International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
***Article Received:** 28 July 2023 **Revised:** 19 September 2023 **Accepted:** 13 October 2023*

Learning to Detect Human Emotions in Digital World by Integrating Ensemble Voting Classifiers

Mr. Shyelendra Madansing Pardeshi [1;2] *, Prof. Dr. Dinesh Chandra Jain [1]
Research Scholar, CSE Department,
[1] Oriental University, Indore, MP, India.
[2] R. C. Patel Institute of Technology, Shirpur, MH, India.
0000-0001-6297-6074
*Pardeshishailendra1184@gmail.com
Head and Professor, CSE Department,
[1] Oriental University, Indore
Madhya Pradesh, India
dineshjain25210@gmail.com

Abstract— Due to the expansion of world of the internet and the quick acceptance of platforms for social media, information is now able to exchange in ways never previously imagined in history of mankind. A social networking site like Twitter offers a forum where people may interact, discuss, as well as respond to specific issues via short entries, like tweets of 140 characters and fewer. Users may engage by utilizing the comment, like and share tabs on texts, videos, images and other content. Although platforms for social media are now so extensively utilized, individuals are creating as well as sharing so much information than shared before, which can be incorrect or unconnected to reality. It is difficult to identify erroneous or inaccurate statements in textual content autonomously and find emotions of people. In this paper, we suggest an Ensemble method for sentiment and emotion analysis. Different textual features of actual and Emotion and sentiment have been utilized. We used a publicly accessible dataset of twitter sentiment analysis that included total 48,247 authenticated tweets out of 23,947 of which were authentic positive texts labeled as binary 0s and 24,300 of which were negative texts labeled as binary 1s. In order to assess our approach, we used well-known (ML) machine learning techniques, these are Logistic Regression (LR), AdaBoost, Decision Tree (DT), SGD, XG-Boost as well as Naive Bayes. In order to get more accurate findings, we created a multi-model sentiment and emotion analyzing system utilizing the ensemble approach and the classifiers stated above. Our recommended ensemble learner method outperforms individual learners, according to an experimental study.

Keywords- Emotion recognition, Sentiment analysis, Machine learning (ML), Social media, NLP, Ensemble.

1. INTRODUCTION

The essential phase in performing emotion analysis is categorizing retrieved data distinct emotion polarity like negative and positive classifications. A variety of emotions may be examined, becoming the subject of the growing disciplines of effective computer vision and text analytics. It has other ways that split feelings based on study subjects. For instance, in political discussions, sentiments or emotions might be classified into angry and satisfied and furious. Text analytics with ambiguous handling may be used to provide for extremely fine outcomes and define emotions like angry, sad, excitement, anxiety and happy. Emotion analysis is an aspect of text extraction that discovers and focuses emotion information gathered from original material, allowing a company to assess the public perception of the entire goods or services while monitoring web discussions. Sentiment classification is undoubtedly the most frequently used text-ordering tool. It determines if incoming motivating factors are negative, neutral or positive. The investigation into internet entertainment sources, on the other hand, is primarily limited to basic opinion evaluation and makes a difference assessment. This really is analogous to beginning to uncover what's beneath and giving up the high-value information that is completely buried. So how would a novice approach the lesser fruit? As just a result, text processing algorithms are more adaptable [1] [2].

M. Z. Asghar, S. Ahmad, et al. It is hard to argue for the detection and classification of activist Twitter messages. Extremist groups have exploited social networking platforms like as Twitter, Facebook, etc. to propagate their messaging and attract new members [9].

Wenyu C and colleagues define Recognition system is an area of classification of sentiment which further includes emotion appraisal and retrieval. With the emergence of web technologies, text categorization and analysis have risen to the top of organizational performance [8, 3].

Aykuth uven et al. state the smart technologies used during everyday life defined as a comprehensive on the customers via sensors linked to it. Data collected is primarily physical

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
***Article Received:** 28 July 2023 **Revised:** 19 September 2023 **Accepted:** 13 October 2023*

information, but applications capture much more, including system resource trends as well as personal preferences [10].

Syal Varun et al. address the problem of recognizing, categorizing, and measuring moods within text almost in any format. It consider English language into consideration. Identifying and categorizing fundamentalist tweets really a contentious issue [6].

Table 2 signifies authenticated metadata represents approach used like text based or other, corpus dataset links, social media platform used like twitter or other and states of emotion-sentiment by researchers.

Emotion assessment in the text, as defined by Khaled Shaalan et al. is a relatively new topic in sentiment classification. Emotion evaluation is the method of analyzing and classifying a message through group based on prior research using feelings qualitative models presented in psychology theory. To recognize emotional states inside writing, emotion evaluation and categorization utilized [4].

