Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 2, 2026

A Hybrid Machine Learning Framework for Multi-Class Twitter Sentiment Analysis Using Logistic Regression and Random Forest

Sami Ullah[1], Amer Taj[2], Alauddin[3], Rida Zarkash[4], Mudassir Ahmad Khan[5] Muhammad Imran Khan Khalil[6]

[1]Department of Computer Science, City University of Science and IT Peshawar 25000, Pakistan
[2]Department of Computer Science, University of Engineering & Technology Peshawar 25000, Pakistan
[3]Department of Computer Science, University of Engineering & Technology Peshawar 25000, Pakistan
[4]Department of Computer Science, University of Engineering & Technology Peshawar 25000, Pakistan
[5]Department of Computer Science, University of Engineering & Technology Peshawar 25000, Pakistan
[6]Department of Computer Science, University of Engineering & Technology Peshawar 25000, Pakistan

DOI: https://doi.org/

Keywords
Machine Learning, Sentiment Analysis, Twitter, Random Forest, Logistic Regression, Support Vector Machines, Naive Bayes

Article History
Received on 20 January 2026
Accepted on 20 February 2026
Published on 27 February 2026

Corresponding Author: *

Sami Ullah
Department of Computer Science, City University of Science and IT Peshawar 25000, Pakistan

Abstract
Sentiment analysis is vital for understanding public opinion, especially on platforms like Twitter, where users express sentiments in real time. It provides insights into trends, behaviors, and opinions, supporting applications in market analysis, social studies, and policymaking. This research focuses on Twitter sentiment analysis to classify tweets into four sentiment categories: Irrelevant, Negative, Neutral, and Positive. It evaluates traditional machine learning Frameworks such as Logistic Regression, Naive Bayes, Random Forest, and Support Vector Machines. The research work highlights the advantages of hybrid approaches in sentiment analysis. It contributes to natural language processing by offering practical applications for extracting actionable insights from social media. Hybrid models are proposed to enhance performance: Logistic Regression-Random Forest (LR-RF). These models leverage feature extraction, contextual embeddings, and ensemble techniques to address challenges like class imbalance and short-text data. The dataset consists of preprocessed tweets, with model evaluation based on accuracy, precision, recall, and F1 score. LR-RF achieved the highest accuracy 92.01\%, after that Random Forest achieve 90.60 %.

INTRODUCTION

The growing popularity of online platforms and social media led to focusing on the role of sentiment analysis in the computational linguistics field [1]. This is mostly attributed to marketing, consumer feedback analysis, as well as product recommendation systems [2]. Sentiment analysis and opinion mining Opinion mining and sentiment analysis is a crucial component of natural language processing (NLP) whose goal is to isolate subjective text, such as opinions, feelings and emotions (i.e. subjective content) [3]. The oxford university press dictionary defines sentiment analysis as the computerized identification and classification of opinions expressed in text, usually to identify whether the authors attitude, in regard to a particular topic or a product, is positive, negative or neutral. Sentiment analysis has become critical in converting opinions and feelings into quantifiable information and making improved decisions. Systems of sentiment analysis deeply evaluate the amount of emotions that are being revealed by the statements that are textual in nature. Sentiment analysis can be performed at three levels: aspect, word and document [4].

Document-level sentiment analysis categorizes a document or a text as a whole to a chosen sentiment category like positive, negative, or neutral. It provides the overview of the document, which is the major attitude to the certain subject that can be traced in the text [5]. It is particularly useful when one needs to extract the essence of the articles, reviews or other essays, cases etc. But it lacks certain drawbacks like it does not understand the analysis of particular types of emotions that are aversive or diametric in various snippets of the text. As an illustration, in a single document one may encounter positive sentiment in one section and negative sentiment in another section, therefore one may filter the information provided at a high level only Such limitation may be crucial in cases of decision-making where sentiments evaluation is more advanced [6].

