International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

Machine Learning Based Twitter Sentiment Analysis and User Influence

Mrs. Ragini Krishna [1] , Dr. Prashanth C.M.[2]
[1]Assistant Professor,
Department of Information Science & Engineering
Sri Krishna Institute of Technology
Bangalore 560090, India
ragini.krishna@gmail.com
[2]Principal
Mangalore Institute of Technology and Engineering
Mangalore – 574225, India
prashanth.ait@gmail.com

Abstract— The use of social media platforms, such as Twitter, has grown exponentially over the years, and it has become a valuable source of information for various fields, including marketing, politics, and finance. Sentiment analysis is particularly relevant in social media analysis. Sentiment analysis involves the use of natural language processing (NLP) techniques to automatically determine the sentiment expressed in a given text, such as positive, negative, or neutral.

In this research paper, we focus on Twitter sentiment analysis and identify the most influential users in a given topic. We propose a methodology based on machine learning techniques to perform sentiment analysis and identify the most influential users on Twitter based on popularity. Specifically, we utilize a combination of NLP techniques, sentiment lexicons, and machine learning algorithms to classify tweets as positive, negative, or neutral. We then employ popularity calculations for each user to identify the top 10 most influential users on a given topic.

The proposed methodology was tested on a large dataset of US airlines tweets which is related to a specific topic i.e. airlines, and the results show that the approach can effectively classify tweets according to sentiment and identify the most influential users. We evaluated the performance of several machine learning algorithms, including Multinomial Naive Bayes, Support Vector Machines (SVM), Decision Trees, Gradient Boosting, logistic regression, AdaBoost, KNN and Random Forest, and found that the logistic regression algorithm has achieved the highest accuracy.

The proposed methodology has several implications for various fields, such as marketing, where sentiment analysis can help companies understand consumer behavior and tailor their marketing strategies accordingly. Moreover, identifying the most influential users can provide insights into opinion leaders in a given topic and help companies and policymakers target their messages more effectively.

Keywords: sentiments, natural language processing, AdaBoost, gradient boosting, Naïve Baye's, Decision Trees, influential user.

I. INTRODUCTION

The use of social media has become a prevalent part of modern society, with platforms like Twitter allowing users to share their thoughts, opinions, and experiences with the world in real-time. However, with the vast amount of content being produced every second, it can be challenging to stand out and gain a significant following.

In recent years, there has been growing interest in measuring social media influence, particularly on Twitter. Influence refers to a user's ability to affect the behavior, attitudes, and opinions of others on the platform. Measuring influence can provide valuable insights into user's behavior and help individuals and organizations make informed decisions on social media marketing, advertising, and content creation strategies.

With the activity levels and the follower count of the users, machine learning models can accurately predict a user's level of influence on the platform. This can help users identify areas for improvement and tailor their content to increase their influence. Machine learning has proven to be a powerful tool for predicting social media influence by analyzing various user-generated data such as the post frequency and engagement.

The motivation behind this research is to contribute to the growing body of knowledge on measuring social network influence and to provide a more accurate and reliable method for predicting Twitter users' influence on a given topic using machine learning. This research can have practical implications for businesses seeking to optimize their social network presence and increase their reach and engagement on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

the platform. Additionally, this research can provide insights into user behavior on social network based on a topic, which can help inform future research in the field.

The paper is organized is the following way: previous works are discussed in section 2, section 3 discusses the research design, and the results are discussed in the section 4 followed by concluding remarks in section 5.

II. LITERATURE REVIEWS

Previous studies have demonstrated the effectiveness of using machine learning algorithms to predict social media influence. These studies have primarily focused on predicting influence on Twitter, as it is one of the most popular social media platforms for content creation and sharing.

The system proposed by Essaidi et al.[1] makes predictions for influential users on Twitter. The various approaches are used by the author in order to predict the most influential users on Twitter. A comparative study of three approaches to find the influential users are: twitter follower and followee ratio(TFF), Tunk rank algorithm and influence score. The influence score is detected by multiplying the count of retweets with the total count of followers and divided by the difference between current and the tweet time. During Tunk rank algorithm, the influence of a user is determined by the expected count of people who will read the retweet and original tweet sent by the influencer itself. The TFF ratio method finds a ratio between count of followers and count of users that it is following. Higher the ratio, higher the influence.