2. RELATED WORK

Figure 1 represents collected dataset of total 48,247 authenticated tweets segmented into positive and negative tweets. Binary 1 consider for positive tweets and binary 0 for negative tweets. After segmentation identified 23,947 are positive tweets and 24,300 are negative tweets [7, 5].

Pre-processing of data in order to better handle the unbalanced data increased the performance of these algorithms [12, 27]. Nordberg, P., et al. developed critical methodologies for automated recognition of emotions serves to establish the foundation for users assistance system meant to assist customers in identifying and avoiding fake information [7, 18]. For dataset, Ashraf, N. et al. developed many classifiers utilizing ML algorithms for the objectives of new allegation and subject categorization [9, 19]. Perez Rosas, et al. [11, 20] focused on the classification of false material automatically through social media. To identify emotions, they offered two more datasets spanning seven unique news categories.

Text mining seems to be a machine intelligence technique that distinguishes between right and wrong lexical objects such as words or phrases and revises them into new concepts. Text analytics is a subfield of extracting information, machine learning methods, retrieval of information and semantics. The volume of published information has expanded considerably since the advent of social networking sites and it continues to grow on a routine basis. People interact digitally by interpersonal conversation, creating or transferring multimedia learning assets via digital streaming platforms, participating in providing ratings, comments and evaluations in blogging and recommender systems [8, 13].

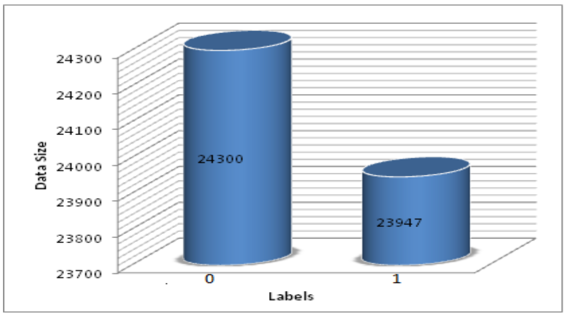


Fig. 1: Visualization of positive and negative texts available in dataset contains 23,947bare positive tweets and 24,300 are negative tweets.

3. MATERIALS AND METHODS

Real-world data is challenging to directly create models with machine learning via it usually includes noise, inadequate data and might come in an unfavorable form. Preparing the data for a machine-learning (ML) model by cleaning it is important in order to increase the model's precision and efficiency [10, 29]. In next covers Feature Extraction, TF_IDF Features CountVectorizer Features and so on.

3.1.1 Feature Extraction:

The extracted features include TF-IDF weights and CountVectorizer. The extracted features are subsequently sent with the edited text.

3.1.2 TF_IDF Features:

The abbreviation Term Frequency is referred as TF and IDF Document Frequency Inverse. This method analyzes the number of words within a collection of information to quantify a word's significance to the document and corpus, a score is often assigned to it. This method is frequently used in purposes for retrieving data and analysis of texts. The information can then be vectorized, which allows us to carry out additional operations like ranking, clustering, and locating the pertinent documents.

$$TF\text{-}IDF = TF(t, d) * IDF(t) \qquad (1)$$

Whereas t mention as number of times the context appears in doc d. The term frequencies-inverse documents frequencies (TF-IDF) represents a measure of statistical significance that assesses the significance of a word to an articles within an ensemble of text documents mention in equation (1). To do this, along a document's word intensity as well as variational document group of documents are multiplied. We need to determine the length of the document and how many word each vocabulary phrase has in order to compute TF. The TF value for a particular document will be zero if the word is missing from it. If the manuscript solely contains similar words, the worst-case situation is that TF will be 1. The final value of the normalized TF value will lie in the range [0, 1] among 0 and 1 [15].

Every document and word has its own TF, thus we can define in equation (2):

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
Article Received: 28 July 2023 Revised: 19 September 2023 Accepted: 13 October 2023

tf(t,d) = t in d count / the amount of words within d (2)

3.1.3 CountVectorizer Features:

Text must be analyzed to eliminate certain terms before it can be used for predictive modeling; this procedure is known as tokenization. Such words must be subsequently encoded to be integer as well floating-point numbers before being utilized as data input for ML methods. This method is referred to as feature extraction [12, 14].

3.2 Emotions and Sentiment Analysis using Machine Learning Classifiers

Decision Tree (DT), XG-Boost (XGB), Extra Trees (ET), Random Forest' (RF), AdaBoost (AB), Logistic Regression' (LR), Support Vector Machine' (SVM), Naive Bayes and Stochastic Gradient Descent (SGD) were the nine machine learning classifiers whose performance we evaluated for the detection of Emotion and sentiment (NB). We designed a multi-model classification model using these nine machine learning classifiers, with ensemble methods being the deciding factor. Each classifier was implemented in Python using the scikit-learn library [5, 17]. Each classifier is discussed briefly below.

