A NOVEL APPROACH FOR TWITTER SENTIMENT ANALYSIS USING HYBRID CLASSIFIER

Kavita Sharma[1], Dr. Harpreet Kaur[2]

[1]Student, Dept. of Computer Science Engineering, DAVIET, Punjab, India
[2]Associate professor, Dept. of Computer Science Engineering, DAVIET, Punjab, India

---

Abstract - *Twitter is a micro blogging site that allows users to communicate and debate their ideas and opinions in 140 characters or less, with no regard for space or time constraints. Every day, millions of tweets on a wide range of topics are sent out. Sentiments or views on various subjects have been identified as a significant feature that characterizes human behavior. Sentiment analysis models think about extremity (good, negative, or nonpartisan), yet in addition feelings and sentiments (irate, satisfied, tragic, and so forth), desperation (critical, not dire), and even aims (intrigued, not intrigued). Therefore, the primary goal of this Endeavour is to examine and analyze Twitter sentiment analysis during important events using a Bayesian network classifier. Also, to implement the principal component analysis (PCA) algorithm for extraction of best features and combining hybrid approach consisting of Linear Regression, Xgboost and Random Forest classifiers. Finally, the results of the trained and tested datasets are based on accuracy, precision, recall and F1-score.*

Key Words*: *Sentiment Analysis, Bayesian Network, Twitter, Principal Component Analysis, Feature Extraction, Linear Regression, Random Forest, XGBoost.

1.INTRODUCTION

Different people from varied areas of life may hold the same perspective on a variety of problems. When these individuals form a group, they are referred to be similar wavelength communities or groupings. That is, same wavelength communities are organizations created on the basis of similar thoughts and feelings expressed by different individuals on a variety of themes. People in critical and intentional teams are basically related by such similar recurrence organizations. Many social network research projects are largely concerned with either analyzing sentiments at the tweet level or studying the features of tweeters in a linked context. People from diverse areas of life may have the same perspective on different problems, but they do not have to be related. Furthermore, automated detection of such implicit groups is useful for a variety of applications [1].

With the assistance of technology, the web has become an extremely valuable place where concepts can be quickly interchanged, online learning, audits for an assistance or item, or movies can be found. It is complex to interpret as well as record the user's emotions since reviews on the internet are accessible for millions of services or products. Sentiments are user feelings about things like products, events, situations, and services that might be excellent, terrific, bad, or neutral. Sentiment analysis [2] is the study of people's emotions and reactions based on online comments. Sentiment analysis is insinuated by an arrangement of names, including evaluation mining, believing mining, appraisal extraction, subjectivity examination, feeling examination, impact assessment, review mining, and others.

SA is a very new area of study that uses Linguistics as well as text analytics, NLP, Computational as well as categorized the polarity of the emotion or opinion to gather subjective information from source material. SA is a language understanding task that employs a computational model to evaluate the user's opinion as well as categorize it as positive, negative, or neutral. The primary goal of SA is to determine a writer's or speaker's attitude toward a given topic. This article's or speaker's attitude could be their assessment, affective state (the author's emotional state when writing), or the intentional emotional communication (intends to produce an emotional impact on the reader). With the assistance of opinion mining, authors can differentiate between low-quality & high-quality content

1.1 Sentiment Analysis in Twitter

An online social networking service is a service or platform that allows individuals who similar interests, hobbies, backgrounds, and real-life relationships to create social networks or social interactions. Twitter is an online person-to-person communication Twitter is an internet based individual-to-individual correspondence and micro blogging administration that enables its customers to send and read text-based communications known as tweets. Tweets are clear, of course, but senders can limit the message conveyance to a certain group. Since its inception in 2012, Twitter has grown to become one among the most often used micro blogging systems, with over 500 million registered users. According to Info graphics Labs5 measurements, 175 million tweets were sent each day in 2012.

Twitter is used by a tremendous number of people to bestow musings, choosing it an enthralling and endeavoring choice for evaluation. When so much consideration is being paid to Twitter, why not screen and develop strategies to investigate these feelings. Twitter has been chosen considering the accompanying purposes.

