International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper | Volume-9, Issue-10, 2022 Regular Edition | E-ISSN: 2349-7084

Machine Learning Based Emotional Sentiment Analysis of Tweet Data Using a Voting Classifier

G. Deepika [1], K. Deepthi Reddy [2]

[1]M.Tech.-A.I. student, Department of CSE, CVR College of Engineering, Vastunagar, Mangalpally, Ibrahimpatnam, T.S., India – 501510.
[2]Assistant Professor, Department of CSE, CVR College of Engineering, Vastunagar, Mangalpally, Ibrahimpatnam, T.S., India – 501510.
Email: deepikag076@gmail.com , deepthi.reddy@cvr.ac.in

Corresponding author : K. Deepthi Reddy
Available online at: http://www.ijcert.org

Received:25/08/2022, Revised: 17/09/2022, Accepted: 10/10/2022, Published: 28/10/2022

Abstract: The introduction of social media and microblogging sites to the World Wide Web was a significant advancement. The website functioned as a place for users to express their opinions and feelings on a variety of problems. The Internet has bloomed into a viable platform for online education, information distribution, and the expression of varied opinions since the dawn of the social networking age. Social media networks contain a multitude of sentiment data in the form of tweets, blogs, status updates, articles, and so on. This study takes advantage of Twitter, the most popular microblogging network. Sentiment analysis is used to derive user views and sentiments from Twitter data (tweets). Some of the seven machine learning models presently employed by the existing system to categorise tweets into happy or sad categories include SVM, DTC, NB, RF, GBM, LR, VC (LR+SGD), and VC (LR+SGD). Following a thorough performance comparison, it was discovered that the voting classifier (LR-SGD) in conjunction with the topic-based information content index (TF-IDF) produces the best results, with an F1 score of 81% and an accuracy of 79%. The suggested system consists of combining the LR, RF, NB, and SVM voting classifiers with the TF-IDF model, which produces 94% accuracy, and the COUNT vectorization model, which yields 95% accuracy. The findings might help governments and companies improve the execution of programmes, goods, and events.

Keywords: Emotion recognition, Opinion mining, Machine Learning, Artificial Intelligence, Text Classification, Sentiment analysis

---

1. Introduction

The best way to get a wide range of information is through micro blog websites. This is due to the fact that everyone expresses their opinions and feelings about everything from current events to the products they use on a daily basis. Foretelling how a reader will feel about a particular word, sentence, or group of documents is the goal of sentiment analysis. It's a tool for decoding the feelings and viewpoints expressed in a social media post or comment. The goal is to get a sense of how the general public feels about various issues. Sentimental analysis is the method used to extract meaningful data from text. Simply put, it is a method for bringing order out of chaos when dealing with large amounts of unstructured data. Customer satisfaction, comments, and ratings can all be gauged with this method. Data from the internet, such as chats, e-mail, pdfs, word files, e-commerce websites, and social networking sites, is also considered part of "unstructured data," which also includes information from internal databases and tables and figures from the organisation. Analytics operations run smoothly on structured data, and so do their results. But when dealing with unstructured data such as that found in e-mail, Twitter, etc.,

In this paper, we analyse Twitter, one of the most widely read micro blogs. Twitter is an online micro-blogging and social-networking platform where users can write short status updates of a maximum length of 140 characters. Through Twitter Sentimental Analysis, we can analyse the mood of the person who tweets, which can help businesses analyse the market and the reviews of their products.

Using a tweet classification system based on term frequency, in-term association, and cosine similarity (TF-IDF and COUNT vectorization), this research evaluates several machine learning models for emotion recognition. To gauge the efficacy of well-known ML classifiers on the Twitter dataset, this study introduces a voting classifier (LR, NB, SVM, and RF). The following are the most important contributions:

- The Twitter dataset is used to evaluate different machine learning classifiers for emotion identification, including the Random Forest (RF), support vector machine (SVM), Naive Bayes (NB), and Logistic Regression (LR).
- When compared to methods that only employ these features, such as TF-IDF and COUNT Vectorization, a voting classifier (VC) that combines RF, SVM, NB, and LR achieved the best results when classifying tweets.

