ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

SENTIMENT ANALYSIS AND AUTOMATIC EMOTION DETECTION ANALYSIS OF TWITTER USING MACHINE LEARNING CLASSIFIERS

S Santhosh Baboo [a]

[a] The Principal & Research Supervisor, P.G & Research Department of Computer Science

D. G. Vaishnav College, Chennai, India

M Amirthapriya*, [b]

[b] The Research Scholar, P.G & Research Department of Computer Science

D. G. Vaishnav College, Chennai, India

ABSTRACT

The development of data analysis algorithms performing sentiment and emotional analysis on social media text is increasing exponentially nowadays. Opinions and expressions used online and in social media are expressed in uniquely different ways, not conforming to the traditional standard. The format of these expressions may include a variety of emotions or feelings. There exist various techniques and methods, in the field of analyzing sentiments and emotions. The majority of the research work focuses mainly on either positivity or negativity of the sentiments only, but not considering the emotions in these expressions. This research focuses not only on the sentiment but further analyses for emotions in these expressions. This paper presents a hybrid rule-based algorithm for creating a fully annotated dataset for five emotions: anger, fear, happiness, sadness, and surprise. The dataset used for this study was a collection of our 45,000 tweets, all related to "Covid19 and India" and the English language. Machine learning classifiers Random Forest, Support Vector Machine, Logistic Regression, Naive Bayes, and Stochastic Gradient Boosting were used to classify sentiments and emotions. In a comparison of all the above classifications, we found that the Support Vector Machine outperforms all the other methods.

Keywords: *Twitter, Sentiment Analysis, Emotion Analysis, Machine Learning, Natural* Language Tool Kit (NLTK), Covid-19 and India.

1. INTRODUCTION

The growth of micro-blogging has resulted in a significant access to text with emotions in recent years. Micro-blogs have a strict length limit, resulting in entirely new types of emotions and inspiring people to express their daily thoughts in real-time. They differ from other textual sources such as blogs, social media, email, and product reviews. In this Internet era, social networking platforms are essential resources for communicating feelings to the entire world. People use audio and video files, images, and texts to express their feelings or opinions. Since text communication via social media is overwhelming, a massive volume of unshaped and unstructured data is shared on the Internet every second. To understand human psychology, the data must be analyzed as quickly as it is generated, which can be done via sentiment analysis, which detects polarity in the text [18]. It evaluates the data as a negative, positive, or neutral sentiment towards a person, a brand, a movie, an event, or about anything.

Sentiment analysis refers to the classification of the sentiments in the text source. In pandemic scenarios, social media sites Facebook, Twitter, YouTube, etc., play a crucial role. Twitter is one of the distinguished, successful, and prevalent social media application where millions of users sharing their thoughts on a variety of things.

Twitter has tended to become a key research area due to the public availability of such information. Tweets are shown publicly on each user's page and can be read and replied to by anybody. Tweets are useful in generating a massive amount of sentiment data on the opinion of the people about different topics. Hence, we developed an automated machine learning sentiment analysis model to compute customer opinions [8]. Sentiment analysis is inadequate at some point and hence we require emotion analysis which determines an individual's emotional state accurately. In text classification, existing research works are mostly on sentiment analysis which classifies the data as either positive or negative, and almost all publicly available datasets were in the English language. Some researches included neutral sentiments too, but analysis of emotions rather than sentiments allows for the extraction of more detailed understandings from the data. A sentiment classification can be two dimensions, where more fine-grained emotions are termed multi-dimensional in the research presented here. English is the language for which most tools and models are aimed at and enhanced. Even those whose native language is not English tends to publish in English. Furthermore, even non-native researchers tend to study the English language and therefore the resources they create are often for English. Our focus is on annotation projection, specifically from English and 'Covid-19 and India' related data, as a simple way of producing quality datasets for the English language. 45,000 tweets from the English language related to "Covid-19 and India" were utilized

to analyze these hybrid rules. This paper explains the machine learning techniques of sentiment and emotion detection and further classification analyses for the text.
In this paper, Section 2 includes the previous related research works of emotion and sentiment analysis. Section 3 discusses the proposed methodology on Tweet's emotion detection, and sentiment analysis. It includes datasets collection, preprocessing of text, data annotation, feature extraction techniques, machine learning classifiers, and performance metrics. Section 4 shows the results of sentiments and emotion analysis, and section 5 concludes the work.

2. REVIEW LITERATURE

Paul Ekman presented the basic six emotion categories anger, disgust, fear, happiness, sadness, and surprise. He described his primary view on the proposed basic emotions [19].

Robert Plutchik proposed an approach for general emotional responses based on psychological observations. He increased the Ekman's basic emotion categories from six to eight by adding the emotions 'anticipation' and 'trust' [20].