Tweets are short texts of length of maximum 140 characters. Every day, millions of thousands of tweets are sent out on a variety of topics in order to share and debate people's ideas and opinions. Twitter Sentiment Analysis (TSA) has been used in a variety of applications. TSA was employed in four primary areas, according to a recent study [3]. They are product reviews, movie reviews, determining political inclination, and stock market forecasting. The literature review part provides a thorough examination. Furthermore, the development of tweets on a variety of topics prompted academics to consider domain-independent solutions to problems such as detecting latent communities based on sentiments, real-world behavior analysis, and so on [4].

Tweets are usually written in a cryptic and informal way. These types of linguistic flexibility in expressions pose additional challenges in analyzing Twitter sentiments.

1.2 Challenges in Sentiment Analysis

Sentiment analysis is concerned with preparing audits, commenting on various persons, and managing them in order to extract any useful information. Various factors impact the SA cycle and must be addressed effectively in order to obtain the final grouping or bunching report. Many of these issues are not thoroughly investigated under the surface [5]:

- *Co-reference Resolution*

This problem is referred to as determining "what does a pronoun or maxim show?" For example, in the line "After watching the film, we went out for supper; it was adequate." What does the term "it" refers to, regardless of whether it refers to a film or food? As a result, after the film analysis is finished, regardless of whether the sentence identifies with motion pictures or food? The expert is concerned about this. This type of problem is most commonly caused by SA that is arranged from a particular point of view.

- *Association with a period*

Because of SA, the time of assessment or audit selection is a key concern. At any one time, a comparable customer or group of clients may have a favourable reaction to an item, and there may be instances where they have a negative reaction. In this sense, it serves as a test for the conclusion analyser at other points in time. This type of problem is common in relative SA.

- *Sarcasm Handling*

Mockery words are words that have the opposite meaning to what they are illuminating. For example, consider the line "What a decent hitter he'll be, he scores zero in every other inning." The positive term "great" has a negative sense of significance in this case. These sentences are tough to find, and as a result, they have an impact on the assessment's evaluation.

- *Domain Dependency*

In SA, the words are primarily utilized as a component for examination. Be that as it may, the importance of the words isn't fixed all through. There are scarcely any words whose implications change from area to space. Aside from that, there exist words that have contrary significance in various circumstances known as contronym. Hence, it is a test to know the setting for which the word is being utilized, as it influences the examination of the content lastly the outcome.

- *Negations*

Negative words in a book can drastically alter the significance of the phrase in which they appear. As a result, these terms should be addressed while reviewing the audits. For example, the words "This is a good book." and "This is not a good book." have the opposite meaning, but when the test is done one word at a time, the outcome may be surprising. The n-gram examination is popular for dealing with this type of situation.

- *Spam Detection*

Surveys are being investigated as part of the sentiment probe. Nonetheless, until date, hardly any subjective analysis has been conducted to determine if the audits are bogus or whether any real people have completed the survey. Numerous individuals with no information on the item or the administration of the organization give a positive audit or negative survey about the administration. This is a lot of hard to check concerning which survey is a phony one and which isn't; that in the end assumes an indispensable part in SA.

The paper is divided into four parts of equal length. The first section gives a review about the function of sentiment analysis in Twitter data. Section 2 goes into detail about literature survey. The third section focuses on the methodology used. Section 4 explains the results obtained. Finally, section 5 summarizes the entire document.

2. LITERATURE SURVEY

This part presents an inside and out writing audit on methods utilized in online media information examination in various domains such as political election, healthcare, and sports. Several attempts had been carried out on Internet-related data for making predictions have been done in different areas. There are various methods which are used by makers for settling through web-based media data fit for heading and considered as fundamental data for assessment.The following is the literature survey related with the Twitter sentiment analysis during important occasions.