The method adopted by Qi *et al*.[2] for twitter sentiment analysis is a hybrid approach based on two approaches lexical based and machine learning approach.

The authors in [3] calculate the sentiments by creating a group of related topics and then compare it with reference to a given topic. They also state that a community of people can have similar sentiments.

The authors in [4][5][6] have used various machine learning algorithms like SVM(support vector machine), logistic regression, decision tree to find the solution to the sentiment analysis.

The sentiments analysis done by [7] and [8] have also considered the significance of semantics or meaning of the sentence to determine the sentiment.

To find the influential communities [10] on social network the emotional behaviour of users were determined by text categorization of the emotional content of the text posted on the social network.

The authors of [9] have used SVM as a method of text categorization and it shows that SVMs performance is better than the currently best performing methods and behaves robustly over a variety of different learning tasks.

The paper [11] discusses the method of finding the opinion which is based on context of the sentences.

The previous works were not able to accurately classify the words due to sparse data and the sarcasm present in the tweets. Thus, the authors in [12] used the hybrid method on the opinion mining technique to overcome the problem of sarcasm and thus achieve higher efficiency in determining the sentiment.

The authors in [13] have done a detailed study of the various methods in identification of influential users. They also state how compliance can affect the influential score of a user in the social network.

In [14] it has been shown that automatic accounts can obtain high influential scores with no intuitive reason and thus fail at distinguishing so-called social capitalists from real, truthful users.

[15] uses a mixed sentiment dictionary concept called HowNet which is claimed to have more accuracy than any other method to find the sentiments on short text.

The authors in [16] state that if a message has to become viral, it has to either contain a very good or sweet thing about their friends or very bad news to the public. This draws the attention of the users and thus the diffusion rate will be high.

The users who promote positive emotions are always active in building relationships on the social network as the users on twitter share emotions with their followers [17] and positive emotions influence the emotional behaviors to user relationships in the network.

Through this work, we are trying to study the behavior of various algorithms like SVM, logistic regression, AdaBoost etc and find which algorithm performs best for the US airline dataset and to find the influential user based on the activity of the user in the network.

III. RESEARCH DESIGN

The research problem for this study is the difficulty in accurately measuring social media influence, particularly on Twitter. While there are various metrics used to gauge influence on the platform, they can be subjective and do not always reflect a user's true impact on the platform. Additionally, the vast amount of data generated on Twitter makes it challenging to identify the most relevant factors that contribute to a user's influence. The primary objective of this research is to develop a machine learning model that can accurately identify the sentiment of text and classify them into positive or negative sentiments and predict Twitter users' influence based on the positive sentiment.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

To achieve this objective, the following sub-objectives have been identified.

- To evaluate the performance of various machine learning algorithms in predicting sentiments of the users.
- To compare the performance of the machine learning models and identify the most effective approach for performing sentiment analysis.
- To minimize the false positive rate using precision score.
- To provide insights into user behavior on social media and how it relates to social media influence.

By achieving these objectives, this study aims to provide a more accurate and reliable method for measuring social media influence, which can have practical implications for individuals and organizations seeking to optimize their social media presence. Additionally, this research can contribute to the growing body of knowledge on social media behavior and aid future research in the field.

Fig. 1 shows the process flow to find the sentiment of tweets posted by users.

