Sentiments Analysis of Covid-19 Vaccine Tweets Using Machine Learning and Vader Lexicon Method

Vishakha Arya[1], Amit Kumar Mishra[2], Alfonso González-Briones[3,4,5]

[1,2]School of Computing- Computer Science & Engineering, DIT University, Dehradun 248001, India

[3]Research Group on Agent-Based, Social and Interdisciplinary Applications (GRASIA), Complutense University of Madrid, 28040 Madrid, Spain;

[4]BISITE Research Group, University of Salamanca. Calle Espejo s/n. Edificio Multiusos I+D+i, 37007, Salamanca, Spain: alfonsogb@usal.es

[5]Air Institute, IoT Digital Innovation Hub, Calle Segunda 4, 37188, Salamanca, Spain
Vishakha.27arya@gmail.com, aec.amit@gmail.com, alfonsogb@ucm.es

KEYWORDS	ABSTRACT
sentiment analysis, VADER lexicon, Twitter, vaccine tweets, Covid-19, classification model	*The novel Coronavirus disease of 2019 (COVID-19) has subsequently named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) have tormented the lives of millions of people worldwide. Effective and safe vaccination might curtail the pandemic. This study aims to apply the VADER lexicon, Text Blob, and machine learning approach: to analyse and detect the ongoing sentiments during the affliction of the Covid-19 pandemic on Twitter, to understand public reaction worldwide towards vaccine and concerns about the effectiveness of the vaccine. Over 200000 tweets vaccine-related using hashtags # Covid Vaccine # Vaccines # CornavirusVaccine were retrieved from 18 August 2020 to 20 July 2021. Data analysis conducted by VADER lexicon method to predict sentiments polarity, counts and sentiment distribution, Text Blob to determine the subjectivity and polarity, and compared with two other models such as Random Forest (RF) and Logistic Regression (LR). The results determine sentiments that public have a positive stance towards a vaccine follows by neutral and negative. Machine learning classification models performed better than the VADER lexicon method on vaccine Tweets. It is anticipated this study aims to help the government in long run, to make policies and a better environment for people suffering from negative thoughts during the ongoing pandemic.*

Vishakha Arya, Amit Kumar Mishra,
Alfonso González-Briones

Sentiments Analysis of Covid-19 Vaccine Tweets Using
Machine Learning and Vader Lexicon Method

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 4 (2022), 507-518
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

507

1. Introduction

In Jan 2020, China confirmed the human transmission of novel coronavirus. The WHO declared the outbreak as a pandemic. Now, it has been more than a year since the corona virus has prevailed in the society (Brunier, A. et al., 2020). In Jan 2021, a new mutated variant of coronavirus identifies in the United Kingdom called B.1.1.7, B.1.351 in South Africa, P.1 in Brazil, B.1.617 in India, B.1.526 in New York and many other mutants reported worldwide (Centre for Disease control and Prevention, 2021). These new variants have a high tendency of spreading more swiftly leads to an increased number of Covid-19 cases means more hospitalization and probably more deaths (Terry, M. 2021). Recently, researchers have reported a unique combination of «three mutant variants» or «Bengal strain» in the SARS-CoV-2 virus detected in India (Cherian, S., Potdar, et al., 2021). It is continuously evolving to elude the human immune system. It is continuously evolving to elude the human immune system. To curtail the COVID-19 pandemic vaccine is very important. As the vaccination program roll-out, people started sharing their views on social media sites like Twitter etc (World Economic Forum, 2020). Analyzing these views or opinions is mandatory to under- stand the sentiments of people globally (Khan, H. et al., 2020). To control the false lies related to vaccines among the public to prevent the spread of the pandemic (Chancellor et al., 2020). Monitoring sentiments of the public on social media sites (Chilwal, B. & Mishra, A. K., 2021) helps the government and other institutes to understand the public behaviour towards the vaccine and make effective policies to roll out immunization programs to mitigate the coronavirus pandemic (Saladino et al., 2020).

Sentiment analysis is one of the ways to analyze the opinions shared by people on Twitter by text analysis (Yazdavar et al, 2018). Social media data is highly unstructured, EmotionAI or affective computing to detect emotion using Natural Language Processing, textual analysis, deep learning, and machine learning models to extract and quantify the posts from social media sites on subjective text (Subhani et al., 2017). By using data mining techniques, extract the data from large databases. By applying natural language processing (NLP), identifying the thoughts and opinions of the text is known as opinion mining or sentiment analysis (Baheti et al., 2019). In this paper, large dataset of covid-19 vaccine related tweets by users globally has been considered (Ritchie, H et al., 2020). To specify these sentiments TextBlob and VADER lexicon methods to determine the polarity, subjectivity, and compound score of the target text (Nguyen et al., 2014). Data analysis was done by the VADER lexicon method to predict sentiments polarity, counts and sentiment distribution, TextBlob to determine the subjectivity and polarity (Aldarwish et al., 2017). It classifies the text as positive, negative, or neutral; also, how much the text is positive or negative (Abd El-Jawad et al., 2018). It further classifies based on polarity as (positive, negative, neutral, compound score) and machine learning classification model such as random forest (RF) and logistic regression (LR) to attain precision, recall, F1-score (suitability), and accuracy to compare the performance with lexicon rule-based method (Raichur et al., 2017). Furthermore, section 2 elaborates the study sites and methodology, section 3 results and experiments, section 4 discussions, and sections 5 conclusions. This study sceptic helps the government and other organization to develop trust in immunization program and the goal of achieving herd immunity.