Alajajian et al. [6] authors proposed and developed a Lexico calorimeter: a web-based, interactive tool for calculating the caloric content of social media and other large-scale texts. Since their Lexico calorimeter is a straight superposition of principled expression scores, they exhibited that they can go past connections to explore what individuals talk about in aggregate detail, just as help in the arrangement and portrayal of how populace scale conditions differ, which discovery strategies can't do.

The primary computational phases in this process, according to Ankit and Saleena [7], are detecting the polarity or attitude of the tweet and then classifying it as positive or negative. The main problem with Twitter sentiment analysis is determining the best sentiment classifier to use in accurately classifying tweets. In this study, an ensemble classifier is constructed, which combines the base learning classifier to build a single classifier with the purpose of improving the sentiment classification approach's performance and accuracy.

As per Snefjella et al. [8], public person generalizations, or thoughts regarding the character qualities of individuals of a country, make a problem. They investigated the positively of national language usage in Study 1B. While concentrating on 2A and 2B, they provided gullible members with the 120 most broadly demonstrative terms and emoticons of every country and requested them to pass judgment on the character from a speculative person who utilizes either symptomatically Canadian or analytically American expressions and emoticons. Therefore, they speculated that public person generalizations might be to some extent established in countries' aggregate language movement.

Goularas and Kamis [9] in this paper the blends of CNN and a sort of RNN known as LSTM networks are assessed and compared. This research upholds the significance of SA by assessing the conduct, benefits, just as limitations of the current strategies utilizing an evaluation technique inside a solitary advancement stage utilizing similar information frame& figuring climate.

Ruz et al. [10] inspect the issue of opinion investigation during significant occasions like normal catastrophes and social developments.They employed Bayesian network classifiers to examine sentiment in two Spanish datasets: the 2010 Chilean earthquake and the 2017 Catalan independence referendum.They used a Bayes factor strategy to automatically regulate the number of edges supported by the training instances in the Bayesian network classifier, resulting in more realistic networks. Given countless preparing occurrences, when contrasted with help vector machines and irregular backwoods, the outcomes uncovered the utility of utilizing the Bayes factor just as its serious forecast yields.

Hasan et al. [11] evaluated public perceptions of a product using Twitter data. This is a project in which BoW as well as TFIDF are often used in tandem to exactly classify positive & negative tweets. Creators found that by using the TF-IDF vectorizer, the exactness of feeling investigation could be definitely improved, just as simulation results exhibit the adequacy of the proposed framework.

Iqbal et al. [12] presents an effective approach for enhancing accuracy as well as flexibility by bridging the gap among lexicon-based as well as machine learning strategies. When compared to much more widely used PCA and LSA based feature reduction methods, suggested approach has up to 15.4 percent higher accuracy over PCA as well as up to 40.2 percent higher accuracy over LSA. The test discoveries exhibited the capacity to precisely quantify public opinions and perspectives on an assortment of themes like psychological warfare, worldwide contentions, social issues, etc.

Karthika et al. [13] explained that the evaluation from the online shopping website flipkart.com is evaluated, as well as the rating is referred to as positive, neutral, or negative based on the attributes of the design. In suggested system, the accuracy, precision, F-measure, as well as recall for both the RF and the SVM approaches are determined, as well as accuracy comparison is made between the two methods. In this case, the Random Forest outperforms the SVM by 97 percent.

Ramasamy et al. [14] The Twitter data set is examined using a classifier based on support vector machines (SVM) with varied settings. The content of a tweet is categorised to determine if it contains factual or opinionated information. The sentiment is partitioned into three classifications: great, negative, and impartial. An essential choice to boost productivity may be made based on this classification and analysis. At the point when the exhibition of SVM outspread piece, SVM direct network, and SVM spiral matrix were looked at, SVM straight framework beat the other SVM models.

As indicated by Martnez-Rojas et al. [15], instant, precise, and powerful usage of accessible data is basic to the effective treatment of crisis circumstances.Twitter, one of the most prominent social networks, has been identified as a source of information that provides crucial real-time data for decision-making. The reason for this work was to lead a methodical writing audit that gives an outline of the current situation with research on the utilization of Twitter in crisis the board, just as difficulties and future examination openings.

