Sentiment Analysis on Covid-19 Vaccination Using Machine Learning Techniques

Ashish A. Bhalerao[1]([✉]), Bharat R. Naiknaware[2], Ramesh R. Manza[3], and Shobha K. Bawiskar[4]

[1] Department of CS and IT, Dr. B.A.M. University, Aurangabad, MH, India
aashish.bhrao@gmail.com
[2] Dr. G. Y. Pathrikar College of CS and IT, MGM University, Aurangabad, MH, India
[3] Department of CS and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MH, India
[4] Government Institute of Forensic Science, Aurangabad, MH, India

Abstract. In this research work, we have performed Machine Learning and Lexicon Based Techniques to identify and analyze user's expression or opinion on covid-19 vaccination from social media platform that is Twitter and acquainted the bulk tweets from 01 June 2021 to August 2021 using various twitter hashtags. Machine learning based Classifiers are used for investigating the evaluation performance of Algorithms. Real time datasets and machine Learning Algorithms are compared with Best Data classification Evaluation based on the size of train data also another approach is to investigating the polarity by using Lexicon Based approach for this Bing Liu Lexicons and Stanford University Lexicons are used. The global pandemic has created the medical emergency and stops the many regular activities. The whole world in the lockdown or quarantine to because Coronavirus disease. Among them, Covaxin, Covishield, Pfizer, Moderna and SputnikVare popular. Universally publicare articulating opinions on protection and success of the vaccines on social media. Research article shows, such tweets are collected from developer Application Management using a Twitter API. Unprocessed tweets are kept and preprocessed through Machine Learning techniques. Users opinion are predicted using a Classifiers Decision Tree, Support Vector Machine, K NN Algorithm and Naïve Bayes. Comparative machine learning classifiers study here comparative analysis is got highest accuracy of 97% for Decision tree with Covaxin dataset, Support vector machine with 94% for SputnikV, Naïve Bayes got highest accuracy of 95 for Covishield dataset and KNN got Highest accuracy of 96% for Covaxin. The Lexicon Based polarity classifies the score into three users opinions, positive, negative, and neutral. Result shows that, Covaxin shows 28.14% positive, 12.5% negative, and 59.36% neutral sentiment. Covishield shows 17.62% positive, 15.04% negative, and 67.34% neutral sentiment. Moderna shows 23.68% positive, 19.28% negative, and 57.02% neutral Pfizer shows 18.28% positive, 34.06% negative, and 47.66% neutral, SputnikV shows 24.62% positive, 14.1% negative, and 61.28% neutral.

Keywords: Machine Learning · Sentiment Analysis · Lexicons · COVID-19 · Covaxin · Covishield · Sputnik

1 Introduction

Data Science is focused on interdisciplinary domains and useful for taking decisions. Effective Vaccines are needed to save lots of lives throughout the worldwide epidemics such as COVID19. The community looks to COVID-19 vaccination progression must be considered sensibly in directive to know the users sentiments and fears to it. To know more about exact information about covid-19 vaccines are from who are taken the vaccination and they are express their opinions. In this research article studied and understands the advantages of social media. Now a Social media has become an important tool for gaining insights about any domain. At the time of Covid-19 pandemic social media applications are playing a key role in users thoughts on various topics sharing. About Vaccination side effects and results confusion is one of the serious issues in realizing herd immunity and suppressing the COVID-19 epidemic. To consider this approach our focus on analyze user opinions on COVID-19 vaccination process. The world face a main corona virus epidemic from the year 2019. The virus infects fast through various ways. All nations lock peoples to avoid the virus. Vaccinations, including Covaxin, Covishield, Pfizer, Moderna, SputnikV have been permitted [1]. This research article, tweet analysis is based on people's opinions about official covid-19 vaccines on social media Twitter. Datasets collected, Covaxin, Covishield, Pfizer, Moderna, SputnikV. These tweets are preprocessed using Machine learning techniques.

In this research article studied the users opinion on Pfizer, Modern, AstraZeneca and Johnson & Johnson. The total posts in each nation for time period of month of Jan 2020 to Apr 2020, May 2020 to Aug 2020 and Sept 2020 to Dec 2020 was plotted [2]. The use of opinion Analysis impacts on each domain like product analysis, Recommendation system, prediction on healthcare and analytics [3].

After declaration of vaccination and governments announces the policy about vaccination. More peoples hesitates about its impact and side effects [4]. In the month of Nov 9, 2020, when the vaccination drive starts and many people are reacting on social media about their effectiveness [5].

2 Survey of Literature

The sentiment analysis is techniques it is used to identifying the users expressions and for that Sentiment Analysis, Machine learning, Natural language processing are popular techniques are effectively used. In this research article here perform the twitter application management and collect real time hashtags discussion on covid-19 vaccination. The twitter general public data collection and preprocessing techniques are applied. Our investigation detected that unigram Sentiment Analysis for all five datasets. Lexicons are used Bing Liu and Sentiment140 are used for interpreting the data. The study is completed on the tweets which are related to the COVID-19 vaccination. Also focused on closest the users approaches of the COVID-19 vaccination process on twitter as a social media platform using machine learning. Maximum of the described sentiments that debated the vaccines effectiveness, security, and the distribution plans of Governments and the plans to safe the dosages for their people. This research analyzed the users opinions since the vaccination drives was started. Logistic Regression classifiers

shows highest correctness was 97.3%, SVM model that shows correctness of 96.26% and MNB model shows correctness of 88%. [1].