2. Material and Methods

2.1. Study sites

Data extracted from Twitter is one of the largest repositories of social media post. Tweets were posted in the English language globally from 18 August 2020 to 20 July 2021 using Twitter API. Then,

Twitter Ids were hydrated or retrieved using Hydrator application to get tweets related to the covid vaccine like «CovidVaccine», «Vaccines», «CoronavirusVaccine», and «COVID19». Total tweets hydrated using related keywords were 207006 collected geospatially.

2.2. Sentiment Analysis

The filtered dataset of tweets related to the covid vaccine tweets pre-processed initially as raw data contains noise, missing values, re-tweets, and outliers (Deshpande et al., 2017). A hybrid model developed for subjective sentiment analysis. A rule-based lexicon Valance Aware Dictionary and Sentiment Reasoner (VADER) keyed to the expression expressed on social media. Text Blob analyses sentiments on subjectivity and polarity. Subjectivity refers to expression of a text given as [0, 1] whereas polarity of text indicated it is positive or negative [-1, 1]. Acquired subjectivity, polarity, counts, compound score, and sentiment distribution combined with the machine learning classification model: Random Forest, and Logistic regression (Shatte et al., 2019) (Rana, M. et al., 2022).

2.2.1. Text pre-processing

For building any machine learning model, data pre-processing is the primary task (Dubey, 2020). Text pre-processing involves removing noise, outliers, conversion to lower case, removing URL, removing punctuations, stemming, lemmatization, stopwords, and normalization. Text encoded into a numeric vector as machine learning understands the numeric format. N-grams, TF-IDF, and count vectorization are some encoding techniques to quantify the text into numeric (Desai et al., 2016). N-grams represent sequencing of n-words in each text or corpus. 1-gram means unigram, 2-gram means bigram, and 3-gram means trigram (Cavazos-Rehg et al., 2016). Figure 1 shows N-gram representation unigram, bigram, and trigram of vaccine tweets.

2.2.2. Lexicon Methods

Covid-19 vaccination tweets (n=2, 00,000) pre-processed using Natural Language Processing (NLP) with Natural Language Toolkit (NLTK) in python and «re» package. To normalize the text noise removal, outliers, text enrichment, and POS tagging improve the efficacy to analyses the sentiments (Bonta, V. et al., 2019).