Rest of the paper is organized as Section 2 describes the related work of the study, Section 3 presents the methodology of the model , section 4 presents results and discussion , finally chapter concludes with section 5 with future scope of the work .

2. Related Work

Through sentiment analysis, companies can learn more about their customers' tastes in goods and services. In addition, it is crucial in reserving information about industries and corporations for use in making entity reviews through interpretation.

Bhumika Gupta et al. [1] use Unigrams, n-grams, an external lexicon, and Term Frequency Inverse Document Frequency (TF-IDF) to determine the average accuracy of various models. The main goal is to train a machine learning model based on the sentiment analysis of the tweets and then test its accuracy. Data collection, text pre-processing, sentiment detection, sentiment classification, training, and model testing are all a part of it. Models in this area have improved over the past decade, reaching efficiencies of 85%-90%. However, data diversity is still missing. Additionally, there are many problems with its practical use because of slang and abbreviations. When the number of classes is made larger, many analyzers struggle to keep up. Additionally, the model's accuracy on other topics has not yet been thoroughly tested.

The proposed ensemble classification system by Ankit et al. [2] has been compared to both the most popular majority voting ensemble classification system and to several classic sentiment analysis techniques. The ensemble classification system is built from several different types of base learners. These include the Naive Bayes classifier, the Random Forest classifier, support vector machines, and logistic regression. Comparisons with both individual classifiers and the widely used majority voting ensemble classifier reveal that the proposed ensemble classifier provides superior performance. Both businesses and consumers can benefit from this method, as it can be used to track how people feel about a product and help people select the best options.

Using the SemEval-2017, Task 4A, 4B, and 4C datasets, Roza H. Hama Aziz et al. [3] propose a weighted majority voting ensemble method to increase the predictive accuracy of sentiment classification. This research demonstrated that the proposed weighted majority voting ensemble scheme can improve sentiment classification performance compared to both the individual classifiers and the simple majority voting classifier. Experiments showed that the proposed system improved the accuracy of predictions on all sentiment-labeled datasets, which was a major finding.

The model was then passed to the Voting Classifier approach of Ensemble Classification techniques, which was developed by Chaudhary Jagrit Varshney et al. [4], who employed three base classifiers: Logistic Regression, Nave Bayes, and Stochastic Gradient Descent. By tallying votes from the top three classifiers, the ensemble classifier determines which classes best fit the data. In this case, the Logistic Regression model performed slightly better than the Voting Classifier, but the Voting Classifier had a higher rate of successful classifications.

According to the work of Anam Yousaf et al. [5], seven different machine learning models—including the support vector machine (SVM), random forest (RF), gradient boosted machine (GBM), logistic regression (LR), decision tree (DT), neural network (NB), and VC—are put through their paces in a series of experiments (LR-SGD). Two feature representation techniques, Tf and TF-IDF, were also used in this research. After testing on the tweet dataset, we found that while all models performed well, our proposed voting classifier VC (LR-SGD) performed best by combining TF and TF-IDF. The measures of precision, F1-score, and Recal were applied.

3. Methodology

To accomplish this, a wide variety of ML techniques were used. Experiments of wide relevance were examined from several angles. While several classifiers were applied on the dataset with varying degrees of success, it was the Voting classifier—an ensemble of RF, SVM, NB, and RF—that yielded the most favourable results. To conduct this study, we obtained Twitter data from the Kaggle data repository. During preprocessing, the dataset is cleaned up of any extraneous data. Following that, we split the data into a training set and an evaluation set. In this case, the ratio between the training set and the test set is 70% to 30%. We then use feature engineering techniques on the training set to further refine it. Common practise in machine learning is to train several classifiers on the training set and then evaluate them on a separate test set. The success of an experiment can be judged by its accuracy, recall, precision, and F1-score.