In the raw dataset 16 attributes where collected and then apply preprocessing for removal of noise and outliers. The users required towards distinguish whether present vaccine can stop the spread of the COVID19 [2].

The results indicate that the Machine learning classification techniques for product-reviews has achieved the maximum classification correctness in comparison with Classification techniques [3].

Vaccine uncertainty slowed due to the few reasons protection, doubt about political forces driving the COVID-19 epidemic, a deficiency of information about the vaccine, confusing content of social media [4].

3   Contribution of Work

Through data Analysis provides the popularity of users on Vaccination and they may able to understand the sentiments about vaccines.

To Show the Effective use of Machine learning and Sentiment Analysis in Medical domain.

Our study emphasized the vital essential for communicating public services with the humanity from diverse social and instructive families in direction to rise the vaccination consciousness and authenticate analysis. Study the various datasets related to Covid-19 Vaccination and identifying the insights in the form of Positive opinions, Negative opinions and Neutral opinions.

4   Methodology

The machine learning common approach is designed it performs the Tweet collection, Tweet Preprocessing, Train the Dataset, Test the Dataset and apply the classifiers and obtain the results (Fig. 1).
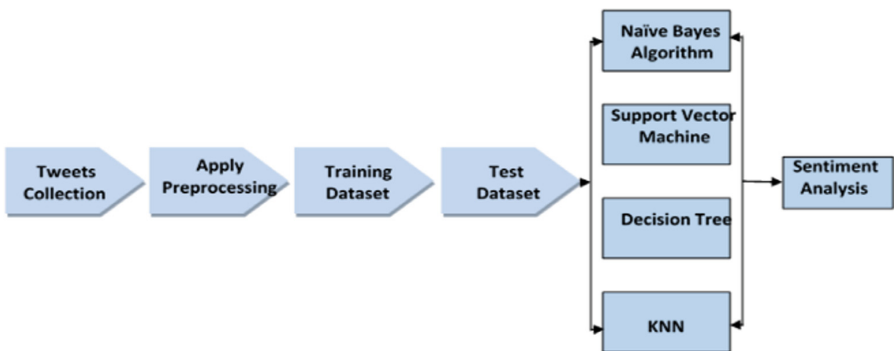

Fig. 1. Methodology of Machine Learning Approach

Table 1. Collection of Tweets

Sr. No	Dataset	Duration	Total Size
1	Covaxin	01 June to August 2021	5000
2	Covishield	01 June to August 2021	5000
3	Pfizer	01 June to August 2021	5000
4	Moderna	01 June to August 2021	3500
5	SputnikV	01 June to August 2021	5000

4.1 Collection of Tweets

Preprocessed dataset is not accessible for Vaccination Analysis so in this research collect twitter datasets from Twitter API and create twitter Management application. For collection of tweets from twitter following are the basic R Packages required library (twitteR), library (OAuth), library (plyr), library (tm) etc. (Table 1).

4.2 Data Preprocessing

Data preprocessing needs because in the proposed work we get raw data with 16 attributes of twitter datasets. Data preprocessing removed the noisy and duplicated data and convert into the quality data. Data Preprocessing Removing the URLs, Data Filtering, Removing Special Characters, Removal of Retweets, Usernames, Remove Punctuations and symbols, Usage of Web links, Hashtags, Tokenization, Exclamation and question marks, Letter Repetition, Negations.

4.3 Train Datasets

In this research work combining the machine learning and lexicon based technique. The Features are created. The Machine Learning classifiers applied for data classifications improved the accuracy with 70:30 Ratio where 70% Train Data and 30% Test Data.

4.4 Test Datasets

In this research here test following twitter dataset. Covaxin, Covishield, Pfizer, Moderna, SputnikV.

4.5 Lexicon Based Approach

In this research work used Stanford University sentiment lexicons, it contains total 1,600,000 from that half of tweets are from Positive and remaining half from Negative Lexicons. Tweets collected and prepared by Stanford University, where the tweets are categorized based on an occurrence of positive and negative score.

