

Temporal analysis and opinion dynamics of COVID-19 vaccination tweets using diverse feature engineering techniques

Shoaib Ahmed¹, Dost Muhammad Khan¹, Saima Sadiq², Muhammad Umer¹, Faisal Shahzad¹, Khalid Mahmood³, Heba Mohsen⁴ and Imran Ashraf⁵

¹ Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

² Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan

³ ICT, Gomal University, Dera Ismail Khan, Pakistan

⁴ Computer Science Department, Future University in Egypt, New Cairo, Egypt

⁵ Information and Communication Engineering, Yeungnam University, Gyeongsan si, South Korea

ABSTRACT

The outbreak of the COVID-19 pandemic has also triggered a tsunami of news, instructions, and precautionary measures related to the disease on social media platforms. Despite the considerable support on social media, a large number of fake propaganda and conspiracies are also circulated. People also reacted to COVID-19 vaccination on social media and expressed their opinions, perceptions, and conceptions. The present research work aims to explore the opinion dynamics of the general public about COVID-19 vaccination to help the administration authorities to devise policies to increase vaccination acceptance. For this purpose, a framework is proposed to perform sentiment analysis of COVID-19 vaccination-related tweets. The influence of term frequency-inverse document frequency, bag of words (BoW), Word2Vec, and combination of TF-IDF and BoW are explored with classifiers including random forest, gradient boosting machine, extra tree classifier (ETC), logistic regression, Naïve Bayes, stochastic gradient descent, multilayer perceptron, convolutional neural network (CNN), bidirectional encoder representations from transformers (BERT), long short-term memory (LSTM), and recurrent neural network (RNN). Results reveal that ETC outperforms using BoW with a 92% of accuracy and is the most suitable approach for sentiment analysis of COVID-19-related tweets. Opinion dynamics show that sentiments in favor of vaccination have increased over time.

Submitted 27 September 2022

Accepted 28 November 2022

Published 10 March 2023

Corresponding authors

Dost Muhammad Khan,

khan.dostkhan@iub.edu.pk

Imran Ashraf,

imranashraf@ynu.ac.kr

Academic editor

Ali Kashif Bashir

Additional Information and
Declarations can be found on
page 23

DOI 10.7717/peerj-cs.1190

© Copyright

2023 Ahmed et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Artificial Intelligence, Data Mining and Machine Learning, Text Mining, Sentiment Analysis

Keywords COVID-19 vaccination, Sentiment analysis, Machine learning, Feature engineering

INTRODUCTION

The COVID-19 outbreak changed the lives of people economically and socially. The global pandemic caused life-threatening fears and anxiety in public and many of such concerns have been shared on social media platforms. Social media platforms continuously spread the latest information globally about viruses and provide help to deal with this situation.

As reported in Statista, 2.9 billion people used social media in 2019 and the number of visitors is expected to exceed 3.4 billion by 2023 (Dixon (2023)). Various surveys have been performed by researchers to observe the social media trends which show the high influence of social media platforms for sharing news and stories. Ofcom (2019) confirms the wide use of social media for news and updates by adults. Similarly, different departments and disease control institutes such as the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) use social networks for providing the latest updates and guidelines about pandemic emergencies. Quarantine, lockdown, and social distancing constraints intensified the use of social platforms globally (Hiscott et al., 2020). Individuals express their emotions and opinions during the different events on this rapidly growing platform (Alamoodi et al., 2020). People rely on updates from social media which makes it an influential channel of communication around the globe.

Despite the appropriate and controlled use of social media by WHO and other public institutions, a tsunami of false information has also been produced on social media, creating a big challenge for information systems. Although the United Nations (UN) warned against the COVID-19 infodemic and spread in February 2020 (Appel et al., 2020), many malicious users deliberately spread confusion, rumors, and fake news on social media platforms. A famous fake claim that went viral in Europe was that 5G weakens the immune system and is a reason for the spread of COVID-19 and people started demolishing towers (Schumaker, Jarmoszko & Labedz, 2016). Many researchers and news reporters highlighted the infodemic issues and discussed case studies to present real information and help people avoid panic. Various posts and ads use the COVID-19 context and mislead the user to install spyware and other cyberattacks. As a result, management and government organizations focus on social media platforms to put a stop to the spread of viral fake news and misinformation. However many platforms are claiming to control this situation by banning harmful content but these platforms were not ready for such information flooding.

Twitter is a famous social media platform where people post their opinion on specific topics in text form called 'tweet' (D'Andrea et al., 2019). A tweet also contains the location information of the user, hashtags, and emoticons that help in sentiment portrayal (Giachanou & Crestani, 2016). Moreover, Twitter is used by government officials to share information about an event or an announcement for the general public (Golbeck, Grimes & Rogers, 2010). Information shared on Twitter has been used in various research works such as analyzing services (Tiwari et al., 2018), sports sentiment (Yu & Wang, 2015), political views (Khatua, Khatua & Cambria, 2020), the sentiment of cancer patients (Crannell et al., 2016) and vaccines (D'Andrea et al., 2019), etc. The subject of vaccination is currently a large debate on social media platforms with respect to questions related to its safety, immunity against the virus, side effects, etc. Studies have been conducted to analyze vaccination hesitation and the effect of social media campaigns (Pedersen et al., 2020; Loft et al., 2020). In general, people show positive and negative opinions on the efficacy of vaccines and the vaccination process itself. Analyzing such opinions from Tweets can help understand the dynamics of vaccination and devise effective policies and social media campaigns to increase vaccination acceptance by the public.

Machine learning approaches have been employed to identify misinformation on social media posts regarding COVID-19. Similarly, public opinions about the COVID-19 vaccination have been studied when two famous vaccines Pfizer and BioNTech were introduced (Cotfas et al., 2021). Mourad et al. (2020) analyzed 800 k tweets and stated that 93% of tweets are misleading about COVID-19 and from non-medical users and real doctors and medical experts contribute less than 1%. In order to develop effective tactics that might lessen anti-vaccination sentiments among various groups, research that can make use of the large amount of data created *via* social media, such as Twitter, will be able to give important information. To identify trends in vaccination tweets on Twitter, one of the first challenges in this context is to create a text categorization system. The enormous volume of data and text-based style make it a difficult process to complete. Using machine learning techniques was a successful strategy used in various research works conducted on Twitter about vaccination sentiment analysis. Likewise, a large number of Tweets are available on social media platforms which can be used to analyze public opinions about vaccination and devise policies accordingly.

The study aims at investigating the impact of different feature approaches regarding the sentiment classification of COVID-19. Although several existing works investigated and explored similar dimensions, the role of various feature engineering approaches is not very well studied. For this purpose, a dataset containing COVID-19 vaccination-related tweets has been collected and analyzed using machine learning models. In the first instance, the dataset is subdivided into five sub-datasets concerning the administered vaccines AstraZeneca, Moderna, Pfizer/BioNTech, Sinopharm, and SputnikV. Each sub-dataset is investigated separately to analyze people's sentiments and a comparison analysis is performed to discuss the trends. Furthermore, opinion dynamics and temporal analysis are also performed. This study uses a large dataset in this regard and performs sentiment analysis using Tweets on COVID-19 vaccination. This study makes the following contributions:

- A machine learning-based framework was developed for sentiment analysis of tweets related to different vaccines for COVID-19. The sentiments of people for different vaccines were analyzed using several models including random forest (RF), gradient boosting machine (GBM), extra tree classifier (ETC), logistic regression (LR), naive Bayes (NB), stochastic gradient descent (SGD), multilayer perceptron ((MLP), convolutional neural network (CNN), bidirectional encoder representations from transformers (BERT), long short-term memory (LSTM), and recurrent neural network (RNN).
- The influence of term frequency-inverse document frequency (TF-IDF), bag of words (BoW), Word2Vec, and feature union of TF-IDF and BoW was investigated regarding the accuracy of models. Since different feature engineering approaches lead to the different classification accuracy of the models, four feature engineering approaches were investigated regarding high accuracy.

- A large dataset containing Tweets on COVID-19 vaccination was used. For performance comparison, the dataset was labeled manually, as well as using TextBlob. Performance was evaluated in terms of accuracy, precision, recall, and F1-score.

The remainder of the article is arranged as follows. Section 2 discusses the related research work along with their used techniques. Section 3 presents the methods and techniques, dataset, and models used for experiments. It also illustrates the proposed framework. Section 4 provides the experimental results while discussions are provided in Section 5. In the end, Section 6 concludes the article.

RELATED WORK

Owing to the increase in the data available on social media platforms, there is a need to address various challenges regarding data shape such as information extraction by data restructuring and selection of appropriate classifiers (*Samuel, Kashyap & Betts, 2018*). Text analysis involves text visualization, exploring syntactic and semantic features, and feature extraction techniques (*Samuel, Kashyap & Betts, 2018; Rustam et al., 2020*). With the wide use of social media platforms, a large number of opinions and reviews are available on review sites, forums, blogs, *etc.* With the help of review-based prediction systems, this unstructured information can automatically be transformed into structured data of public opinions. This structured data can later be used to find the sentiments about specific applications, products, services, and brands and serves as a piece of important information for product and service refinement.

Twitter data has been widely explored by previous researchers over the years regarding topic modeling, information retrieval, product positioning, and analysis of psychological conditions. Text analysis using tweets has been performed in many types of research such as opinion mining (*Naseem et al., 2019*), aggression detection (*Sadiq et al., 2021*), content mining (*Majumdar & Bose, 2019*), and topic detection related to COVID-19 (*Garcia & Berton, 2021*). Analysis regarding COVID-19 tweets has been performed covering different perspectives such as COVID-19 detection (*Castiglione et al., 2021b*), the role of the internet of things (IoT) to control COVID-19 spread (*Castiglione et al., 2021a*), productivity analysis (*Shoukat et al., 2021*), and effect on mental health (*Sohail et al., 2021*), *etc.* A French company's customer feedback has been analyzed on approximately seventy thousand tweets in *Pépin et al. (2017)*. The authors apply frequency-based feature extraction techniques and topic modeling is done using the Latent Dirichlet Allocation (LDA) method. The authors utilized linguistic and psychological features to explore emotions in social media posts of different languages (*Jain, Kumar & Fernandes, 2017*).