3.2.1 Emotions and Sentiment Analysis Using Machine Learning Ensemble Methods

3.2.1.1 Ensemble Voting Classifiers

Voting ensemble is commonly used for task classification because of its capacity to mix several different learning algorithms which were trained with the Complexity entire dataset. [35]. Each approach predicts a hypothetical portion of information and that estimate is considered a voting in support of class picked by the structure. Following the prediction of each model, the final predictions are based upon that majority of votes for a particular class [32]. Voting collective algorithms are easier to implement than boosting & bagging. As previously discussed, bagging algorithms generate a number of datasets at random sampling and replacing all of the dataset's data. A model is then trained using each dataset, and the outcome is an amalgamation of the results from every model. A generic model which can accurately classify the issue is created in the instance of boosting by developing a number of models successively, each one learning from the one before it by enhancing weights for points that were incorrectly categorized [16, 24]. Voting ensemble, in contrast, combines a number of distinct models to produce classification results that agree with the majority's overall prediction. In this case, it breaks down a conceptual framework in to the two or more sub-models. There are five in total. The ensemble technique is used to incorporate predictions from each sub-model. It is a meta-classifier that, through a majority vote, determines whether two machine learning classifiers are conceptually equivalent or dissimilar. We use ensemble technique to predict last class name, that corresponds to the categorical variable has categorization methods frequently forecast.. We predicted the class label y using formula and also the majority vote of the each classification model Cj [23, 27].

3.2.1.2 Decision Tree (DT)

One of the categorization techniques that is most frequently used is tree-based modeling. It is quite effective and also has classification accuracy that is equivalent to different learning strategies. The tree of decisions has structure that represents knowledge-based classification algorithm. It uses a classification of decision trees paradigm that is simple to understand. The approach assesses each practical data divides test as well as selects the one the particular that provides the most new knowledge [21].

3.2.1.3 Boosting Ensemble Classifiers

Another well-liked ensemble method for converting ineffective models into effective learners is boosting. The final prediction is based on the outcomes of the majority vote of each tree in a forest of randomized trees that has been educated for that purpose. This strategy enables poor learners to incrementally and correctly identify data points that are typically misclassified. To categorize a specific issue, initial equally weighted coefficients are applied to all data points. The weighted coefficients change in the subsequent rounds, going up in order to properly categorized data points & down for ones that were misclassified [31]. Each successive tree created in a round gains knowledge by properly identifying data points that have been incorrectly classified in earlier rounds, hence reducing the errors from previous round and improving overall accuracy. Over fitting to the training data, which can result in inaccurate predictions for cases that have not yet been observed, is a significant issue with boosting ensemble [25, 33]. There are numerous boosting techniques for solving categorization and issues with regression. For purpose of classification in our trials, we employed the AdaBoost [34] and XGBoost [35] algorithms.

3.2.1.4 Logistic Regression' (LR)

The logistical regression (LR) technique is used as it provides an easy method for categorizing problems either basic to various categories. Because it allows for categorizing content using an extensive feature collection that produces an outcome as binary (true, incorrect or true article) [27]. While several parameters were tested before obtaining the maximum levels of We performed parametric tuning in order to acquire optimal result for every piece of dataset so as to improve performance from the LR model.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
***Article Received:** 28 July 2023 **Revised:** 19 September 2023 **Accepted:** 13 October 2023*

3.2.1.5 Stochastic Gradient Descent' (SGD)

This SGD approach is a straightforward but effective method for optimizing linear classification algorithms as well as regression coefficients under convex loss functions, such as those employed with SVM and LR. It have been used in machine-learning over some time, however it is only now gaining momentum in the context of huge processing. Large-scale as well as unstructured algorithms issues, which frequently arise in text classification and text mining, have been effectively addressed using SGD [26]. In essence, the SGD algorithm uses a simple SGD method of learning that supports different Incentives and classifications loss function. Scikit learn includes the SGDClassifier plugin for SGD classifier modulation. [28].

3.2.1.6 Voting Ensemble Classifiers

The vote ensemble is commonly used for the classification of data since its capacity to mix multiple learning algorithms that were trained with the difficulty level and complete dataset [37]. Every model predicts an outcome for a subset of the collected scores and that results can be considered to be a 'vote' towards support of the group predicted by the system. The resultant prognostication is made a majority vote in favour of a class after early technique has predicted the outcome [32]. Comparing voting ensemble to bagging & boosting algorithms, the former is easier to implement. As previously stated, bagging techniques generate numerous data sets at random selecting as well as updating all of the set of data. A model is then trained using each dataset, and the outcome is an amalgamation of outcomes between every technique. Ultimately the case of increasing, different models are trained successively, with every model learning on the previous model by increasing weights over erroneously classified information, resulting in an overall system able to correctly categorizing this problem. [29].

