International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# Emotion Recognition by Textual Tweets Classification Using Voting Classifier LR-SGD

**[1] Bharath Patil I, [2] Dr Shankaragowda B B**

[1]Sem MCA Student, Department of MCA, BIET, Davanagere
[2]Associate Professor and HOD, Department of MCA, BIET, Davanagere

**ABSTRACT:** The proliferation of user-generated content on social media has made opinion mining an arduous job. As a microblogging platform, Twitter is being used to collect views about products, trends, and politics. Sentiment analysis is a technique used to analyze the attitude, emotions and opinions of different people towards anything, and it can be carried out on tweets to analyze public opinion on news, policies, social movements, and personalities. By employing Machine Learning models, opinion mining can be performed without reading tweets manually. Their results could assist governments and businesses in rolling out policies, products, and events. Seven Machine Learning models are implemented for emotion recognition by classifying tweets as happy or unhappy.With an in-depth comparative performance analysis, it was observed that proposed voting classifier(LR-SGD) with TF-IDF produces the most optimal result with 79% accuracy and 81% F1 score. To further validate stability of the proposed approach on two more datasets, one binary and other multi-class dataset and achieved robust results.

Automatic emotion recognition, pattern recognition and computer vision have become significantly important in Artificial Intelligence lately with applications is a wide range of areas. Recently, social media platforms such as Twitter have generated enormous amounts of structured, unstructured and semi-structured data. One of the most recent example is COVID-19 infodemic that shows misinformation in social media can be far more important and devastating than a disaster such as a pendemic.

There is a need to analyze to accurately assign sentiment classes on a large scale. To perform such tasks, accurate NLP techniques and machine learning (ML) models for text classification are required. Twitter provides an opportunity to its users to analyze its data on a large and broader point of view. Efficient methods are important to automatically label text data due to its noisy nature. In the past many studies have been performed on Twitter sentiment classification [1]. As Twitter is very fast and an efficient micro-blogging examination that facilitates the end users to transmit small posts are said to be tweets. Twitter is a highly demanding app in the world and is a successful platform in social media[2].

Free account can be created by using Twitter that can provide an enormous audience potential. With the purpose of business and marketing, Twitter can be proved as the best platform, through which one can get in touch with very rich and famous personalities like stars and celebrities, so their purchasing can be very charming for them as well as for advertisers. Using Twitter, every celebrity is linked with fans as well as to grant a communication to followers. Such a platform is one of the superlative approaches for lovers as well [3]. But, it has a short note range; only 140 letters for each post and it can type a post or link on the website since it has no cost and also open as the advertisements as well. There is no problem with clusters of personal ads which are similar to other social networking sites. It is quick because as a tweet is posted on Twitter, the public who is subsequent to respective business will get it without delay.

Companies and advertisers can compose utilization of this source to check the diverse operational point of views which are very considerable. With help of this, they will obtain an immediate response from their followers. Remarkably, a lot of businesses with the intention of purchase, Twitter followers increase their deals. Twitter facilitates the followers by making them identify regarding fresh

business, products, services, websites, blogs, eBooks etc. Consequently, Twitter clients might tick lying on link and also optimistically endow in a manufactured goods or examine the products presented and to get share in profit. It is extremely effortless to utilize as people can follow to get the news and updates, as organizations can tweet or re-tweet, they can mark favorite or selected people to send the tweets, also know how to propel the posts plus to b able to endow their money and instance through it. Academy, Industry, super bowls and Grammy Awards of such major Sports and Entertainment events generate a lot of buzz in the global world by using it [5].

Competition is rising among different products on Twitter. People love to express their feelings about a particular product on social networks like twitter. Product owners are ready to spend more money on social media platforms to better advertise their products and to generate more revenue. When a person shares experience about a product, it helps the owner to change their market strategy, selling schemes, and improving the quality. Customer reviews serve as a feedback to the owners or manufacturers too .The data generated in such a way is of large amount and requires an analysis expert team to classify the customer sentiment from the reviews [4], [5]. Experts can make a human error in sentiment analysis, therefore it requires machine learning and ensemble learning classifiers to accurately classify the sentiment of the customers. This study compare various machine learning models for emotion recognition by tweet classification using Tf and TF-IDF. This research presents a voting classifier (LR- SGD) and aims to estimate the performance of famous ML classifiers on twitter datasets. The key contributions are as follows: _

Machine learning-based classifiers including support vector machine (SVM), Decision Tree Classifier (DTC), Naive Bayes (NB), Random Forest (RF), Gradient Boosting Machine (GBM) and Logistic Regression (LR) trained on Twitter dataset are compared for emotion recognition [6].