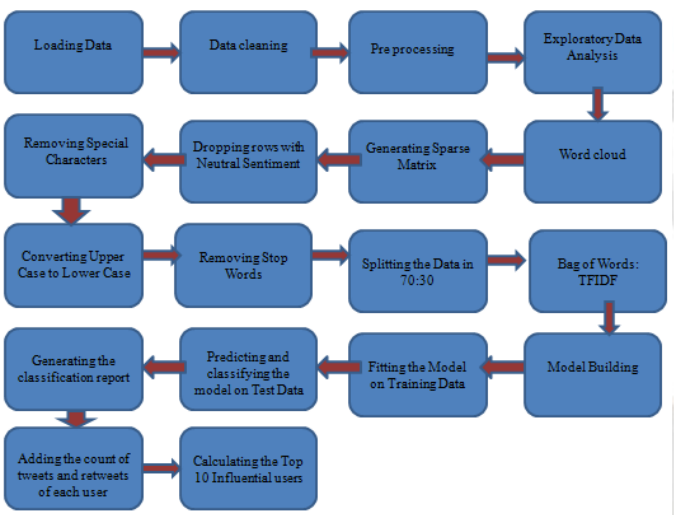

Figure 1: Process flow of Twitter sentiment analysis model

A. Data Cleaning: During data cleaning, null values and duplicate tweets are removed from data. The percentage of null values is calculated for each variable and those variables which have more than 90% of null values are removed from the dataset. The percentages of null values in dataset are shown in Fig 2. below:

```
Percentage null or na values in df
Out[110]:
tweet_id                      0.00
airline_sentiment             0.00
airline_sentiment_confidence  0.00
negativereason               37.31
negativereason_confidence    28.13
airline                       0.00
airline_sentiment_gold       99.73
name                          0.00
negativereason_gold          99.78
retweet_count                 0.00
text                          0.00
tweet_coord                  93.04
tweet_created                 0.00
tweet_location               32.33
user_timezone                32.92
dtype: float64
```

Figure 2: Percentage of null values in dataset

Therefore, ‘tweet_coord', 'airline_sentiment_gold', 'negativereason_gold’ variables are removed from the data as more than 90% of information is not available in the variables.

***B. Exploratory Data Analysis*:**

The exploratory data analysis is performed on features to obtain the understanding of the dataset. The data collected is between 16th Feb 2015 and 24th Feb 2015. The highest number of tweets was sent on 22nd Feb 2015. The highest numbers of tweets are available for United airlines which indicates the highest popularity of the airline. It was observed during EDA that US Airways, United and American airlines have highest dissatisfied customers whereas Virgin airlines have least negative sentiments and the customers are mostly dissatisfied due to customer service for all the airlines.

The word cloud is built to know the frequency of words in the tweets as shown below. Fig.3 shows the word cloud of the words contributing to the positive sentiment. Fig. 4 shows the word cloud of the words contributing to the negative sentiment.

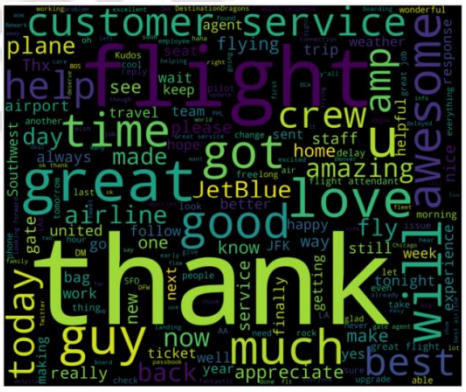

Figure 3: Word cloud for positive sentiment

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

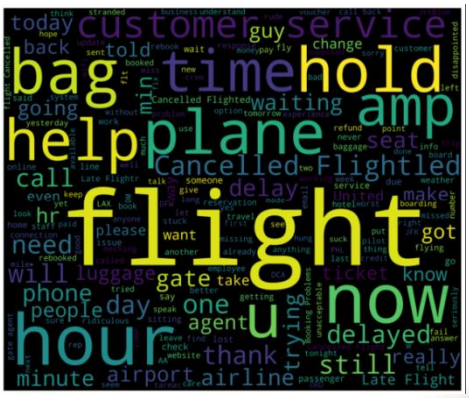

Figure 4: Word cloud for negative sentiment

The purpose of building a word cloud is to figure out the words with maximum frequency in the tweet. It is evident from the word cloud in fig. 3 that some words occurs more in tweets with positive sentiments i.e. thank, great, good, flight, whereas the tweets with negative sentiments as shown in fig. 4 have common query words like help, time, hold, cancelled, hour etc. These words can be used by the model to identify the positive and negative sentiments for airlines.