According to Oztürk and Ayvaz [16], they researched popular attitudes and feelings concerning the Syrian refugee issue. Turkish opinions were deemed significant since Turkey has received the greatest number of Syrian refugees, and Turkish tweets offered information that reflected popular image of a refugee-hosting nation firsthand. In contrast with the proportion of positive opinions in Turkish tweets, which made up 35% of every Turkish tweet, the extent of uplifting outlooks towards Syrians and exiles in English tweets was perceptibly lower.

As per SoYeop Yoo et al. [17], in view of the importance and social extensive impact of electronic media objections, a couple of assessments are being coordinated to study the substance given by customers.They proposed Polaris, a framework for surveying and estimating clients' emotive directions for occasions inspected progressively from gigantic web-based media substance, in this article, and showed the aftereffects of starter approval work.They display both trajectory analysis and sentiment analysis so that people may gain information quickly. They also improved sentiment analysis and prediction accuracy by employing the most recent deep-learning technology.

Gautam and Yadav [18] presented a bunch of AI approaches with semantic examination for recognizing sentences and item assessments dependent on Twitter information.The main purpose is to use a pre-labeled Twitter dataset to analyse a large number of reviews.The naïve byes technique, which outperforms maximum entropy and SVM, is subjected to the unigram model, which outperforms it on its own.When the semantic research WordNet is tracked by the previously indicated technique, the precision rises to 89.9 percent, up from 88.2 %.The training data set, as well as WordNet for review summarization, may be enlarged to improve the feature vector associated sentence identification process.

Pagolu et al. [19] demonstrated that there is a solid relationship between a business stock value rise/fall and public considerations or sentiments concerning that organization communicated on Twitter through tweets.The key contribution is the creation of a sentiment analyser that can determine the type of sentiment contained in a tweet. Positive, negative, and neutral tweets are separated into three groups. At first, favourable feelings were assessed or sentiments expressed by the public on Twitter regarding a firm would be reflected in its stock cost.

Hao et al. [20] presented a visual investigation of Twitter timeseries that consolidates opinion and stream examination with geo-and time sensitive intelligent perceptions for dissecting true Twitter information streams.Visual sentiment techniques are utilized to post-buy web study information and entertainment mecca Twitter information, finding captivating examples of client remark.Today's visual analysis tools (for example, SAS JMP, Vivisimo, Polyanalyst, and others) mostly give feedback on evaluations via yes/no questions, numeric ratings, and direct remarks.

To do fine-grained sentiment analysis, Zainuddin et al. [21] developed a new hybrid technique for a Twitter aspect-based sentiment analysis. This study looked at association rule mining (ARM) in part-of-speech (POS) patterns that was supplemented with a heuristics combination for recognising explicit single and multi-word features. A rule-based method with feature selection is also included in the system for recognising sentiment phrases. The experiments with multiple classification algorithms also indicated that the unique hybrid sentiment classification produced appropriate findings using Twitter datasets from diverse topics. Results exhibited that the new mixture opinion characterization strategy, which coordinated disclosures from the viewpoint-based feeling classifier technique, was viable, beat the earlier benchmark feeling grouping strategies by 76.55, 71.62, and 74.24 percent, individually.

Larsen et al. [22] provided an overview of the We Feel system, which collects and categorises emotional tweets on a worldwide scale in real-time. 2.7310 9 tweets were examined over a 12-week timeframe. A variety of studies were carried out to highlight possible applications of the data, including finding normal diurnal and weekly patterns in emotional expression and utilising these to detect key events. Correlations were also found between emotional tweets and anxiety and suicide indices, showing the possibility for the creation of social media-based measures of population mental well-being to supplement current data sets.

Zainuddin et al. [23] proposed a feature selection approach based on PCA for determining the most relevant collection of characteristics for emotion categorization depending on aspects. A feature selection technique for twitter viewpoint put together opinion arrangement based with respect to PCA is utilized, which is joined with Sentiwordnet dictionary based methodology that is consolidated with SVM learning system. Tests utilizing the Disdain Wrongdoing Twitter Opinion (HCTS) and Stanford Twitter Feeling (STS) datasets uncover exactness paces of 94.53 and 97.93 percent, separately.