VADER lexicon method is a tool used for labelling according to the semantic orientation of a text positive, negative, and neutral. It classifies the intensity of a text, how much positive or negative. It classified sentiment analysis into three steps: subjective, semantic, and polarity. Sentiment classification is done on subjective text rather than objective text. The subjective text defines the opinion, thoughts, emotions of a person help for semantic classification, and objective text carries relevant information about the topic. SentimentIntensityAnalyzer() creates an object sid.polarity_ scores() provides dictionary gives polarity of a text classifies into positive, negative, neutral, and compound score. Sentiment scores were accredited to the text according to their semantic orientation if compound score > 0.5 assigns as «positive», compound score < -0.5 assigns as «negative», and compound score = 0 assigns as «neutral». Table 1 shows sentiment polarity of tweets assigning their scores and values relat- ed to semantic orientation. The compound score lexicon calculated by aggregating all sentiment scores positive and negative sentiments then normalized between -1(negative) to +1 (positive). Equation (1) intimates the «Y» sum of all polarity words. Alpha (α) is a constant word.
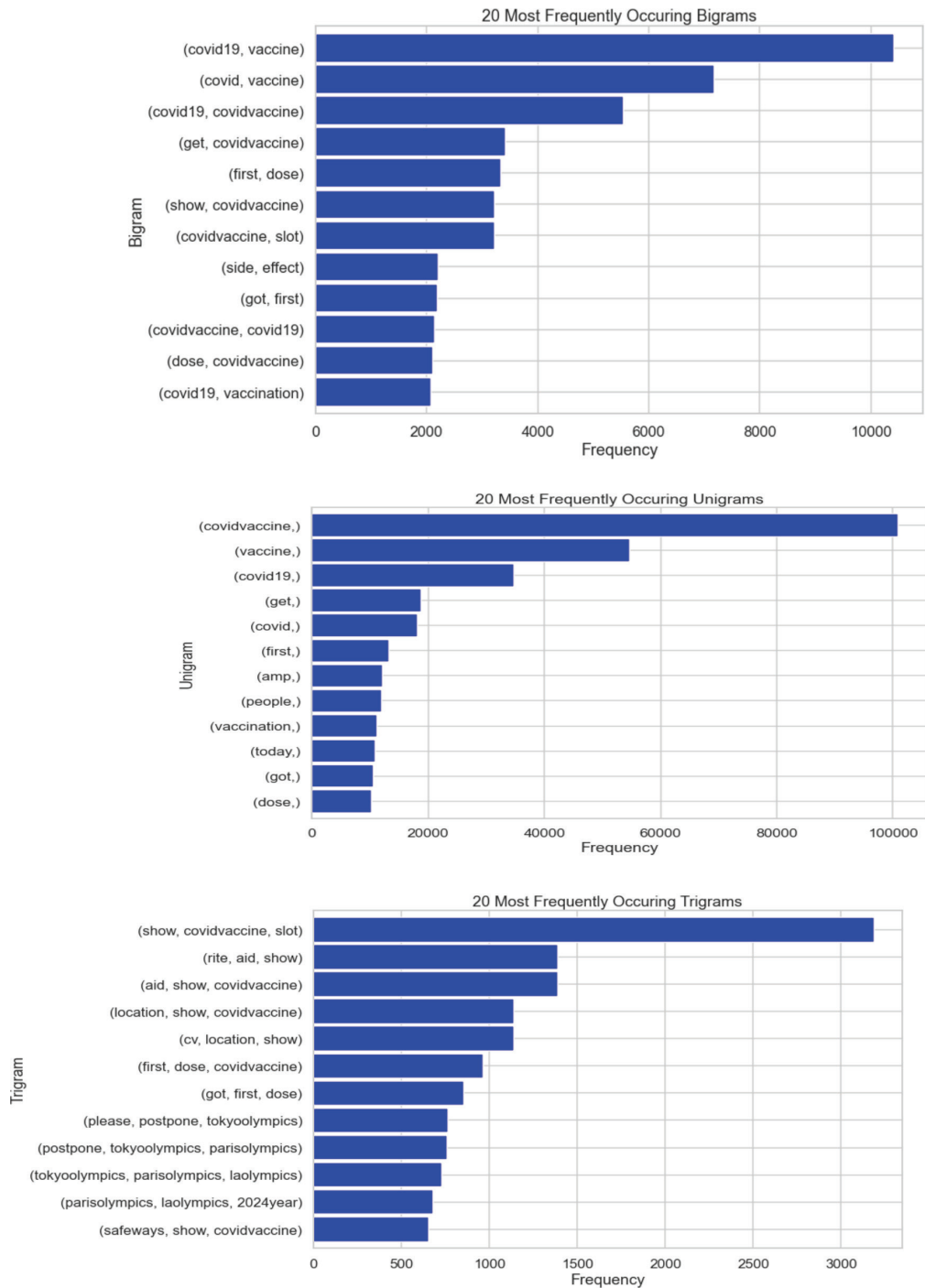
Figure 1. Unigram, Bigram, and Trigrams representation of covid-19 vaccine tweets

Vishakha Arya, Amit Kumar Mishra,
Alfonso González-Briones

Sentiments Analysis of Covid-19 Vaccine Tweets Using
Machine Learning and Vader Lexicon Method

$$\text{Compound Score} = \frac{Y}{\sqrt{Y^2 + \alpha}} \tag{1}$$

By leveraging TextBlob (Gbashi S et al., 2021) library sentiment scores can be extracted from a text or document. Subjectivity and polarity of a text obtained with Text Blob. Polarity assigns positive and negative ranges from -1 to 1 (negative to positive). Subjectivity tells the intensity of a text thoughts, emotions, and expressions ranges from 0 to 1 (objective to subjective). Table 2 shows the subjective and polarity of vaccine tweets.

Table 1. Sentiment Polarity of COVID-19 vaccine tweets pos, neg, neu, compound score and its value

Text	Positive	Negative	Neutral	Compound	Value
australia manufacture vaccine give citizens	0.231	0.000	0.769	0.5106	Positive
Coronavirus vaccine coronavaccine covid vaccine	0.420	0.000	0.580	0.4404	Positive
deaths due affected countries red rpiryani sh...	0.000	0.211	0.789	-0.1531	Negative
michellegrattan conversationedu passes leaders	0.000	0.000	1.000	0.0000	Neutral
protection could last 6 months click read mode.	0.000	0.000	1.000	0.000	Neutral

Table 2. Subjectivity and polarity of vaccine tweets

Text	Subjective	Polarity
vaccines r wonderful protecting viruses exist	1.0	1.0
whipclyburn dearth warrant lack funding inovio	1.0	0.3
realdonaldtrump gonna go worst president time	-1.0	1.0
boring life its okay tono tbe okayep16 apple boy	-1.0	1.0
best luck russia hope vaccine works russianvac	1.0	0.3

Sentiment score or polarity of tweets represented using different graphs to improve the visualization and understanding. Sentiments counts of tweets depicts people have more positive attitude towards vaccine shown in the graph below Fig. 2(a). Correlation matrix represents the relationship between all the sentiment scores and compound score of vaccine tweets. This represents the linear relationship Fig. 2(b). Distribution of sentiment score positive, negative, and neutral tweets shown in Fig. 2(c).