Also used Bing Liu sentiment lexicons. In the Sentiment Lexicon method, the Bing Liu Dataset having 2,006 and 4,783 Positive and Negative lexicons respectively in this dataset around 6789 words (Fig. 2).
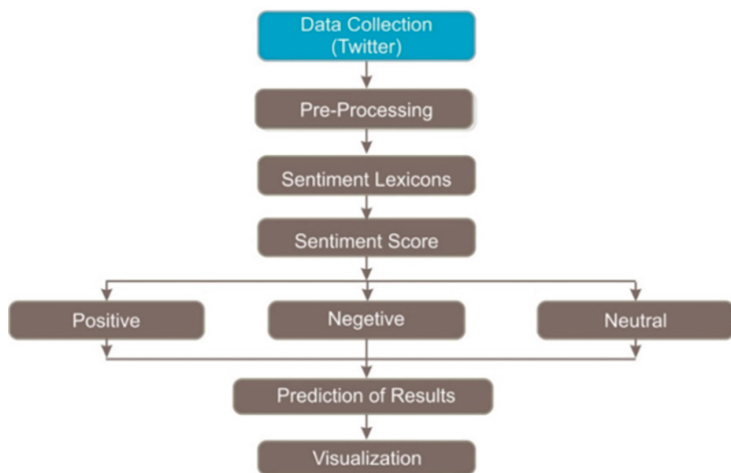
Fig. 2. Lexicon Based Approach

Table 2. Peoples Opinion Analysis on Vaccination

Name of Vaccination	Positive	Negative	Neutral	Total
Covaxin	1407	625	2968	5000
Covishield	881	752	3367	5000
Moderna	829	675	1996	3500
Pfizer	914	1703	2383	5000
SputnikV	1231	705	3064	5000

4.6 Sentiment Score

The Key insights Opinions of users to produce opinion of each Sentence. Sentiment Score = sum (pos.matches) – sum(neg.matches) (Table 2).

In the above Opinion Mining of Covaxin Twitter Dataset total of 5000 tweets are extracted, in which 1407 tweets are categorized as positive, 625 as negative and remaining 2968 are categorized as neutral tweets. In Covishield Dataset total of 5000 tweets are extracted, in which 881 tweets are categorized as positive, 752 as negative and remaining 3367 are categorized as neutral tweets. In Moderna Dataset total of 3500 tweets are extracted, in which 829 tweets are categorized as positive, 675 as negative and remaining 1996 are categorized as neutral tweets. In Pfizer Dataset total of 5000 tweets are extracted, in which 914 tweets are classified as positive, 1703 as negative and remaining 2383 are categorized as neutral tweets. In SputnikV Dataset total of 5000 tweets are extracted, in which 1231 tweets are classified as positive, 705 as negative and remaining 3064 are classified as neutral tweets. The overall opinion analysis of vaccination is most of the data shows the neutral opinion about vaccination but in another approach of peoples when
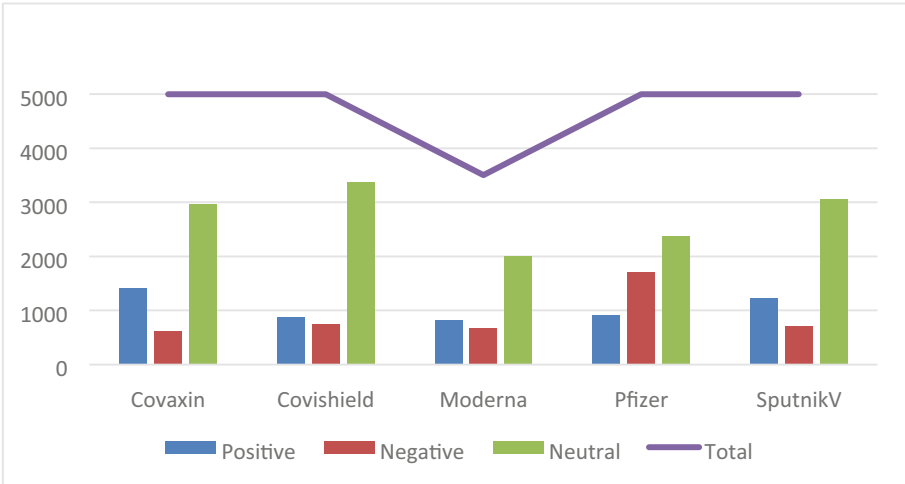
Fig. 3. Users Opinion Analysis on Vaccination

Table 3. Users Opinion Analysis on Vaccination

Name of Vaccination	%Positive	%Negative	%Neutral
Covaxin	28.14	12.5	59.36
Covishield	17.62	15.04	67.34
Moderna	23.68	19.28	57.02
Pfizer	18.28	34.06	47.66
SputnikV	24.62	14.1	61.28

we consider here most positive expressions of peoples on Covaxin and most negative expressions on Pfizer according to twitter data (Fig. 3 and Table 3).