Pansy Nandwani and Rupali Verma [18] provided reviews on existing sentiment analyses, and emotion detection models for the text. They also discussed the challenges faced during the processes.

Maryam Hasan et al. [16] created and assessed a supervised machine learning system for automatically classifying emotions in text streams. Two major tasks offline and online categorization were incorporated by the approach. For categorizing emotion throughout the offline task, a system named 'Emotex' was formed to generate models. For the classification of real-time tweet's emotion tracking, a 2-stage framework named 'EmotexStream' was created in the task. Their experiments showed that almost 90% of emotions in text messages were accurately categorized by the created models.

Jitendra Kumar Rout et al. [21] employed supervised and unsupervised algorithms for the sentiment classification of Twitter data. For classification, a lexicon was produced and utilized. For determining the score, a Google search engine was utilized by the model. For the sentiment detection in disparate datasets, ML algorithms were implemented. The sentence-level categorization with unigram presence and POS features were found to be the most precise level when compared to the other features used.

Kashfia Sailunaz and Reda Alhajj [12] identified and examined emotions and sentiment in Twitter posts. The collection of tweets and their corresponding replies for certain topics was done. A dataset with Twitter text, sentiments and emotions, user information, etc., was created. For detecting sentiment along with emotion as of tweets, the dataset was applied. Centered on different user-centered and tweet-centered parameters, the user's replies and influence scores were estimated. Lastly, for generating generalized along with personalized suggestions for users, the latter information was employed grounded upon their Twitter activity.

Suboh M Alkhushayni et al. [23] gathered a tweets dataset that was specified at least one of '7' fundamental emotions. A collection of 42,000 tweets was encompassed by the dataset with a balanced existence of every emotion. A lexicon of about 40,000 words was created from this collection each of them was related to a weighted vector equivalent to one of the emotions. Then, in these cleaned tweets, disparate techniques for detecting emotion were executed and assessed. Lexically-centered classification along with supervised ML-centered classification was incorporated by these techniques. Lastly, the assessment of an ensemble technique was done which comprised numerous multi-class classifiers that were trained on the lexicon's unigram features. This evaluation illustrated that every other tested method was outperformed by the ensemble method when tested on existent datasets along with the dataset formed for this study.

Fereshteh Ghanbari-Adiv and Mohammad Mosleh [6] introduced an ensemble classifier that comprised '1500' of Multilayer Perceptron, k-Nearest Neighbor, along with Decision Tree classifiers. It was capable of systematically distinguishing disparate emotions among regular and irregular sentences with appropriate accuracy. Furthermore, for tuning the basic classifier's parameters, Tree-structured Parzen Estimator was utilized. For calculating the method, '3' disparate sets of ISEAR, OANC, along with CrowdFlower were employed that comprised regular along with irregular sentences. In the recognition of regular along with irregular sentences, the ensemble classifier's accuracies were 99.49 as well as 88.49%, correspondingly as exhibited by the results.

Jurek et al. 2015 [11] proposed a new method by the normalization function, for calculating sentiment value. Their method was more precise than the conventional summation and mean function.
Singh et al. 2021 [22] performed sentiment analysis on Twitter data using the BERT model. The data collection is based on two categories, such as the location of tweets and tweets related to India. The tweets were collected when the pandemic was high and people were in the panic situation around the world. So the dataset has lot of negativity about covid-19.
László Nemes and Attila Kiss [13] developed a model using RNN, for the prediction of various emotions on tweets. They classified the texts into more intense sentiment categories as strongly positive, weakly positive, strongly negative, and weakly negative. The comparisons for classification were made against TextBlob and RNN. They resulted that the RNN model was able to categorize even based on small details to make a decision.
Maria Krommyda et al. [14] produced a fully annotated dataset by using Plutchik's eight basic emotions. They presented an algorithm that considered the emoji in the text and utilized them as the main indicators of the expressed emotion.

Amrita Mathur et al. [3] analysed the Twitter data from all over the world for realizing the physiological health of people during the COVID-19. The analysis classified emotions as eight basic types based on the emojis on the expressions.

Nourah Alswaidan and Mohamed El Bachir Menai [17] surveyed the research for emotion in text for their implicit and explicit recognition. The study displayed that both the hybrid and learning-based approaches using traditional text representation outperform the other approaches. Also, the survey highlighted the effects of NLP tasks, part-of-speech tagging, and parsing on the performances of the proposed approaches.

Ahmad Fakhri Ab. Nasir [2] developed an emotion recognition and prediction system based on text. The machine learning techniques Decision Tree, k-Nearest Neighbor, Multinomial Naïve Bayes, and Support Vector Machine were investigated for the classification. Centred upon Ekman's six basic emotions, the model was developed. The ISEAR dataset was utilized to test all models and resulted that the Multinomial Naïve Bayes classifier resulting in an average accuracy of 64.08%.

