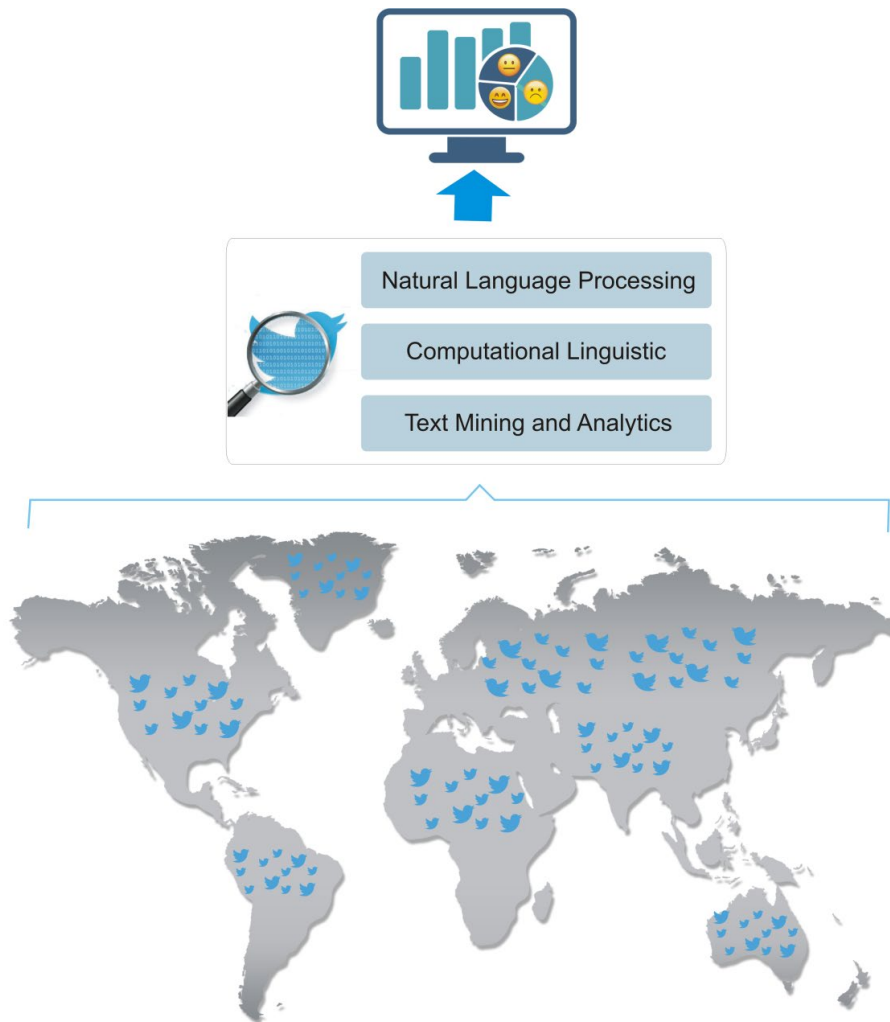
Figure 1. Sentiment analysis through worldwide twitter content

devices and online browsers, as well as maintained in numerous formats, due to the advent of social networking sites.

Twitter enables companies to contact consumers privately. There are so many Twitter statistics, although, that marketers can find it harder to give priority to references that may affect their firm. For that reason, sentiment analysis has become a significant tool in digital marketing tactics that automatically monitors emotional discussion on social networking sites [8-10].

The battle about vaccination advancements, availability, effectiveness, and side effects is continuing, and it pervades media headlines and Twitter domains daily. As internet users, though, the visibility is restricted. As a result, the purpose of this research is to broaden the perspective on the situation of the worldwide epidemic through the use of Twitter data. It would be practically difficult for humans to read and comprehend anything that has been tweeted regarding COVID-19 vaccinations. However, we may look into an incredibly complicated and wide-ranging issue using NLP tools such as textual series of information, sentiment analysis, as well as word cloud visualizations. This research work makes utilization of the COVID-19 Vaccines Tweets. This research study was accompanied using Tweepy, a Python tool that allows users to browse the Twitter API after correctly creating a Twitter Developer profile and obtaining access credentials.

The remaining part of the work is divided into the sections listed below. Section *Related Works* discusses similar studies in the field of sentiment analysis of Twitter data. Section *Data Retrieval and Methodology* discusses the proposed methodology for sentiment analysis. Section *Results and Discussion* delves into the details of data collection and sentiment analysis, as well as the research performed on the data-set utilizing sentiment analyzers with classifiers. Section *Limitations* addresses the limitations and acquired levels of accuracy. Section *Conclusion* draws this research work to a conclusion.

RELATED WORKS

Sentiment Analysis is an area of text mining research that is still evolving. The computerized processing of text's viewpoints, feelings, and objectivity is considered in this type of research. The study of recognising and compiling subjective statements or other non-factual representations of textual contents defining people's ideas, sentiments, or feelings employing medical blog writings has accelerated in recent years. In essence, several previous works based on sentiment analysis are presented below.

Chakraborty et al. [11] highlighted the inability of tweets with all accounts associated with COVID-19 and World Health Organization (WHO) in directing people throughout this

pandemic crisis. Two types of tweets collected during pandemics are discussed. In the first case, around 23,000 re-tweeted tweets were favour throughout the period 1 Jan 2019 – 23 March 2020, with the finding that the max 35 tweets show neutral or negative feelings. The studies show that while the majority of people in the 40 population tweeted positively concerning COVID-19, nets were busy re-tweeting bad tweets as well as WordCloud, and calculations with the word frequency in tweets did not find relevant words.

Praveen et al. [12] used machine-learning approaches to determine whether or not there has been a shift in general public opinion to the digital tracking of contacts in different months of crisis and to gain knowledge of general public feelings towards contact tracking. This research also favours the public's important questions concerning digital disease surveillance.