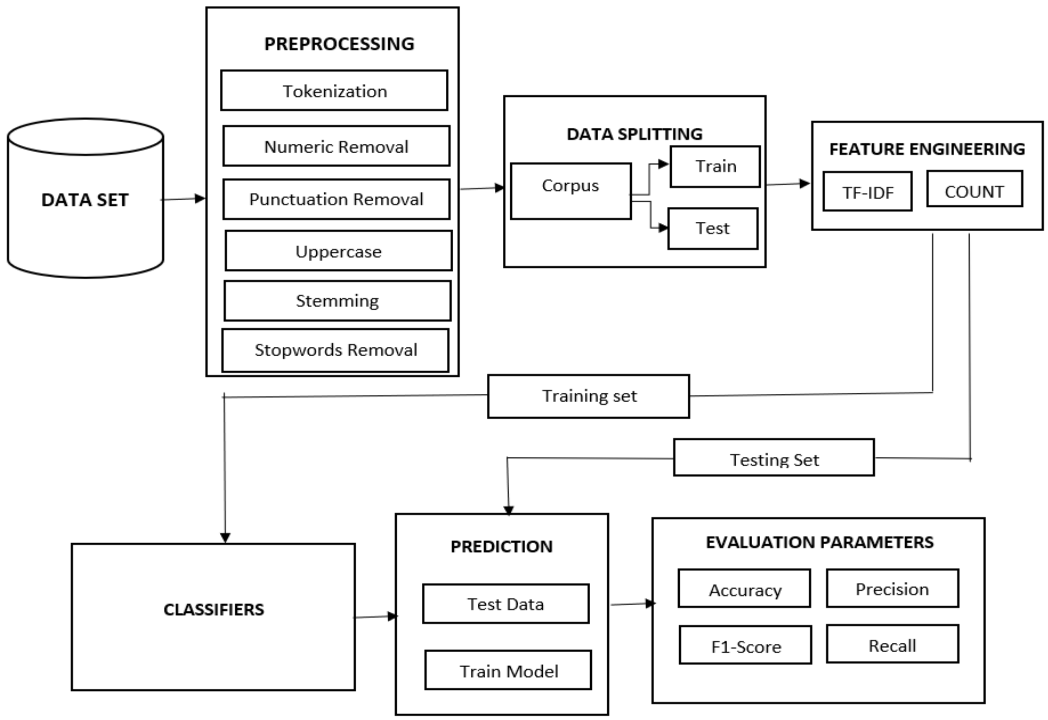

Figure 1. The proposed model design

3.1 Data Visualization

The breakdown of class sizes is displayed in Figure 2. According to the study, 7.02 percent of tweets are positive, while 92.98 percent are negative. This exemplifies how data visualisation may be used to discover previously unseen patterns within a dataset and learn more about it by graphically representing the data's properties.

3.2 Data Pre-Processing

Text pre-processing improves the prediction accuracy of the model by performing the following steps: tokenization, case-conversion, stop word removal, and removal of irrelevant punctuation. Datasets contain unnecessary data in raw form that can be unstructured or semi-structured, increasing the training time of the model and possibly degrading its performance.

3.3 Feature Extraction

Supervised machine learning classifiers need textual data in vector form to get trained on it, so in this work, we use TF-IDF and count vectorization techniques to convert textual features into vector form. This not only allows us to train our classifiers on the data, but also to make predictions based on the data.

3.4 TF-IDF

Term frequency (TF) is a measurement of how frequently a given term appears within a given document. Because the length of each document varies, TF can be calculated by taking into account the fact that a given term will appear more frequently in longer documents than in shorter ones.

$$TF(t) = \frac{No.of\ times\ term\ t\ shows\ in\ a\ document}{\text{Total no.of terms inside document}} \quad (1)$$

Every term is measured equally when TF is computed, but it is known that convinced terms, like "is," "of," and "that," can show up many more times except for containing small prominence, so frequent terms need to be weighed down as level up except

$$IDF(t) = \log(\mathrm{e}) \frac{\text{Total No.of documents}}{\text{No. of documents through term t in it}} \quad (2)$$

The frequency with which an expression (term, word) appears in a report is measured using a statistic called term frequency (TF), which is useful for data recovery.