Twitter data has been also used for tracking and analyzing crisis situations during epidemics (*Ye et al., 2016*). Sentiment analysis on Twitter data related to healthcare is carried out regarding the postnatal behavior or depression of new mothers to find their emotions, language style, and social involvement (*De Choudhury, Counts & Horvitz, 2013*). The authors highlighted government policies during the pandemic and performed topic modeling using multi-lingual Twitter data (*Chun et al., 2020*). Similarly, study *Garcia-Gasulla & Suzumura (2020)* analyzed the growth of sinophobia during pandemics

from Twitter data. The study concludes that depression during a pandemic is mainly caused by unemployment, fear of death, and inactive staying at home.

Researchers are exploring tweets from different perspectives using the expressed sentiments toward the COVID-19 pandemic. Tweets from twenty days of March 2020 are collected from Europe and analyzed for the impact of COVID-19 disease spread (*Alhajji et al., 2020*). The authors applied different unsupervised machine-learning models to explore COVID-19-related textual data. Tweets sentiment analysis is done using Naive Bayes with the topic modeling using the LDA in *Prabhakar Kaila & Prasad (2020)*. Similarly, TextBlob and the natural language processing toolkit (NLTK) library are used for the same purpose by *Kaur & Sharma (2020)*. The authors investigated the impact of COVID-19 symptoms on quarantine in *Pastor (2020)*.

Along the same lines, sentiments of the public in China related to COVID-19 are explored by researchers in *Han et al. (2020)*. They divided posts into general seven categories and thirteen subcategories based on topics. *Radwan & Radwan (2020)* discussed that panic caused by COVID-19 by posts on social media is inevitable and spread with more speed than COVID-19 itself. The study further states that public behavior, sentiments, and rumors need to be investigated quickly by experts to assist authorities in taking action accordingly. Similarly, the study analyzed the emotions of the general public using the data from the discussion forum to conclude that Twitter posts have the highest influence on the behavior of people (*Hanson et al., 2013*).

Lexicon-based, machine learning, and hybrid techniques are mostly used by researchers for polarity analysis. Lexicon techniques include sentiment lexicons like SentiWordnet (*Baccianella, Esuli & Sebastiani, 2010*), VaderSentiment (*Hutto & Gilbert, 2014*), and sentiment140 (*Mohammad, Kiritchenko & Zhu, 2013*) consisting of words and polarity score. The sentiment lexicons are utilized with semantic approaches, which commonly take negations and booster words into account, to accomplish polarity identification. *Hutto & Gilbert (2014)* proposed Vader, a simple rule-based model that incorporates a sentiment lexicon as well as syntactic and grammatical rules. The authors demonstrate that the suggested model performs better than a single human rater. The authors demonstrate that Vader gives a higher performance on the datasets gathered from Twitter, Amazon reviews, and NYT editorials when compared to traditional machine learning models.

In order to track the dynamics of emotions in the first few months after the public learned about COVID-19, *Kaur, Kaul & Zadeh (2020)* utilized data taken from Twitter. The IBM Watson Tone Analyzer was used to extract and analyze a total of 16,138 tweets. In all three months analyzed in the article, more negative tweets were sent than neutral or good ones, as was to be expected. Deep learning and transfer learning models have been employed for cross-domain sentiment analysis. Deep learning models' capability of transferability improves the performance and avoids overfitting (*Cao et al., 2021*). The adversarial training model is proposed in *Dai et al. (2022)* to transfer sentiments across domains. Authors applied decision boundaries in cross-domain sentiment analysis (*Fu & Liu, 2022*). Authors mine keywords and applied feature engineering techniques to explore patterns (*Asgarnezhad, Monadjemi & Aghaei, 2022*). Aspect-level sentiment analysis has

been performed using an adaptive SVM model and Twitter dataset (Liu et al., 2022). Du et al. (2022) applied a gated attention model for sentiment classification.

The opinions of the general population on 11 chosen topics using Latent Dirichlet Allocation were examined by Xue et al. (2020) utilizing data from COVID-19 tweets. The findings are consistent with earlier research on COVID-19 that claims that the coronavirus epidemic has a major influence on people's psychological states, according to the authors, who also determined that fear is the most prevalent emotion across all of the themes they investigated.

Despite the above-mentioned studies, the sentiments related to different vaccines and vaccination acceptability is an under-investigated area and requires further research. This study presents a detailed analysis in this regard by obtaining the data for different vaccines and a separate analysis for each vaccine.

MATERIALS AND METHODS

This section discusses the dataset, feature engineering techniques, machine learning, deep learning models, and proposed methodology used to analyze the COVID-19 vaccine-related sentiments on manually labeled Twitter data.

Dataset description

This study uses the dataset, 'COVID-19 All Vaccines Tweets', which was obtained from the Kaggle repository (Preda, 2022). The dataset contains tweets related to COVID-19 vaccines. Twitter data are used for the following reasons:

- For social media platforms like Facebook, users need to be friends with each other before they can follow others because it is based on friendship pattern (Stieglitz & Dang-Xuan, 2013). In contrast, there are no such restrictions on Twitter; anyone can follow others according to their interest.
- Soft policy of Twitter for developers to access their data. The 'no friendship' pattern of Twitter makes it more vulnerable to spreading misinformation rapidly. Such platforms need more attention to using automatic detection methods to moderate their discussions.

Tweets were related to five vaccines including 'AstraZeneca', 'Moderna', 'Pfizer/BioNTech', 'Sinopharm' and 'Sputnik V'. The dataset was labeled manually. Tweets classified into the former class present in favor opinions of users, while tweets under the latter class present negative comments the users regarding COVID-19 vaccination. The dataset was divided into two classes: 'Against' and 'In favor'. A few sample reviews are given in Table 1. Table 2 shows the distribution of labels.

Feature engineering techniques

Feature engineering techniques are used to extract appropriate information from raw data to train machine learning models (Bocca & Rodrigues, 2016). The feature engineering process is required for machine learning models and their performance is affected by the choice of the feature engineering method (Heaton, 2016). This process converts the data

Table 1 Sample reviews for tweets related to COVID-19 vaccination.

Vaccine	Tweet	Sentiment
Sinopharm	The vaccine manufacturers have said that their formulas are effective against the new variant.	In favor
	Feeling pain in my shoulder after getting first dose of vaccine, not recommended	Against
AstraZeneca	Do not take the vaccine.	Against
	Good morning. I had my COVID vaccination yesterday—feeling fine!	In favor
SputnikV	Good grief. This is just pure evil.	Against
	#SputnikV #COVID-19 Russian vaccine is created to last 24 years effective	In favor
PfizerBioNTech	#PfizerBioNTech COVID vaccine is not safe whilst breastfeeding	Against
	COVID vaccine you getting it #COVIDVaccine #Pfizer/BioNTech	In favor
Moderna	While the world has been on the wrong side of history this year, hopefully the biggest vaccination effort ever.	Against
	There have not been many bright days in 2020, but here are some of the best #Moderna.	In favor

Table 2 COVID-19 vaccination tweets labeling using VADER and TextBlob.

Vaccine name	VADER		TextBlob		Manual labeling	
	In favor	Against	In favor	Against	In favor	Against
AstraZeneca	1,393	1,980	1,408	1,965	1,768	1,605
Moderna	9,206	13,178	9,209	13,175	13,739	8,645
PfizerBioNTech	2,038	2,906	2,052	2,892	2,975	1,969
Sinopharm	2,053	2,519	2,084	2,488	2,557	2,015
SputnikV	2,977	5,526	2,972	5,531	5,718	2,785

into a feature vector, suitable to train the models. In this work, three feature engineering techniques BoW, TF-IDF, Word2vec, and feature union (TF-IDF+BoW) are used. The advantages and disadvantages of feature engineering techniques are presented in [Table 3](#).

Bag of words

The BoW is a simple and widely used technique to extract features from raw text. It is easy to implement and is mostly used in text categorization and language modeling. It uses CountVectorizer for feature extraction by considering term occurrences in the form of a matrix ([Eshan & Hasan, 2017](#)). Each feature or word in a matrix is assigned a value according to its number of occurrences in the *corpus* ([Hu, Downie & Ehmann, 2009](#)).

Term frequency-inverse document frequency

TF-IDF is another commonly used technique for feature extraction from raw text data. It is mostly used in textual information retrieval and text classification ([Yu, 2008](#)). In contrast to simple term count in BoW, TF-IDF also assigns weights to each word regarding its importance. It was done using inverse document frequency along with term frequency ([Robertson, 2004](#)). Important terms are represented with higher weight values. It can be calculated using

Table 3 Advantages and disadvantages of feature representation technique.

Technique	Type	Advantages	Disadvantages
TF	Vectorization technique	<ul style="list-style-type: none"> –Calculate the frequency of a document’s most frequently used term. –Count the number of times each word appears. 	<ul style="list-style-type: none"> –The issue with using raw word frequency data is that adding relevance does not make usage more proportionate.
TF-IDF	Vectorization technique	<ul style="list-style-type: none"> –Quickly compare documents for similarities. –Calculate the frequency of each distinct term in a text as well as the entire <i>corpus</i>. –Weight is inversely correlated with word frequency within texts and directly correlated with word frequency inside documents. –Stop words like is, a, <i>etc.</i> have less impact than uncommon words. 	<ul style="list-style-type: none"> –Enormous vector size. –Position and its co-occurring phrases are not taken into consideration. –Do not take semantics and context into account. –Sparsity problem. –It is ineffective to distinguish polysemy terms and compare similarities between synonyms.
BoW	Vectorization technique	<ul style="list-style-type: none"> –Simple and easy to use-offers feature representation of free-form text for NLP tasks. –Words to vectors mapping 	<ul style="list-style-type: none"> –Large vocabulary makes it challenging to train the model. –Sparsity matrix. –Our vocabulary would expand if the new phrases included new terms, which would also lengthen the vectors.
Word2Vec	Prediction based technique	<ul style="list-style-type: none"> –Works on words’ probability. –Map words to target vectors. –CBOW predicts the words’ probability and skip-gram determines the words’ context. 	<ul style="list-style-type: none"> –Large-sized vocabulary make the model difficult to train on Word2Vec. –Consider word similarities. –CBOW Take polysemy words’ average, separate vectors are used to present skip-gram.

$$W_{i,j} = [1 + \log(tf_{i,j})] \times \left[\log \left(\frac{N}{df_i} \right) \right] \quad (1)$$

where N is the total number of documents, $TF_{i,j}$ represents term frequency in document and $df_{f,t}$ is the document containing term t .