Chen et al. [13] proposed a new Twitter sentiment analysis methodology with a lot of emoji emphasis. First of all, they learned two-way embedding under positively and negatively Twitter posts, and afterward, they learned to classify the sentiments using the two-sense emoji embeddings using a long short-term memory (LSTM) network based on interest. They showed that the two-way integrating is efficient at identifying emojis' feel-conscious embedding and outperforming cutting edge favoured.

Reddy and Reddy [14] categorized tweets into positive and negatively feeling, but they used dispersed words and phrases to categorize tweets rather than standard approaches or pre-processing text information. They utilized LSTM, Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN). LSTM and CNN are used to spread visual words whereas the ANN is utilized for spread sentence presentation. They also proposed the finest and best techniques to produce spread sentiment lexicons for analysis of sentiments using the current approaches.

Hasan et al. [15] proposed a hybrid technique involving a sentiment analysis and machine learning. Furthermore, they also compared sentiment methodological approaches in the analysis of political beliefs using supervised machine-learning procedures such as Naïve Bayes as well as support vector machines (SVM).

Xue et al. [16] discussed how Twitter users talk about COVID-19 and how they feel about it. They used machine learning approaches to evaluate Tweets about coronavirus. A maximum of 11 salient important subjects are identified and then classified into ten subtexts. Their Sentiment Analysis results indicate that fear about the coronavirus's uncertain nature dominates all issues. They also examined the study's significance and shortcomings.

Sanders et al. [17] investigated a collection of over one million tweets gathered from March to July 2020 to depict popular attitudes regarding mask use during the COVID-19 outbreak. They used NLP, segmentation, and sentiment analysis methods to arrange tweets about mask-wearing under high-level patterns, then use automatic text synthesis to convey narratives for every category. They investigated topic clustering depending on mask-related Twitter data, which provides revealing insights about public opinions of COVID-19 and preventative strategies. They also observed a great spike in the amount and intensity of mask-related tweets. Furthermore, the reported analysis pipeline can be used by the medical community to conduct qualitative assessments of public reaction to health intervention strategies in real-time.

Gupta et al. [18] addressed mining Indian residents' sentiments regarding the Indian leadership's statewide lockdown imposed to minimize the pace of Coronavirus transmission. The sentiment analysis of tweets provided by Indian citizens was conducted in this work utilizing NLP and machine learning models. Data was gathered from Twitter with the help of the Tweepy API, with the TextBlob as well as Valence Aware Dictionary for Sentiment Reasoning (VADER) lexicons, and formatted with the Python NLP tools. The data were classified using eight distinct classifiers. Using the LinearSVC classifier and unigrams, the study attained the maximum accuracy of 84.4 percent. According to the findings of this study, the majority of Indian residents accept the Indian government's decision to impose a lockdown amid the corona outbreak.

Das and Kolya [19] proposed a novel strategy for attaining sentiment assessment accuracy on posts on Twitter concerning Coronavirus and future case increase forecasting with the help of a deep neural network. They build a big tweet collection just from Coronavirus tweets. They also divided the data into the training set and test set. Further, they performed polarity categorization and predictive analytics in parallel. Furthermore, they presented a statistically informed forecast for the growth of Coronavirus cases in the future.

Gambino et al. [20] addressed numerous samples of opinions produced by social media users to investigate Sentiment Analysis tasks such as Sentiment Polarity as well as Emotion Mining. The majority of these datasets are centered on the writers' point of view; that is, the post posted by a user is evaluated to ascertain the stated attitude on it. They presented a dataset centered on the users' points of view. The produced database containing news stories from three newspapers as well as the spread of six specified emotions experienced by Twitter readers of the articles. This dataset was created to investigate how six emotions are conveyed by Twitter users following reading the news report. They also presented some findings from a machine learning algorithm that was used to forecast the dispersion of emotions in previously unknown news stories.

The news and discussion about COVID-19 vaccines are making their way around social media platforms. As a result, throughout numerous outbreak-related occurrences, these social media channels are witnessing and conveying a variety of views, ideas, and feelings about vaccines. This huge amount of data are an excellent resource for computational scientists and researchers in studying people's reactions to COVID-19 vaccines. As a result, analyzing these feelings will offer astonishing results. There is a need for a study that looked at COVID-19 vaccines responses using sentiment analysis. This paper used COVID-19 vaccine-related tweets to present an overview of the public's reactions to current vaccination drives by using thematic sentiment and emotional analysis, and demographics interpretation to people. Further, experiments were carried out for sentiment analysis in order to uncover fresh information about the effect of location and gender.

DATA RETRIEVAL AND METHODOLOGY

Twitter is considered as a great data mine [21-23]. Usually, every user's tweets are public and comprehensive, unlike some

other social media. This is a major advantage in attempting to collect a large quantity of data for the sentiment analysis. Twitter data is quite particular, too. Twitter's API enables users to make complicated searches, such as retrieving a particular topic in every tweet over the last twenty minutes or extracting non-retreated tweets from a specified user. An analysis of how an enterprise is received by the broader public may be a straightforward application of this. One can pick up the last 2000 tweets based on any keyword that references the research objective and use a sentiment analysis program. Twitter data can be a wide gateway to the broader public's thoughts and how it is treated. This can deliver effective outcomes together with the accessibility and reasonable rate restriction of Twitter's API. For information gathering from Twitter API, a Tweepy Python package was used. Tweepy is a Python module open-source that allows users to utilize Python's Twitter API extremely conveniently. Tweepy offers several Twitter models as well as API endpoints classes and methods, handling transparently different optimization data, like data encoding and decoding, HTTP requests, outcome pagination, Oauth validation, rate limitation, and streaming. The Twitter API allows leverage of nearly all of the features via Tweepy. Through keyword search, hashtag, timing, trend, or geo-localization, Tweepy offers the proper data recovery [24-26]. This research is, therefore, aimed at data gathering all across the globe. Although there are several Twitter API constraints, successive efforts have been undertaken to obtain as many posts as feasible. In Google Colab, all the tasks were done. Google Colab is a freely accessible cloud service for Free GPU [27]. We have developed deep learning coding with popular libraries including Keras, TensorFlow, PyTorch, OpenCV, and others. Twitter API was built using the properties and supplied the credentials into Tweepy when the library was imported.