A voting classifier (VC) designed to classify tweets which combines LR and SGD and outperformed using TF-IDF.

The proposed model stability is further validated by applying it on two different datasets, one binary dataset (containing hatred or non hatred classes) and other multi-class dataset (containing product reviews having 1 to 5 ratings).

**EXISTING SYSTEM**

❖ Sarlan *et al.* [2] established a sentiment analysis through extracting number of tweets with the help of prototyping and the results organized customers' views via tweets into positive and negative. Their research divided into two phrases. The first part is based on literature study which involves the Sentiment analysis techniques and methods that nowadays are used. In the second part, the application necessities and operations are described preceding to its development.

❖ In another research Alsaeedi and Zubair Khan [3] analyzed various kinds of sentiment analysis that is applied on to Twitter dataset and its conclusions. The distinct approaches and conclusions of algorithm performance were compared. Methods were used which were supervised ML based,, lexicon- based, ensemble methods. Authors used four methods that were Twitter sentiment Analysis using Supervised ML Approaches; Twitter sentiment Analysis using Ensemble Approaches. Twitter sentiment Analysis is using lexicon based Approaches.

❖ Lexicon based approaches have been explored by many researchers for emotion classification. Bandhakavi *et al.*

[4] performed emotion-based feature extraction using domain specific lexicon generation.

**Disadvantages**

➢ The existing model which is ensemble of LR and SGD is not applied on both dataset and the results.

➢ Voting Classifier(VC) is not a cooperative learning which engages multiple individual classifiers. **P**

**PROPOSED SYSTEM:**

❖ In the proposed system, different techniques have been used for methodology in ML for its objectives. Versatile experiments were examined using different methods and techniques. Multiple classifiers applied on the dataset, but the Voting classifier is an ensemble of Logistic Regression and Stochastic Gradient Descent

outperforms than all other ML models in terms of accuracy, recall, precision and F1-score.

❖ Twitter dataset used in this experiment is scrapped from Kaggle repository. First the dataset is pre- processed by removing unwanted data. Then, the data was split into two sets: training set and testing set. The training set was given the percentage of 70% while the test set portion is 30%. After that feature engineering techniques are applied on the training set. Multiple machine learning classifiers are trained on the training set and tested using the test set. The evaluation parameters used in this experiment are: (a) Accuracy (b) Recall (c) Precision (d) F1-score.

**Advantages**

➢ The proposed system presents a voting classifier (LR-SGD) and aims to estimate the performance of famous ML classifiers on twitter datasets.

➢ Data Visualization helps to understand the hidden patterns lying inside the dataset. It helps to qualitatively get more details about the dataset by visualizing the characteristics of the attributes.

**IMPLEMENTATION**

**Service Provider**

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as
Login, Train and Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, Predict Emotion From Data Set Details, Find Emotion Prediction Ratio on Data Sets, Download Trained Data Sets, View Emotion Prediction Ratio Results,
View All Remote Users.

**View and Authorize Users**

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

**Remote User**

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, POST TWEET DATA SETS, SEARCH AND PREDICT EMOTION, VIEW YOUR PROFILE.