Word2Vec

The Word2Vec model extracts the idea of similarity between words or items, such as semantic similarity, synonym identification, concept classification, selectional preferences, and analogies. In word embedding, words that have the same meaning are represented similarly, which is a learned representation for text (Egger, 2022). One of the major advances in deep learning for difficult natural language processing tasks may be attributed to this method of encoding words and documents. Word embeddings are n-dimensional distributed representations of text. These are necessary for resolving the majority of NLP issues.

Feature union

The methodology used for feature union is presented in Fig. 1. Features are extracted using TF-IDF and BoW separately and are joined to enlarge the feature vector.

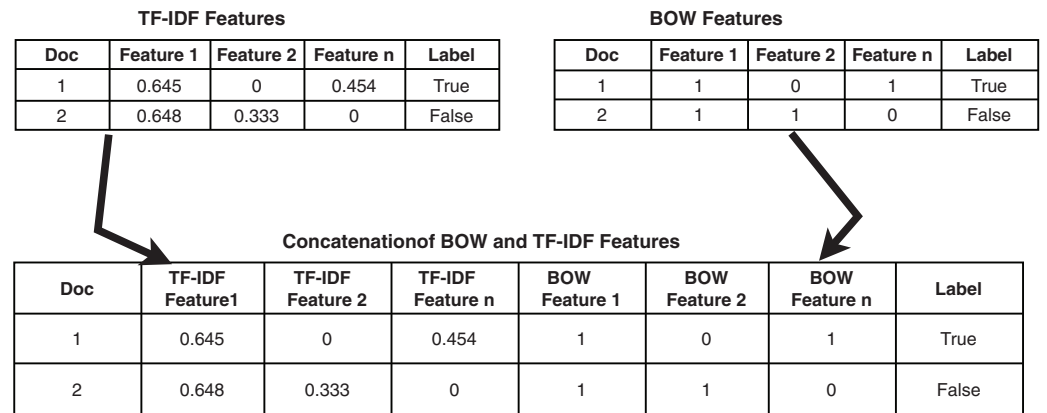


Figure 1 Methodology adopted for feature union.

Full-size DOI: 10.7717/peerj-cs.1190/fig-1

Models

This research takes advantage of various machine learning models (such as ensemble learning classifiers, regression-based models, and probability-based models), and deep learning models to classify tweets related to COVID-19. This study considers the use of the following classifiers for sentiment analysis of COVID-19 vaccine-related tweets as presented in Table 4. All classifiers are implemented using Sci-kit library (Pedregosa et al., 2011).

Proposed methodology

This section discusses the proposed methodology to investigate the COVID-19 vaccine-related sentiments from Twitter data.

Figure 2 shows the architecture of the proposed framework. At first, the Twitter dataset related to COVID-19 is divided into five subsets according to vaccine types AstraZeneca, Moderna, Pfizer/BioNTech, Sinopharm, and SputnikV. The dataset contains the highest number of tweets regarding these vaccines. Each subset is analyzed individually and each tweet is classified as 'In Favor' or 'Against'. Data goes through preprocessing steps like stopwords removal, number removal, special character removal, lemmatization, and tokenization. The dataset is labeled with the help of graduate students from the artificial intelligence department. Each vaccine subset is assigned to three students. For labeling, the following criteria are used:

- Three students label the data separately,
- A label is assigned, if at least two of the annotators agree,
- In case of a different label for a tweet from each annotator, the tweet is dropped.

After labeling, datasets are prepared to train the machine learning models. Dataset is split into training and testing in the ratio of 70% and 30%, respectively. Then feature engineering techniques are applied to both training and test sets. This study uses BoW, TF-IDF and their union (BoW+TF-IDF) with supervised machine learning models to

Table 4 Description of machine learning and deep learning models.

Reference	Model	Description
<i>Breiman (2001)</i>	RF	RF is one of the meta-estimators that integrate aggregation of a number of decision trees (DT) in order to provide improved efficacy and reduced over-fitting of the framework. It works by fitting DT classifiers on a number of samples of the input data. Afterward, it averages the results obtained from each DT classifier thus working as an ensemble learner.
<i>Friedman (2001)</i>	GBM	GBM is an ensemble model that develops an additive model in an optimized manner by the integration of a loss function. It works in an iterative manner that optimizes the error rate at each iteration by using the loss function. The purpose of the gradient boosting algorithm is to specify the outcomes of the target variable for the next model to lessen the prediction error.
<i>Sharaff & Gupta (2019)</i>	ETC	ETC works similarly to the RF model and a tree-based model. It is also known as an extremely randomized tree and unlike RF, it does not use bootstrap data, it builds trees from the actual data samples. It was proposed to build trees by considering the numeric input and selecting optimal cut-point to avoid variance at each node which reduces the computational complexity.
<i>Boyd, Tolson & Copes (1987)</i>	LR	LR works on a probability-based model and is used for classification tasks. It uses a logistic function for the modeling of binary variables. LR utilizes the correlation coefficient which is the measure of the relationship between the target variable and the independent variable.
<i>Pérez, Larrañaga & Inza (2006)</i>	NB	NB is based on 'Bayes' theorem which works on the assumption of independent features. It focuses on the prior probability and posterior probability of a target label in the dataset. Its supposition of considering feature independence is unrealistic for actual data. It shows robust results on large-sized and complex data having multiple classes.
<i>Gardner (1984)</i>	SGD	SGDC works on a one-versus-all technique. It is an optimization algorithm and finds the best suitable features or parameters among predicted and actual target values (<i>Gardner, 1984</i>). It gives good results on the large-sized dataset and uses a maximum sample at each iteration. It is sensitive regarding hyperparameter tuning.
<i>Kocyigit, Alkan & Erol (2008)</i>	MLP	MLP has significant characteristics with respect to classification such as it is easy and simple to implement. MLP performs well on the small-sized training set. MLP consists of mainly three layers that are hidden layers, the input layer, and the output layer.
<i>Krizhevsky, Sutskever & Hinton (2012)</i>	CNN	CNN is a deep neural network and efficiently learns features with the help of pooling layers, non-linear activation, dropout layer, and most importantly convolution layers. It was first developed for image-based tasks such as image categorization and image segmentation. End-to-end training makes CNN more efficient.
<i>Škrlić et al. (2019)</i>	RNN	Recurrent neural networks are a type of artificial neural network in which connections between nodes can form a cycle, allowing the output of certain nodes to influence input to other nodes in the same network in the future. This enables it to display temporal dynamic behavior.
<i>Staudemeyer & Morris (2019)</i>	LSTM	An artificial neural network called long short-term memory is utilized in deep learning and artificial intelligence. LSTM features feedback connections as opposed to typical feedforward neural networks. LSTM may analyze complete data sequences in addition to single data points.
<i>Yang & Cui (2021)</i>	BERT	Google has created a transformer-based machine learning method for pre-training natural language processing called Bidirectional Encoder Representations from Transformers (BERT). It is an open-source model and pre-trained on a large volume of data and often performs well.

select features from COVID-19 vaccine-related tweets. Models are optimized using several hyperparameters which are fined tuned, as shown in [Table 5](#).

Machine learning models are trained using three settings of feature extraction methods and then test data is used for performance evaluation in terms of accuracy, precision, recall, and F1-score. The following equations are used for performance evaluation metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

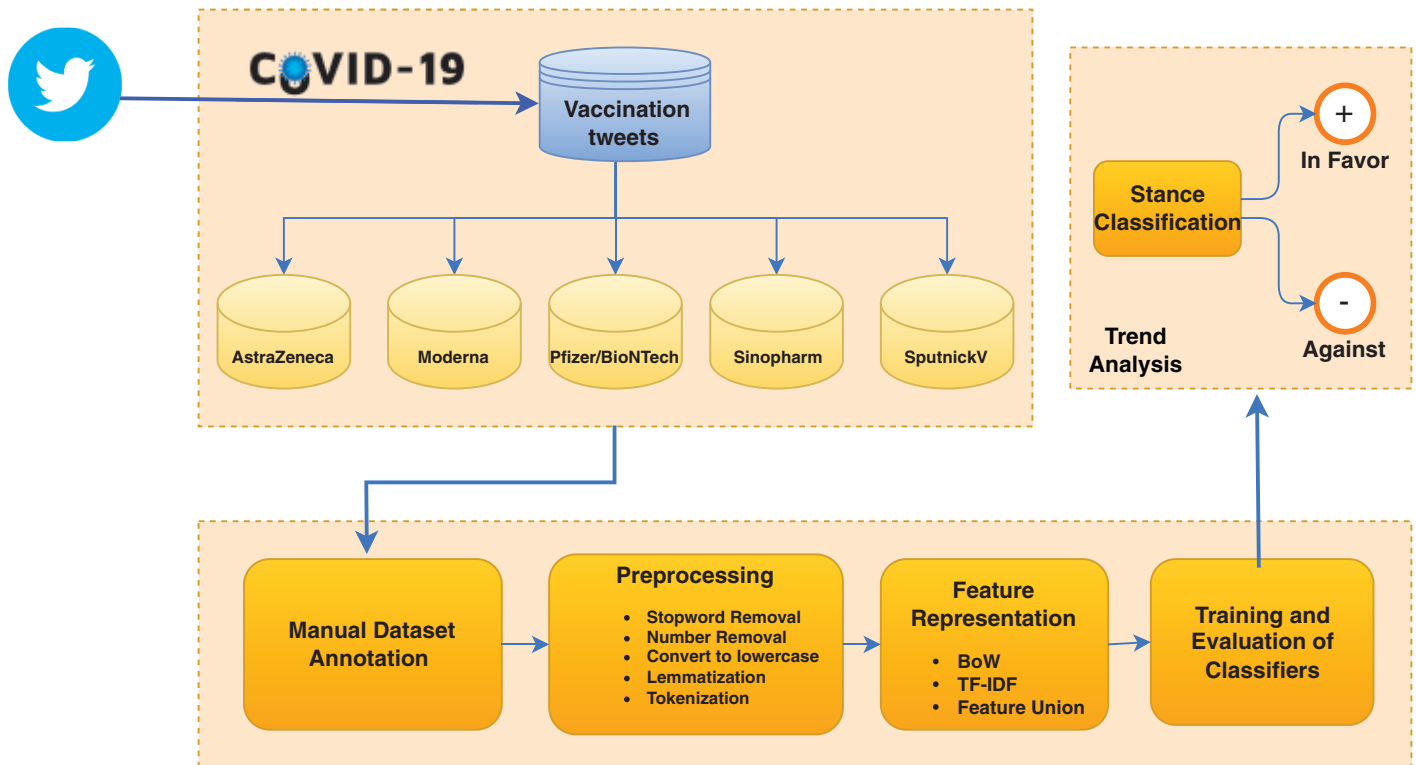


Figure 2 The architecture of the proposed framework.