Anjaria and Guddeti [24] fostered a crossover procedure to removing assessment from Twitter information using immediate and roundabout attributes dependent on directed classifiers dependent on SVM, Innocent Bayes, most extreme Entropy, and Counterfeit Neural Organizations. SVM beats any remaining classifiers in trial information, with a greatest effective expectation precision of 88% for US Official Decisions in November 2012 and a most extreme forecast exactness of 58% for Karnataka State Gathering Races in May 2013.

It is a very challenging task. In spite of its broad use in logical exploration, a systemic discussion has emitted as of late about the representativeness and legitimacy of Twitter tests. Two principal questions emerge: regardless of whether Twitter clients address society and whether Twitter tests address Twitter correspondence overall. When capturing tweets for sentimental analysis, the dataset also includes redundant data such as stop words, special characters etc. which causes irregularities in classification of sentiments out of them. Few more limitations are described in above research papers are as follows:

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 02 | Feb 2022 www.irjet.net p-ISSN: 2395-0072

The drawback using Bayesian networks classifiers for classification is that there is no commonly accepted way for generating networks from data. It is up to the user to create the network and ensure that it works properly. It struggles with high-dimensional data. The hazard emerges during the data collection process, where re-tweets are created, resulting in an increase in size that is unrelated to the results. The vulnerability can be mitigated by employing filter commands that filter the data [25].

- In certain studies, the size of the dataset had a direct influence on performance of the SVM algorithm. Increased training data would have improved accuracy [14].
- There is a requirement to segregate material concerning a single entity and then assess emotion toward it. Because some tweets contain neither positive nor negative emotion, the primary focus should be on the research of neutral tweets. A basic bag-of-words technique can categorize it as neutral; nevertheless, it conveys a distinct attitude for both entities in the sentence, which can lower the accuracy rate [12].
- The key constraint of the latest study work is that the majority of the work focuses on English material, although Twitter has many different users from all over the world. The majority of cutting-edge Twitter sentiment analysis techniques only work with English tweets. It is critical for developing a time-independent vocabulary that is realistic and practicable for real-world applications [25].
- Emoticons and irregular words, on the other hand, are widespread in Internet content written in a variety of languages. Emojis are viewed as comments since they promptly express one's mentality. Shockingly, they arrive in an assortment of shapes. and sizes; most research manually choose a smiley (e.g. :-) ") for labelling. They neglect to examine many additional numbers such as \<3 " (heart, signifies love) (heart, means love). We need to find additional potential emoticons in different forms on Twitter because they are language-independent and useful for multilingual analysis. Then again, sporadic words are ordinarily seen when clients need to preserve keystrokes or when the length of a message is restricted[26].
- There are likewise a few specialized issues in the past paper, for example, working on the sack or-words model for feeling investigation for higher accuracy. On account of Twitter, the most common tweet length is 140 characters, which runs from 13 to 15 words by and large.These tweets contain misspellings, informal language, various slangs, opinions about an entity, and symbolic words, which present numerous challenges in processing and analysing the data. As a result, it is critical to extract the best features while retaining any meaningful words that can add value to the text [17].

So, in this research work we will filter the redundant data and extract the features of data using principal component analysis (PCA). After that we will train and test our data to analyse the tweets on various aspects using hybrid approach.

3. PROPOSED METHODOLOGY

Step 1: Reading the dataset for tweets and its sentiments

Social media creates a large quantity of sentiment data in many formats, such as tweet id, status updates, reviews, author, content, tweet type, and tweet status update.For this application, I evaluated the suggested algorithm's performance using a Twitter dataset of the 2010 Chilean Earthquake.