Further, feature extraction using (Term frequency and Inverse document frequency) TF-IDF transform the text into a vector (Bania R. K., 2020). It reduces the occurrence of words in a corpus. Feature extraction improves the efficacy of the model. Count Vectorization converts or encodes text into vector format. It counts the number of occurrence of words in text or a document (Das et al., 2018).

Vishakha Arya, Amit Kumar Mishra,
Alfonso González-Briones

Sentiments Analysis of Covid-19 Vaccine Tweets Using
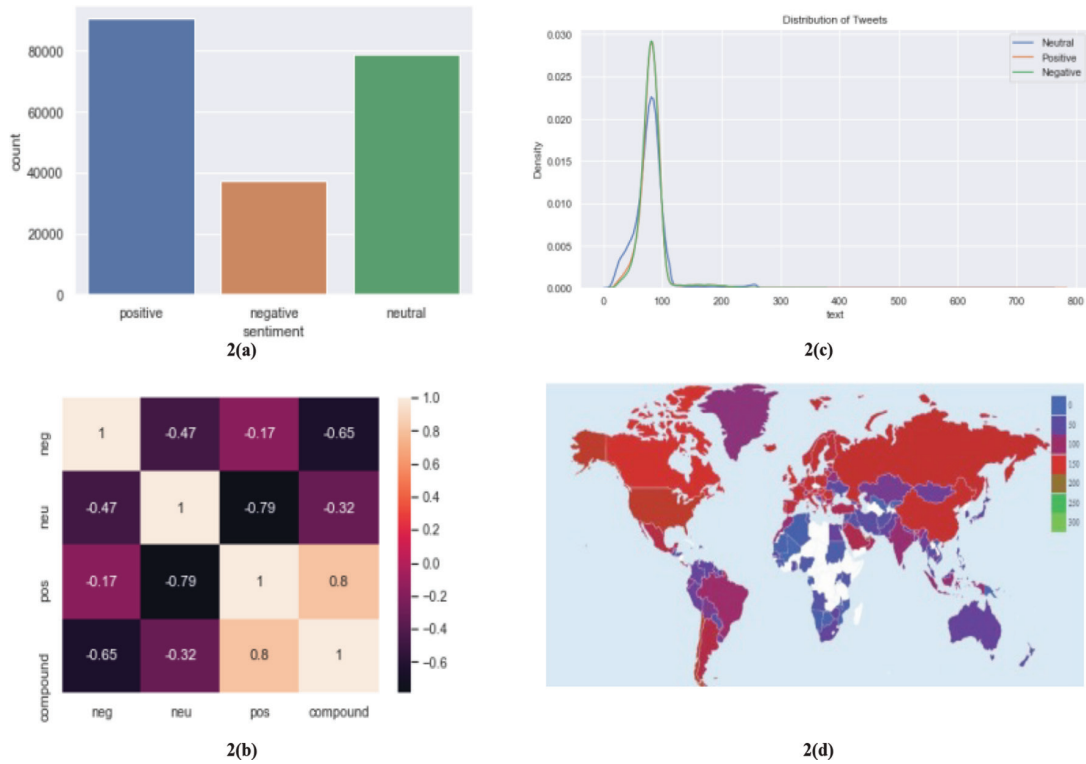Machine Learning and Vader Lexicon Method

*Figure 2. (a). Sentiment counts of vaccination tweets positive, negative, neutral.
(b). Correlation matrix of sentiment score pos, neg, neu, and compound of tweets.
(c). Sentiments distribution of tweets positive, negative, and neutral. (d). Geospatial mapping
of Covid-19 vaccination doses per million of coronavirus disease globally*

Source: https://ourworldindata.org/covid-vaccinations

2.3. Machine Learning Models

Building ML model for sentiment analysis after pre-processing is an important task (Khan et al., 2014). Machine Learning helps to extract insights for making accurate decision. Labelled datasets as positive, negative, or neutral fed into the model. To predict the sentiments of covid-19 vaccine tweets, Random Forest and Logistic Regression implemented to the dataset (Troussas et al., 2013) (Arya, V., & Mishra, A., 2021).