3. METHODOLOGY

In this paper, we present a corpus collected from Twitter which was annotated for each tweet with basic emotions such as anger, fear, happiness, sadness, and surprise. A hybrid rule-based algorithm, using Natural Language Processing techniques was utilized to annotate the corpus. The proposed model was to train a classifier that automatically discovers the sentiments and emotions in tweets. We classified the tweets according to the emotions expressed, using the machine learning classifiers Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest, and Stochastic Gradient Boosting. The performance of the classifiers used in our corpus was evaluated with metrics precision, recall, f1-score, and accuracy. Figure 1 depicts the methodology for analysing Twitter sentiment and emotion using machine learning classifiers.The methodology involves the subsequent steps:

- Tweets Data Collection
- Preprocessing of Data
- Data Annotation for Sentiments and Emotions
- Feature Extraction for the Annotated Dataset
- Sentiment Analysis
- Emotion Analysis
- Performance Evaluation

3.1 Tweets Data Collection

To collect the needed data, a Twitter developer account with Academic Research Access was created, enabling to register an application, which was coded using Python. The Python script is used with the Tweepy library implementing the Twitter API V2 for data download. For data download, a Twitter search for "Covid-19 and India" was conducted, with filtering rules requiring that the tweet be written in English and that no retweets be included. Responses to tweets, quotes, and retweets were removed from the dataset to attain the accurate prediction. The tweets are downloaded with the columns "id", "created_at", "author_id", "lang", and "text" through the application, and store in "JSON" object format for further processing.

3.2 Preprocessing of Data

Most of the real-world data are open to uncertain, noisy, and missing data. Applying data mining methods on this uncertain and noisy data will not fetch quality results. Data preprocessing is the only way to get quality data.

Hence as part of preprocessing, the data in raw format is transformed using the below-mentioned steps to clean the "Noise" out of the Tweets. The preprocessing involves the following steps:

- All text in tweets was changed to its lowercase.
- Removed all the punctuations and numeric from the text.
- Removed all the URL, HTML, and hyperlinks from the text.
- Removed all the emojis from the text data.
- Removed all the hashtags from the text.
- Abbreviations, acronyms, and short words were expanded.
- Stopwords (list by NLTK) were removed from the text.
- Negations were carried out for the text data (For example \s'n't will be replaced as '\snot' and '\lln't' will be '\will not' and so on.
- Reduced the excess spaces to make the parsing of the sentence accurate.
- Tokenization was carried out, in which a sentence or a text is separated into single pieces of text which are known as tokens.
- Stemming of words was carried out for each word in the sentence.
- Lemmatization of each word in the tweet reduces the word to its base or root word.

The Pre-processed dataset is then saved in another dataset using the panda's library so that the integrity of the original downloaded dataset is maintained.

3.3 Data Annotation for Sentiments and Emotions

The NLTK library is extensively used to analyze the collected tweets as part of the annotation process. The tweets are annotated for sentiments and emotions using the Python script. The script determines the polarity and subjectivity of the sentence in the tweet and determines if the sentiment expressed is either "Positive" or "Negative". Whilst analyzing the Emotions in a tweet, the process considers multiple strategies and uses lemmatization, which matches with the five categories of emotions such as anger, fear, happiness, sadness, and surprise, for each word in the tweet after the pre-processing is completed. The dataset is separated as two parts as training set and test set with 80% and 20% of data respectively. Now both the datasets are ready for further processing.

3.4 Feature Extraction for the Annotated Dataset

The process of transforming data into features that may be employed in a machine learning model is known as feature extraction. Typically, machine learning algorithms are programmed with numeric values. Machine Learning models are unable to understand a text, so vectorization is needed in NLP. Vectorization is the process of converting each message or tweet (list of tokens now) into vectors in a matrix format. The text or word is now mapped into numerical vectors feature called Bag of Words representation. These features can be extracted by using Count Vectorizer and TF-IDF. Count vectorizer is a straightforward method that extracts features by counting the number of times a word or a token appears in a given document. Term Frequency-Inverse Document Frequency (TF-IDF) is a weighting technique of a term or a word in the document or corpus [2].