Finding the sentiments of a tweet involves the following stages:

- **Pre-processing**: The neutral sentiment data points are removed because they do not indicate the positive or negative sentiment towards an airline.
- **Removal of stopwords:** Stop words are those words in the text which do not add any meaning to the sentence and their removal will not affect the processing of text for the defined purpose. They are removed from the vocabulary to reduce noise and to reduce the dimension of the feature set. After cleaning the tweets, the dataset is split into training and test set in the ratio of 70:30.

- **Bag of words:** This method does not take into account the positioning of the words, and only tokenizes if a given string contains the word or not. We can do this with Keras' tokenizer, through the texts_to_matix() method. This gives us a matrix, with each column corresponding to a given word, and each row contains an indicator, showing if the text for this row contained that word or not.

- **Tokenisation:** In tokenisation, we focus on preserving the order of the words. We can use the texts_to_sequences() method of Keras' tokeniser to achieve this. This method generates a sequence of integers, with each unique integer corresponding to the word used at this point in the Tweet. The number of unique integers is defined by the corpus size of the tokeniser, which is determined when defining the tokeniser. If one of the words used is not in the corpus, it is ignored. Usually we choose a relatively large corpus, so we only miss out on information from rare words that we would struggle to learn the meaning of anyway.

To find the influence of users based on a positive sentiment for a given topic (US Airline) on the social network like twitter involves two stages:

- **Sentiment analysis:** Machine learning algorithm using six classification algorithms (Decision Tree, Support vector Machine, K-Nearest Neighbour, Random Forest, Logistic Regression, Naive Bayes, Ada Boost and Gradient Boosting). The number of estimators for random forest model are 200 and 5 nearest neighbours are selected for KNN algorithm. The SVM model has used RBF (Radial Basis Function) kernel and C parameter has value 0.25. The Adaboost classifier had used Decision tree as base estimator with a depth of 18 and number of estimators as 10. The gradient boosting model had used 100 number of estimators, learning rate of 1 and maximum depth of 4. The F-Score, Accuracy, Precision, and Recall are considered as performance metrics.

- **Influencer Ranking:** A ranking system will be created using a condition based on number of followers and positive or negative sentiment towards the influencer. In the proposed method, the top 10 influencers were identified by popularity of the user. The popularity of each user is determined by calculating the total number of tweets and total count of retweets. The user with highest number of total tweets is assigned as highest in the popularity assessment. Another method for ranking can be based on time difference between a tweet and retweet where minimum response time will be assigned higher rank as compared to those tweets with more response time.

IV. RESULTS AND DISCUSSION

A. Logistic Regression: The logistic regression has achieved an overall accuracy of 91%. The classifier is able to classify the 91% of negative sentiments correctly among all samples of negative sentiments. It is able to correctly predict the negative sentiments with a precision of 98%. Fig. 5 shows the above claimed results.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

LogisticRegression Accuracy Score : 90.56%

	precision	recall	f1-score	support
negative	0.98	0.91	0.94	2981
positive	0.60	0.90	0.72	472
accuracy			0.91	3453
macro avg	0.79	0.90	0.83	3453
weighted avg	0.93	0.91	0.91	3453

Figure 5: Results from logistic regression

B. Multinomial Naïve Baye's: The Naïve Baye's algorithm has achieved an accuracy of 84% in classifying the classes. The classifier is able to classify the negative sentiment classes correctly with an accuracy of 83% and 100% precision. The model has higher percentage of false negatives and 0 false positives as no positive sentiment is wrongly classified as negative sentiment. Fig. 6 shows the above claimed results.

MultinomialNB Accuracy Score : 83.78%

	precision	recall	f1-score	support
negative	1.00	0.83	0.91	3307
positive	0.21	1.00	0.34	146
accuracy			0.84	3453
macro avg	0.60	0.92	0.63	3453
weighted avg	0.97	0.84	0.88	3453

Figure 6: Results from Multinomial Naïve Baye's

C. Decision Tree: The Decision Tree model can classify the classes with an accuracy of 85%. The classifier is able to classify the negative sentiments classes correctly with an accuracy of 85%. The recall score and precision for negative sentiments, of the model are 90% and 92% respectively indicating higher false negatives as compared to false positives. Fig. 7 shows the above claimed results.