Subsequently, other categories were created. During data processing the non-English keyword data collection, stop words, @, #, RT, emojis are eliminated and the phrase is tacit. The tweets were therefore sorted by the COVID-19, CORONAVIRUS, CORONA, Vaccine, Vaccination, COVID-19 Vaccines, SARS-CoV-2, Vaccination Refusal, Pandemic, Vaccination drive keywords. The data collected was stored in CSV file format and processed with Textblob, which was used in the Sentiment Analysis package.

The TextBlob library was invoked after the data was available to perform Sentiment Analysis. TextBlob is popular text data processing package for Python (2 and 3). It offers a simple API for digging into standard NLP activities including speaking part, extracting noun sentences, feelings, classification, interpretation, and more [28]. The data gathering focuses mostly on Twitter data as well as timestamps. Sentiment analysis has a defined technique, beginning with the collection of data and subsequently the identification of information. In the phase of feeling polarity and subjectivity, the ultimate decision is made. TextBlob uses the Naïve Bayes (NB) categorization model following the TextBlob documentation [29]. NB classification has been taught to recognize the valence of collected tweets on NLTK (Natural Language ToolKit).

NB is the statistical technique using the Bayes theorem to calculate the dispersion of feelings over the dataset. But NB analyses any text into a bag of phrases meaning that words' placements are not taken into consideration. The equation of Bayes to determine the probability of feeling is:

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(\text{features}|\text{label})}{P(\text{features})}$$

Where P (label) is the previous possibility of a label, the previous possibility P (features |label) is that a particular set of features is categorized as a label, while P(features) the probability distribution is that a particular feature set is established. After this function, the sentiments were categorized as "positive," "negative" and "neutral". **Figure 2** shows the proposed framework used in this paper for COVID-19 vaccines sentiment analysis through Twitter content.
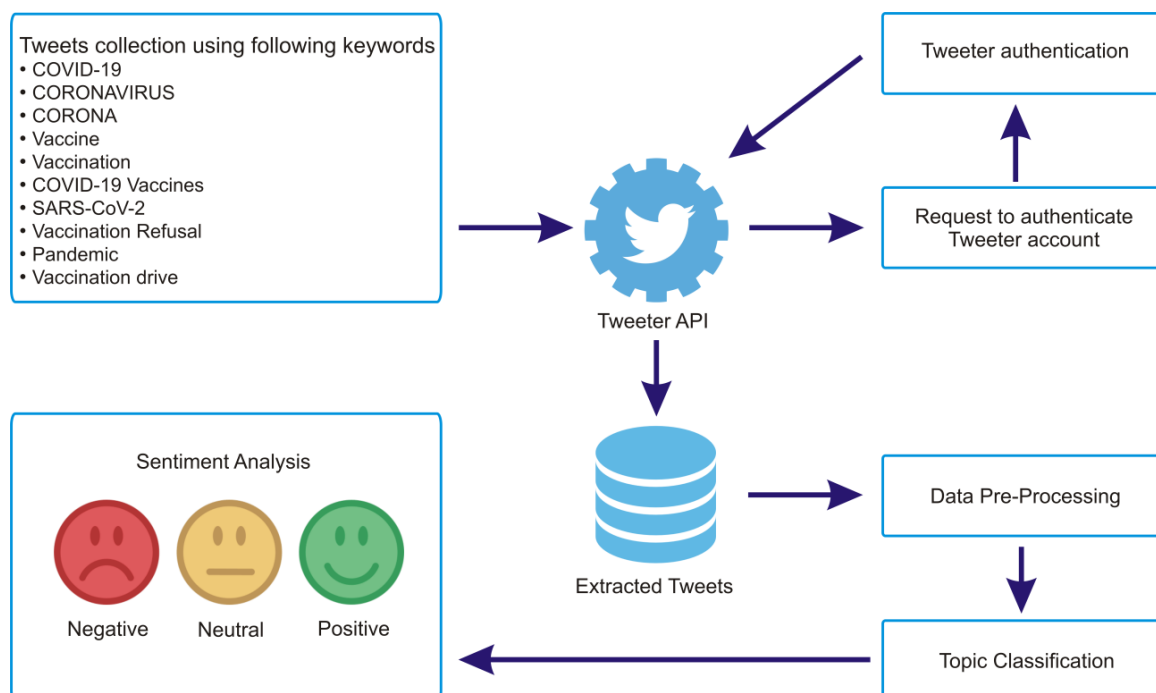