Sentence-level sentiment analysis: This approach relies on the sentiment analysis of each specific sentence of the text and the assignment of the sentiment to that sentence [9]. It is more specific than Document-level opinion mining, and comes in handy when sentiments are to be detected, which might be contained in a small portion of the document. Indicatively, in a movie review, only a single sentence can be used to express positive feelings by the writer about the way actors performed and another sentence can be used to express negative feelings about the screen play [7][8]. The difference in the belongings to the past and the current tags as well as the sentence-level analysis chooses these patterns well, such that one can effortlessly follow the changes in the sentiment of the text [10]. However, we still have possibilities to lose concurrency in terms of the ways the multiple sentences are interrelated to form a general contribution to the positive or negative mood of the text. Aspect-level sentiment analysis: It is the most detailed form of sentiment analysis as it takes into account the opinion about the attributes of some objects. As an example, a product review transmits data regarding the opinion created by a consumer about the price, quality, or even shape of a specific product [11]. The approach is observable in the analysis of customer feedbacks particularly because it can guide companies to their points of strength and their points of weakness in terms of products or services. The aspect level also considers the overall data focusing on the selected attributes and provides the information

that cannot be disclosed at the document level or sentence level analysis.

In order to address the drawbacks experienced in the earlier researches, this research employs a broad spectrum of models belonging to various families, i.e., machine learning architectures to conduct sentiment analysis on the selected Twitter dataset. Unlike other popular approaches, the models used in this work are LR, RF, NB and SVM, which are supposedly more inclined to assess the contextual and particularly, the language peculiarities inflicting the multiple levels of the language complexity [12]. The study undertakes contextualization of its embeddings to pick up subtle changes in sentiment. Moreover, it is mentioned that the problems of imbalanced datasets, sarcasm, irony can be addressed with such models and ensemble techniques. This longer method stabilizes the task and makes it more precise and connects the classical directions and the burden of the new large-scale datasets of social media. Further, the comparative evaluations of these models also are present in the proposed study and will rely on the performance metrics: accuracy, precision, recall, and F1-score to further analyze the efficiency of these models. Thus, the presented work will attempt to introduce new insights into the realm of sentiment analysis and address these long-standing problems, which will make it a meaningful contribution to the domain [13].

Literature Review

Zhao et al. [14] thoroughly analyzed different text pre-processing strategies on Twitter SA based on five datasets namely Twitter Sentiment Dictionary, Twitter SA dataset, Sentiment140, the Twitter US Airline Sentiment and the Stanford Large Movie Review Dataset. They tested tokenization, stemming, lemmatization, stop-word removal, spell-checking, and negation processing and acronym expansion on four classifiers: SVM, NB, RF and LR. The running of all the experiments illustrated that, the removal of URLs, stop words and numbers did not bring significant changes, whilst, the negation replacement and acronym expansion had a substantial effect on the improvement of the accuracy. The RF classifier shows the best result with accuracy of 85.6\%. The analysis also draws attention to the requirement of applying appropriate pre-processing methods depending on different classifiers to study sentiments on the twitting platform.

The relevant contributions of the work of Rodrigues et al. [15] to the area of social media analysis are twitter spam detection and SA. In this case, the research employs ML and DL methods to categorize the tweets into spam and non-spam as well as the SA of those tweets. The techniques used by the authors are SVM and RF in supervised learning and LSTM networks in DL. To analyze the effectiveness of the research, the tweeters Spam Dataset was applied by the researchers indicating that such approaches can be effective in handling big data of the Twitter platform. The results clearly show that the proposed approach could identify 92.5 Percent of the spam and perform sentiment classification. By highlighting the importance of addressing the challenges of spam detection and SA in the SNS context, the research shows how the performance of the two could be enhanced by integrating both ML and DL into the social media analysis tool.

Safari et al. [16] provides a valuable breakdown of the state-of-art, future opportunities, and existing and potential drawbacks of using NLP and AI in analyzing emotion and personality. The paper provides a survey of text-based approaches which can be applied in various fields, such as psychology, marketing, health sciences, and also describes some systems of emotion and personality recognition, such as Emotion Recognition in Conversations (ERC) dataset and Personality Dataset of the MyPersonality project. The strengths and weaknesses of the approaches considered here are also described, which contributes to the

state of knowledge concerning the peculiarities and possible problems associated with emotion and personality identification based on NLP tools. Although there is no actual quantification of the performance, the research emphasizes the gains that one can acquire due to the adoption of the NLP methodologies, which can be applied toward the ascertainable accuracies, in the practical issues, therefore, advancing the further evolution of the field.