DecisionTreeClassifier Accuracy Score : 85.43%

	precision	recall	f1-score	support
negative	0.92	0.90	0.91	2784
positive	0.62	0.65	0.63	669
accuracy			0.85	3453
macro avg	0.77	0.78	0.77	3453
weighted avg	0.86	0.85	0.86	3453

Figure 7: Results from Decision Tree classifier

D. Random Forest classifier: The Random Forest model can identify the classes with an overall accuracy of 90%. The classifier is able to classify the negative sentiment classes correctly with an accuracy of 90% where only 10% of negative sentiments are wrongly predicted as positive sentiment. The precision is 97% indicating that only 3% of positive sentiments are wrongly predicted as negative sentiment. Fig. 8 shows the above claimed results.

RandomForestClassifier Accuracy Score : 89.52%

	precision	recall	f1-score	support
negative	0.97	0.90	0.94	2945
positive	0.60	0.84	0.70	508
accuracy			0.90	3453
macro avg	0.79	0.87	0.82	3453
weighted avg	0.92	0.90	0.90	3453

Figure 8: Results from Random Forest classifier

E. KNN: The KNN model can identify the classes with an overall accuracy of 88%. The classifier is able to classify the negative sentiment classes correctly with an accuracy of 92% and the precision is 92% indicating that only 8% of positive sentiments are wrongly predicted as negative sentiment and vice-versa. Fig. 9 shows the above claimed results.

KNeighborsClassifier Accuracy Score : 87.75%

	precision	recall	f1-score	support
negative	0.92	0.92	0.92	2746
positive	0.70	0.70	0.70	707
accuracy			0.88	3453
macro avg	0.81	0.81	0.81	3453
weighted avg	0.88	0.88	0.88	3453

Figure 9: Results from KNearest Neighbour

F. SVC: The SVM model can identify the classes with an accuracy of 80%. The classifier is able to classify the negative sentiment classes correctly with an accuracy of 80% where 20% of negative sentiments are wrongly predicted as positive sentiment. The precision is 100% indicating that no positive sentiment is wrongly predicted as negative sentiment. Fig. 10 shows the above claimed results.

SVC Accuracy Score : 80.02%

	precision	recall	f1-score	support
negative	1.00	0.80	0.89	3437
positive	0.02	1.00	0.04	16
accuracy			0.80	3453
macro avg	0.51	0.90	0.47	3453
weighted avg	1.00	0.80	0.88	3453

Figure 10: Results from Support Vector Machine

G. Gradient Boosting: The Gradient Boosting model can identify the classes with an overall accuracy of 87%. The recall and precision of the model is 91% and 94% respectively indicating the higher count of false negatives than false positives. Fig. 11 shows the above claimed results.

GradientBoostingClassifier Accuracy Score : 87.29%

	precision	recall	f1-score	support
negative	0.94	0.91	0.92	2838
positive	0.62	0.72	0.67	615
accuracy			0.87	3453
macro avg	0.78	0.81	0.79	3453
weighted avg	0.88	0.87	0.88	3453

Figure 11: Results from Gradient Boosting

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

H. AdaBoost classifier: The Ada Boost model can identify the classes with an overall accuracy of 87%. The classifier is able to classify the negative sentiment classes correctly with an accuracy of 91% where only 9% of negative sentiments are wrongly predicted as positive sentiment. The precision of the model is 93% indicating that only 7% of positive sentiments are wrongly predicted as negative sentiments. Fig. 12 shows the above claimed results.

```
AdaBoostClassifier  Accuracy Score : 87.46%
              precision    recall  f1-score   support

    negative       0.93      0.91      0.92      2790
    positive       0.66      0.71      0.68       663

    accuracy                           0.87      3453
   macro avg       0.80      0.81      0.80      3453
weighted avg       0.88      0.87      0.88      3453
```

Figure 12: Results from AdaBoost classifier

The performance comparison of all the above algorithms which are considered for study is stated in the table 1 below.