Figure 2. Framework for COVID-19 vaccines sentiment analysis through Twitter content
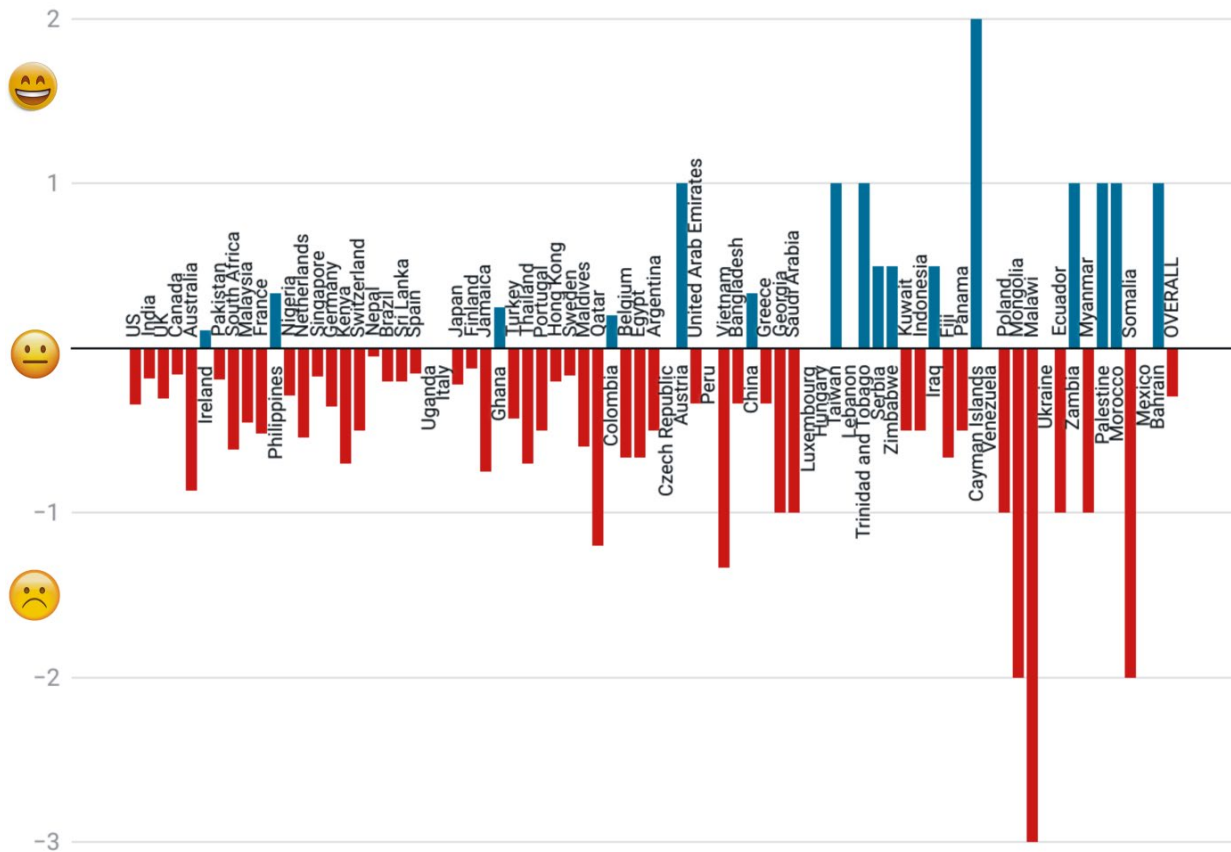
Figure 3. Country wise COVID-19 vaccines sentiment analysis

Let $S = (s_1, s_2)$ be the gender class, and $F = (f_1, f_2, \ldots .f_n)$ are features, according to Bayes's theorem:

$$P(s|F) = \frac{P(s)P(F|s)}{P(F)}$$

The Naïve Bayes assumption is that:

$$\hat{P}(F|s) = \prod_{i=1}^{n} \hat{P}(fi|s)$$

The core assumption is that every feature remains distinct. This assumption is very powerful, and it also works extremely well.

RESULTS AND DISCUSSION

From May 15, 2021, to June 25, 2021, a significant proportion of 820,000 tweets were gathered. A large percentage of people on Twitter needed to know if the available vaccine could prevent the growth of COVID-19. The surge at the beginning of May, on the other hand, was due to the news of accomplishment in the deployment of numerous vaccinations in the world. The study's findings are presented in two stages. The sentiments of tweets from around the world are addressed in the first phase. **Figure 3** depicts the sentiments of tweets made by persons all around the world for which the research was undertaken. As per the findings shown in **Figure 3**, the tweets from Cayman Island, Austria, Taiwan, Trinidad & Tobago, Zambia, Palestine, Morocco, and Bahrain represented the most positive sentiments. Uganda, Italy, the Czech Republic, Peru, Luxembourg, Hungary, Lebanon, Venezuela, Ukraine, and Mexico, on the other hand, had about balanced positive and negative attitudes. Malawi, on the other hand, had a higher proportion of people tweeting with favourable views. According to Voice of America News, Malawi officials say the country is swiftly running out of coronavirus vaccinations as confirmed infections reach over 35,000 and 1,200 fatalities in a third wave of the pandemic. The shortage came only weeks after Malawi trashed over 20,000 out-dated doses, which were caused in part by vaccine reluctance [30]. **Table 1** and **Figure 3** shows the worldwide sentiment findings for COVID-19 vaccination.

Table 1. Worldwide sentiment findings for COVID-19 vaccination

Country	Positive Sentiment	Negative Sentiment	Average Sentiment
US	1.5049	1.8464	-0.3415
India	1.6059	1.7919	-0.1860
UK	1.6146	1.9194	-0.3048
Canada	1.5741	1.7315	-0.1574
Australia	1.5385	2.4038	-0.8654
Ireland	1.6389	1.5278	0.1111
Pakistan	1.5135	1.7027	-0.1892
South Africa	1.4839	2.0968	-0.6129
Malaysia	1.5161	1.9677	-0.4516
France	1.2963	1.8148	-0.5185