In this research, we are using TF-IDF for feature extraction, where Term Frequency (TF) denotes the count of a term or a word that appears in a tweet and Inverse Document Frequency (IDF) represents the weight of the words in terms of their frequency. Since every tweet is in various lengths, there is a possibility or a chance that a word would appear several times in the data frame. Hence TF-IDF is calculated as below:

$$\mathrm{TF} = \frac{\text{No.of times a word occurs in a document}}{\text{No.of words in that document}} \qquad (1)$$

$$\mathrm{IDF} = \frac{\text{No. of documents}}{\text{No. of documents containing the words}} \qquad (2)$$

$$\mathrm{TFIDF} = \mathrm{TF} * \mathrm{IDF} \qquad (3)$$

3.5 Sentiment Analysis

Sentiment analysis is the process of mining opinions, views, reviews, and emotions from various forms of data like text, audio, videos, tweets, and other social media data through Natural Language Processing. Sentiment analysis categorizes the tweets as positive or negative. In this paper, we analyze the sentiment of the tweets from our downloaded dataset using five machine learning classifiers along with Term Frequency- Inverse Document Frequency (TF-IDF). Then the performance of these classifiers is evaluated using the metrics such as precision, recall, f1-score, and accuracy. We have implemented sentiment analysis that takes the tweets and then determines the subjectivity and polarity level for each tweet.

The Python Textblob library which acts as a wrapper for implementing the NLTK library [4] was used to do sentiment assessment on the tweet's corpus. Textblob has some basic features of Natural Language Process fundamentals. For each tweet, we are assigning subjectivity and polarity using Textblob. The sentiment subjectivity score was ranged from 0 to 1, where 1 represents positive, with 0 representing negative sentiment. Each tweet's polarity was determined by assigning a score ranging from -1 to 1 based on the terms used in the tweet [7]. A negative score (ranges from -1 to 0) represents the negative sentiment and a positive score (ranges from 0 to 1) represents the positive sentiment.