The study [17] gives a comprehensive review of the ML models applied to SA and emotion recognition, with details on the common issues in the field and possible solutions. They point out the sophistication of SA tasks by reviewing the problems of model interpretability, domain adaptation, and data scarcity. It analyses certain approaches such as feature engineering, DL architectures and ensemble methods to address these issues. Further dataset-specific experiments are SVM, Decision Trees, NB and LR on IMDB and SEMEVAL.

Abimbola et al. [18] is devoted to the reliability and the reproducibility of the research results not only presupposing the clear set of standards to evaluate the dataset but also offering the clear guidelines to build models. The legal SA is described as applying document-level approach with CNN-LSTM. This model is an ensemble of CNN and LSTM networks, which are adapted to work with legal documents, which include temporal information regarding a case. The significance of the suggested method can, thus, be explained through the perspective of its capability to retrieve the sentiments embedded in the legal texts supported by the outcomes of the experiments on the selected datasets. In this paper, the literature is surveyed in order to outline the efficacy of the document-level model, CNN-LSTM on LSA-based SA and the further ways to enhance it.

This Singh et al. [19] study investigates the results of sentiment classification following various preprocessing steps such as noise removal, stemming, and tokenization. The authors perform theoretical analysis to explain the impact of preprocessing on the outcome of SA. Further, the review evaluates some of the methods through which performance enhancement is achievable by selecting and implementing appropriate preprocessing methods. On the Twitter data, the second large dataset is referred to as Sentiment140, known as one of the finest datasets to use in solving SA tasks, the findings indicate that the best preprocessing techniques can improve the classification accuracy by up to 10\%. The review is significant in understanding the authors and our bodies of knowledge about text preprocessing in SA and provides insights that are badly needed by researchers and practitioners in the field.

To improve the decision-making process and content quality in the dynamic digital media ecosystem the study considered SA in digital media material through various NLP techniques and ML algorithms [20]. The major results showed that BERT has exceptional precision of 94.56, which proves the effectiveness of transfer learning. Moreover, they achieved significant findings, where the RF approach received 70.82 accuracy and the LSTM model showed the best results in Recall, Precision, and F1 measures. DL methods, namely the RNN-LSTM model demonstrated an enormous potential with 96 accuracy, making LSTM and BERT contenders to lead tweet SA. The researcher, most likely, presented the results of the experiment that showed the degree to which their approach to identifying sentiment worked reliably across a diverse range of digital media sources.

In their research, Versha et al. [21] conduct a thorough SA of the Twitter-obtained data, utilizing its NLP strategy and ML models to classify the sentiments expressed in the tweets. The contributed work also pays attention to the peculiarities of the given peculiarities to the specifics of Twitter data analysed, along with the suggested approach to features extraction, sentiment classification, and visualisation. The authors apply various learning models,

including SVM and NB, on the Sentiment140 data set that is SA-optimized in the Twitter context. The results indicate that the optimal model trained on mean accuracy of the proposed methodologies of approximately 85 percent indicates effectiveness of the proposed models in determining sentiment in the Twitter data. It is practical to consider the efficacy and viability of corresponding NLP and ML methods to the sentiment classification in the SM environment and thus complement the research results of separate studies of sentiment classification in the SM setting.

Research Methodology

This research work compares several machine learning (ML) algorithms for sentiment analysis on Twitter with special reference to sentiment patterns in the area of online education. A systematic approach is used to filter relevant information from Twitter using the approach described in Figure 1. To gather tweets, the relevant hashtags concerning online education are used to filter tweets to make them relevant to the study. The data is cleaned where example, similar contents which appear many times, punctuations such as @ #, / * are removed to improve the quality of data. Pre-processing is performed on text data with the help of methods like TF-IDF and Vectorization that prepare the data in a format suitable to be fed to ML models. Sentiment classification is performed using four models: They include, Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression. The efficiency of such models is measured using such parameters as Accuracy, Precision, Recall, etc. Thus, this approach allows for a complete examination of sentiment in the discussion of online education in Twitter, which can be of real interest for educators, researchers, and both national and regional policymakers.