In the above Opinion Mining of Covaxin Twitter Dataset total of 5000 tweets are extracted, in which 28.14% tweets are categorized as positive, 12.5% as negative and remaining 59.36% are categorized as neutral tweets. In Covishield Dataset total of 5000 tweets are extracted, in which 17.62% tweets are categorized as positive, 15.04% as negative and remaining 67.34% are classified as neutral tweets. In Moderna Dataset total of 3500 tweets are extracted, in which 23.68% tweets are categorized as positive, 19.28% as negative and remaining 57.02% are categorized as neutral tweets. In Pfizer Dataset total of 5000 tweets are extracted, in which 18.28% tweets are categorized as positive, 34.06% as negative and remaining 47.66% are categorized as neutral tweets. In SputnikV Dataset total of 5000 tweets are extracted, in which 24.62% tweets are categorized as positive, 14.1% as negative and remaining 61.28% are categorized as neutral tweets (Fig. 4).
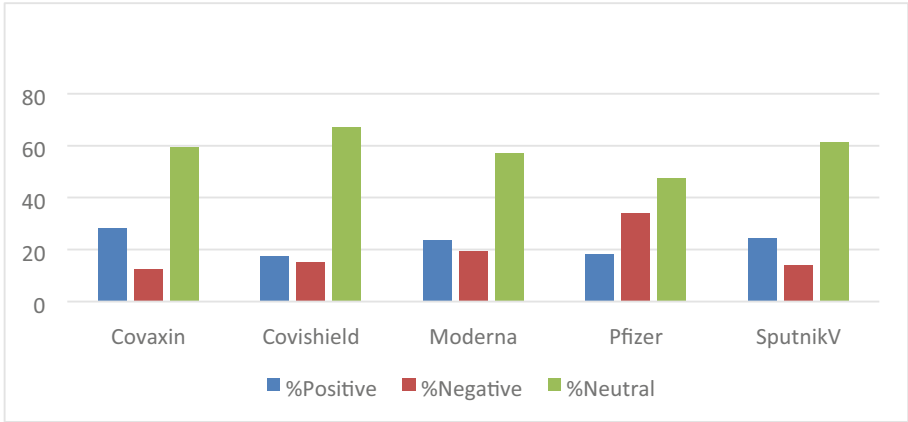
Fig. 4. % of Peoples Opinion Analysis on Covid-19 Vaccination

4.7 Machine Learning Classifiers

For this research work, apply the four machine Learning classifiers for performance evaluation.

Support Vector Machine Classifiers
The classifier represents two classed of labeled data i.e. Positive and Negative training dataset in high dimensional space and classify through Bias boundary. Several probable bias that can distinct train dataset that fits to dissimilar classes, purpose of the SVM classifier is to select the one which splits them with the main likely gap.

The training datasets are nearby to the SVM hyperplane, i.e., which are situated on the boundary, is known as support vectors.

$$D(P) = P \times Q + R \tag{1}$$

Naive Bayes Classifier
Naïve Bayes is a classification algorithm that scores how well each point belongs to each class based on the feature. It has been used since of its easy in both training and classifying step. It is a probabilistic classifier and can learn the pattern of exploratory a set of sentences that have been characterized. It relates the contents with the list of words to categorize the sentence to their right class.

Naive Bayes used with the following consideration

$$\gamma(\alpha|\beta) = \gamma(\alpha) * \gamma(\beta|\alpha)/\gamma(\beta) \tag{2}$$

Decision Tree
The Classifier is based on partition resolution distance based on the characteristic separations into subspaces till the last results is received. Classified demonstration is exact

Table 4. Result Analysis for Support Vector Machine classifiers

Dataset	Accuracy	Precision	Recall	F-Score
Covaxin	93	62	46	51
Covishield	89	60	43	47
Pfizer	92	62	59	59
Moderna	90	49	54	57
SputnikV	94	65	53	63

normal to the human problematic solving process. Result of tree Classifier applied by following expression

$$\text{Info(D)} = \sum \text{pi log2(pi)} \tag{3}$$

K-Nearest Neighbor

K-Nearest Neighbor (K-NN) algorithm is a technique for classifying data based on nearest class that are neighboring to the data. For calculating space is called as Euclidean Distance.

$$D(x, p) = \sqrt{(x - p)} \tag{4}$$

5 Result Interpretation Using ML Based Approach

In this section, we determined the results using various machine learning approach.

5.1 Result Analysis for Support Vector Machine Classifiers

In Fig. 5, the SVM classifier is applied on five datasets and got the highest accuracy of 94% with the SputnikV dataset and the lowest accuracy of 89% with the Covishield dataset. The results are obtained according to our experimental datasets (Table 4).

5.2 Result Analysis for Naive Bayes Classifiers

In Fig. 6, the Naïve Bayes classifier is applied on five datasets and got the highest accuracy of 95% with the Covishield dataset and the lowest accuracy of 87% with the Moderna dataset. The results are obtained according to our experimental datasets (Table 5).

5.3 Result Analysis for Decision Tree Classifiers

In Fig. 7, the Decision Tree classifier is applied on five datasets and got the highest accuracy of 97% with the Covishield dataset and the lowest accuracy of 88% with the Pfizer dataset. The results are obtained according to our experimental datasets (Table 6).
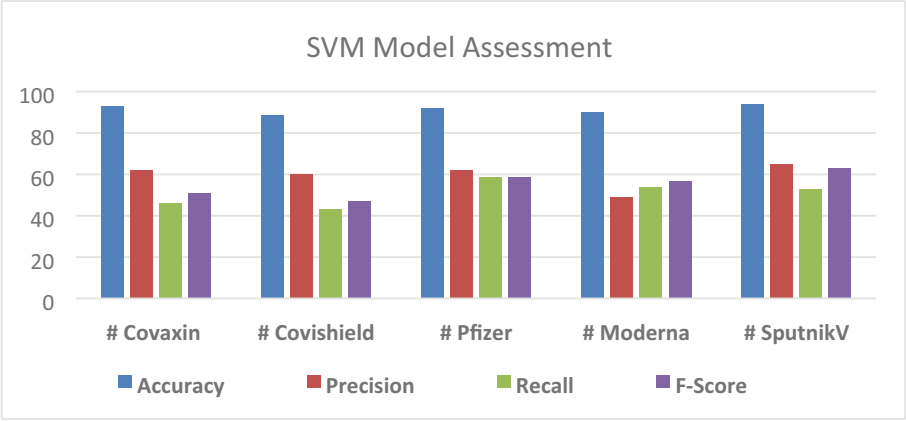
Fig. 5. Support Vector Machine Model Assessment