Full-size DOI: 10.7717/peerj-cs.1190/fig-2

Table 5 Hyperparameter setting of learning models.

Classifiers	Parameters
RF	n_estimator=200, max_depth=30, random_state=52
ETC	n_estimator=200, max_depth=30, random_state=52
GBM	n_estimator=200, max_depth=30, random_state=52, learning_rate=0.1
LR	penalty='l2', solver='lbfgs'
NB	alpha=1.0, binarize=0.0
SGD	penalty='l2', loss='log'
MLP	Dense (neurons=300), dense (neurons=200), dense (neurons=100), activation='relu', dropout (0.5), optimizer='adam', softmax(2)
CNN	Conv (7, @64), Max pooling (2x2), Conv (7, @64), GlobalMax pooling (2x2), Dropout (0.5), Dense (32 neurons), optimizer='adam', Softmax (2)

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1-score = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively, and extracted from the confusion matrix.

RESULTS

This study compared the result of machine learning models by combining each model with feature extraction techniques and their union. Machine learning models are implemented using scikit-learn (*Hackeling, 2017*) in Python. Experiments have been performed in different settings and the best parameters are selected following the grid search approach. Algorithms have been evaluated by applying TF-IDF, Word2Vec, BoW, and Feature Union (TF-IDF+ BoW) techniques.

Experimental results of machine learning models

A set of experiments are performed using the TF-IDF, Word2vec, Feature Union, and BoW features with selected machine-learning models on all five divisions of the dataset. Different features are used on the manually labeled dataset and accuracy results are presented in [Table 6](#). Results reveal that the best-performing model is ETC using BoW on all sub-datasets. ETC using BoW achieved 92% accuracy for the Sinopharm sub-dataset. SGD has also shown good results and ranked second on this dataset using BoW for sentiment analysis of tweets. In the case of the Moderna dataset, SGD using BoW has shown the highest result in terms of accuracy value with 90% which is equal to the value achieved by ETC. The results show that ETC outperforms other models using BoW on all sub-datasets for sentiment analysis.

Class-wise results achieved on the manually labeled dataset are presented in [Table 7](#) separately using the BoW feature with which machine learning models have shown the highest accuracy. It can be observed that ETC outperformed other models with 86% accuracy using BoW on the AstraZeneca dataset. ETC has also achieved the highest scores for precision and F1-score for both the 'In favor' and 'Against' classes. The highest recall is achieved by GBM and LR for 'Against' class each with 97% and by ETC for In favor class with 72%. In the case of the Moderna dataset, the highest accuracy result has been achieved by ETC and SGD with 90% accuracy each. The highest precision for the 'Against' class is 91% by ETC and SGD. For the 'In favor' class, the highest precision is 91% by both NB and LR. The highest recall is 97% for the 'Against' class by NB.

Similarly, results of classifiers with the Pfizer dataset reveal that ETC surpassed other models with 90% accuracy and 92% F1-score for the positive class and 87% F1-score for the negative class using BoW. For the Sinopharm dataset, results show that the highest results in terms of accuracy and precision, and F1-score for the positive and negative classes are achieved by ETC. SGD obtains the second-best results using TF-IDF for sentiment analysis on the Pfizer dataset. The lowest results have been achieved by NB. Subsequently, it can be seen that ETC using BoW has shown the highest results in terms of accuracy and F1-score on the SputnikV dataset. The highest precision for the negative class is 92% achieved by ETC. The highest recall for the negative class is 98% and it is achieved by NB.

Table 6 Accuracy of models with different features using manually labeled dataset.

Dataset	Model	TF-IDF	Word2Vec	Feature union	BoW
AstraZeneca	RF	83%	80%	81%	85%
	ETC	84%	79%	82%	86%
	GBM	79%	70%	74%	82%
	LR	84%	77%	81%	83%
	NB	78%	74%	79%	78%
	SGD	84%	78%	81%	85%
Moderna	RF	87%	83%	85%	88%
	ETC	89%	85%	88%	90%
	GBM	85%	82%	83%	84%
	LR	88%	84%	89%	89%
	NB	84%	80%	81%	81%
	SGD	88%	84%	89%	90%
Pfizer	RF	86%	84%	86%	88%
	ETC	88%	86%	85%	90%
	GBM	86%	87%	86%	86%
	LR	87%	85%	88%	88%
	NB	85%	87%	86%	83%
	SGD	87%	86%	88%	89%
Sinopharm	RF	90%	89%	88%	90%
	ETC	89%	86%	88%	92%
	GBM	84%	85%	84%	90%
	LR	87%	86%	88%	89%
	NB	84%	84%	86%	81%
	SGD	87%	84%	89%	91%
SputnikV	RF	87%	89%	84%	90%
	ETC	89%	87%	88%	90%
	GBM	86%	87%	86%	87%
	LR	89%	81%	87%	88%
	NB	83%	78%	85%	81%
	SGD	89%	83%	90%	89%

Experimental results of deep learning models

For a fair comparison, experiments are also performed using deep learning models and results are presented in Table 8. Two feed-forward deep learning models including convolutional neural network (CNN) (Kamath, Liu & Whitaker, 2019) and multilayer perceptron (MLP) (Tang, Deng & Huang, 2015) and three RNN-based deep learning models including bidirectional encoder representations from Transformers (BERT) (Yang & Cui, 2021), LSTM (Staudemeyer & Morris, 2019), and recurrent neural network (RNN) (Škrlj et al., 2019) are used in the experiments. Experimental results reveal that deep learning models using word2Vec have not shown better results in comparison with machine learning models using TF-IDF, BoW, and feature union.

Table 7 Experimental results of the manually labeled dataset using BoW features.

Dataset	Model	Class	Prec.	Recall	F1-score	Accuracy
AstraZeneca	RF	Against	82%	96%	88%	85%
		In favor	92%	66%	77%	
	ETC	Against	84%	95%	89%	86%
		In favor	90%	72%	80%	
	GBM	Against	79%	97%	87%	82%
		In favor	93%	58%	72%	
	LR	Against	80%	96%	87%	83%
		In favor	91%	63%	74%	
	NB	Against	75%	95%	84%	78%
		In favor	86%	50%	64%	
	SGD	Against	83%	95%	88%	85%
		In favor	89%	69%	78%	
Moderna	RF	Against	88%	94%	91%	88%
		In favor	89%	79%	84%	
	ETC	Against	91%	93%	92%	90%
		In favor	89%	86%	87%	
	GBM	Against	81%	96%	88%	84%
		In favor	90%	64%	75%	
	LR	Against	89%	95%	92%	89%
		In favor	91%	81%	86%	
	NB	Against	78%	97%	86%	81%
		In favor	91%	57%	70%	
	SGD	Against	91%	94%	92%	90%
		In favor	90%	85%	87%	
Pfizer	RF	Against	86%	95%	90%	88%
		In favor	91%	76%	83%	
	ETC	Against	91%	93%	92%	90%
		In favor	88%	86%	87%	
	GBM	Against	83%	96%	89%	86%
		In favor	93%	69%	79%	
	LR	Against	87%	96%	91%	88%
		In favor	92%	77%	84%	
	NB	Against	80%	96%	87%	83%
		In favor	91%	63%	75%	
	SGD	Against	89%	94%	91%	89%
		In favor	90%	82%	85%	
Sinopharm	RF	Negative	87%	98%	92%	90%
		In favor	96%	81%	88%	
	ETC	Against	89%	97%	93%	92%
		In favor	96%	85%	90%	
	GBM	Against	87%	97%	92%	90%

Table 7 (continued)

Dataset	Model	Class	Prec.	Recall	F1-score	Accuracy
SputnikV	LR	In favor	96%	81%	88%	89%
		Against	87%	97%	91%	
	NB	In favor	95%	81%	88%	81%
		Against	756%	98%	85%	
	SGD	In favor	95%	61%	74 %	91%
		Against	88%	96%	92 %	
	RF	In favor	94%	85%	89 %	90%
		Against	90%	96%	93%	
	ETC	In favor	90%	78%	83%	90%
		Against	92%	95%	93%	
	GBM	In favor	88%	81%	84%	87%
		Against	86%	97%	91%	
	LR	In favor	91%	65%	76%	88%
		Against	88%	96%	92 %	
	NB	In favor	89%	72%	79%	81%
		Against	79%	98%	87 %	
	SGD	In favor	89%	44%	59%	89%
		Against	91%	93%	92%	
		In favor	85%	79%	82%	

Results indicate that the best performance of the BERT model is on the Sinopharm dataset where its accuracy is 85%. On the other hand, the LSTM model shows the best performance on the Sinopharm dataset with an 83% accuracy while the precision, recall, and F1-score each is 82%, 80%, and 81% respectively. RNN performed the best using the tweets regarding the Sinopharm vaccination and achieve an accuracy of 80%, precision of 85%, recall of 82%, and F1-score of 83%. Results from these models show lower accuracy as compared to the proposed approach.

Results of K-fold cross-validation

This study validates the proposed approach by performing a 10-fold cross-validation. Results of 10-fold cross-validation for the proposed model are given in Table 9. Experiments results are provided with respect to each of the five vaccines considered in this study. Results demonstrate the proposed approach shows better results with cross-validation as well.

Experimental results using COVID-19 VAERS dataset

To prove the effectiveness of the proposed approach, we performed additional experiments using another manually annotated dataset. For this purpose, the COVID-19 VAERS dataset is used which is publicly available on Kaggle. It is a benchmark dataset that contains adverse events reported after COVID-19 vaccination (Garg, 2021). It has a total of 5,351 event reports. This study utilized the multi-class classification problem using the

Table 8 Result of deep learning classifiers using Word2Vec features.