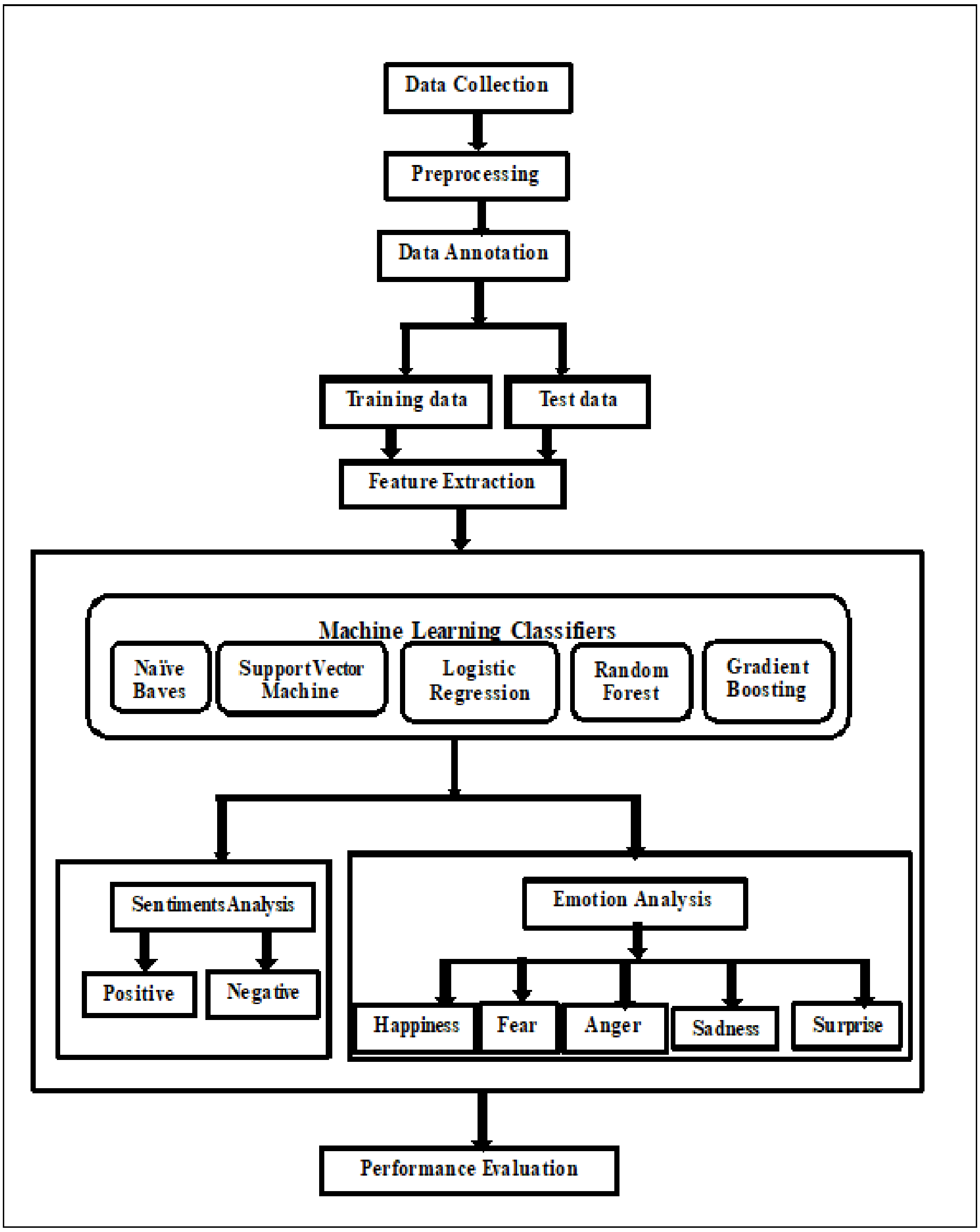

Fig. 1 The methodology of Twitter sentiment and Emotion analysis

3.6 EMOTION CLASSIFICATION

Classification is the process of categorizing or structuring documents or sentences into a predefined set of categories. There are two types of classifications, the first one is rule-based and the second one is machine learning-based approaches. In this paper, we are classifying the Twitter dataset with the emotions anger, fear, happiness, sadness, and surprise using five machine learning classifiers.

3.6.1 Support Vector Machine Classifier

Support vector machine is the most effective binary classifier, with the main goal of finding the optimal separating hyper-plane which has the maximum margin to both sides. The idea is to find a hypothesis that can promise the lowest true error. Following figure 2 shows the complete understanding of the process of the SVM classifier.

SVM algorithm supports to define the decision boundary and the best boundary is also known as a hyperplane. The points in the figure 2 (blue and green points) are called support vectors and the distance between the hyperplanes and the support vectors is called a margin. SVM aims to maximize the margin, and the hyperplane with the maximum margin is termed as the optimal hyperplane [1].

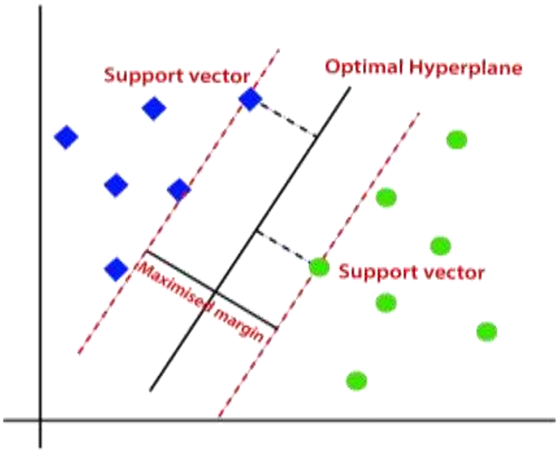

Fig. 2 Support Vector Machine algorithm in machine learning

3.6.2 Naïve Bayes

Naïve Bayes is one of the simplest machine learning classification techniques which is based upon the Bayes' Theorem. The Naïve Bayes classifier is a fast, accurate, and dependable method that achieves high accuracy on huge datasets. Naive Bayes classifiers have been mainly used for text analysis and text classification problems. The Naïve Bayes is described as two terms: Naïve and Bayes. The Bayes' theorem is given as [9]:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \qquad (4)$$

where P (A|B) denotes the posterior probability of hypothesis A, on observed event B. P (B|A) is the likelihood probability given that the probability of a hypothesis is true. P (A) is the probability of the hypothesis prior to observing the evidence, also known as prior probability, and P (B) is the probability of the evidence, also known as marginal probability.

3.6.3 Logistic Regression

Logistic regression is one of the easy and extensively used supervised machine learning algorithms. When the target variable is definite, logistic regression is utilized. Based on the probability concept, Logistic Regression is a predictive analysis that predicts the output value as Yes or No, 0 or 1, True or False, etc. Based on the categories, LR can be classified as Binomial, Multinomial, and Ordinal. A logistic regression model predicts the function of X as P(Y=1). LR is used for various classification problems like cancer cell detection, prediction of diabetics and cancer detection, etc.

Figure 3 shows the linear relationship among dependent and independent variables. The value of variable y (which is dependent) increases similarly when the value of variable x (that is independent) increases. The red line in figure 3 is denoted as the best fit straight line [10].

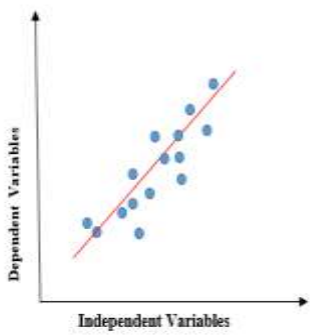

Fig. 3 Linear Regression algorithm in machine learning

3.6.4 Random Forest

Random Forest is the strongest methods among all machine learning algorithms that is mainly used for and regression and classification problems. The Random Forest is a set of decision trees built by the random forest classifier using randomly selected training data. The majority voting of different decision trees will be the final class of the test object. Since many decision trees are combined, it reduces the noise and gives more accurate results [15]. The following figure 4 shows steps involved in the random forest machine learning algorithm.