Table 5. Result Analysis for Naive Bayes Classifiers

Dataset	Accuracy	Precision	Recall	F-Score
Covaxin	92	53	47	51
Covishield	95	57	43	52
Pfizer	90	61	54	56
Moderna	87	54	56	49
SputnikV	93	62	51	61
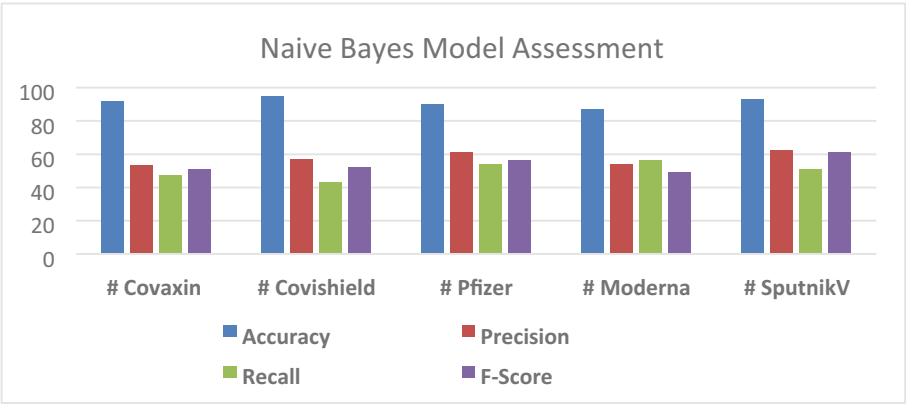
Fig. 6. Naive Bayes Classifiers Model Assessment

Table 6. Result Analysis for Decision Tree Classifiers

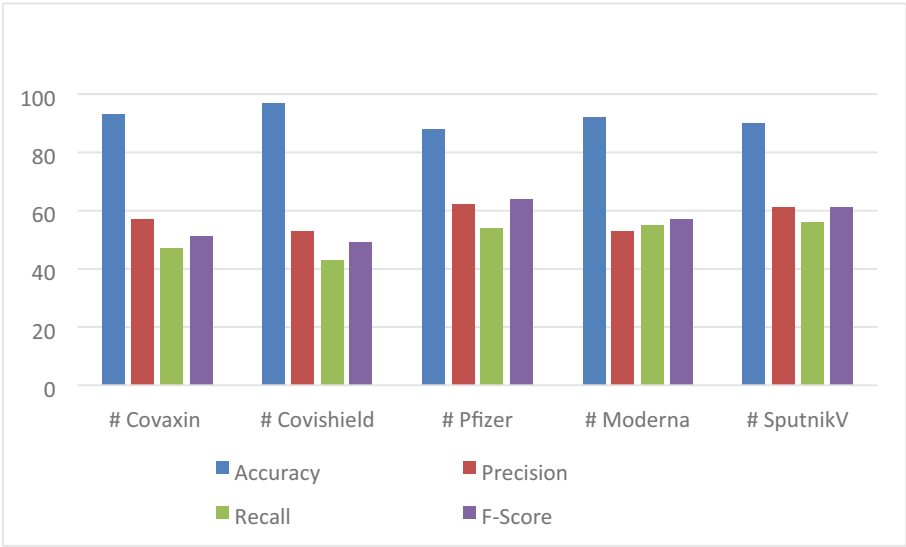| Dataset	Accuracy	Precision	Recall	F-Score
Covaxin	93	57	47	51
Covishield	97	53	43	49
Pfizer	88	62	54	64
Moderna	92	53	55	57
SputnikV	90	61	56	61

Fig. 7. Decision Tree Classifiers Model Assessment

5.4 Result Analysis for KNN Classifiers

In Fig. 8, the KNN classifier is applied on five datasets and got the highest accuracy of 96% with the Covaxin dataset and the lowest accuracy of 89% with the SputnikV dataset. The results are obtained according to our experimental datasets (Table 7).