**Fig.1. Home page.**


**Fig.2. user details login page**

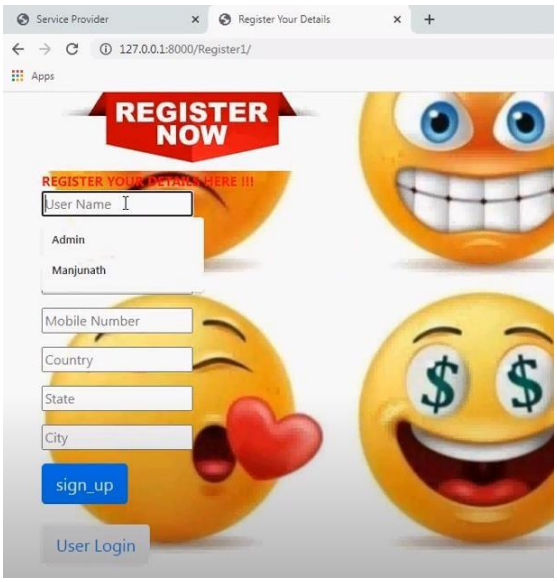


**Fig.3. Registration page.**


**Fig.4. Dataset details.**

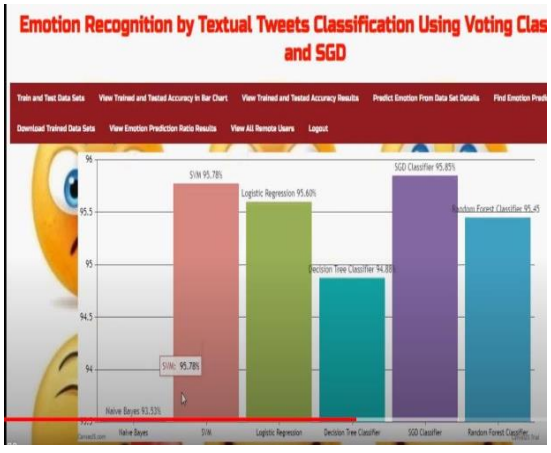


**Fig.5. Output graphs with different algorithms.**

**Fig.6. Output results.**

## CONCLUSIONS

This paper proposed a novel combination of LR and SGD as a voting classifier for emotion recognition by classifying tweets as happy or unhappy. Our experiments showed that one can improve the performance of models by recognizing patterns efficiently and through effective averaging combination of models. Experiments are conducted to test seven machine learning models that are; (1) SVM, (2) RF, (3) GBM, (4) LR,

(5) DT, (6) NB and (7) VC(LR-SGD).

This study also employed two feature representation techniques TF and TF- IDF. The results showed that all models performed well on tweet dataset but our proposed voting classifier VC(LR-SGD) outperforms by using both TF and TF- IDF among all. Proposed model achieves the highest results using TF- IDF with 79% Accuracy, 84% Recall and 81% F1-score. The proposed model is further validated on two more dataset and achieved robust results. The future work will compare more feature engineering techniques and explore more combinations of ensemble models to improve the performance. In addition, new techniques will be investigated to deal with sarcastic comments.

## REFERENCES

[1] N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, ``Tweet sentiment analysis with classi_er ensembles," Decis. Support Syst., vol. 66, pp. 170_179, Oct. 2014.

[2] C. Kariya and P. Khodke, ``Twitter sentiment analysis," in Proc. Int. Conf. Emerg. Technol. (INCET), Jun. 2020,
pp. 212_216.

[3] 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.

[4] A. Bandhakavi, N.Wiratunga, D. Padmanabhan, and S. Massie, ``Lexicon based feature extraction for emotion text classi_cation," Pattern Recognit. Lett., vol. 93, pp. 133_142, Jul.

2017.

[5] J. Capdevila, J. Cerquides, J. Nin, and J. Torres, ``Tweet-SCAN: An event discovery technique for geo-located tweets," Pattern Recognit. Lett., vol. 93,
pp. 58_68, Jul. 2017.

[6] T. Alsinet, J. Argelich, R. Béjar, C. Fernández, C. Mateu, and J. Planes,
``An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships," Pattern Recognit. Lett., vol. 105, pp. 191_199, Apr. 2018.

[7] W. Chen, Y. Zhang, C. K. Yeo, C. T. Lau, and B. S. Lee, ``Unsupervised rumor detection based on users' behaviors using neural networks," Pattern Recognit. Lett., vol. 105, pp. 226_233, Apr. 2018.

[8] 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.

[9] R. Xia, C. Zong, and S. Li,

``Ensemble of feature sets and classi_cation algorithms for sentiment classi_cation," Inf. Sci., vol. 181, no. 6,
pp. 1138_1152, Mar. 2011.

[10] 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.