Table 1 (continued). Worldwide sentiment findings for COVID-19 vaccination

Country	Positive Sentiment	Negative Sentiment	Average Sentiment
Philippines	1.7500	1.4167	0.3333
Nigeria	1.3810	1.6667	-0.2857
Netherlands	1.4545	2.0000	-0.5455
Singapore	1.4348	1.6087	-0.1739
Germany	1.7059	2.0588	-0.3529
Kenya	1.3500	2.0500	-0.7000
Switzerland	1.7143	2.2143	-0.5000
Nepal	1.7368	1.7895	-0.0526
Brazil	1.4000	1.6000	-0.2000
Sri Lanka	1.6000	1.8000	-0.2000
Spain	1.6923	1.8462	-0.1538
Uganda	1.7000	1.7000	0.0000
Italy	1.3333	1.3333	0.0000
Japan	1.6667	1.8889	-0.2222
Finland	1.6250	1.7500	-0.1250
Jamaica	1.0000	1.7500	-0.7500
Ghana	1.3750	1.1250	0.2500
Turkey	1.4286	1.8571	-0.4286
Thailand	1.2000	1.9000	-0.7000
Portugal	1.3750	1.8750	-0.5000
Hong Kong	1.7000	1.9000	-0.2000
Sweden	1.8333	2.0000	-0.1667
Maldives	1.8000	2.4000	-0.6000
Qatar	1.0000	2.2000	-1.2000
Colombia	1.6000	1.4000	0.2000
Belgium	1.3333	2.0000	-0.6667
Egypt	1.0000	1.6667	-0.6667
Argentina	1.7500	2.2500	-0.5000
Czech Republic	2.0000	2.0000	0.0000
Austria	2.0000	1.0000	1.0000
United Arab Emirates	1.0000	1.3333	-0.3333
Peru	1.6667	1.6667	0.0000
Vietnam	1.6667	3.0000	-1.3333
Bangladesh	2.0000	2.3333	-0.3333
China	1.6667	1.3333	0.3333
Greece	1.0000	1.3333	-0.3333
Georgia	1.0000	2.0000	-1.0000
Saudi Arabia	1.5000	2.5000	-1.0000
Luxembourg	1.0000	1.0000	0.0000
Hungary	1.0000	1.0000	0.0000
Taiwan	2.0000	1.0000	1.0000
Lebanon	1.5000	1.5000	0.0000
Trinidad and Tobago	2.0000	1.0000	1.0000
Serbia	1.5000	1.0000	0.5000
Zimbabwe	3.0000	2.5000	0.5000
Kuwait	1.5000	2.0000	-0.5000
Indonesia	1.0000	1.5000	-0.5000
Iraq	1.5000	1.0000	0.5000
Fiji	1.0000	1.6667	-0.6667
Panama	2.0000	2.5000	-0.5000
Cayman Islands	3.0000	1.0000	2.0000
Venezuela	1.0000	1.0000	0.0000
Poland	1.0000	2.0000	-1.0000
Mongolia	2.0000	4.0000	-2.0000
Malawi	1.0000	4.0000	-3.0000
Ukraine	1.0000	1.0000	0.0000
Ecuador	1.0000	2.0000	-1.0000
Zambia	2.0000	1.0000	1.0000
Myanmar	1.0000	1.0000	-1.0000
Palestine	2.0000	1.0000	1.0000
Morocco	2.0000	1.0000	1.0000
Somalia	1.0000	3.0000	-2.0000
Mexico	1.0000	1.0000	0.0000
Bahrain	2.0000	1.0000	1.0000
OVERALL	1.5610	1.8550	-0.2940

Figure 4 shows the overall global findings of COVID-19 vaccination sentiment analysis. To collect data such as user emotions, attitudes and values in all commercial sectors categorization and analysis as per demographic information like gender age as well as geographical location is crucial. Especially because the circumstances and the requirement can vary by gender, it is vital to determine the gender for comprehensive policy-making. This could also enable users to see which issues men and women discuss more and which services men and women prefer and reject [31,32]. Knowledge of this information is vital for sentiment analysis research as it can be utilised to improve the process and personalization. **Figure 5** shows male gender-based global findings of COVID-19 vaccination sentiment analysis. **Figure 6** shows the female gender-based global findings of COVID-19 vaccination sentiment analysis. Comparing to both findings male very strong positive sentiment is 0.22% and very strong negative sentiment is 0.33% while in case of female gender the very strong positive sentiment is 0.00% and very strong negative sentiment is 0.81%.

In the second phase of the discussion, the tweets were organized into word clouds to analyze what words have been frequently used by twitters users of different countries and also, what emotions were behind these words. As it can be seen in **Figure 7**, words like vaccine, covid-19, pandemic, vaccination, corona, sars-cov-2, virus, India, health, need, government, dose, death, against, vaccinated were very repeatedly used by the Twitter users of every country. Considering that COVID19 started in Wuhan, China, China received a significant number of references in many of these tweets. The words Largest Vaccination Drive used more extensively by Indians, but these words were related to the feelings of Aggression and Happiness. This tendency in India could be connected to the fact that India's vaccination drive is the largest vaccination campaign in the world. Furthermore, there were a significant number of references of a political figure throughout all of the tweets evaluated from various countries.