3.3 Performance Metrics

Analyze effectiveness of the methods used a variety of measures. The majority of them are built upon the confusion-matrix. A classification model's effectiveness on a testing sample with four variables true positive, false positive, false negative and true negative is displayed a table called the confusion matrix.

3.3.1 System evaluation and experimental results

A total of 48,247 verified tweets were gathered and split into good and negative tweets. Consider using binary 1 for positive tweets as well as binary 0 for negative tweets. After segmentation, 23,947 positive tweets and 24,300 negative tweets were discovered. The classifiers split the dataset into training and testing halves in an 80:20 ratio [16].

Metrics of Evaluation:

Since that we are developing a system for classification in this case, a prediction of true for an item that was actually false (a false positive) can have adverse impacts. In the majority of instances, an excellent accuracy rating suggests that the model is effective. In a similar vein a prediction of false for an article that contains factual information). precisely an outcome, we used three extra metrics which account for recall, precision and F1-scores for the wrongly classified observation [7, 18]. At below equation 3, 4, 5 and 6 a TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative. Equation 3 used to find accuracy, equation 4 used to find precision, equation 5 used to find recall and equation 6 used to find F1score for analysis purpose.

'Accuracy' = (TP+TN)/(TP+FP+FN+TN) (3)

'Precision' = TP/(TP+FP) (4)

'Recall' = TP/(TP+FN) (5)

'F1 Score' = 2*(Recall * Precision)/ (Recall +precision) (6)

3.3.2 Classifier Performance Evaluation Using the Emotion Dataset:

The accuracy, recall, precision and F1scores of the classifier on the Kaggle Emotion and sentiment dataset are shown in Table 1 and Fig. 2.

The Tree of Decision (DT) classifier achieved 66.18 % accuracy, 71 % precision, 66 % recall, and 74% on the F1 scale. The accuracy, recall, precision and F1scores of the Linear Regression (LR) classifier were 85.4 %, 85 %, 85%, and 85 % respectively [22].

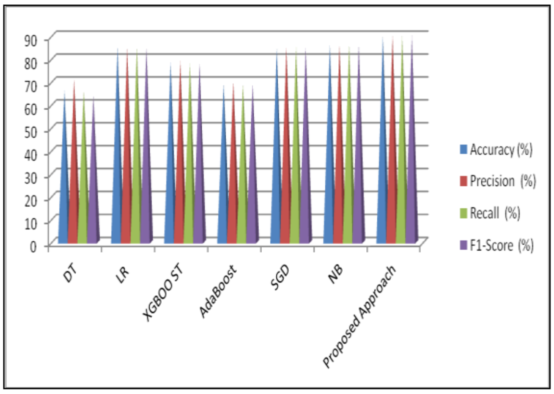


Fig 2: Analysis of classifier performance on dataset

TABLE 1: CLASSIFIER PERFORMANCE EVALUATION USING THE TWITTER SENTIMENT DATASET

	DT	LR	XG-BOOST	Ada-Boost	SGD	NB	Proposed Approach

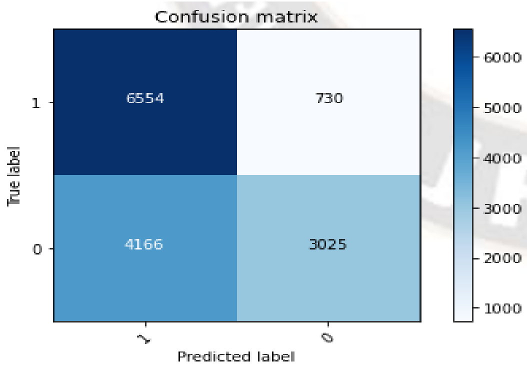
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
***Article Received:** 28 July 2023 **Revised:** 19 September 2023 **Accepted:** 13 October 2023*

Accuracy (%)	66.18	85.4	78.43	69.05	84.64	86.11	89.89
Precision (%)	71	85	79	70	85	86	90
Recall (%)	66	85	78	69	85	86	90
F1-Score (%)	64	85	78	69	85	86	90