3.5 Count Vectorization

An excellent tool is CountVectorizer, part of Python's scikit-learn library. It takes a text and generates a vector based on the frequency (count) of each word in the text as a whole (for use in further text analysis).

CountVectorizer makes a matrix for each document, where each unique word is a column and each text excerpt is a row. The value of each cell is the number of times that word appears in that excerpt.

PROPOSED MODELS FOR TWEETS SENTIMENT CLASSIFICATION

This section will discuss the classifiers that were used to categorise tweets. Figure 2 depicts the proposed data and workflow of this study. These algorithms included Support Vector Machines (SVM), Naive Bayes (NB), Random Forest (RF), Logistic Regression (LR), and a Voting Classifier (SVM, NB, RF, LR).

a. Support Vector Machines (SVM)

Information (shown as a vector) has been arranged in kind to achieve this goal. Next, the boundary is classified into two training sets by strategy. This is how the support vector machine (SVM) performs well as sentiment analysis.

b. Naive Bayes (NB)

The Innocent Bayes classifier is a probabilistic classifier that applies Bayes' Theorem with strong independence between the highlights, making it suitable for clustering with discrete highlights (e.g., word-means text characterization). They are extremely flexible.

c. Random Forest (RF)

RF is a tree-based classifier in which the trees are generated randomly based on the input vector. It uses random features to create multiple decision trees to make a forest, and then predicts the class labels of the test data by averaging the votes of all the trees.

d. Logistic Regression (LR)

In LR, class probabilities are estimated based on the output, such that if the input is from class X with probability x and from class Y with probability y, then the predicted output class is X, otherwise it is Y. Insight is a method used in logistics to show how likely it is that one group or the other will happen, such as top/bottom, white/black, up/down, positive/negative, happy/unhappy.

e. Voting Classifier(VC)

The Voting Classifier (VC) is a meta-classifier for joining similar or hypothetically exceptional ML classifiers for order through greater part-throwing a voting form, and it can achieve better performance than a single classifier by combining their predictions.

4. Evaluation Metrics

In classification tasks, ML models are evaluated using many standard metrics, including accuracy, recall, precision, and F1-score.

Confusion Matrix:

The confusion matrix, also known as the error matrix, is a table used to describe the model's performance on the set of test data. Each row of the matrix shows the instances in a predicted class, while each column shows the instances in an actual class (or vice versa).

- True Positive (TP): Observation is Positive, and is predicted to be positive.
- False Negative (FN): Observation is positive, but predicted negative
- True Negative (TN): Observation is negative, and is predicted to be negative.
- False Positive (FP): Observation is negative, but is predicted positive.

Accuracy:

The formula for accuracy is accuracy/Classification Rate, where accuracy is the fraction of correct predictions relative to the total number of input samples, and classification rate is the fraction of input samples correctly classified.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (3)$$

Recall:

A high recall indicates that the class is being correctly recognised and is defined as the ratio of the total number of correctly classified positive examples to the number of positive examples.

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

Precision:

Precision is the Total positive divided Total number of positive prediction (TP+FP)

$$\text{Precision} = \frac{TP}{(TP+FP)} \qquad (5)$$

F1-Score:

As a measure of a test's accuracy, the F1-score takes into account both its precision and recall.

$$\text{F-1 Score}=2\,\frac{(\text{Recall}*\text{Precision})}{(\text{Recall}+\text{Precision})} \qquad (6)$$

5. Experiment Results And Discussion

5.1 Dataset

There are many contradictory tweets in the dataset. Tweets in English are included in the final dataset, which is called "Sentiment Analysis of Hatred Speech on Twitter Data" and contains 31962 records. Each record is marked with a 1 or 0 to show whether it has a positive or negative emotional polarity. Table 1 contains features and description of each feature.