Table 1: A comparison of different ML models

Performance Metrics	Accuracy	Recall	Precision	F1-score
Logistic Regression	91%	91%	98%	94%
Multinomial Naïve Baye's	84%	83%	100%	91%
Decision Tree	85%	90%	92%	91%
Random Forest	90%	90%	97%	94%
KNN	88%	92%	92%	92%
SVM	80%	80%	100%	89%
Gradient Boost	87%	91%	94%	92%
Ada Boost	87%	91%	93%	92%

A comparative study of the efficiency of proposed system with paper proposed by Qi et al. [2] is tabulated below in table 2.

A comparison of proposed model with respect to twitter sentiment model in existing research paper as displayed in Table 2. It is observed that proposed Logistic Regression model had achieved better performance on following metrics: F1-score, accuracy, recall, and precision than the Support Vector Machine model described in paper [2]. The logistic regression model has higher precision of 98% specifying the lesser false alarm rate and better recall of 91% specifying the lesser count of false detection of negative sentiments as positive. Also, the model was evaluated on other metrics i.e. accuracy and F1-score parameters where the highest accuracy and high F1 score of 91% and 94% is achieved respectively. The overall performance of the proposed model is better than SVM based model discussed in paper [2].

Table 2: A comparison of proposed framework with previous research work

Performance Metrics	Proposed Model	Paper [2] Qi *et al.*
	Logistic Regression	Support vector machine
Accuracy	91%	71%
Recall	91%	-
Precision	98%	-
F1-score	94%	-

A comparison has been made for the accuracy achieved by various models for classification of tweets and it was observed that logistic regression has achieved highest accuracy as compared to all other models. Fig. 13 shows that Logistic Regression has the highest accuracy.

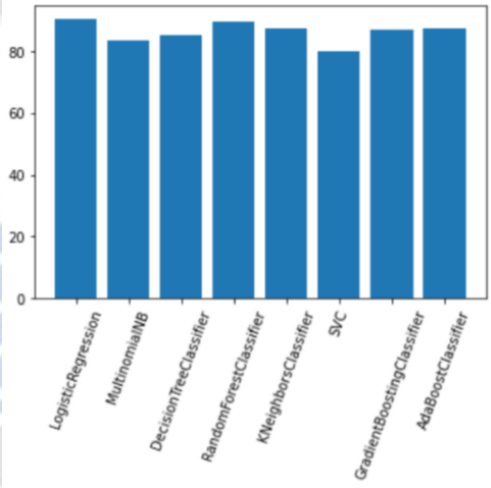

Figure 13: A comparison of accuracy of all ML models

Twitter user influence: The most influential users with respect to airlines are identified based on the popularity. The popularity is detected by calculating the total number of tweets and retweets by each user. The higher the count of tweets and retweets indicates more influence the user has on the topic.

Fig.14 shows the top 10 influential users according to the US airline dataset.

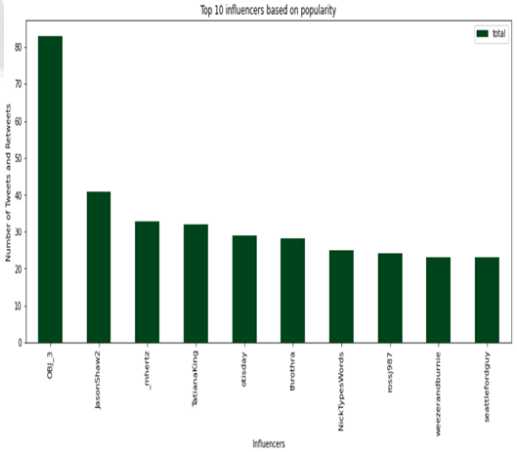

Figure 14: Top 10 influential twitter users

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7192
***Article Received**: 24 April 2023 **Revised:** 14 June 2023 **Accepted:** 25 June 2023*

V. CONCLUSION

The sentiment analysis is a useful tool for understanding the public opinion on Twitter. Machine learning algorithms can be used effectively for sentiment analysis on Twitter, with high levels of accuracy and precision.