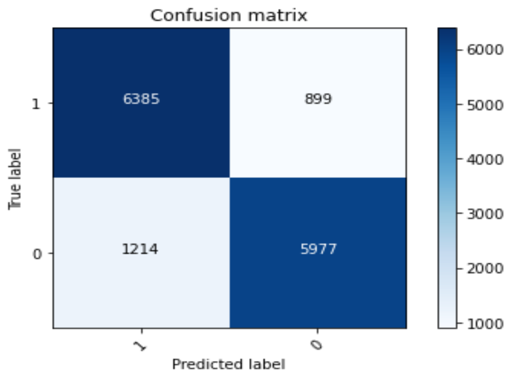
A 78.43 % accuracy', 78 % recall, 79 % precision'and 78 % F1-score' were attained by the XGBoost classifier. The AdaBoost classifier achieved 69.05 % accuracy, 70% precision, 69% recall, and 69% F1-score.The Stochastic Gradient Decent (SGD) classifier achieved 84.64 % accuracy, 85 % precision, 85 % recall, and 85 % F1-score.The Naive Bayes (NB) classifier achieved an F1-score of 86.11 %, accuracy of 86 %, precision of 86 %, and recall of 86 % [19].

Our proposed multi-model classifier, which relies on ensemble technique, achieved 89.89% accuracy, 90% precision, 90% recall as well as 90% F1-score. In comparison the mentioned classifiers, it was discovered that our proposed approach produced results that were more precise and stable [33].

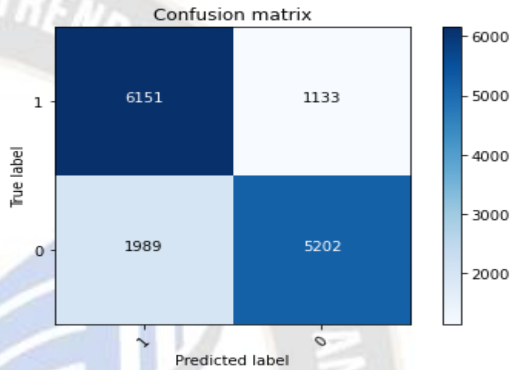
An method's efficacy in the area for learning by machine, especially the difficulty with respect to statistical categorization, can be visualized using a particular table structure known as a matrix of confusion, also known as an error matrix. The Confusion Matrix for the following methods is shown in Figure 3: (i) the Decision Tree (DT), (ii) logarithmic regression (LR), (iii) X G Boost (XGB), (iv) the AdaBoost (AB), (v) the stochastic gradient descent (SGD), (vi) naive Bayesian (NB), and (vii) suggested technique via collective technique [26, 35].