Table 1. Dataset Specification

Features	Description
Label	Class label of the tweet.
Sentiment Text	Original Text posted by the user

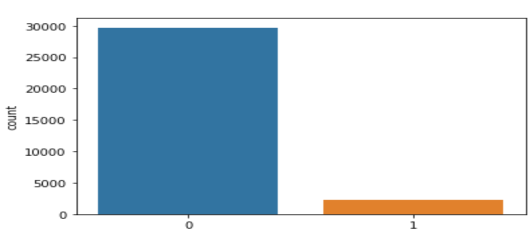

Figure 2. Count Plot Showing Class-Wise Data Distribution

5.2 Experiment Results And Discussion

A Voting Accuracy of 94% is achieved by a Voting Classifier using TF-IDF as an ensemble of Random Forest, Support Vector Machine, Naive Bayes, and Logistic Regression. Classification using TF-IDF features is evaluated with respect to accuracy, recall, precision, and F1-score, as shown in Table 2. Voting classifiers using TF-IDF achieved an accuracy value of 94%, with NB achieving the highest accuracy value of 95%. Results from all classifiers and a comparison of them using TF-IDF features are shown in Figure 3. The highest accuracy is achieved by combining a Count Vectorizer with a Voting Classifier, which is an ensemble of a Support Vector Machine, Naive Bayes, and Logistic Regression. Classification results using Count Vectorizer features are shown in Table 3 along with their associated Accuracy,

Recall, Precision, and F1-score. The highest accuracy was achieved by the voting classifier in conjunction with the Count Vectorizer (95%), followed by the LR and NB (94%). The highest accuracy was achieved by LR, at 74%, followed by the Voting classifier, at 57%. The highest recall was 79%, which was also achieved by SVM. The highest f1-score was 67%, which was achieved by the Voting Classifier. The results of all the classifiers are shown and compared in Figure 4. Utilizing the Count Vectorizer tool The Voting Classifier has the highest accuracy (95%) of all the classifiers. Maximum accuracy is achieved when a voting classifier with Count Vectorizer is combined with Random Forest, Support Vector Machine, Naive Bayes, and Logistic Regression.

TF-IDF Vectorizer

Table 2. The Output Of All Machine Learning Classifiers Is Based On Tf-Idf Features.

MODELS	ACCURACY	PRECISION	RECALL	F1-SCORE
RF	88%	57%	71%	46%
SVM	89%	63%	80%	51%
NB	95%	65%	44%	55%
LR	92%	63%	80%	61%
VC	94%	57%	66%	61%

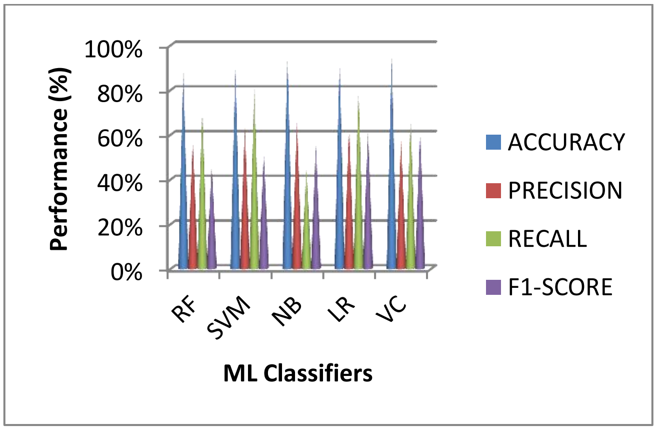

Figure 3. Classification Result Comparision Of All Machine Learning Models Using TF-IDF Features

Table 3: The outcome of every machine learning model's classification using the features from COUNT VECTORIZER