LIMITATIONS

COVID-19 is a persistent worldwide problem, and the vaccination debate is likewise worldwide. The virus outbreak has demonstrated how interconnected the globe has become, as well as vaccination has therefore become a worldwide problem which a nation cannot attain a specific level of vaccination among its citizenry, it faces a significant risk of disruption as well as virus mutation; as a result, it will be challenging for the country to reclaim its positions in the international economy, as well as global cooperation, will be required to overcome the deadly virus. As a result, the pandemic's economic consequences and vaccine research are critical concerns [33-35]. The current study included sentiment and opinion analysis of huge tweets about COVID-19 vaccines. The Twitter interface utilized in this study could be a beneficial strategy for public health awareness reinforcing vaccine acceptance while decreasing vaccine hesitancy and resistance. Ultimately, identifying vaccination feelings and perspectives can assist public health officials in reinforcing good judgment and remarks within positive postings while refuting aggressive language spreading false information within unfavorable posts. Furthermore, public health officials may be prepared to use Twitter and other media channels to raise positive
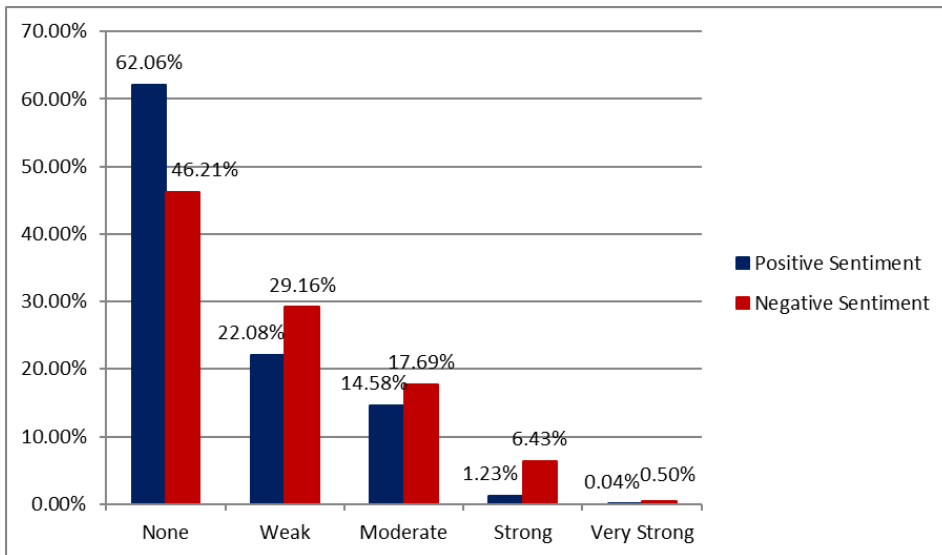
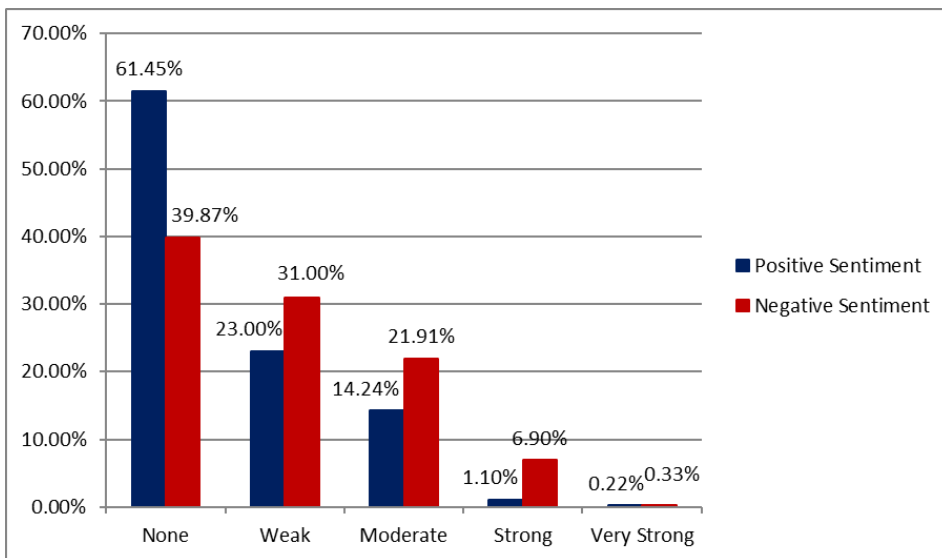**Figure 4.** Overall global findings of COVID-19 vaccination sentiment analysis

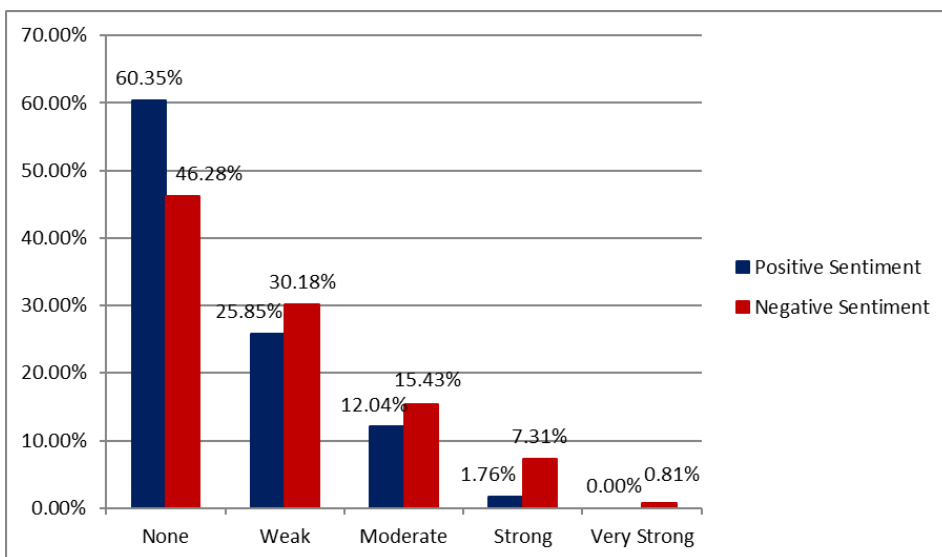**Figure 5.** Male gender-based global findings of COVID-19 vaccination sentiment analysis