2.3.1. Random Forest

Random Forest also called "Random decision forest" ensembles learning aggregating multiple decision trees together to get more accurate and precise result. It is used for both classification and regression. It reduces over fitting as its hyper parameter return good results. Random forest takes prediction from each tree which enhances its accuracy rather than relying on one decision tree. It is a multiclass problem that performs well with both numeric and absolute features. Perform voting for each result with the highest vote as the result (Arya, V. et al., 2022).

*Vishakha Arya, Amit Kumar Mishra,
Alfonso González-Briones*

Sentiments Analysis of Covid-19 Vaccine Tweets Using
Machine Learning and Vader Lexicon Method

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 4 (2022), 507-518
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

512

2.3.2. Logistic Regression

LR is a probabilistic model used for binary or dichotomous classifications such as yes/no, malignant/benign, and pass/fail. Logistic function or sigmoid function used to map values between 0 and 1 provides a constant output. It is sub divided into three parts: binomial, multinomial or ordinal. In binomial LR, prediction based on dependent variable as 0 and 1. Multinomial LR based on multiclass prediction having more than one outcome, and lastly ordinal LR outcome or prediction is not in order.

$$\text{Sigmoid function (x)} = \frac{1}{1 + e^{-n}} \tag{2}$$

3. Experiments and Results

The lexicon methods implemented using NLTK library in python, and machine learning classification models implemented using Scikit-learn library in python (M. Rathi et al., 2018).

VADER lexicon and Text Blob tools used for classifying sentiment of a text as positive, negative, neutral, compound, subjectivity, and polarity. Once installing vadersentiment call the object SentimentIntensityAnalyser(), then polarity obtained by using polarity_scores() as positive, negative, neutral, or compound. Compound polarity or score indicates the aggregate of all the polarities range from -1 to +1. If the compound score >0.5 assigns "pos", compound< 0.5 assigns "neg", and compound = 0 assigns "neu". Fig. 3 shows graphical representation of sentiment polarity of tweets pos, neg, neu, and compound. TextBlob is an NLP library used in python for processing text. Text Blob analyzes the polarity and subjectivity of a text which indicates the positive, negative, and intensity. Polarity assigns the text as positive and negative ranges from 1 to -1. Subjectivity assigns the text as subjective and objective ranges from 1 to 0. Fig.4 graph shows the subjectivity and polarity of tweets.

The machine learning classification models split dataset for train_test_validation in ratio 80:20 implies 80% retained for training and 20% for testing (Gyeongcheol Cho et al, 2019). Evaluation measures used for the classification models are precision, recall, f1-score, and accuracy. For multiclass classification, it categorizes into three: positive, negative, and neutral. True positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are evaluation measures used for building models. The confusion matrix can be map for both binary classification and multi-class classification. Table 3 illustrates the confusion matrix of Random Forest, Logistic Regression, and VADER lexicon for pos, neg, and neu their precision, recall, F1-score, accuracy, macro average and weighted average. For qualitative evaluation of all measures as macro-average and weighted average precision, recall, and f1-score measured used to evaluate train models for multi-class classification. The confusion matrix of machine learning classification model Random Forest and Logistic regression shown in Fig. 5.

Macro average – It is the average of precision and recall of all classes. In this equation, precision noted as Pr, recall as Re and macro avg. as M.

$$\text{Macro avg. precision (MP)} = (\text{Pr of c1} + \text{Pr of c2})/2 \tag{3}$$

$$\text{Macro avg. Recall (MR)} = (\text{Re of c1} + \text{Re of c2})/2 \tag{4}$$

Macro-average F1-score –F1 score also define as "Suitability". Harmonic mean concatenates the precision and recall and state balance between both.

$$\text{F1- score} = 2 *(\text{Pr}*\text{Re})/(\text{Pr}+\text{Re}) \tag{5}$$

Vishakha Arya, Amit Kumar Mishra,
Alfonso González-Briones

Sentiments Analysis of Covid-19 Vaccine Tweets Using
Machine Learning and Vader Lexicon Method

Weighted average- It is the aggregation of all the precisions scores. In this equation, precision denoted as Pr and weighted average as Wg.

$$\text{Weighted avg.} = \frac{(TPr1 + TPr2 + TPr3 + ... + Tprn)}{\text{Sum of all classes}} \quad (6)$$

Accuracy – It states the overall score of the classifier. Summation of true positive (TP) and true negative (TN) divides by sum of all evaluation measures (TP+TN+FP+FN).