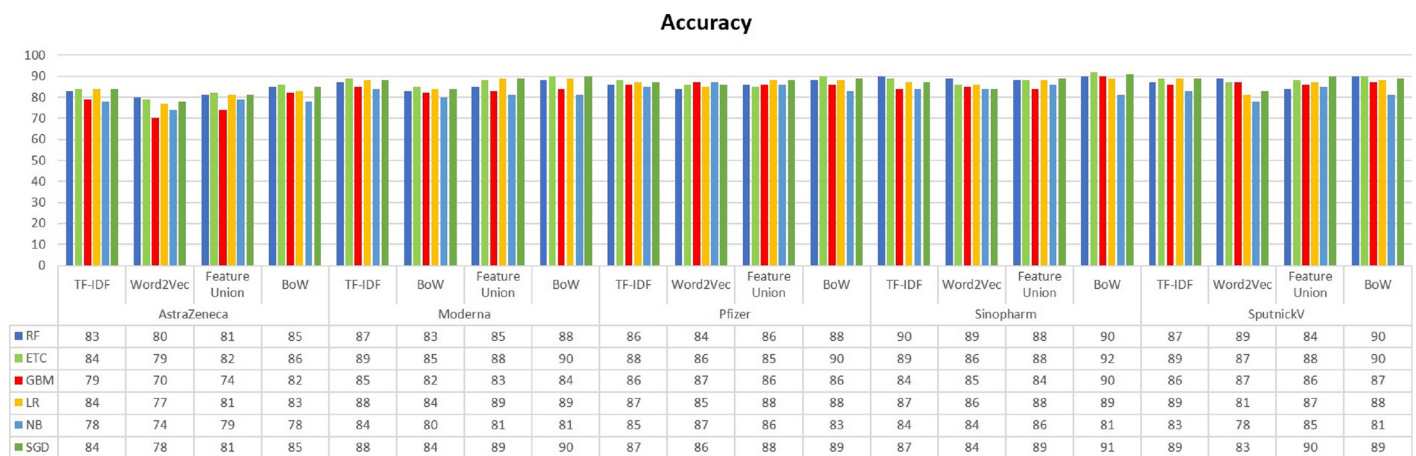
Dataset	Model	Precision	Recall	F1-score	Accuracy
AstraZeneca	CNN	78%	76%	77%	80%
	MLP	81%	85%	83%	81%
	BERT	80%	82%	81%	79%
	LSTM	79%	82%	81%	78%
	RNN	76%	79%	78%	74%
Moderna	CNN	80%	88%	86%	82%
	MLP	82%	89%	85%	81%
	BERT	81%	85%	83%	84%
	LSTM	80%	82%	81%	81%
	RNN	77%	81%	79%	80%
Pfizer	CNN	84%	82%	83%	79%
	MLP	81%	84%	82%	81%
	BERT	84%	80%	82%	83%
	LSTM	77%	79%	78%	81%
	RNN	80%	78%	79%	80%
Sinopharm	CNN	84%	79%	83%	87%
	MLP	80%	80%	80%	82%
	BERT	87%	85%	86%	85%
	LSTM	82%	80%	81%	83%
	RNN	85%	82%	83%	80%
SputnikV	CNN	77%	84%	81%	88%
	MLP	76%	79%	78%	85%
	BERT	84%	84%	84%	83%
	LSTM	81%	79%	80%	80%
	RNN	80%	77%	79%	80%

Table 9 Results of 10-fold cross-validation on all datasets with the best performing model ETC and BoW features.

Fold number	Astrazeneca	Moderna	Pfizer	Sinopharm	SputnikV
1st-fold	0.913	0.922	0.911	0.902	0.902
2nd-fold	0.898	0.913	0.909	0.911	0.923
3rd-fold	0.869	0.911	0.914	0.923	0.914
4th-fold	0.863	0.901	0.913	0.937	0.931
5th-fold	0.854	0.933	0.912	0.922	0.911
6th-fold	0.852	0.913	0.923	0.912	0.878
7th-fold	0.858	0.921	0.866	0.871	0.882
8th-fold	0.899	0.908	0.912	0.909	0.923
9th-fold	0.889	0.919	0.882	0.902	0.927
10th-fold	0.891	0.942	0.892	0.911	0.933
Average	0.8786	0.9183	0.9034	0.9100	0.9124

Table 10 Results of machine learning models using BoW on the COVID-19 VAERS dataset.

Model	Accuracy	Precision	Recall	F1-score
RF	0.9681	0.94	0.97	0.95
ETC	0.9701	0.94	0.97	0.96
GBM	0.9581	0.94	0.96	0.95
LR	0.9601	0.94	0.97	0.96
NB	0.9681	0.95	0.97	0.96
SGD	0.9701	0.94	0.97	0.96

**Figure 3** Accuracy result comparison of machine learning models.

Full-size DOI: 10.7717/peerj-cs.1190/fig-3

‘recovered’, ‘not covered’, and ‘recovery status unknown’ classes of the dataset. For complete details of the dataset, the readers are referred to [Saad et al. \(2022\)](#). We implemented the proposed approach on this dataset and the results of multiclass classification are presented in [Table 10](#). Results indicate that the proposed approach shows superior performance on the manually labeled dataset as well. Of the used models, ETC and SGD show the best performance with a 0.97 accuracy score while the precision, recall, and F1-scores are 0.94, 0.97, and 0.96, respectively. In addition, the performance of other models is marginally different.

DISCUSSION

Performance comparison of classifiers using different feature representation techniques has been carried out on five subsets of datasets based on the COVID-19 vaccine type. The impact of TF-IDF, BoW, and their union (TF-IDF+BoW) has been investigated in tweets to determine the trend of public opinion about COVID-19 vaccines. Comparative analysis in terms of accuracy, precision, recall, and F1-score has been presented separately.

[Figure 3](#) presents the accuracy comparison of classifiers on all five datasets. It can be noted that RF, GBM, ETC, LR, and SGD have achieved the highest accuracy score on every

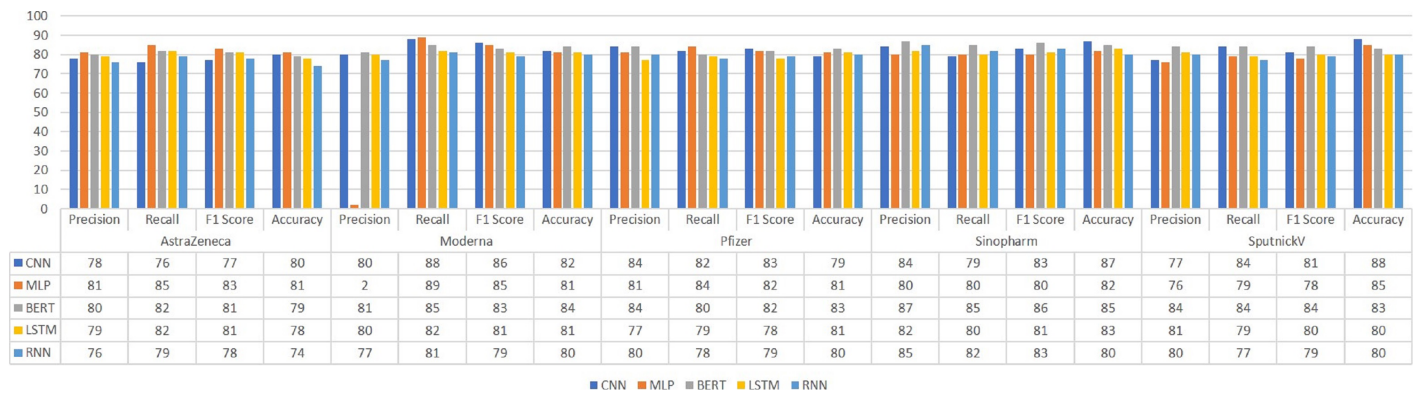


Figure 4 Comparison of deep learning models using Word2Vec features.

Full-size DOI: 10.7717/peerj-cs.1190/fig-4

dataset when trained using BoW features. The highest accuracy score is 92% which is achieved by ETC using BoW features. Classifiers using Word2Vec have achieved low results as compared to BoW and feature unions. Besides the highest accuracy, overall performance using BoW features is better. Despite the simple occurrence count, BoW features often show better results as compared to other complex feature engineering approaches.

Figure 4 illustrates the precision, recall, F1-score, and accuracy results comparison of the deep learning models. The highest precision score has been achieved by RNN on the Sinopharm sub-dataset with an 85% score. SGD has achieved the highest precision using BoW and feature union on the Moderna dataset. ETC has achieved the highest precision score using BoW on Sinopharm and SputnikV datasets. BERT has achieved the highest recall on the Moderna dataset and the highest F1-score on the Sinopharm dataset. Deep learning models MLP and BERT have achieved the highest 85% accuracy score which is lower than the results achieved with ETC and BoW.

From the above discussion, it is clear that the classifiers show the highest result regarding the sentiment classification of vaccine-related tweets when trained using the BoW features. Overall, ETC has shown the highest results on all datasets. Randomization and optimization features make ETC more efficient in text classification by reducing bias and variance. TF-IDF considers the importance of words and assigns weights accordingly while BoW is a simple and flexible technique and only considers the frequency of unique terms. The feature union of both techniques contains redundant features, also increases the training time, and did not improve the performance of the models. Deep learning models often provide high accuracy on large-sized datasets but with more training time. But in the case of sentiment analysis of COVID-19-related tasks, the deep learning model did not achieve robust results. ETC in combination with BoW is the most suitable approach for the sentiment analysis of COVID-19 vaccine-related tweets.

In this study, the average accuracy score on all datasets is 0.9111 which is comparable to the scores of other studies such as 0.8177 using NB by Villavicencio et al. (2021), 0.9059 using LSTM and 0.9083 by BiLSTM by Alam et al. (2021). However, this study uses 20,967

tweets that are larger than those used in other studies like 993 tweets by [Villavicencio et al. \(2021\)](#). [Alam et al. \(2021\)](#) used the same dataset as ours but employed complex deep-learning models and achieved lower results. All these things proved the superiority of the proposed approach.

Deep learning models do not perform well compared to machine learning models used in this study. First of all, the size of the dataset is small and not enough for deep learning models to get a good fit. Second, the data is sparse which leads to poor performance of deep learning models; that is the reason word2vec features also do not perform well using machine learning models. Third, deep learning models need to tune a large number of hyperparameters which require a large-sized dataset. For the current study, the small-sized dataset is not enough to produce good results using deep learning models.

This research has some limitations like the tweets utilized in this study represented just 1% of daily tweets, therefore they might not be an accurate representation of all tweets. In addition, model fine-tuning was restricted to some parameters only due to a lack of resources needed for training; additional parameters were not tweaked. If the tuning had been done more extensively, the performance of these models may have been enhanced much further.