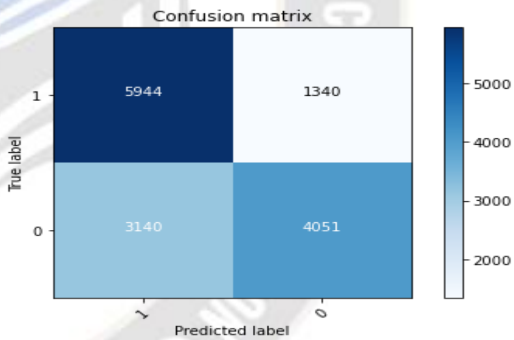
(i) Decision Tree



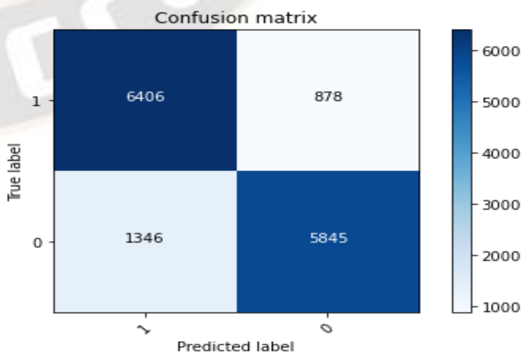
(ii) Logistic Regression



(iii) XGBOOST

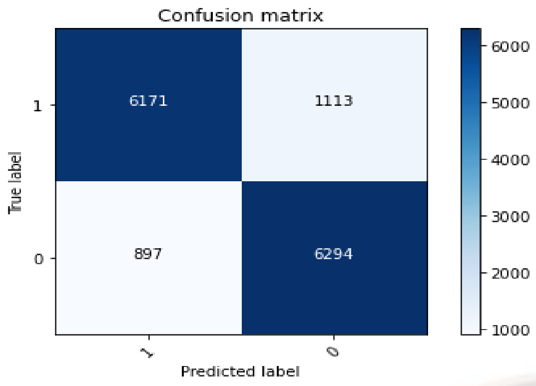


(iv) AdaBoost

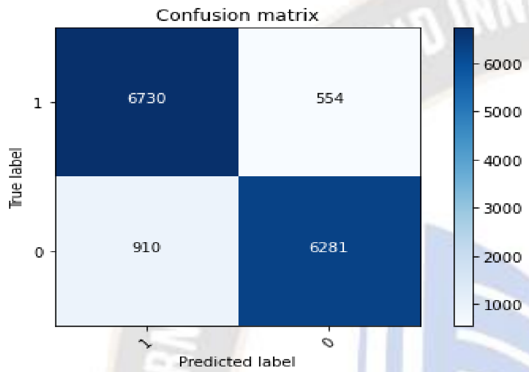


(v) SGD Classifier

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
Article Received: *28 July 2023* ***Revised:*** *19 September 2023* ***Accepted:*** *13 October 2023*



(vi) Naive Bayes Classifier



(vii) Recommended approach using ensemble method

Fig. 3: Confusion Matrix for (i) Decision Tree (DT), (ii) Logistic Regression (LR) (iii) XGBoost (XGB) (iv) AdaBoost (AB) (v) Stochastic Gradient Descent (SGD) (vi) Naive Bayes (NB) and (vii) recommended method using ensemble method.

A binary categorization system's diagnosing skills are illustrated visually by its ROC (Receiver operating characteristic curve) curved when its detection level is altered. A recipient's operational characteristics curve (i.e. ROC-curve) represents a graph which displays an algorithm's classification performance, across all categorization levels. The one illustrates a pair of variables: The percentage of true positives. The percentage of false positives.

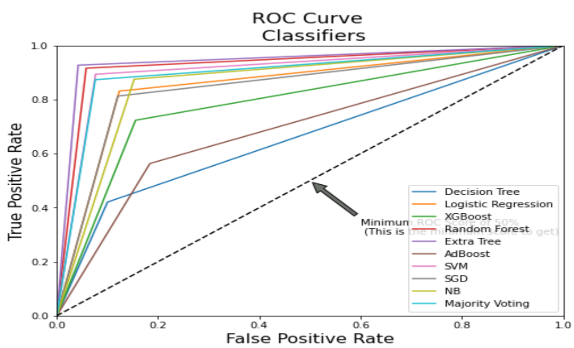


Fig. 4: ROC curve for depicts the effectiveness of a model for classification

Figure 4 visualize ROC curve for above used classifiers. It depicts in comparison to the aforementioned classifiers. it was discovered that our proposed approach produced results that were more accurate and stable [25, 34].

4. CONCLUSION

With the rise of social media, chances to increase the variety of emotions or thoughts that individuals honestly share about themselves as well as their competition have emerged. Humans in the modern world are continuously sharing their emotions, feelings and sentiments via social media. His personal feelings, thoughts, sentiments, as well as other emotions, through social networks like Twitter or other excellent places to begin exploring public sentiment. In this study, we compare several classifiers with different feature sets to reach the predicted results. To get more appropriate results, we constructed a emotion detection system with several models utilizing the Ensemble approach which mentioned classifiers and attributes. The experimental results show that our proposed strategy delivered more accurate and reliable outcomes. In future possible to increase classifiers and parameters for increase accuracy.

5. ACKNOWLEDGEMENTS

I want to express the deepest appreciation to Prof. Dr. Dinesh Chandra Jain, my research mentor, for their continuous support, guidance along with insightful criticisms of this study.
I would also like to thank Dr. R. B. Wagh, Prof. V. D. Punjabi, Dr. D. R. Patil and Prof. T. M. Pattewar for their help in maintaining my development on course.
Finally, Want to express gratitude to my parents, wife & entire family for their support and encouragement throughout my education.

REFERENCES

[1] Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. IEEE Access: Practical Innovations, Open Solutions, 8, 181074–181090.

[2] Shyelendra M. Pardeshi, Dr. Dinesh Chandra Jain (2021). A Study of Sentiment Analysis from Text through Social Networking Sites. Webology (ISSN: 1735-188X), Volume 18, Number 4, 2021, 666-675.

[3] Referred lexical data from link https://www.lexalytics.com/technology/sentiment-analysis.

[4] Said A. Salloum, Mostafa Al-Emran, Azza Abdel Monem & Khaled Shaalan (2017). A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives. Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 1, 127-133.

[5] Shyelendra M. Pardeshi, Dr. Dinesh Chandra Jain, et al.(2022). Predicting Public Opinion on Current Events through Twitter Trend Analysis. High Technology Letters. ISSN NO: 1006-6748. Volume 28, Issue 4, 2022, 454-460.