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (7)$$
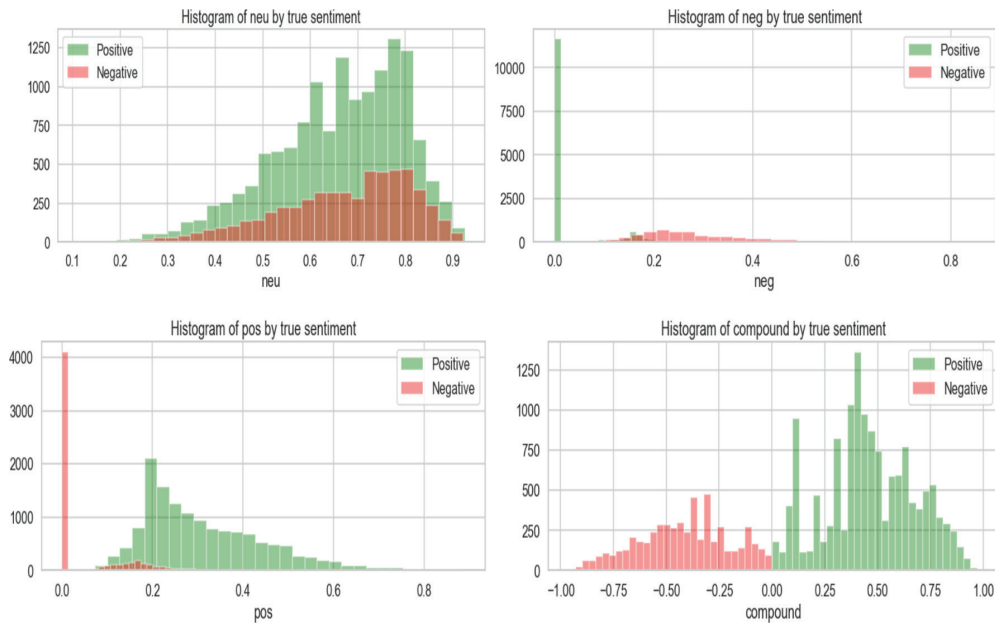

Figure 3. Sentiment polarity of tweets neutral, negative, positive, and compound of vaccine tweets
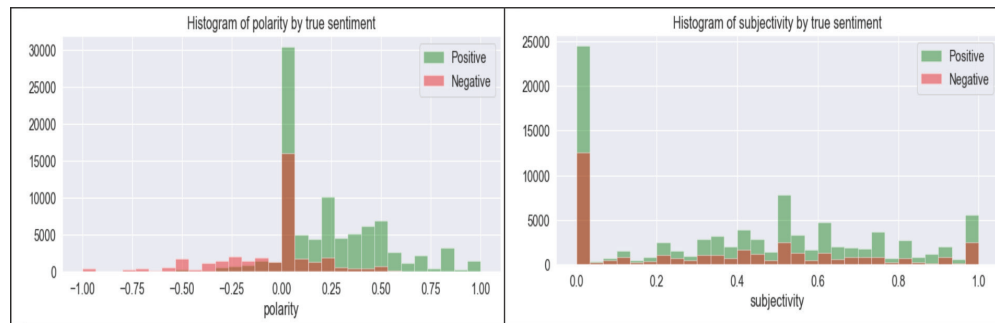

Figure 4. Subjectivity and Polarity of covid-19 vaccine tweets positive and negative of vaccine tweets

Random Forest and Logistic Regression achieved an accuracy of >90% shows supervised learning models performs better than unsupervised learning Vader Lexicon attain an accuracy of 64%. RF best suited for classification and regression. It can manage accuracy even over missing values and can handle complex and large dataset. LogisticRegression imported in python from sci-kit-learn moreover, input to the model in the two-dimensional array. Once training and testing data prepared with n_gram (1,1) transforming using countVectorizer create an instance of logistic regression. Parameters defined for the model random_state=0, max_iteration=100, n_estimator=100, keeping C=1. Multiclass classification model evaluates precision, recall, and f1-score for negative, positive, neutral, macro average, and weighted average. The model performed well obtained an accuracy of 92%.

In Random Forest, importing RandomForest from scitkit learn parameters random_state=0, n_estimator=100, test_size=0.20, bootstrap=True, keeping n_jobs=1.Removing noise, outliers and missing values enhance the performance of the model later split data into train and test. Evaluating all the measures precision, recall, f1-sore, macro avg, and weighted avg acquired an accuracy of 90%. Comparing to supervised machine learning models lexicon method performed low obtained an accuracy of 64%. Keeping, vader_polarity as compound, target as positive and negative, calculate the evaluation matrix and macro avg, and weighted avg. The VADER lexicon rule-based method performed least compared to other models' logistic regression (LR) and random forest (RF).
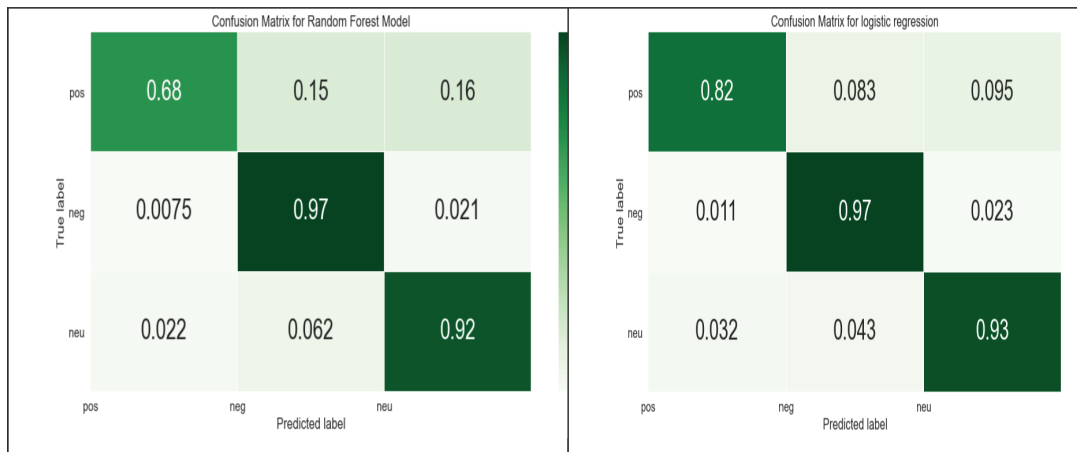