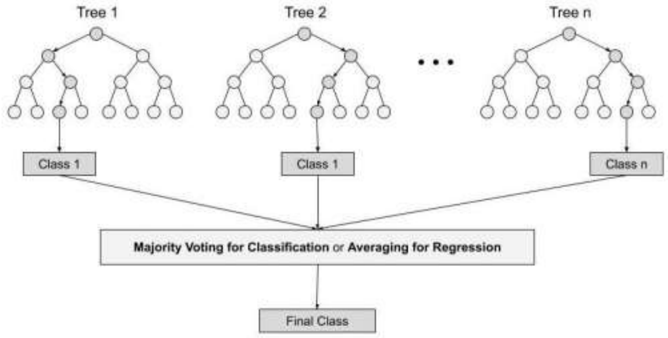

Fig. 4 Functionality of random forest algorithm in machine learning

3.6.5 Stochastic Gradient Boost

Stochastic Gradient Boosting algorithm has been used to solve wide-ranging, sparse machine learning tasks including text categorization and natural language processing. The efficiency and simplicity of Stochastic Gradient Descent are its main benefits. Below figure 5 shows the decision boundary of an SGD Classifier, equivalent to a linear SVM. Like other classifiers, SGD has two arrays (X, Y) where X has a shape of n_samples and n_features for the training samples. Y has a shape of n_samples, holding the target values for the training samples. In SGD, the gradient of the cost function of a single instance for each iteration is found instead of the sum of the gradient of the cost function of all the examples [5].

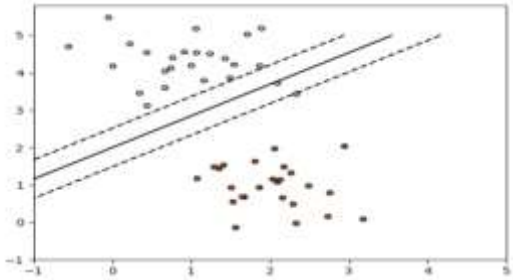

Fig. 5 Stochastic Gradient Boost algorithm in machine learning

3.6.6 Performance Analysis

The performance of the classifiers has been calculated using the information retrieval metrics accuracy, precision, recall, and f1-score.

The precision is estimated as (Eq.5):

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

Where TP is the total number of sentences correctly classified to a category, and FP is the total number of sentences incorrectly classified to a category.

The Recall is estimated as (Eq.6)

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})} \qquad (6)$$

Where FN is the number of sentences that were not classified at all and TN is the numbers of sentences marked as being in a particular category and were not.

The f1-score is evaluated as in (Eq.7):

$$\text{F1-score} = \frac{\text{Precision} \times \text{Recall} \times 2}{(\text{Precision+Recall})} \qquad (7)$$

The accuracy is evaluated as in (Eq.8):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

4. RESULTS

For sentiment and emotion classification for Twitter data, the attained outcomes of the proposed machine learning model are proffered. The proposed model is applied in Python, and to examine the presented methodology, the dataset is utilized on our Twitter dataset. The proposed model was evaluated by the classification metrics accuracy, f1-score, precision, and recall.

Figure 6 shows the positive and negative sentiments of the collected dataset. From the results, it was shown that most tweets have the positive sentiment and few tweets only have negative sentiment towards information related to 'Covid-19 and India'.

Table 1 describes the results of the performance metrics of the classifiers based on the sentiments. From the results shown in table 1, it was noticed that for the weighting scheme TF-IDF, 95, 87, 93, 92, and 91 accuracy values are attained by the proposed SVM, NB, RF, LR, and SGB algorithms respectively.

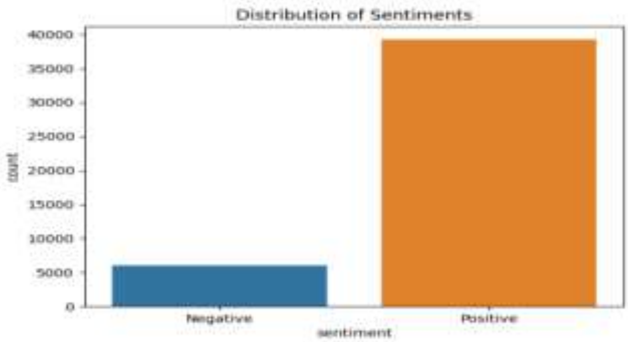

Fig. 6 Sentiment Analysis of tweets

Table 1 Machine learning classifier's performance for sentiments in tweets

CLASSIFIER	METRICS				
	Sentiment	**Accuracy**	**Precision**	**Recall**	**F1-Score**
SVM	Positive	0.95	0.96	0.98	0.97
	Negative		0.88	0.72	0.79
NB	Positive	0.87	0.87	1.00	0.93
	Negative		1.00	0.05	0.09
RF	Positive	0.93	0.94	0.99	0.96
	Negative		0.87	0.57	0.69
LR	Positive	0.92	0.92	0.99	0.96
	Negative		0.90	0.47	0.62
SGB	Positive	0.91	0.91	0.99	0.95
	Negative		0.81	0.99	0.53