[6] Salminen, J., Hopf, M., Chowdhury, S. A., Jung, S.-G., Almerekhi, H., & Jansen, B. J. (2020). Developing an online hate classifier for multiple social media platforms. Human-Centric Computing and Information Sciences, 10(1). doi: 10.1186/s13673-019-0205-6.

[7] Nemes, L., & Kiss, A. (2021). Social media sentiment analysis based on COVID-19. Journal of Information and Telecommunication, 5(1), 1–15

[8] F. A. Acheampong, Nuno Mensah and Wenyu (2020). Defines Text based emotion detection: challenges, advances and opportunities. Engineering term Reports, 2(7).

[9] S. Ahmad, M. Asghar, F. M. Alotibi,& Awan, (2019). Defines Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Human-Centric Computing and Information Sciences, 9(1).

[10] David Zimbra, Santa Clara University ,Ahmed Abbasi, University Of Virginia ,Daniel Zeng, University Of Arizona ,Hsinchun Chen, University of Arizona (2018, March). The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation, ACM Transactions on Management Information Systems.

[11] Gaind, B., & Padgalwar, V. S. A. (Eds.). (May 17-18, 2018). Emotion Detection and Analysis on Social Media. Retrieved from http://www.gjesr.com/ICRTCET-18.html).

[12] Dharmaraj Patil, Tareek Pattewar, Shailendra Pardeshi1, et al. (2022). Learning to Detect Phishing Web Pages Using Lexical and String Complexity Analysis. doi: 10.4108/eai.20-4-2022.173950

[13] Wang, Y., Kim, K., Lee, B., & Youn, H. (2018). Word clustering based on POS feature for efficient twitter sentiment analysis. Human-Centric Computing and Information Sciences, 8(1)doi:10.1186/s13673-018-0140-y

[14] Kilimci, Z. H., Güven, A., Uysal, M., & Akyokus, S. (2019). Mood detection from physical and neurophysical data using deep learning models. Complexity, 2019, 1–15.

[15] Liu, X., Wang, M., & Huet, B. (2016). Event analysis in social multimedia: a survey. Frontiers of Computer Science, 1-14.

[16] Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. Data Mining and Knowledge Discovery, 24(3), 478-514.

[17] Al-Emran, M., & Malik, S. I. (2016). The Impact of Google Apps at Work: Higher Educational Perspective. International Journal of Interactive Mobile Technologies (iJIM), 10(4), 85-88.

[18] Al-Emran, M. N. H. (2014). Investigating Students' and Faculty members' Attitudes towards the Use of Mobile Learning in Higher Educational Environments at the Gulf Region.

[19] Al-Emran, M., & Shaalan, K. (2015, August). Learners and educators attitudes towards mobile learning in higher education: State of the art. In Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on (pp. 907-913). IEEE.

[20] Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A. & Tziritas, N. (2015). A survey on text mining in social networks. The Knowledge Engineering Review, 30(02), 157-170.

[21] Berry Michael, W. (2004). Automatic Discovery of Similar Words. Survey of Text Mining: Clustering, Classification and Retrieval", Springer Verlag, New York, 200, 24-43.

[22] Salloum, S. A., Al-Emran, M., & Shaalan, K. (2016). A Survey of Lexical Functional Grammar in the Arabic Context. Int. J. Com. Net. Tech, 4(3).

[23] Al Emran, M., & Shaalan, K. (2014, September). A Survey of Intelligent Language Tutoring Systems. In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on (pp. 393-399). IEEE.

[24] Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. Annals of Operations Research, 236(1), 197-213.

[25] Vipul Devendra Punjabi, Dr. Rajesh Kumar Shukla (2021). A Study on Fake News Recognition And Detection Methods Using Machine Learning Techniques. Webology (ISSN: 1735-188X), Volume 18, Number 5, 2021, 742-754.

[26] Yang, L., Geng, X., & Liao, H. (2016). A web sentiment analysis method on fuzzy clustering for mobile social media users. EURASIP Journal on Wireless Communications and Networking, 2016(1), 1.

[27] Robinson, R., Goh, T. T., & Zhang, R. (2012). Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. Electronic Commerce Research, 12(3), 301-330.

[28] Wang, Y., Kim, K., Lee, B., & Youn, H. (2018). Word clustering based on POS feature for efficient twitter sentiment analysis. Human-Centric Computing and Information Sciences.

[29] Referred Plutchik wheel of emotions from Wikimedia Commons, the free media repository Source: https://commons.wikimedia.org/wiki/File:Plutchik-wheel.svg.

[30] Wong, "Almost all the traffic to fake news sites is from facebook, new data show," 2016. D. M. J. Lazer, M. A. Baum, Y. Benkler et al., ") e science of fake news," Science, vol. 359, no. 6380, pp.1094–1096, 2018.