Figure 5. Confusion Matrix of Random Forest Classifier, and Logistic Regression of classification models

Table 3. Confusion Matrix of Random Forest, Logistic Regression, and VADER lexicon

Model		Precision	Recall	F1-score	Accuracy
Logistic Regression	Negative	0.89	0.82	0.85	
	Neutral	0.92	0.97	0.94	
	Positive	0.94	0.93	0.93	0.92
	Macro Avg.	0.92	0.90	0.91	
	Weight Avg.	0.92	0.92	0.92	

(continued)

Table 3. Confusion Matrix of Random Forest, Logistic Regression, and VADER lexicon (continued)

Model		Precision	Recall	F1-score	Accuracy
Random Forest	Negative	0.91	0.68	0.78	
	Neutral	0.87	0.97	0.92	
	Positive	0.92	0.92	0.92	0.90
	Macro Avg.	0.90	0.86	0.87	
	Weight Avg.	0.90	0.90	0.89	
VADER lexicon	Negative	1.00	1.00	1.00	
	Positive	0.34	1.00	0.50	0.64
	Macro Avg.	0.45	0.67	0.50	
	Weight Avg.	0.52	0.64	0.55	

4. Discussions

Most of the people globally have affected due to Covid-19 maybe it emotional, mental, social, and financial. The results obtained by the VADER lexicon method, are less than the machine learning classification models, including parameters precision, recall, and f1-score of tweets. In this analysis, Covid-19 vaccine tweets deduced from the Twitter database pre-processed with NLTK package using different Python libraries, the sentiment polarity of each text determined as pos, neg, neu, and compound score. Based on the compound score classified as positive, negative, and neutral. The stance of tweets after analysis observed to have positive over neutral and negative stance. The machine learning approach has outperformed lexicon methods Vader and TextBlob. The machine learning classification model implemented on the vectorized tweets and target. The Logistic Regression shows the highest accuracy in sentiment analysis of vaccine tweets with 92% compared to Random Forest acquired an accuracy of 90%. VADER lexicon rule-based method achieved an accuracy of 64%. RF and LR show better results on unseen tweets and classify them into three: positive, negative, and neutral. However, LR is also a good choice for text analysis.

5. Conclusion

With the proliferation of social media has become challenging to analyze the sentiments of data. This paper retrieved vaccination tweets from the Twitter database, pre-processed the tweets to remove noise and outliers, and classify them into positive, negative, or neutral. NLTK is a python library package used for the analysis of sentiments from the text. Machine learning classification models implemented to the pre-processed dataset employed Logistic Regression with an accuracy of 92%. Random Forest Classifier with TF-IDF vectorizer with an accuracy of 90% and VADER lexicon method achieved an accuracy of 64%. Logistic regression performed well over the vaccination tweets. The aim of this paper to evaluate the sentiments of the public vaccination of COVID-19 tweets, at an early phase helps the governments to make intervening policies to regulate immunization program. Evaluation of vaccination tweets provides an idea of public mindset and limiting fake news related to vaccination. This study sceptic helps the government and other organization to develop trust in immunization

program and the goal of achieving herd immunity. Analyzing mental health during covid-19 impacting most of the countries, and the number of beds in hospitals worldwide. The future work, the dataset needs to be embedded in the different classification model to analyze the stance of sentiments. Another language should be incorporated while analyzing the sentiments from social media.

References

Abd El-Jawad, M. H., Hodhod, R., & Omar, Y. M. (2018, December). Sentiment analysis of social media networks using machine learning. In *2018 14th international computer engineering conference (ICENCO)* (pp. 174-176). IEEE.

Aldarwish, M. M., & Ahmad, H. F. (2017, March). Predicting depression levels using social media posts. In *2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS)* (pp. 277-280). IEEE.