Step 2: Pre-processing of data for filtration of redundant data

The Twitter dataset utilized in this review work is now partitioned into two classes, negative and positive extremity, empowering feeling examination of the information clear and permitting the impact of numerous boundaries to be assessed. Irregularity and excess are normal in crude information with extremity. The accompanying focuses are remembered for tweet preprocessing:

- Undo all URLs (e. g. www.xyz.com), hash tags (e. g. #topic), targets (@username)
- Make the spellings; repeating characters must be handled.
- All of the emoticons should be replaced with their relevant feelings.
- Remove all punctuation, symbols, or digits.
- Remove Stop Words
- Acronyms should be expanded
- Remove Non-English Tweets

Step 3: Extracting the features from data using principal component analysis

We extract elements of processed datasets using the feature extraction approach and for that I will be using principal component analysis (PCA).

Step 4: Creating training and testing data

Generating randomly sample of 70% training data and 30% of testing data.

Step 5: Training of classifier using data

Supervised learning is a useful approach for dealing with classification challenges. Training the classifier facilitates subsequent predictions for unknown data.

Step 6: Testing of data using the designed network model

For dataset testing, a hybrid technique is used, with Straight relapse and Arbitrary Backwoods classifiers used.

Step 7: Calculation of results

In this step I will be graphically evaluating the results by showing which model is best suitable for this dataset.

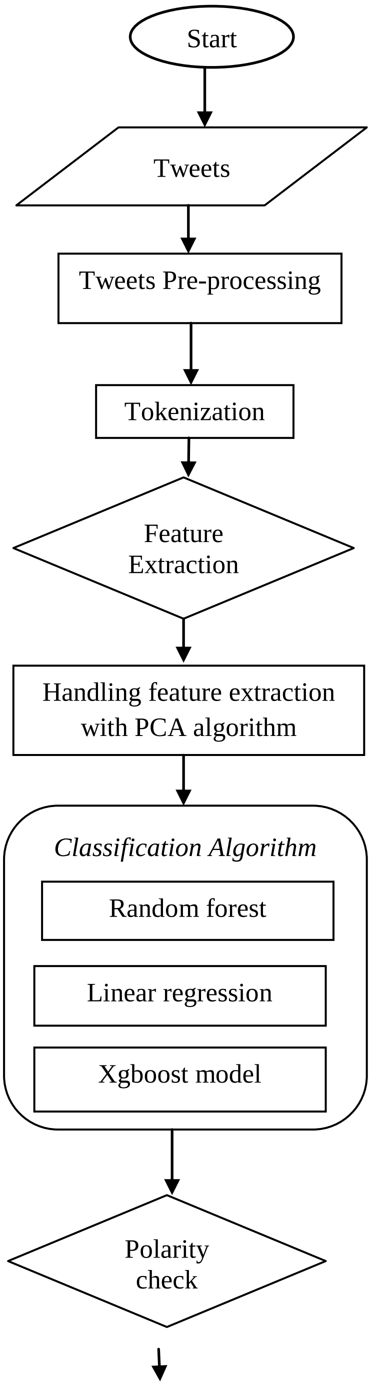

Figure 1. Flowchart of Proposed Work

4. RESULTS AND DISCUSSIONS

- *Precision*

The fraction of correctly predicted positive observations to all anticipated positive observations is defined as precision.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

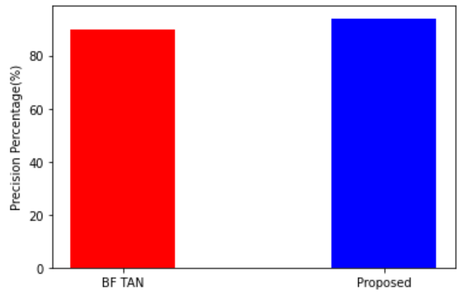

Figure 2 . Precision graph

- *Recall*

The recall is explained as the proportion of accurately predicted positive observations to the total number of observations in the class.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

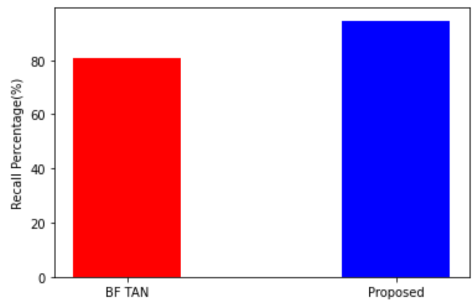

Figure 3. Recall for Proposed Technique

Figure 2 depicts the precision and recall graphs for two classifiers. In compared to the BF TAN classifier, the recommended hybrid methodology produces the best outcomes. BF TAN has the lowest value.