Different feature sets and machine learning algorithms can yield different results in sentiment analysis. Therefore, careful selection of features and algorithms is critical to achieving accurate results. In the proposed system Decision Tree, Support vector Machine, K-Nearest Neighbour, Random Forest, Logistic Regression, Naive Bayes, Ada Boost and Gradient Boosting models are used. Among them, logistic regression model achieved highest accuracy of 91% with recall score of 91%, precision of 98% and F1-score of 94%.

User influence on Twitter can be measured using a combination of metrics such as the number of followers, retweets, and mentions. In the proposed system, the user influence is identified by measuring the popularity of each user. The popularity is calculated by obtaining the total count of tweets and retweets for each user. The user influence is positively correlated with sentiment polarity, meaning that users with more positive sentiment tend to have more influence on Twitter.

In future, various feature selection techniques like random forest, XgBoost can be used to improve accuracy. The hyper-parameter tuning to obtain the optimal hyper-parameters is another aspect which can be explored to improve the efficiency of model.

REFERENCES

[1] Essaidi, Abdessamad, Dounia Zaidouni, and Mostafa Bellafkih. "New method to measure the influence of Twitter users." 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS). IEEE, 2020.

[2] Qi, Yuxing, and Zahratu Shabrina. "Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach." Social Network Analysis and Mining 13.1 (2023): 31.

[3] Bhatnagar, Sarvesh, and Nitin Choubey. "Making sense of tweets using sentiment analysis on closely related topics." Social Network Analysis and Mining 11.1 (2021): 44..

[4] Gupta, Bhumika, et al. "Study of Twitter sentiment analysis using machine learning algorithms on Python." International Journal of Computer Applications 165.9 (2017): 29-34.

[5] Hasan, Ali, et al. "Machine learning-based sentiment analysis for twitter accounts." Mathematical and computational applications 23.1 (2018): 11.

[6] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).

[7] Thomas Wilson, Andrew Evans, Alejandro Perez, Luis Pérez, Juan Martinez. Integrating Machine Learning and Decision Science for Effective Risk Management. Kuwait Journal of Machine Learning, 2(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/208

[8] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." 2014 Seventh international conference on contemporary computing (IC3). IEEE, 2014.

[9] Mahato, M. K. ., Seth, S. ., & Yadav, P. . (2023). Numerical Simulation and Design of Improved Optimized Green Advertising Framework for Sustainability through Eco-Centric Computation. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 11–17. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2502

[10] Navigli, Roberto. "Word sense disambiguation: A survey." ACM computing surveys (CSUR) 41.2 (2009): 1-69.

[11] Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." Machine Learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[12] Kanavos, Andreas, et al. "Emotional community detection in social networks." Computers & Electrical Engineering 65 (2018): 449-460.

[13] Ding, Xiaowen, Bing Liu, and Philip S. Yu. "A holistic lexicon-based approach to opinion mining." Proceedings of the 2008 international conference on web search and data mining. 2008.

[14] Khan, Farhan Hassan, Saba Bashir, and Usman Qamar. "TOM: Twitter opinion mining framework using hybrid classification scheme." Decision support systems 57 (2014): 245-257.

[15] Krishna, Ragini, and C. M. Prashanth. "Identifying influential users on social network: an insight." Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019, Volume 1. Springer Singapore, 2020.

[16] Danisch, Maximilien, Nicolas Dugué, and Anthony Perez. "On the importance of considering social capitalism when measuring influence on Twitter." BESC 2014-International Conference on Behavioral, Economic, and Socio-Cultural Computing. IEEE, 2014.

[17] Zhang, Yaocheng, et al. "MoSa: A modeling and sentiment analysis system for mobile application big data." Symmetry 11.1 (2019): 115.

[18] Hansen, Lars Kai, et al. "Good friends, bad news-affect and virality in twitter." Future Information Technology: 6th International Conference, FutureTech 2011, Loutraki, Greece, June 28-30, 2011, Proceedings, Part II. Springer Berlin Heidelberg, 2011.

[19] Tago, Kiichi, and Qun Jin. "Influence analysis of emotional behaviors and user relationships based on Twitter data." Tsinghua Science and Technology 23.1 (2018): 104-113.