Analogizing altogether, for the weighting scheme, the highest accuracy value is yielded by the SVM algorithm and the least accuracy value is attained by the NB algorithm. Tables 2-6 show the results of accuracy, precision, recall, and f1-scores of the classifiers for emotions anger, fear, happiness, sadness, and surprise. From the results it was shown that the emotion 'anger' reached the highest accuracy among other emotions expressed. Finally, it was validated as of the outcomes that the Twitter data's emotions are categorized much precisely whilst utilizing the TF-IDF weighting scheme together with the proposed Machine Learning Model.

Figure 7 shows the accuracy values of five machine learning classifiers on Twitter sentiment analysis. Figure 8 demonstrates the accuracy values of five machine learning classifiers on emotions happiness, fear, anger, sadness, and surprise for tweets.

Table 2 Results of naïve bayes classifier for emotions

Emotions	Accuracy	Precision	Recall	F1-Score
Happiness	0.81	0.99	0.10	0.19
Anger	0.88	1.0	0.01	0.03
Surprise	0.84	0.94	0.63	0.76
Sadness	0.83	0.88	0.65	0.75
Fear	0.71	0.70	0.99	0.82

Table 3 - Results of support vector machine classifier for emotions

Emotions	Accuracy	Precision	Recall	F1-Score
Happiness	0.98	0.99	0.93	0.96
Anger	0.98	0.99	0.91	0.95
Surprise	0.98	0.99	0.95	0.97
Sadness	0.97	0.99	0.94	0.97
Fear	0.96	0.97	0.96	0.97

Table 4 - Results of logistic regression classifier for emotions

Emotions	Accuracy	Precision	Recall	F1-Score
Happiness	0.94	0.99	0.70	0.82
Anger	0.95	0.99	0.61	0.76
Surprise	0.94	0.99	0.85	0.91
Sadness	0.93	0.98	0.82	0.89
Fear	0.91	0.91	0.96	0.93

Table 5 - Results of random forest classifier for emotions

Emotions	Accuracy	Precision	Recall	F1-Score
Happiness	0.79	0.0	0.0	0.0
Anger	0.87	0.0	0.0	0.0
Surprise	0.61	0.0	0.0	0.0
Sadness	0.61	0.0	0.0	0.0
Fear	0.68	0.68	1.0	0.81

Table 6 - Results of gradient boost classifier for emotions

Emotions	Accuracy	Precision	Recall	F1-Score
Happiness	0.79	0.0	0.0	0.0
Anger	0.87	0.0	0.0	0.0
Surprise	0.61	0.0	0.0	0.0
Sadness	0.61	0.0	0.0	0.0
Fear	0.68	0.68	1.0	0.81

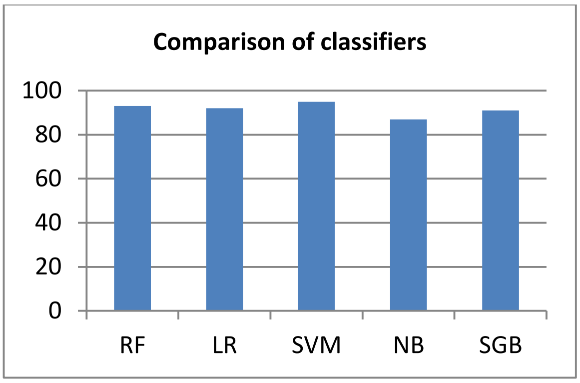

Fig. 7 The accuracy comparison of different classifiers for sentiment analysis

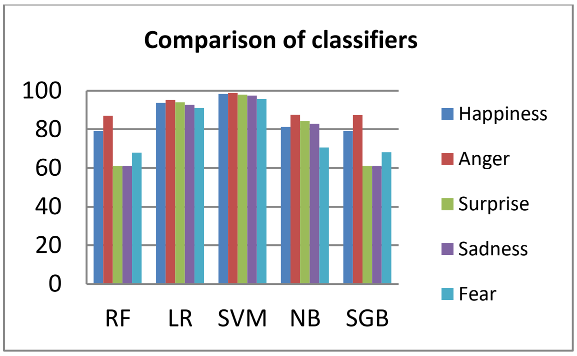

Fig. 8 The accuracy comparison of different classifiers for emotion analysis

5. Conclusion

The aim of this research was to examine people's sentiments and emotions during the Covid-19 epidemic in India. For the Twitter data's emotional prediction, a Twitter dataset, which can visualize the people's opinions for a specific domain is developed in this paper. We examined the sentiments and emotions using different machine learning approaches in the proposed model**.** The accuracy results are yielded above 85% by the classifiers for sentiment analysis whilst utilizing the proposed term weighting scheme (TF-IDF). For emotional analysis, the SVM classifier attained the highest accuracy values (95% approximately) for the proposed emotions.