Arya, V., & Mishra, A. (2021). Machine learning approaches to mental stress detection: a review. *Annals of Optimization Theory and Practice*, *4*(2), 55-67. https://doi.org/10.22121/aotp.2021.292083.1074

Arya, V., Mishra, A. K. M., & González-Briones, A. (2022). Analysis of sentiments on the onset of COVID-19 using machine learning techniques.

Baheti, R., & Kinariwala, S. (2019). Detection and analysis of stress using machine learning techniques. *Int. J. Eng. Adv. Technol*, *9*, 335-342.

Bania, R. K. (2020). COVID-19 public tweets sentiment analysis using TF-IDF and inductive learning models. *INFOCOMP Journal of Computer Science*, *19*(2), 23-41.

Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon-based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1-6.

Brunier, A., & Drysdale, C. (2020). COVID-19 disrupting mental health services in most countries, WHO survey. *World Heal Organ*.

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in human behavior*, *54*, 351-357.

Centre for Disease control and Prevention, 2021. Variants of the virus that causes COVID-19. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, *3*(1), 1-11.

Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., ... & Panda, S. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. bioRxiv 2021: 2021.04. 22.440932.

Chilwal, B., & Mishra, A. K. (2021). Extraction of Depression Symptoms From Social Networks. *The Smart Cyber Ecosystem for Sustainable Development*, 307-321.

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S. H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation*, *16*(4), 262.

Das, B., & Chakraborty, S. (2018). An improved text sentiment classification model using TF-IDF and next word negation. *arXiv preprint arXiv:1806.06407*.

Desai, M., & Mehta, M. A. (2016, April). Techniques for sentiment analysis of Twitter data: A comprehensive survey. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 149-154). IEEE.

Vishakha Arya, Amit Kumar Mishra, Alfonso González-Briones

Sentiments Analysis of Covid-19 Vaccine Tweets Using Machine Learning and Vader Lexicon Method

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 4 (2022), 507-518
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

517

Deshpande, M., & Rao, V. (2017, December). Depression detection using emotion artificial intelligence. In *2017 international conference on intelligent sustainable systems (iciss)* (pp. 858-862). IEEE.

Dubey, A. D. (2020). Twitter sentiment analysis during COVID-19 outbreak. *Available at SSRN 3572023*.

Gbashi, S., Adebo, O. A., Doorsamy, W., & Njobeh, P. B. (2021). Systematic delineation of media polarity on COVID-19 vaccines in Africa: computational linguistic modeling study. *JMIR medical informatics*, *9*(3), e22916.

Khan, H., Srivastav, A., & Mishra, A. K. (2020). Use of classification algorithms in health care. In *Big Data Analytics and Intelligence: A Perspective for Health Care*. Emerald Publishing Limited.

Khan, M., Rizvi, Z., Shaikh, M. Z., Kazmi, W., & Shaikh, A. (2014). Design and implementation of intelligent human stress monitoring system. *International Journal of Innovation and Scientific Research*, *10*(1), 179-190.

Nguyen, T., Phung, D., Dao, B., Venkatesh, S., & Berk, M. (2014). Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, *5*(3), 217-226.

Raichur, N., Lonakadi, N., & Mural, P. (2017). Detection of stress using image processing and machine learning techniques. *International journal of engineering and technology*, *9*(3), 1-8.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., ... & Roser, M. (2020). Coronavirus pandemic (COVID-19). *Our world in data*.

Rana, M., Rehman, M. Z. U., & Jain, S. (2022, February). Comparative Study of Supervised Machine Learning Methods for Prediction of Heart Disease. In *2022 IEEE VLSI Device Circuit and System (VLSI DCS)* (pp. 295-299). IEEE.

Rathi, M., Malik, A., Varshney, D., Sharma, R., & Mendiratta, S. (2018, August). Sentiment analysis of tweets using machine learning approach. In *2018 Eleventh international conference on contemporary computing (IC3)* (pp. 1-3). IEEE.

Saladino, V., Algeri, D., & Auriemma, V. (2020). The psychological and social impact of Covid-19: new perspectives of well-being. *Frontiers in psychology*, 2550.

Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, *49*(9), 1426-1448.

Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., & Malik, A. S. (2017). Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access*, *5*, 13545-13556.

Terry, M. (2021). UPDATED comparing COVID-19 vaccines: timelines, types, and prices. *BioSpace. April 23*.

Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013, July). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In *IISA 2013* (pp. 1-6). IEEE.

World Economic Forum, 2020. https://www.weforum.org/agenda/2020/08/covid-19-coronavirus-mental-health-well-being-countries/

Yazdavar, A. H., Mahdavinejad, M. S., Bajaj, G., Thirunarayan, K., Pathak, J., & Sheth, A. (2018, June). Mental health analysis via social media data. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 459-460). IEEE.

Vishakha Arya, Amit Kumar Mishra, Alfonso González-Briones

Sentiments Analysis of Covid-19 Vaccine Tweets Using Machine Learning and Vader Lexicon Method

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 4 (2022), 507-518
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

518