- *Accuracy*

Accuracy is the measurement used to evaluate which model plays out awesome at perceiving relationships and examples between factors in a dataset dependent on the info, or preparing, information.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}}$$

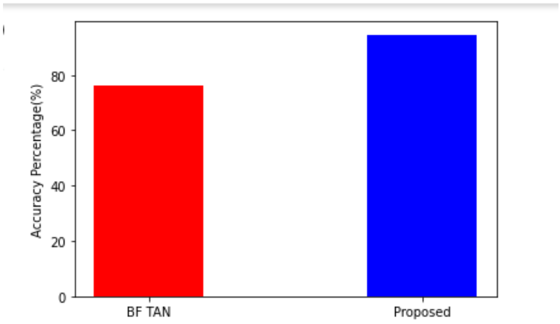

Figure 4. Accuracy Graph

Figure 4 depicts two graphs: one depicting the accuracy of the different ML algorithms mentioned, and the other is % comparison graph for accuracy, precision, recall, and F1-score. The accuracy graph clearly illustrates that the suggested approach has the highest accuracy as compared to BF TAN.

- *F1-Score*

F1 is determined by accuracy and recall. Subsequently, both false positives and false negatives are considered in this score.

$$\text{F1-Score} = 2* \frac{(Recall*Precision)}{(Recall+Precision)}$$

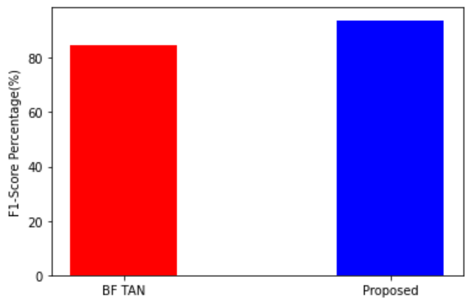

Figure 5. F1-Score

Figure 4 presents the F1 score results of Gaussian Naïve Bayes classifier and the proposed classifier.The accompanying chart clearly illustrates that the suggested strategy outperforms the other classifier discussed.

5. CONCLUSION AND FUTURE SCOPE

This study effort concludes the whole suggested study on Twitter spam streaming analysis and grouping of emotion. A feature extraction approach based on a blend of linear regression, random forest, and principal component analysis (PCA) is presented for extracting specific feature sets for boosting classification accuracy and using machine learning classifiers to expose spam tweets. When compared to other existing works, the simulation results suggest that the proposed work has a higher detection ratio. When employing a vast quantity of data, the results achieved in this suggested study show a very big difference in terms of precision, recall, accuracy and F1-score when compared to other classifiers. Furthermore, this hybrid technique is applied for sentiment classification of tweets with modest alterations in the suggested algorithm in terms of positive and negative tweets, resulting in good classification accuracy.

REFERENCES

[1] Al-Qurishi, M., Hossain, M. S., Alrubaian, M., Rahman, S. M. M. and Alamri, A., "Leveraging analysis of user behavior to identify malicious activities in largescale social networks,"IEEE Transactions on Industrial Informatics, vol. 14, no. 2, pp. 799–813, 2018.

[2] Bouazizi, M. and Ohtsuki, T., "A pattern-based approach for multi-class sentiment analysis in twitter,"IEEE Access,vol. 5, pp. 20617–20639, 2017.

[3]Qamar, A. M., & Ahmed, S. S., "Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies," International Journal of Advanced Computer Science and Applications, vol. 8, no. 1, pp. 395-401, 2017.