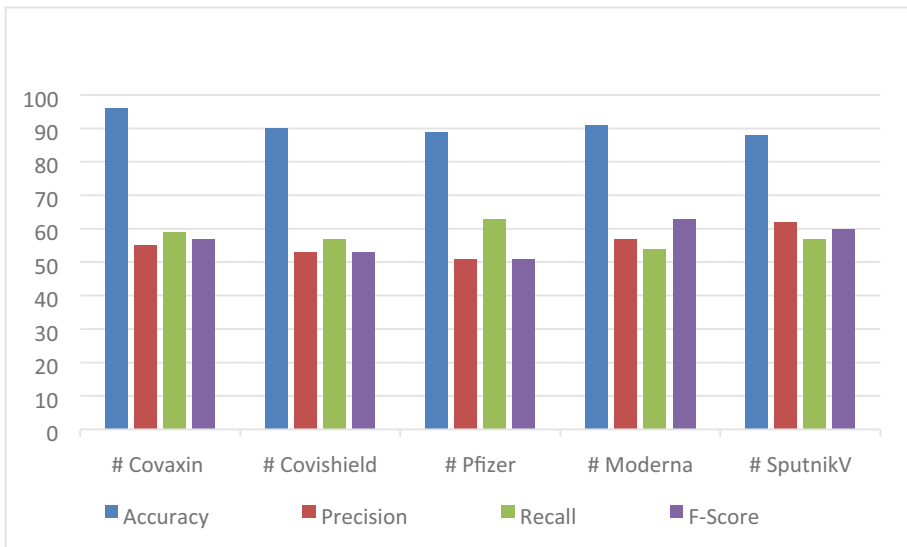### 5.5 Comparative Performance Analysis of Machine Learning Classifiers

See Table 8 and Fig. 9.

5.6 Comparative Result Analysis (Benchmarking)

See Table 9.

Table 7. Result Analysis for KNN Classifiers

Dataset	Accuracy	Precision	Recall	F-Score
Covaxin	96	55	59	57
Covishield	90	53	57	53
Pfizer	89	51	63	51
Moderna	91	57	54	63
SputnikV	88	62	57	60

Fig. 8. KNN Classifiers Model Assessment

Table 8. Comparative Analysis of Machine Learning Classifiers

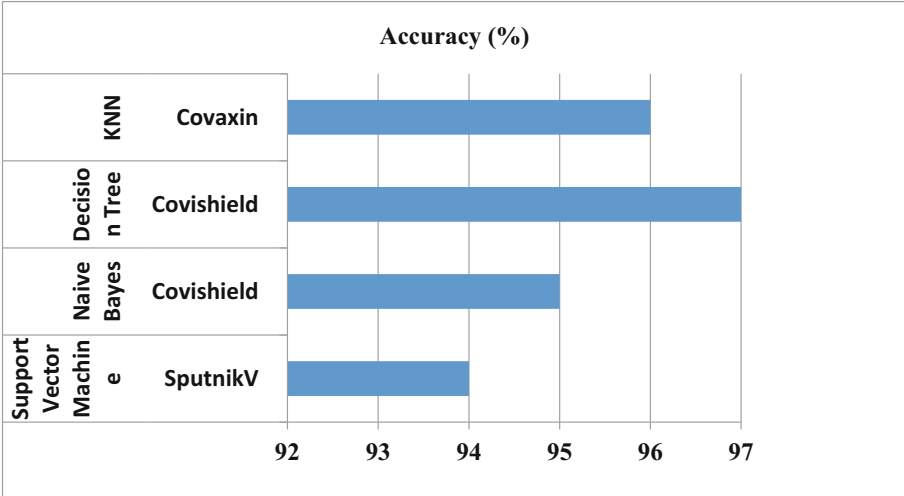| Classifiers	Dataset	Accuracy (%)
Support Vector Machine	SputnikV	94
Naive Bayes	Covishield	95
Decision Tree	Covishield	97
KNN	Covaxin	96

Fig. 9. Comparative Analysis of Machine Learning Classifiers

Table 9. Results comparison with other published work

Author	Purpose	Datasets	Techniques	Results Existing
Reem El-Deeb et al.	Public Perception of COVID-19 Vaccination	Amazon Comprehend module	K-means	Identifying the public perception
Gutti Gowri Jayasurya et al.	Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter	Lexicon-based TextBlob and Vader	logistic regression TF-IDF class	91.925% 92%
Our proposed work	Prediction and popularity of vaccination	Covaxin Covishield Pfizer Moderna SputnikV	Support Vector Machine Naive Bayes Decision Tree KNN	SVM-94 NB-95 DT-97 KNN-96

6 Dataset Links

For dataset collection https://developer.twitter.com/en/apps/13293847.

7 Future Work

The researcher have scope to work with sarcasm data detection of social media text analysis and also they may work with prediction of Covid-19 waves based on Vaccination data.

8   Conclusions

In this paper, we collected data from twitter and then apply preprocessing for data exploration, classification. Data Preprocessing involves the Removing URLs, Data Filtering, Removing Special Characters, Removal of Retweets, Usernames, Remove Punctuations and symbols, Usage of Web links, Hashtags, Tokenization, Exclamation and question marks, Letter Repetition, Negations.

Machine learning classifier used and studied the comparative analysis between KNN, Support Vector Machines, Naïve Bayes, Decision Tree algorithms for data classification.

Finding shows that Decision Tree classifier for Covishield dataset has achieved the highest 97% accuracy with compared to Naïve Bayes, Support Vector Machine, KNN classification methods. Support Vector Machine has lowest Accuracy with 94% for SputnikV.