Performance of manually vs TextBlob annotated dataset

For analyzing the impact of data annotation from different techniques including TextBlob and manual annotation, experiments are performed with both datasets separately. [Table 11](#) shows the accuracy of all the models using both datasets. The results indicate that the performance of the models is marginally better when TextBlob annotated dataset is used. Although manual annotation is considered the best for machine learning models, their performance is better using the TextBlob dataset. Since the models and TextBlob follow a similar mechanism of objectivity, it is possible that TextBlob makes mistakes similar to machine learning models. Also, since the machine learning models work on features, TextBlob may be providing more correlated features to the models for training which increases their performance. The performance using VADER annotation is inferior to TextBlob.

Trend analysis and future directions

The distribution of stances into two categories: 'Against' and 'In favor' is presented in [Fig. 5](#) which considers the tweets according to vaccine types. It can be observed that the 'Against' is the dominant stance found in the total tweets. On the entire dataset, the number of tweets for the 'Against' stance is 60% while only 40% belongs to the 'In favor' stance. It also shows that the tweets related to Moderna and SputnikV vaccines have the highest number of 'against' stances as compared to other vaccines.

The results suggest that the trends against and in favor of vaccines have similar patterns indicating that the majority of tweets contain negative sentiments regarding the vaccination process. The highest number of tweets 'In favor' is for 'Sinopharm' where 46% favor this vaccine. Similarly, the SputnikV vaccine has the highest number of 'Against' tweets, which are 65%.

Table 11 Accuracy of models with TextBlob and manually labeled dataset using BoW features.

Dataset	Model	TextBlob	VADER	Manual annotation
AstraZeneca	RF	87%	85%	85%
	ETC	89%	87%	86%
	GBM	86%	85%	82%
	LR	87%	87%	83%
	NB	81%	79%	78%
	SGD	88%	85%	85%
Moderna	RF	90%	87%	88%
	ETC	91%	91%	90%
	GBM	84%	82%	84%
	LR	91%	89%	89%
	NB	85%	82%	81%
	SGD	92%	88%	90%
Pfizer	RF	85%	86%	88%
	ETC	90%	88%	90%
	GBM	85%	85%	86%
	LR	88%	87%	88%
	NB	85%	86%	83%
	SGD	89%	89%	89%
Sinopharm	RF	90%	88%	90%
	ETC	92%	91%	92%
	GBM	90%	87%	90%
	LR	89%	86%	89%
	NB	81%	78%	81%
	SGD	91%	88%	91%
SputnikV	RF	92%	89%	90%
	ETC	92%	91%	90%
	GBM	86%	86%	87%
	LR	90%	85%	88%
	NB	80%	77%	81%
	SGD	91%	91%	89%

This study also performs an analysis of people's opinions with time. People's opinion changes with time. [Figure 6](#) presents how sentiments of tweets vary with time. In favor sentiments are in green color and Against sentiments are in red. Fluctuation or Variation of sentiments can be seen clearly in [Fig. 6](#). The 'In favor' sentiments are at their peak on March 21 at the end of the final trial of vaccines, while the 'Against' sentiments are at a high rate in the mid of April 21 and May 21.

Furthermore, we have analyzed keywords and themes in each sentiment class that are in favor and against. [Table 12](#) presents that 'In Favor' of vaccine tweets are related to hope, support and happiness. On the other hand, 'Against' is related to fear, anger, and disappointment.

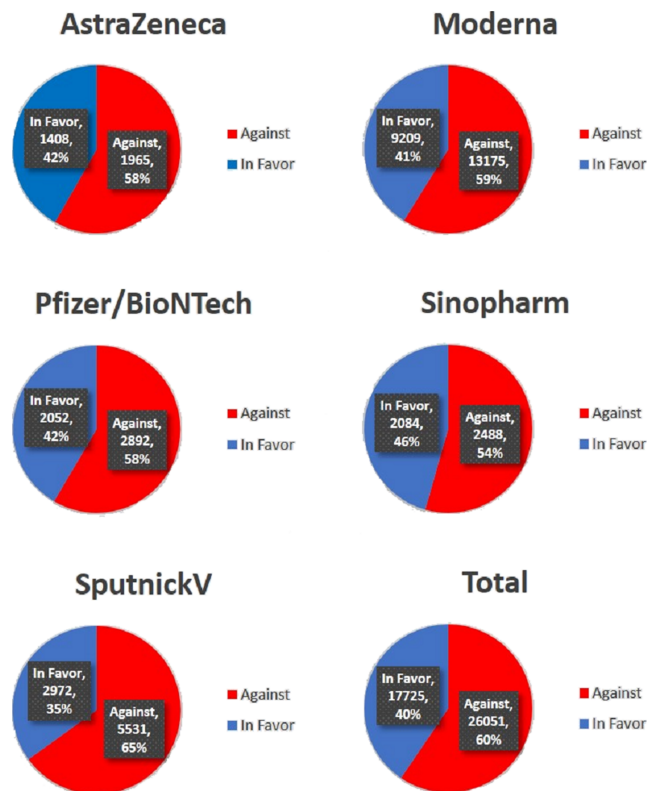


Figure 5 Comparison of sub-datasets regarding sentiment trends.

Full-size DOI: 10.7717/peerj-cs.1190/fig-5

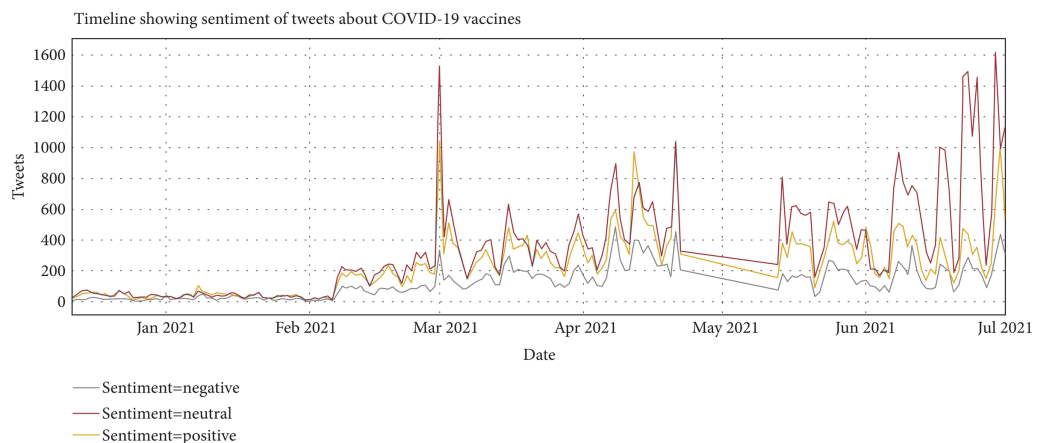


Figure 6 Variation of sentiments over time.

Full-size DOI: 10.7717/peerj-cs.1190/fig-6

Potential future directions of this work could be to analyze public trends regarding more vaccine types by employing natural language processing techniques. Moreover, government and other relevant agencies should provide more detail about the effectiveness and advantages of vaccines to increase the trust of the general public. Public trends for vaccines regarding gender and age groups can also be analyzed in the future. Twitter has been a source for a large number of research studies for sentiment analysis. However, the

Table 12 Keywords and themes in favor and against polarities.

Sentiment polarity	Theme	Keywords
In favor	Hope	Innovators, fitness, favouring, fortunes, adventures, confident, overcoming, efficiency, winnings, productivity, fascinating
	Happiness	Congratulates, thrill, appeal, cheerful, motivator
	Support	Aid, kindness, pardoned, wished, truthfulness, greatness, devoting, appreciating, consulted, facilitate, propel, assistance
Against	Fear	Adverse, torturous, hoard, wrecks, poisoned
	Anger	Disagreeing, shooting, angering, aggravated, outrage, terrifies, crazies, hates
	Disappointment	Stealing, misguide, misleads, fool, scammed, impossible, discouraged, blackmail, poisoned, pocketed, disregarding, slanders

Table 13 The acronyms used in this manuscript.

Acronyms	Definition
ANN	Artificial neural network
BERT	Bidirectional encoder representations from transformers
BoW	Bag of words
CDC	Centre of disease control
CNN	Convolutional neural network
COVID-19	Coronavirus disease of 2019
ETC	Extra tree classifier
GBM	Gradient boosting machine
IoT	Internet of things
LDA	Latent Dirichlet allocation
LR	logistic regression
LSTM	long short-term memory
MLP	Multilayer perceptron
NB	Naive Bayes
RF	Random forest
RNN	Recurrent neural network
SGD	Stochastic gradient descent
TF-IDF	Term frequency-inverse document frequency
UN	United nations
WHO	World health organization

probability of fake and biased tweets can not be ignored. Several studies point out a higher ratio of false/biased news as high as 25%, especially for the political campaigns ([Bovet & Makse, 2019](#); [Vosoughi, Roy & Aral, 2018](#); [Shao et al., 2018](#)). However, a recent study reveals that the ratio of biased tweets or misinformation related to COVID-19 is approximately 3.29% ([Sharma et al., 2020](#)). Considering this ratio of biased tweets, for the

current study we do not handle the biased tweets aspect and intend to incorporate it in the future study. The acronyms used in the manuscript are presented in [Table 13](#).

This study has several limitations. Since it is based on the manual feature engineering approach BoW in combination with the machine learning model and has inherent limitations of BoW. BoW neglects word semantics and the use of grammar. It simply provides the terms' occurrence and ignores the semantic importance of terms. In the future, we will explore more feature engineering techniques to improve the sentiment analysis task.

CONCLUSION

This research explores the opinion dynamics related to the COVID-19 vaccine by performing sentiment analysis on different vaccine-related tweets. The dataset is divided into five subsets and investigated separately to get deep insights and quantitative assessment. For sentiment analysis, several machine learning models coupled with four feature representation techniques (TF-IDF, BoW, Word2Vec, and feature union) have been compared. The results show that the ETC with BoW has the highest accuracy of 92% for sentiment analysis of COVID-19-related tweets. Predominantly, the performance of models is better when used with the BoW features. Deep learning models tend to show poor performance with the current dataset, as compared to machine learning models. Empirical and trend analysis of COVID-19 vaccine-related tweets reveals that the spread of unreliable and misinformation is increasing on social media platforms. Deep insights show that 60% are against vaccines. Furthermore, it was observed that a large number of 'Against' are for the SputnikV vaccine, followed by Moderna. Temporal analysis indicates that the ratio of 'In favor' sentiments for COVID-19 vaccination has been elevated over time. For future work, the analysis regarding age groups and gender can be incorporated for vaccination-related trend analysis.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

Imran Ashraf is an Academic Editor for PeerJ.