[4] Kharde, V. A., & Sonawane, S. S., "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, vol. 139, no. 11, pp. 5-15, Apr. 2016.

[5]Vohra, S., & Teraiya, J., "Applications and Challenges for Sentiment Analysis: A Survey," International Journal of Engineering Research & Technology (IJERT), vol. 2, no. 2, pp. 1-5, Feb. 2013.

[6] Alajajian, S.E., Williams, J. R., Reagan, A.J., Alajajian, S. C., Frank, M. R. and Mitchell L., "The Lexicocalorimeter: Gauging public health through caloric input and output on social media", vol. 12, no. 2, 2017.

[7] Ankit & Saleena, N., "An ensemble classification system for twitter sentiment analysis", Elsevier, 2018.

[8] B., Snefjella,ISchmidtke, D., & Kuperman, V., "National character stereotypes mirror language use: A study of Canadian and American tweets", vol. 13, no. 11, 2018.

[9]Goularas, D., & Kamis, S., "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications, 2019.

[10] Ruz, G. A., Henriquez, P. A., Mascareno, A., "Sentiment analysis of twitter data during critical events through Bayesian network classifiers," Elsevier, 2019.

[11]Hasan, M. R., Maliha, M., & Arifuzzaman, M., "Sentiment Analysis with NLP on Twitter Data," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, 2019.

[12] Iqbal, F., Maqbool, J., Fung, B. C. M., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. K., "A Hybrid Framework for Sentiment Analysis using Genetic Algorithm based Feature Reduction," IEEE Access, 2019.

[13] Karthika, P., Murugeswari, R., & Manoranjithem, R. "Sentiment Analysis of Social Media Network Using Random Forest Algorithm", 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2019.

[14] Ramasamy, L. K., Kadry, S., Yunyoung Nam, Y., & Meqdad., M. N., "Performance analysis of sentiments in twitter dataset using SVM model", International Journal of Electrical and Computer Engineering, 2021.

[15] Martínez-Rojas, M., Pardo-Ferreira, M. C., & Rubio-Romero, J. C., "Twitter as a tool for the management and analysis of emergency situations: A systematic literature review," Elsevier, 2018.

[16] Öztürk, N., and Ayvaz, S., "Sentiment Analysis on Twitter: A Text Mining Approach to the Syrian Refugee Crisis", Telematics and Informatics, 2017.

[17] SoYeop Yoo, JeIn Song, & OkRan Jeong, "Social media contents-based sentiment analysis and prediction system," Elsevier, 2018.

[18] Gautam, G., & Yadav, D., "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," 2014 Seventh International Conference on Contemporary Computing (IC3), 2014.

[19]Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B., "Sentiment analysis of Twitter data for predicting stock market movements," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 1345-1350, 2016.

[20] Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U., Keim, D. A., Haug, L.-E., & Hsu, M.-C., "Visual sentiment analysis on twitter data streams," 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 277-278, 2011.

[21] Zainuddin, N., Selamat, A., & Ibrahim, R., "Hybrid sentiment classification on twitter aspect-based sentiment analysis," Applied Intelligence, vol. 48, pp. 1218-1232, 2017.

[22] Larsen, M. E., Boonstra, T. W., Batterham, P. J., O'Dea, B., Paris, C., & Christensen, H., "We Feel: Mapping Emotion on Twitter," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1246–1252, 2015.

[23] Zainuddin, N., Selamat, A., & Ibrahim, R., "Twitter Feature Selection and Classification Using Support Vector Machine for Aspect-Based Sentiment Analysis," Lecture Notes in Computer Science, pp. 269–279, 2016.

[24]Anjaria, M., & Guddeti, R. M. R., "Influence factor-based opinion mining of Twitter data using supervised learning," 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), 2014.

[25]Wongkar, M., & Angdresey, A., "Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019.

[26]Umar, S., Maryam, Azzhar, F., Malik, S., & Samdani, G., "Sentiment Analysis Approaches and Applications: A Survey," International Journal of Computer Applications, vol. 181, no. 1, pp. 1-9, Jul. 2018.