Figure 6. Female gender-based global findings of COVID-19 vaccination sentiment analysis
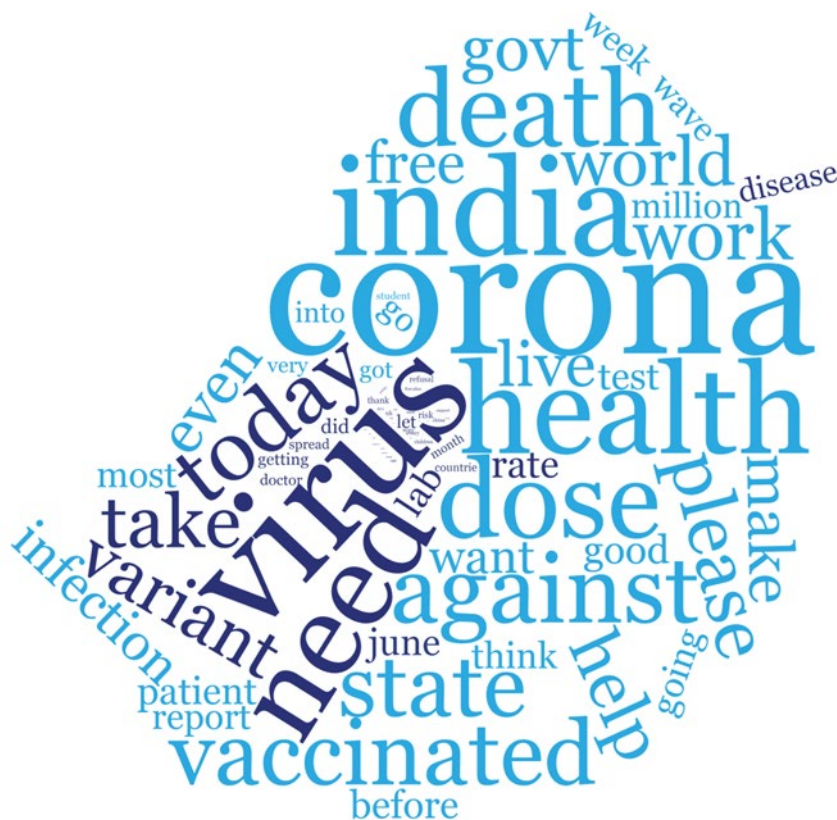
Figure 7. Top trending words over Twitter amid COVID-19

messaging, minimize negative and oppositional messages, and pro-actively ban anti-vaccination profiles such as bots to promote and improve vaccination coverage. Several limitations hampered this research. To begin, we acknowledge that the data collection does not include all tweets on the COVID-19 vaccination. To assure the accuracy of recorded tweets for this preliminary work, this paper used a limited Twitter search strategy that excluded phrases like "shot(s)," "immunization," and "infection." Furthermore, given the number of tweets evaluated, Twitter's powerful search feature limits us to only a meaningful proportion of all tweets. Second, this research used innovative systems to analyze attitudes and emotions in tweets that were not about health care, which may have affected our research. Third, tweets about the COVID-19 vaccination might have been reported or deleted by Twitter for providing inaccurate information, but this research didn't have access to that specific situation to see how it might have impacted our data collection. Ultimately, because this research paper only focused on tweets in English and therefore unable to ascertain users' local language other the English, this research constrained in concluding specific tweets which are not in the English language.

CONCLUSION

Ultimately, the findings of this investigation are encouraging. Although there are many social networking conversations is about vaccination. Many of the mainstream discourses are about whether vaccination helps protect lives and is medically beneficial and safe. Vaccination, according to popular belief, prevents disease. This study emphasizes this significant discovery. Based on the present age of the internet and social media, industry and public officials around the world must involve the public proactively and continuously in risk awareness and interaction, guided by monitoring social media conversation and sentiment. This study demonstrates a favourable trend in global health perspectives, as evidenced by the analysis of the role of comprehensive science and research in vaccination. Tweets were generally negative in tone, with a rising lack of trust. Fear may have remained as the prevailing feeling, increasing concerns about the desire to get the COVID-19 vaccination, and clusters of negative sentiment evolved as a result. The application of new technologies like machine learning, NLP, sentiment analysis, as well as text analysis may speed the development of understanding popular sentiment in the scenario of the pandemic. In the future, such NLP capabilities would be used to participate in customized communications depending on user preferences and sentiments.

REFERENCES

1. Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. 2015; In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432). https://doi.org/10.18653/v1/d15-1167   PMid:25656747 PMCid:PMC5706456

2. Dzikovska M, Steinhauser N, Farrow E, Moore J, Campbell G. BEETLE II: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. Int J Artif Intell Edu. 2014;24(3):284-332. https://doi.org/10.1007/s40593-014-0017-9

3. Allan J, Aslam J, Belkin N, Buckley C, Callan J, Croft B, Zhai C. Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, university of massachusetts Amherst. 2002. In ACM SIGIR Forum (Vol. 37, No. 1, pp. 31-47). New York, NY, USA: ACM. https://doi.org/10.1145/945546.945549

4. Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy. IEEE Trans Pat Anal Mach Intell. 2018;41(2):423-43. https://doi.org/10.1109/TPAMI.2018.2798607 PMid:29994351

5. Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Trans Pat Anal Mach Intell. 2013;35(8):1798-828. https://doi.org/10.1109/TPAMI.2013.50 PMid:23787338

6. Hardeniya N, Perkins J, Chopra D, Joshi N, Mathur I. Natural language processing: python and NLTK. Packt Publishing Ltd. 2016.