MODELS	ACCURACY	PRECISION	RECALL	F1-SCORE
RF	90%	62%	69%	51%
SVM	92%	71%	79%	60%
NB	94%	69%	61%	62%
LR	94%	74%	75%	65%
VC	95%	57%	70%	67%

Count Vectorizer

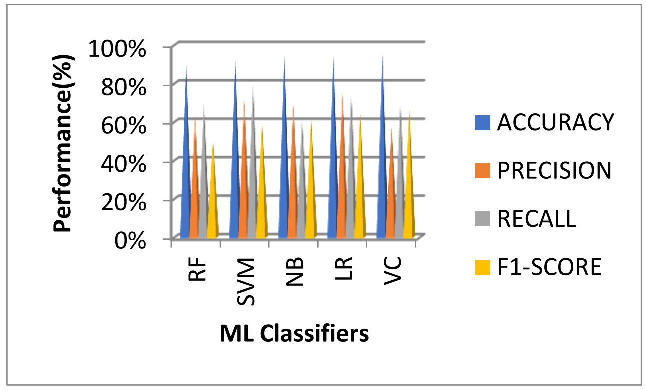

Figure 4. Classification performance evaluation across various machine learning models based on COUNT Vectorizer features.

Table 4. Comparative Analysis: Comparison of the proposed model to other, already-existing models for analysing Twitter sentiment

Author & Year	**Title**	**Techniques**	**Accuracy**
1) Bhumika Gupta and Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani, 2017	Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python	DAN2 SVM Bayesian Logistic Regression Naïve Bayes Random Forest Classifier Neural Network Maximum Entropy Ensemble classifier	86.06% 85.0% 74.84% 66.24% 87.5% 89.93% 90.0% 90.0%
2) Ankita and Nabizath Saleena,2018	An Ensemble Classification System for Twitter Sentiment Analysis	Naive Bayes Random Forest Support Vector Machine Logistic Regression Majority Voting Proposed Ensemble	72.81% 65.59% 72.91% 71.35% 71.14% 73.33%
3) Roza H. Hama Aziz and Nazife Dimililer,2020	Twitter Sentiment Analysis using an Ensemble Weighted Majority Vote Classifier	Logistic Regression (LR) Stochastic Gradient Descent (SGD) Naïve Bayes (NB)	87.2% 86.9% 86.84%

		Random Forest (RF) Decision Tree (DT) Support Vector Machine (SVM) Simple Majority Voting Weighted Majority Voting	86.3% 82.6% 85.3% 87.83% 90.8%
4) Chaudhary Jagrit Varshney , Dr. Ashish Sharma and Dhirendra Prasad Yadav,2020	Sentiment Analysis using Ensemble Classification Technique	SGD Classifier Naïve Bayes Logistic Regression Ensemble Voting Classifier	77.53% 77.50% 80.29% 79.89%
5) Anam Yousafi , Muhammad umer, Saima Sadiq, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara , and Michele Nappi,2021	Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)	***TF-IDF features:*** Random Forest Support Vector Machine Naive Bayes Logistic Regression Voting Classifier (LR-SGD)	 84% 88% 93% 88% 94%
6) Proposed work	Emotional Sentiment Analysis Of Tweet Data Using A Voting Classifier	***TF-IDF features:*** Random Forest Support Vector Machine Naive Bayes Logistic Regression Voting Classifier (LR) ***Count Vectorization :*** Random Forest Support Vector Machine Naive Bayes Logistic Regression Voting Classifier	 88% 89% 95% 92% 94% 90% 92% 94% 94% 95%

Ensemble methods for analysing Twitter sentiment are shown in Table 4. After implementing the ensemble techniques given by Bhumika Gupta et al. [1], the performance of their ensemble classifier using TF-IDF improved to 90% accuracy. The ensemble classifier developed by Ankit et al. [2], which used the Bag-of-words feature selection technique, demonstrated an accuracy of 73.33 percent. As demonstrated by Roza H. Hama Aziz et al. [3], their Weighted Majority Voting ensemble classifier using TF-IDF attained an accuracy of 90.8%. An accuracy of 79.95% was attained by the Ensemble classifier developed by Chaudhary Jagrit

Varshney et al. [4]. The ensemble classifier using the TF and TF-IDF feature selection approach demonstrated 94% accuracy, as demonstrated by the work of Anam Yousaf et al. [5].