[31] S. A. Garc´ıa, G. G. Garc´ıa, M. S. Prieto, A. J. M. Guerrero, and C. R. Jim´enez, ")e impact of term fake news on the scientific community scientific performance and mapping in web of science," Social Sciences, vol. 9, no. 5, 2020.

[32] V. P´erez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," 2017, https://arxiv.org/ abs/1708.07104.

[33] scikit-learn Machine Learning in Python, available online, https://scikit-learn.org/stable/

[34] Kumari, S. (2021). NoFake at CheckThat! 2021: fake news detection using BERT. arXiv preprint arXiv:2108.05419.

[35] Dong, X., & Qian, L. (2022). Integrating Human-in-the-loop into Swarm Learning for Decentralized Fake News Detection. arXiv preprint arXiv:2201.02048.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8513
***Article Received:** 28 July 2023 **Revised:** 19 September 2023 **Accepted:** 13 October 2023*

TABLE 2: METADATA REPRESENTS APPROACH USED, CORPUS, SOCIAL MEDIA PLATFORM AND STATES OF EMOTION-SENTIMENT BY RESEARCHERS

Sr. No.	Authors	Approach	Corpus	Social Media Platform	Emotions/Sentiments
1	Francisca Adoma Acheampong,ChenWen yu, Henry Nunoo-Mensah [8]	Text based	1. https://www.kaggle.com/shrivastava/isears-dataset 2.alt.qcri.org/semeval2017/task4/index.php?id=download-the-full-training-data-for-semeval-2017-task-4 3.https://github.com/JULIELab/EmoBank 4.http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html 5.http://people.rc.rit.edu/~coagla/affectdata/index.html 6. https://www.aclweb.org/anthology/I17-1099/ 7. https://www.crowdflower.com/wp-content/uploads/2016/07/text_emotion.csv 8.http://web.eecs.umich.edu/~mihalcea/downloads.html$delimiter%20%22005C317%20$#GroundedEmotions	Twitter, News Headlines, YouTube comments , facebook	Love, optimism, submission, awe, disapproval, remorse, contempt, aggressiveness. happiness, sadness, fear, anger, surprise, and so on
2	Ali Shariq Imran , (Member, Ieee), Sher Muhammad Daudpota , Zenun Kastrati , And Rakhi Batra [1]	Text based	Trending Hashtag # Data, Kaggle Dataset, SENTIMENT140 DATASET, EMOTIONAL TWEETS DATASET,	Twitter	Joy, surprise, sad, fear, anger and disgust
3	Zhenpeng Chen, Yanbin Cao, And Huihan Yao, Xuan Lu,Xin Peng,Hong Mei And Xuanzhe Liu [26]	Emoji based	Https://Github.Com/Sentimoji/Sentimoji JIRA Dataset, Stack Overflow Dataset, Code Review And Java Library	Twitter	Positive-love, joy, Negetive-sad, fear,anger, Discard-surprise
4	Joni Salminen* , Maximilian Hopf, Shammur A. Chowdhury, Soon-Gyo Jung1, Hind Almerekhi4 And Bernard J. Jansen [6]	Text based	1 Structured repository of regionalized, multilingual hate speech: https ://hateb ase.org/. 2 https ://githu b.com/t-david son/hate-speec h-and-offen sive-langu age. 3 https: //githu b.com/zeera kw/hates peech . 4 https: //githu b.com/t-david son/hate-speec h-and-offen sive-langu age. 5 https: //githu b.com/ben-aaron 188/ucl_aca_20182 019. 6 https: //githu b.com/punya joy/Hatem iners -EVALI TA. 7 https: //githu b.com/jing-qian/A-Bench mark-Datas et-for-Learn ing-to-Inter vene-in-Onlin e-Hate-Speec h. 8https ://githu b.com/UCSM-DUE/IWG_hates peech public. 9 https: //githu b.com/aitor -garci a-p/hate-speec hdatas et. 10 https: //githu b.com/pinke shbad jatiy a/twitter-hates peech. 11 https: //bitbu cket.org/ceshwar/bag-of-communities/src/maste r/.Currently known as Figure-Eight. 12 https: //www.perspectiv eapi.com.	Twitter, YouTube, Reddit and Wikipedia	Positive and negative sentiment
5	NEMES, Rodrigo Masaru Ohashi[7]	Text and Emoji based	http://sentiment140.com/ fetched from the official API, Using the library tweepy	Twitter	Sadness, fear, anger, joy also used Plutchik's Wheel of emotions
6	Asghar, Fahad M. Alotaibi and Irfanullah Awan [9]	Text based	Twitter streaming API to scrap tweets	Twitter	Word embedding