7. Jackson P, Moulinier I. Natural language processing for online applications: Text retrieval, extraction and categorization (Vol. 5). John Benjamins Publishing. 2007. https://doi.org/10.1075/nlp.5(1st)

8. Kwon ES, Sung Y. Follow me! Global marketers' Twitter use. Journal of Interactive Advertising. 2011;12(1):4-16. https://doi.org/10.1080/15252019.2011.10722187

9. Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Generation Computer Systems. 2021;115:279-94. https://doi.org/10.1016/j.future.2020.08.005

10. Oueslati O, Cambria E, HajHmida MB, Ounelli H. A review of sentiment analysis research in Arabic language. Fut Gen Comp Sys. 2020;112:408-30. https://doi.org/10.1016/j.future.2020.05.034

11. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Appl Soft Comp. 2020;97:106754. https://doi.org/10.1016/j.asoc.2020.106754 PMid:33013254 PMCid:PMC7521435

12. 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 Per Comp Comm. 2020. https://doi.org/10.1108/IJPCC-09-2020-0121

13. Chen Y, Yuan J, You Q, Luo J. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In Proceedings of the 26th ACM international conference on Multimedia. 2018 (pp. 117-125). https://doi.org/10.1145/3240508.3240533

14. Reddy DM, Reddy DN. Twitter Sentiment Analysis using Distributed Word and Sentence Representation. arXiv preprint arXiv:1904.12580. 2019.

15. Hasan A, Moin S, Karim A, Shamshirband S. Machine learning-based sentiment analysis for twitter accounts. Math Comp App. 2018;23(1):11. https://doi.org/10.3390/mca23010011

16. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PloS one. 2020;15(9):e0239441. https://doi.org/10.1371/journal.pone.0239441 PMid:32976519 PMCid:PMC7518625

17. Sanders AC, White RC, Severson LS, Ma R, McQueen R, Paulo HCA, Bennett KP. Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. medRxiv, 2020-08. 2021. https://doi.org/10.1101/2020.08.28.20183863

18. Gupta P, Kumar S, Suman RR, Kumar V. Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter. IEEE Trans Comp Soc Sys. 2020. https://doi.org/10.1109/TCSS.2020.3042446

19. Das S, Kolya AK. Predicting the pandemic: sentiment evaluation and predictive analysis from large-scale tweets on Covid-19 by deep convolutional neural network. Evol Intell. 2021:1-22. https://doi.org/10.1007/s12065-021-00598-7 PMid:33815622 PMCid:PMC8007226

20. Gambino OJ, Calvo H, García-Mendoza CV. Distribution of emotional reactions to news articles in twitter. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 2018.

21. Ansari MTJ, Baz A, Alhakami H, Alhakami W, Kumar R, Khan RA. P-STORE: Extension of STORE methodology to elicit privacy requirements. Arabian J Sci Eng. 2021;46:8287-310. https://doi.org/10.1007/s13369-021-05476-z

22. Arias M, Arratia A, Xuriguera R. Forecasting with twitter data. ACM Transactions on Intelligent Systems and Technology (TIST). 2014;5(1):1-24. https://doi.org/10.1145/2542182.2542190

23. Bing L, Chan KC, Ou C. Public sentiment analysis in Twitter data for prediction of a company's stock price movements. In 2014 IEEE 11th International Conference on e-Business Engineering (pp. 232-239). IEEE. https://doi.org/10.1109/ICEBE.2014.47

24. Salhi DE, Tari A, Kechadi MT. Using e-reputation for sentiment analysis: Twitter as a case study. Int J Cloud App Comp (IJCAC). 2021;11(2):32-47. https://doi.org/10.4018/IJCAC.2021040103

25. W Alhakami, Binmahfoudh A, Baz A, Alhakami H, Ansari MTJ, Khan RA. Atrocious Impinging of COVID-19 Pandemic on Software Development Industries. Comp Sys Sci Eng. 2021;36(2):323-338. https://doi.org/10.32604/csse.2021.014929

26. Sui Z. Hierarchical text topic modeling with applications in social media-enabled cyber maintenance decision analysis and quality hypothesis generation (Doctoral dissertation, The Ohio State University). 2017.

27. Canesche M, Bragança L, Neto OPV, Nacif JA, Ferreira R. Google Colab CAD4U: Hands-on cloud laboratories for digital design. In 2021 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1-5). IEEE. https://doi.org/10.1109/ISCAS51556.2021.9401151

28. Kunal S, Saha A, Varma A, Tiwari V. Textual dissection of live Twitter reviews using naive Bayes. Procedia Comp Sci. 2018;132:07-313. https://doi.org/10.1016/j.procs.2018.05.182

29. Loria S. textblob Documentation. Release 0.15, 2. 2018.

30. Masina L. Malawi facing COVID vaccine shortage after burning expired doses Lameck Masina. Voice of America. 2021. Available at: https://www.voanews.com/covid-19-pandemic/malawi-facing-covid-vaccine-shortage-after-burning-expired-doses

31. Sullivan CF. Gendered cybersupport: A thematic analysis of two online cancer support groups. Journal of health psychology. 2003;8(1):83-104. https://doi.org/10.1177/1359105303008001446 PMid:22113903

32. Park S, Woo J. Gender classification using sentiment analysis and deep learning in a health Web forum. Applied Sciences. 2019;9(6):1249. https://doi.org/10.3390/app9061249

33. Lyu JC, Le Han E, Luli GK. COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis. Journal of Medical Internet Res. 2021, 23(6): e24435. https://doi.org/10.2196/24435 PMid:34115608 PMCid:PMC8244724

34. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. Journal of communication in healthcare. 2021;14(1):12-9. https://doi.org/10.1080/17538068.2020.1858222

35. Marcec R, Likic R. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioTech and Moderna COVID-19 vaccines. Postgraduate Medical Journal. 2021, https://doi.org/10.1136/postgradmedj-2021-140685 PMid:34373343 PMCid:PMC8354810