Author Contributions

- Shoaib Ahmed conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Dost Muhammad Khan conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Saima Sadiq conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Muhammad Umer conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

- Faisal Shahzad performed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Khalid Mahmood performed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Heba Mohsen performed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Imran Ashraf performed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code and data is available at GitHub: <https://github.com/MUmerSabir/CovidVaccine>; MUmerSabir. (2022). MUmerSabir/CovidVaccine: PeerJ Covid Vaccine (PeerJVaccine). Zenodo. DOI 10.5281/zenodo.7298734.

REFERENCES

- Alam KN, Khan MS, Dhruva AR, Khan MM, Al-Amri JF, Masud M, Rawashdeh M. 2021. Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data. *Computational and Mathematical Methods in Medicine* 2021(1):1–15 DOI 10.1155/2021/4321131.
- Alamoodi AH, Zaidan BB, Zaidan AA, Albahri OS, Mohammed KI, Malik RQ, Almahdi EM, Chyad MA, Tareq Z, Albahri AS, Hameed H, Alaa M. 2020. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review. *Expert Systems with Applications* 167:114155 DOI 10.1016/j.eswa.2020.114155.
- Alhajji M, Al Khalifah A, Aljubran MJ, Alkhalifah M. 2020. Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19. Available at <https://www.preprints.org/manuscript/202004.0031/v1>.
- Appel G, Grewal L, Hadi R, Stephen AT. 2020. The future of social media in marketing. *Journal of the Academy of Marketing Science* 48(1):79–95 DOI 10.1007/s11747-019-00695-1.
- Asgarnezhad R, Monadjemi SA, Aghaei MS. 2022. A new hierarchy framework for feature engineering through multi-objective evolutionary algorithm in text classification. *Concurrency and Computation: Practice and Experience* 34(3):e6594 DOI 10.1002/cpe.6594.
- Baccianella S, Esuli A, Sebastiani F. 2010. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
- Bocca FF, Rodrigues LHA. 2016. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture* 128(3):67–76 DOI 10.1016/j.compag.2016.08.015.
- Bovet A, Makse HA. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10(1):1–14 DOI 10.1038/s41467-018-07761-2.
- Boyd CR, Tolson MA, Copes WS. 1987. Evaluating trauma care: the triss method. *Journal of Trauma and Acute Care Surgery* 27(4):370–378 DOI 10.1097/00005373-198704000-00005.
- Breiman L. 2001. Random forests. *Machine Learning* 45(1):5–32 DOI 10.1023/A:1010933404324.

- Cao Z, Zhou Y, Yang A, Peng S. 2021.** Deep transfer learning mechanism for fine-grained cross-domain sentiment classification. *Connection Science* **33(4)**:911–928
DOI [10.1080/09540091.2021.1912711](https://doi.org/10.1080/09540091.2021.1912711).
- Castiglione A, Umer M, Sadiq S, Obaidat MS, Vijayakumar P. 2021a.** The role of internet of things to control the outbreak of COVID-19 pandemic. *IEEE Internet of Things Journal* **8(21)**:16072–16082 DOI [10.1109/JIOT.2021.3070306](https://doi.org/10.1109/JIOT.2021.3070306).
- Castiglione A, Vijayakumar P, Nappi M, Sadiq S, Umer M. 2021b.** COVID-19: automatic detection of the novel coronavirus disease from CT images using an optimized convolutional neural network. *IEEE Transactions on Industrial Informatics* **17(9)**:6480–6488
DOI [10.1109/TII.2021.3057524](https://doi.org/10.1109/TII.2021.3057524).
- Chun SA, Li AC-Y, Toliyat A, Geller J. 2020.** Tracking citizen's concerns during COVID-19 pandemic. In: *The 21st Annual International Conference on Digital Government Research*. 322–323.
- Cotfas L-A, Delcea C, Roxin I, Ioanăș C, Gherai DS, Tajariol F. 2021.** The longest month: analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access* **9**:33203–33223 DOI [10.1109/ACCESS.2021.3059821](https://doi.org/10.1109/ACCESS.2021.3059821).
- Crannell WC, Clark E, Jones C, James TA, Moore J. 2016.** A pattern-matched Twitter analysis of US cancer-patient sentiments. *Journal of Surgical Research* **206(2)**:536–542
DOI [10.1016/j.jss.2016.06.050](https://doi.org/10.1016/j.jss.2016.06.050).
- Dai K, Li X, Huang X, Ye Y. 2022.** SentATN: learning sentence transferable embeddings for cross-domain sentiment classification. *Applied Intelligence* **52(15)**:1–14
DOI [10.1007/s10489-022-03434-2](https://doi.org/10.1007/s10489-022-03434-2).
- De Choudhury M, Counts S, Horvitz E. 2013.** Predicting postpartum changes in emotion and behavior via social media. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3267–3276.
- Dixon S. 2023.** Number of social network users worldwide from 2017 to 2027. Available at <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>.
- Du Y, Liu Y, Peng Z, Jin X. 2022.** Gated attention fusion network for multimodal sentiment classification. *Knowledge-Based Systems* **240(2)**:108107 DOI [10.1016/j.knosys.2021.108107](https://doi.org/10.1016/j.knosys.2021.108107).
- D'Andrea E, Ducange P, Bechini A, Renda A, Marcelloni F. 2019.** Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications* **116(2)**:209–226 DOI [10.1016/j.eswa.2018.09.009](https://doi.org/10.1016/j.eswa.2018.09.009).
- Egger R. 2022.** Text representations and word embeddings. In: *Applied Data Science in Tourism*. Cham: Springer, 335–361.
- Eshan SC, Hasan MS. 2017.** An application of machine learning to detect abusive Bengali text. In: *2017 20th International Conference of Computer and Information Technology (ICCIT)*. Piscataway: IEEE, 1–6.
- Friedman JH. 2001.** Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* **29(5)**:1189–1232 DOI [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451).
- Fu Y, Liu Y. 2022.** Domain adaptation with a shrinkable discrepancy strategy for cross-domain sentiment classification. *Neurocomputing* **494**:56–66 DOI [10.1016/j.neucom.2022.04.092](https://doi.org/10.1016/j.neucom.2022.04.092).
- Garcia K, Berton L. 2021.** Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing* **101(1)**:107057
DOI [10.1016/j.asoc.2020.107057](https://doi.org/10.1016/j.asoc.2020.107057).
- Garcia-Gasulla D, Suzumura T. 2020.** What are we depressed about when we talk about COVID-19: mental health analysis on tweets using natural language processing. In: *Artificial*

- Intelligence XXXVII: 40th SGA International Conference on Artificial Intelligence, AI 2020, Cambridge, UK, December 15–17, 2020, Proceedings*. Vol. 12498. Cham: Springer Nature, 358.
- Gardner WA. 1984.** Learning characteristics of stochastic-gradient-descent algorithms: a general study, analysis, and critique. *Signal Processing* **6(2)**:113–133
DOI [10.1016/0165-1684\(84\)90013-6](https://doi.org/10.1016/0165-1684(84)90013-6).
- Garg A. 2021.** COVID-19 world vaccine adverse reactions. kaggle. Available at <https://www.kaggle.com/datasets/ayushggarg/covid19-vaccine-adverse-reactions?select=2021VAERSDATA.csv>.
- Giachanou A, Crestani F. 2016.** Like it or not: a survey of Twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)* **49(2)**:1–41 DOI [10.1145/2938640](https://doi.org/10.1145/2938640).
- Golbeck J, Grimes JM, Rogers A. 2010.** Twitter use by the US congress. *Journal of the American Society for Information Science and Technology* **61(8)**:1612–1621 DOI [10.1002/asi.21344](https://doi.org/10.1002/asi.21344).
- Hackeling G. 2017.** *Mastering machine learning with scikit-learn*. Birmingham: Packt Publishing Ltd.
- Han X, Wang J, Zhang M, Wang X. 2020.** Using social media to mine and analyze public opinion related to COVID-19 in China. *International Journal of Environmental Research and Public Health* **17(8)**:2788 DOI [10.3390/ijerph17082788](https://doi.org/10.3390/ijerph17082788).
- Hanson CL, Cannon B, Burton S, Giraud-Carrier C. 2013.** An exploration of social circles and prescription drug abuse through Twitter. *Journal of Medical Internet Research* **15(9)**:e189 DOI [10.2196/jmir.2741](https://doi.org/10.2196/jmir.2741).
- Heaton J. 2016.** An empirical analysis of feature engineering for predictive modeling. In: *SoutheastCon 2016*. Piscataway: IEEE, 1–6.
- Hiscott J, Alexandridi M, Muscolini M, Tassone E, Palermo E, Soultioti M, Zevini A. 2020.** The global impact of the coronavirus pandemic. *Cytokine & Growth Factor Reviews* **53(2)**:1–9 DOI [10.1016/j.cytogfr.2020.05.010](https://doi.org/10.1016/j.cytogfr.2020.05.010).
- Hu X, Downie JS, Ehmann AF.** Lyric text mining in music mood classification. Zenodo. DOI [10.5281/zenodo.1416790](https://doi.org/10.5281/zenodo.1416790).
- Hutto C, Gilbert E. 2014.** Vader: a parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media* **8(1)**:216–225 DOI [10.1609/icwsm.v8i1.14550](https://doi.org/10.1609/icwsm.v8i1.14550).
- Jain VK, Kumar S, Fernandes SL. 2017.** Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *Journal of Computational Science* **21(1)**:316–326 DOI [10.1016/j.jocs.2017.01.010](https://doi.org/10.1016/j.jocs.2017.01.010).
- Kamath U, Liu J, Whitaker J. 2019.** Convolutional neural networks. In: *Deep Learning for NLP and Speech Recognition*. Cham: Springer, 263–314.
- Kaur S, Kaul P, Zadeh PM. 2020.** Monitoring the dynamics of emotions during COVID-19 using Twitter data. *Procedia Computer Science* **177**:423–430 DOI [10.1016/j.procs.2020.10.056](https://doi.org/10.1016/j.procs.2020.10.056).
- Kaur C, Sharma A. 2020.** Twitter sentiment analysis on coronavirus using TextBlob. EasyChair. Available at <https://easychair.org/publications/preprint/Fd5m>.
- Khatua A, Khatua A, Cambria E. 2020.** Predicting political sentiments of voters from Twitter in multi-party contexts. *Applied Soft Computing* **97(1)**:106743 DOI [10.1016/j.asoc.2020.106743](https://doi.org/10.1016/j.asoc.2020.106743).
- Kocyyigit Y, Alkan A, Erol H. 2008.** Classification of EEG recordings by using fast independent component analysis and artificial neural network. *Journal of Medical Systems* **32(1)**:17–20 DOI [10.1007/s10916-007-9102-z](https://doi.org/10.1007/s10916-007-9102-z).
- Krizhevsky A, Sutskever I, Hinton G. 2012.** ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems (NIPS 2012)*. 25.