The proposed method obtained 94% accuracy using the TF-IDF model's ensemble voting classifier and 95% using the count vectorization method. With other methods like Random Forest, Support Vector Machine, Naive Bayes, and Logistic Regression, a Voting Classifier built with Count Vectorizer may achieve the highest levels of accuracy.

6. Conclusion and Future Work

This work proposes a unique combination of FR, SVM, NB, and LR as a voting classifier for determining if a tweet is happy or sad. Seven machine learning models (SVM, RF, LR, NB, and VC) are tested and compared in experiments (RF, SVM, NB, LR). The authors also employed the utilisation of the TF-IDF and the Count Vectorizer feature representation methods. Testing on the Twitter dataset showed that all models did well, but our suggested voting classifier VC (RF, SVM, NB, LR) did the best when using both TF-IDF and Count Vectorizer. The suggested model employs a 95% effective count vectorizer to optimise performance. Future studies should compare different feature engineering approaches and investigate other combinations of ensemble models to further improve performance. To improve precision, we plan to use a more extensive dataset that takes emoticons and localizations into consideration.

References

[1] Bhumika Gupta and Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani ,"Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python,". In International Journal of Computer Applications (0975 – 8887) Volume 165 – No.9, May 2017

[2] Ankita and Nabizath Saleena, ‘‘An Ensemble Classification System for Twitter Sentiment Analysis,’’ in International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

[3] Roza H. Hama Aziz and Nazife Dimililer, ‘‘Twitter Sentiment Analysis using an Ensemble Weighted Majority Vote Classifier, Third International Conference on Advanced Science and Engineering (ICOASE2020)

[4] Chaudhary Jagrit Varshney , Dr. Ashish Sharma and Dhirendra Prasad Yadav, ‘‘Sentiment Analysis using Ensemble Classification Technique,’’ IEEE ,2020

[5] Anam Yousaf , Muhammad Umer , Saima Sadiq , Saleem Ullah, Seyefali Mirjalili(Senior Member, IEEE), Vainhav Rupapara , and Michele Nappi, (Senior Member, IEEE) , "Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)",IEEE,2021

[6] N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, ‘‘Tweet sentiment analysis with classifier ensembles,’’ Decis. Support Syst., vol. 66, pp. 170–179, Oct. 2014.

[7] C. Kariya and P. Khodke, ‘‘Twitter sentiment analysis,’’ in Proc. Int. Conf. Emerg. Technol. (INCET), Jun. 2020, pp. 212–216.

[8] A. Alsaeedi and M. Zubair, ‘‘A study on sentiment analysis techniques of Twitter data,’’ Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 2, pp. 361–374, 2019.

[9] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, ‘‘Lexicon based feature extraction for emotion text classification,’’ Pattern Recognit. Lett., vol. 93, pp. 133–142, Jul. 2017

[10] H. Hakh, I. Aljarah, and B. Al-Shboul, ‘‘Online social media-based sentiment analysis for us airline companies,’’ in New Trends in Information Technology. Amman, Jordan: Univ. of Jordan, Apr. 2017.

[11] R. Xia, C. Zong, and S. Li, ‘‘Ensemble of feature sets and classification algorithms for sentiment classification,’’ Inf. Sci., vol. 181, no. 6, pp. 1138–1152, Mar. 2011.

[12] M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, ‘‘A novel stacked CNN for malarial parasite detection in thin blood smear images,’’ IEEE Access, vol. 8, pp. 93782–93792, 2020.