People's mental and physical health management are more important in the pandemic situation caused due to Covid-19. Since most people were expressed their feelings on social media during this pandemic, our model uses Twitter posts by people all around the world and information regarding "Covid-19 and India". This analysis can be used to realize the mental stability of the people against coronavirus.

To execute the textual data's sentiment and emotional classification, this work can be enhanced by establishing an ML design that fully automates the detection and analysis in the forthcoming future.

REFERENCES

[1] A, Ashwitha & Gowda, Shruthi & R, Shruthi & Upadhyaya, Makarand & Ray, Abhra & C, Manjunath. (2020). Sarcasm detection in natural language processing. Materials Today: Proceedings. 37. 10.1016/j.matpr.2020.09.124.

[2] Ahmad Fakhri Ab. Nasir, Eng Seok Nee, Chun Sern Choong, Ahmad Shahrizan Abdul Ghani, Anwar P. P. Abdul Majeed, Asrul Adam and Mhd. Furqan, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 769 012022.

[3] Amrita Mathur Purnima Kubde Sonali Vaidya, "Emotional analysis using twitter data during pandemic situation COVID-19", Fifth International Conference on Communication and Electronics Systems, 10-12 June, Coimbatore, India, 2020.

[4] Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc.

[5] D. Gupta, V. Kushwaha, A. Gupta and P. K. Singh, "Deep Learning based Detection of Water Bodies using Satellite Images," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498442.

[6] Fereshteh Ghanbari-Adivi and Mohammad Mosleh, "Text emotion detection in social networks using a novel ensemble classifier based on parzen tree estimator (TPE)", Neural Computing and Applications, vol. 31, no. 4, pp. 1-13, 2019.

[7] Hasan A, Moin S, Karim A, Shamshirband S. Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical and Computational Applications*. 2018; 23(1):11. https://doi.org/10.3390/mca23010011

[8] https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/

[9] https://www.javatpoint.com/machine-learning-naive-bayes-classifier

[10] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_logistic_regression.htm

[11] Jurek A, Mulvenna MD, Bi Y (2015) Improved lexicon-based sentiment analysis for social media analytics. Secur Inform 4(1):1–13.

[12] Kashfia Sailunaz Reda Alhajj, “Emotion and sentiment analysis from twitter text”, Journal of Computational Science, vol. 36, pp. 1-42, 2019.

[13] László Nemes & Attila Kiss (2021) Social media sentiment analysis based on COVID-19, Journal of Information and Telecommunication, 5:1, 1-15, DOI: 10.1080/24751839.2020.1790793.

[14] M. Krommyda, A. Rigos, K. Bouklas and A. Amditis, "Emotion detection in Twitter posts: a rule-based algorithm for annotated data acquisition," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2020 pp. 257-262. doi: 10.1109/CSCI51800.2020.00050

[15] M. S. Munir Prince, A. Hasan and F. M. Shah, "An Efficient Ensemble Method for Cancer Detection," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934817.

[16] Maryam Hasan, Elke Rundensteiner and Emmanuel Agu, “Automatic emotion detection in text streams by analyzing Twitter data”, International Journal of Data Science and Analytics, 2018, Doi:10.1007/s41060-018-0096-z.

[17] N. Alswaidan and Mohamed El Bachir Menai, “A Survey of State-of-The-Art Approaches for Emotion Recognition in Text”, Knowledge and Information Systems, Vol. 62, pp. 2937-2987, 2020.

[18] Nandwani P, Verma R. A review on sentiment analysis and emotion detection from text. Soc Netw Anal Min. 2021;11(1):81. doi: 10.1007/s13278-021-00776-6. Epub 2021 Aug 28. PMID: 34484462; PMCID: PMC8402961

[19] P. Ekman, “A methodological discussion of nonverbal behavior,” The Journal of Psychology, vol. 43, no. 1, pp. 141–149, 1957.

[20] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In Theories of emotion, pages 3–33. Elsevier

[21] Rout, J.K., Choo, K., Dash, A.K., Bakshi, S., Jena, S.K., & Williams, K.L. (2018). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, *18*, 181-199.

[22] Singh M, Jakhar AK, Pandey S (2021) Sentiment analysis on the impact of coronavirus in social life using the Bert model. Soc Netw Anal Min 11(1):1–11.

[23] Suboh M Alkhushayni, Daniel C Zellmer, Ryan J DeBusk and Dua Alzaleq, “Text emotion mining on twitter”, IOP SciNotes, vol. 1, no. 3, pp. 1-9, 2020.

.