- Liu M, Zhou F, He J, Chen K, Zhao Y, Sun H. 2022.** Self-attention networks and adaptive support vector machine for aspect-level sentiment classification. *Soft Computing* 26:1–14 DOI 10.1007/s00500-022-06793-7.
- Loft LH, Pedersen EA, Jacobsen SU, Søborg B, Bigaard J. 2020.** Using Facebook to increase coverage of HPV vaccination among Danish girls: an assessment of a Danish social media campaign. *Vaccine* 38(31):4901–4908 DOI 10.1016/j.vaccine.2020.04.032.
- Majumdar A, Bose I. 2019.** Do tweets create value? A multi-period analysis of Twitter use and content of tweets for manufacturing firms. *International Journal of Production Economics* 216(2):1–11 DOI 10.1016/j.ijpe.2019.04.008.
- Mohammad SM, Kiritchenko S, Zhu X. 2013.** NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. *ArXiv preprint*. DOI 10.48550/arXiv.1308.6242.
- Mourad A, Srour A, Harmanai H, Jenainati C, Arafeh M. 2020.** Critical impact of social networks infodemic on defeating coronavirus COVID-19 pandemic: Twitter-based study and research directions. *IEEE Transactions on Network and Service Management* 17(4):2145–2155 DOI 10.1109/TNSM.2020.3031034.
- Naseem U, Khan SK, Razzak I, Hameed IA. 2019.** Hybrid words representation for airlines sentiment analysis. In: *Australasian Joint Conference on Artificial Intelligence*. Cham: Springer, 381–392.
- Ofcom. 2019.** News consumption in the U.K.. Available at <https://www.ofcom.org.uk/> (accessed 5 November 2021).
- Pastor CK. 2020.** Sentiment analysis of Filipinos and effects of extreme community quarantine due to coronavirus (COVID-19) pandemic. Available at <https://ssrn.com/abstract=3574385>.
- Pedersen EA, Loft LH, Jacobsen SU, Søborg B, Bigaard J. 2020.** Strategic health communication on social media: insights from a Danish social media campaign to address HPV vaccination hesitancy. *Vaccine* 38(31):4909–4915 DOI 10.1016/j.vaccine.2020.05.061.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. 2011.** Scikit-learn: machine learning in python. *Journal of Machine Learning Research* 12(October):2825–2830.
- Prabhakar Kaila D, Prasad DA. 2020.** Informational flow on Twitter–corona virus outbreak–topic modelling approach. *International Journal of Advanced Research in Engineering and Technology (IJARET)* 11(3):128–134.
- Preda G. 2022.** COVID-19 all vaccines tweets. Available at <https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets>.
- Pépin L, Kuntz P, Blanchard J, Guillet F, Suignard P. 2017.** Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets. *Computers & Industrial Engineering* 112(5):450–458 DOI 10.1016/j.cie.2017.01.025.
- Pérez A, Larrañaga P, Inza I. 2006.** Supervised classification with conditional Gaussian networks: increasing the structure complexity from Naive Bayes. *International Journal of Approximate Reasoning* 43(1):1–25 DOI 10.1016/j.ijar.2006.01.002.
- Radwan E, Radwan A. 2020.** The spread of the pandemic of social media panic during the COVID-19 outbreak. *European Journal of Environment and Public Health* 4(2):em0044 DOI 10.29333/ejeph/8277.
- Robertson S. 2004.** Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of Documentation* 60(5):503–520 DOI 10.1108/00220410410560582.

- Rustam F, Mehmood A, Ahmad M, Ullah S, Khan DM, Choi GS. 2020.** Classification of Shopify app user reviews using novel multi text features. *IEEE Access* **8**:30234–30244
DOI [10.1109/ACCESS.2020.2972632](https://doi.org/10.1109/ACCESS.2020.2972632).
- Saad E, Sadiq S, Jamil R, Rustam F, Mehmood A, Choi GS, Ashraf I. 2022.** Predicting death risk analysis in fully vaccinated people using novel extreme regression-voting classifier. *Digital Health* **8**:20552076221109530 DOI [10.1177/20552076221109530](https://doi.org/10.1177/20552076221109530).
- Sadiq S, Mehmood A, Ullah S, Ahmad M, Choi GS, On B-W. 2021.** Aggression detection through deep neural model on Twitter. *Future Generation Computer Systems* **114**:120–129
DOI [10.1016/j.future.2020.07.050](https://doi.org/10.1016/j.future.2020.07.050).
- Samuel J, Kashyap R, Betts S. 2018.** Strategic directions for big data analytics in E-commerce with machine learning and tactical synopses: propositions for intelligence based strategic information modeling (SIM). *Journal of Strategic Innovation and Sustainability* **13**(1):99–106.
- Schumaker RP, Jarmoszko AT, Labeledz CS Jr. 2016.** Predicting wins and spread in the premier league using a sentiment analysis of Twitter. *Decision Support Systems* **88**:76–84
DOI [10.1016/j.dss.2016.05.010](https://doi.org/10.1016/j.dss.2016.05.010).
- Shao C, Hui P-M, Wang L, Jiang X, Flammini A, Menczer F, Ciampaglia GL. 2018.** Anatomy of an online misinformation network. *PLOS ONE* **13**(4):e0196087
DOI [10.1371/journal.pone.0196087](https://doi.org/10.1371/journal.pone.0196087).
- Sharaff A, Gupta H. 2019.** Extra-tree classifier with metaheuristics approach for email classification. In: *Advances in Computer Communication and Computational Sciences*. Cham: Springer, 189–197.
- Sharma K, Seo S, Meng C, Rambhatla S, Liu Y. 2020.** COVID-19 on social media: analyzing misinformation in Twitter conversations. *ArXiv preprint*. DOI [10.48550/arXiv.2003.12309](https://doi.org/10.48550/arXiv.2003.12309).
- Shoukat SJ, Afzal H, Mufti MR, Sohail MK, Khan DM, Akhtar N, Hussain S, Ahmed M. 2021.** Analyzing COVID-19 impact on the researchers productivity through their perceptions. *Computers, Materials & Continua* **67**(2):1835–1847 DOI [10.32604/cmc.2021.014397](https://doi.org/10.32604/cmc.2021.014397).
- Škrlić B, Kralj J, Lavrač N, Pollak S. 2019.** Towards robust text classification with semantics-aware recurrent neural architecture. *Machine Learning and Knowledge Extraction* **1**(2):34
DOI [10.3390/make1020034](https://doi.org/10.3390/make1020034).
- Sohail MK, Akhtar N, Muhammad D, Afzal H, Mufti MR, Hussain S, Ahmed M. 2021.** Analyzing COVID-2019 impact on mental health through social media forum. *Computers, Materials, & Continua* **67**(3):3737–3748 DOI [10.32604/cmc.2021.014398](https://doi.org/10.32604/cmc.2021.014398).
- Staudemeyer RC, Morris ER. 2019.** Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. *ArXiv preprint* DOI [10.48550/arXiv.1909.09586](https://doi.org/10.48550/arXiv.1909.09586).
- Stieglitz S, Dang-Xuan L. 2013.** Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining* **3**(4):1277–1291
DOI [10.1007/s13278-012-0079-3](https://doi.org/10.1007/s13278-012-0079-3).
- Tang J, Deng C, Huang G-B. 2015.** Extreme learning machine for multilayer perceptron. *IEEE Transactions on Neural Networks and Learning Systems* **27**(4):809–821
DOI [10.1109/TNNLS.2015.2424995](https://doi.org/10.1109/TNNLS.2015.2424995).
- Tiwari P, Yadav P, Kumar S, Mishra BK, Nguyen GN, Gochhayat SP, Singhk J, Prasad M. 2018.** Sentiment analysis for airlines services based on Twitter dataset. In: *Social Network Analytics: Computational Research Methods and Techniques*. Amsterdam: Elsevier, 149.
- Villavicencio C, Macrohon JJ, Inbaraj XA, Jeng J-H, Hsieh J-G. 2021.** Twitter sentiment analysis towards COVID-19 vaccines in the Philippines using Naïve Bayes. *Information* **12**(5):204
DOI [10.3390/info12050204](https://doi.org/10.3390/info12050204).

- Vosoughi S, Roy D, Aral S. 2018.** The spread of true and false news online. *Science* **359(6380)**:1146–1151 DOI [10.1126/science.aap9559](https://doi.org/10.1126/science.aap9559).
- Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. 2020.** Public discourse and sentiment during the COVID-19 pandemic: using latent Dirichlet allocation for topic modeling on Twitter. *PLOS ONE* **15(9)**:e0239441 DOI [10.1371/journal.pone.0239441](https://doi.org/10.1371/journal.pone.0239441).
- Yang Y, Cui X. 2021.** BERT-enhanced text graph neural network for classification. *Entropy* **23(11)**:1536 DOI [10.3390/e23111536](https://doi.org/10.3390/e23111536).
- Ye X, Li S, Yang X, Qin C. 2016.** Use of social media for the detection and analysis of infectious diseases in China. *ISPRS International Journal of Geo-Information* **5(9)**:156 DOI [10.3390/ijgi5090156](https://doi.org/10.3390/ijgi5090156).
- Yu B. 2008.** An evaluation of text classification methods for literary study. *Literary and Linguistic Computing* **23(3)**:327–343 DOI [10.1093/lc/fqn015](https://doi.org/10.1093/lc/fqn015).
- Yu Y, Wang X. 2015.** World cup 2014 in the Twitter world: a big data analysis of sentiments in US sports fans' tweets. *Computers in Human Behavior* **48(1–2)**:392–400 DOI [10.1016/j.chb.2015.01.075](https://doi.org/10.1016/j.chb.2015.01.075).