COVID-19 Vaccination dataset wise machine learning model evaluation performance studied and got highest and lowest results of Machine learning classifiers. The Support Vector Machine SputnikV dataset got highest accuracy with 94% and Covishield dataset got lowest accuracy with 89%, The Naïve Bayes got highest accuracy for Covishield dataset with 95% and lowest accuracy with 87% for Moderna dataset, The Decision tree got highest accuracy for Covishield dataset with 97% and lowest accuracy with 88% for Pfizer dataset, The KNN got highest accuracy for Covaxin dataset with 96% and lowest accuracy with 88% for SputnikV dataset.

In Lexicon Based approached Sentiment polarity classification here total 23500 tweets taken for result analysis and predict the vaccination opinions on SputnikV, Covishield, Covishield, Covaxin, Pfizer datasets. Overall here identify the Neutral opinions on Vaccinations. In other side when we focused on positive and negative opinions here Covaxin is more positive compare with all other vaccination datasets according twitter discussion of users insights and negative opinions on Pfizer vaccination datasets.

Data Availability Statement.   The data used to support the findings of this study are included in the article.

Conflicts of Interest.   The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. F. M. Javed Mehedi Shamrat1 et al. "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm" Indonesian Journal of Electrical Engineering and Computer Science Vol. 23, No. 1, July 2021, pp. 463–470 ISSN: 25024752, https://doi.org/10.11591/ijeecs.v23.i1.pp463-470.
2. Samira Yousefinaghani, Rozita Dara et al. "An analysis of COVID-19 vaccine sentiments and opinions on Twitter" International Journal of Infectious Diseases 108 (2021) Page no-256–262

3. Tomas PRANCKEVIČIUS et al. "Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification" Baltic J. Modern Computing, Vol. 5 (2017), No. 2, 221–232. https://doi.org/10.22364/bjmc.2017.5.2.05.

4. Janessa Griffith et al. "COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework" Journal of Medical Internet Research Med Internet Res 2021 vol. 23 issue. 4 e26874, p. 1 to 10.

5. Liviu-Adrian Cotfas, CameliaDelcea, et al. "The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics From Tweets in the Month Following the First Vaccine Announcement" IEEE Access (Volume: 9) Page no 33203–33223. Electronic ISSN: 2169–3536. https://doi.org/10.1109/ACCESS.2021.3059821

6. Achmad Bayhaqy et al. "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes" 2018 International Conference on Orange Technologies (ICOT) date-23–26 Oct. 2018 Nusa Dua, Bali, Indonesia. Page no-1–6. Electronic ISBN:978-1-53867319-5 USB ISBN:978-1-5386-7318-8

7. AnanyaSarker, Md. ShahidUzZaman, Md. Azmain Yakin Srizon "Twitter Data Classification by Applying and Comparing Multiple Machine Learning Techniques" International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-7, Issue-6, November 2019

8. Anuja P Jain, Padma Dandannavar "Application of Machine Learning Techniques to Sentiment Analysis" 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) 21–23 July 2016 Bangalore, India. Page no-628–632. Electronic ISBN: 978-1-5090-2399-8. https://doi.org/10.1109/ICATCCT.2016.7912076

9. Seongik Park, Yangon Kim "Building Thesaurus Lexicon using Dictionary-Based Approach for Sentiment Classification" IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA) 8–10 June 2016 Towson, MD, USA, page no-1–6. Electronic ISBN: 978-1-5090-0809-4. https://doi.org/10.1109/SERA.2016.7516126.

10. Luiz F. S. Coletta et al. "Combining Classification and Clustering for Tweet Sentiment Analysis" Brazilian Conference on Intelligent Systems 18–22 Oct. 2014 Sao Paulo, Brazil. Page no-210–215. Electronic ISBN: 978-1-4799-5618-0. https://doi.org/10.1109/BRACIS.2014.46

11. MoonisShakeel, VikramKarwal "Lexicon-based Sentiment Analysis of Indian Union Budget 2016–17" International Conference on Signal Processing and Communication (ICSC) 26–28 Dec. 2016 Noida, India. Page no-299–302 Electronic ISBN: 978-1-5090-2684-5. https://doi.org/10.1109/ICSPCom.2016.7980595.

12. Mr.Jeevankumar P. Jain et al. "Opinion Analysis of Text-based on Three Domain Classification" International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) 9–10 Sept. 2016 Pune, India. Page No-713–717. Electronic ISBN: 978-1-5090-2080-5. https://doi.org/10.1109/ICACDOT.2016.7877680.

13. Wilfrido J. Paredes García et al. "Polarity classification of political tweets in Mexico using a lexicon-based approach" International Engineering Congress (CONIIN) 14–19 May 2018 Queretaro, Mexico, Mexico. Page no-1–7. Electronic ISBN: 978-1-5386-7018-7. https://doi.org/10.1109/CONIIN.2018.8489802.

14. ArdhianEkawijana, HeriHeryono "Composite Naive Bayes Classification and Semantic Method to Enhance Sentiment Accuracy Score" 4th International Conference on Cyber and IT Service Management 26–27 April 2016 Bandung, Indonesia. Page No-1–4. Electronic ISBN: 978-1-4673-8443-8. https://doi.org/10.1109/CITSM.2016.7577591

15. Naeem A. Mahoto, SeharGul Khan and Sadaquat Ali Ruk "A Lexicon-based Method to Determine Subjectivity of Unstructured Data" 5th International Multi-Topic ICT Conference (IMTIC) 25–27 April 2018 Jamshoro, Pakistan. Page No-1–7. Electronic ISBN: 978-1-5386-6594-7. https://doi.org/10.1109/IMTIC.2018.8467271.

16. A.Haripriya, SantoshiKumari "Real-Time Analysis of Top Trending Event on Twitter: Lexicon Based Approach" 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 3–5 July 2017 Delhi, India. Page No-1–4. Electronic ISBN: 978-1-5090-3038-5. https://doi.org/10.1109/ICCCNT.2017.8204123

17. Ms. FarhaNausheen, Ms. SayyadaHajera Begum "Sentiment Analysis to Predict Election Results Using Python" 2nd International Conference on Inventive Systems and Control (ICISC) 19–20 Jan. 2018 Coimbatore, India. Page No-1259–1262. Electronic ISBN: 978-1-5386-0807-4. https://doi.org/10.1109/ICISC.2018.8399007

18. NehaUpadhyay, Prof. Angad Singh "Sentiment Analysis on Twitter by using Machine Learning Technique" International Journal for Research in Applied ScienceEngineeringTechnology (IJRASET)ISSN: 2321–9653Volume 4 Issue V, May 2016 page no-488–494.

19. Snehal L. Rathod, Sachin N. Deshmukh "Sentiment Analysis Using SVM and Maximum Entropy" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056Volume: 03 Issue: 08 | Aug-2016 page no-453–458.

20. BhagyashriWagh, J. V. Shinde, N. R. Wankhade "Sentimental Analysis on Twitter Data using Naive Bayes" International Journal of Advanced Research in Computer and Communication Engineering ISSN (Online) 2278–1021Vol. 5, Issue 12, December 2016 page no-316–319.

21. Durgesh Patel, SakshiSaxena, ToranVerma "Sentiment Analysis using Maximum EntropyAlgorithm in Big Data" International Journal of Innovative Research in Science, Engineering and Technology ISSN (Online): 2319–8753Vol. 5, Issue 5, May 2016 page no-8355–8361.

22. AnkurGoel, JyotiGautam, Sitesh Kumar "Real-Time Sentiment Analysis of Tweets UsingNaive Bayes" 2nd International Conference on Next Generation Computing Technologies (NGCT-2016) Dehradun, India 14–16 October 2016. page no-257–261. Electronic ISBN: 978-1-5090-3257-0. https://doi.org/10.1109/NGCT.2016.7877424.

23. MohitMertiyaAshima Singh "Combining Naive Bayes and Adjective Analysis for sentiment Detection on Twitter" International Conference on Inventive Computation Technologies (ICICT) Coimbatore, India. page no-1–6 Electronic ISBN: 978-1-5090-1285-5. https://doi.org/10.1109/INVENTIVE.2016.7824847.

24. Dr. S. Vijayarani, Ms. J. Ilamathi, Ms. Nithya "Preprocessing Techniques for Text Mining - An Overview" International Journal of Computer Science & Communication Networks Vol 5(1) ISSN: 2249–5789 page no-7–16.

25. I. Hemalatha, Dr. G. P Saradhi Varma, Dr. A. Govardhan "Preprocessing the Informal Text for efficient Sentiment Analysis" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)ISSN 2278–6856Volume 1, Issue 2, July–August 2012 page no-58–61.

26. Huma Parveen, Prof. Shikha Pandey" Sentiment Analysis on Twitter Data-set using Naïve Bayes Algorithm" 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) 21–23 July 2016 Electronic ISBN: 978-1-5090-2399-8. https://doi.org/10.1109/ICATCCT.2016.7912034Bangalore, India page no-416–419.

27. RabiaBatool, Asad Masood Khattak, JahanzebMaqbool and Sungyoung Lee "Precise Tweet Classification and Sentiment Analysis" IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS) Niigata, Japan16–20 June 2013. page no-461–466. Electronic ISBN: 978-1-4799-0174-6. https://doi.org/10.1109/ICIS.2013.6607883.

28. P. BavithraMatharasi, Dr. A. Senthilrajan "Sentiment Analysis of Twitter Data using Naïve Bayes with Unigram Approach" International Journal of Scientific and Research Publications ISSN 2250–3153 Volume 7, Issue 5, May 2017 page no-337–341.

29. M. Vadivukarassi, N. Puviarasan and P. Aruna "Sentimental Analysis of Tweets Using Naive Bayes Algorithm" World Applied Sciences Journal ISSN 1818–4952 Volume 35 (1) 2017. page no-54–59. https://doi.org/10.5829/idosi.wasj.2017.54.59.