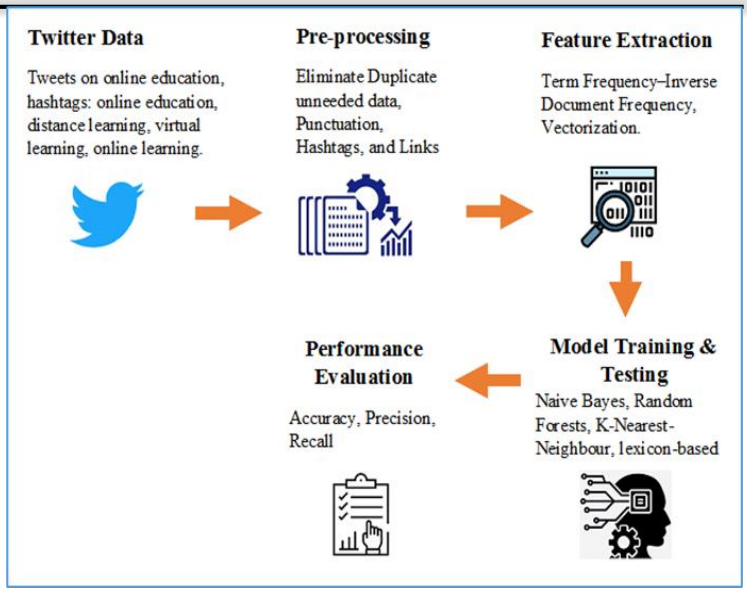

Figure 1: A Systematic Approach used in to filter relevant information from Twitter

The dataset used in this study is taken from the Kaggle which have four classes and 75000 records. This English language dataset consists of four attributes: tweeted, entity, target, and content. Descriptions of the dataset is presented in Table 3.1

Data Preprocessing

Pre-processing includes text normalization to remove special characters, punctuation, and digits, case conversion to lowercase, tokenization to segment text into individual words or tokens, and lemmatization and stemming to reduce words to their base or root form [22].

Feature Extraction

Feature extraction methods that were employed in this study include the One-Hot Encoding method and the TF-IDF method that was utilized to transform textual data into data structures that could be processed by machine learning model algorithms. The process used by One-Hot Encoding transformed text into binary vectors that presence or absence of words to assist other algorithms like Logistic Regression and SVM for data processing. TF-IDF, on the other hand, measured word importance by

focusing on terms that would occur frequently in some of the tweets but would rarely occur in the entire dataset. For instance, it was seen that when analyzing tweets to do with "machine learning", the importance given to words like "learning" was higher than that given to words like "the". All these techniques helped in making the mode more effective in achieving high levels of performance in the classification of sentiment patterns.

Preprocessing

Techniques like TF-IDF (Term Frequency-Inverse Document Frequency) are applied. These methods convert textual data into numerical features by calculating the importance of words within the dataset, which is essential for traditional ML models like Logistic Regression, Random Forest, or SVM.

Stop Words Removal

The first operation that is performed in the context of the pre-processing for text documents, the removal stops words that do not contribute to the SA. The words "is," "the" and "a" etc are removed from the dataset. This step simplified the data by eliminating any words that are not sentiment-bearing, leaving only the most informative words such as amazing or terrible. The exclusion of stop words helped focus on more precise and sentiment-carrying words, which positively impacted the reliability of the SA results.

Special Character Removal

The special character removal is, eliminating special characters, which entailed the removal of unnecessary elements such as punctuation, special symbols, hyperlinks, numbers, and other non-alphanumeric characters. These elements were deemed to be extraneous to the SA, as they do not contribute to the understanding of the underlying sentiment and instead introduce noise into the data. The analysis was able to focus on the core linguistic features that convey sentiment, thereby enhancing the overall quality of the results.

Lemmatization

Lemmatization, reduces words to their root or base form, standardizing word usage across tweets. By using lemmatization, words like "running" are converted to "run," and "better" is changed to "good." This process ensures that different forms of a word are treated as the same concept, enhancing the model's ability to recognize sentiment patterns consistently.

Tokenization

Tokenization breaking down each tweet into individual words or tokens. For example, a tweet such as "I love this product!" is divided into the tokens ["I," "love", "this" ,"product"]. Tokenization enables the model to analyze words separately and recognize meaningful patterns, allowing the sentiment of each word to be independently assessed.

Performance evaluation

Any research project's principal purpose is to analyze models. Some typical assessment measures must be taken into account. In this study, assessment measures used are accuracy, recall, F1-score and precision. These measures can be calculated as:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

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

$$F1 - \text{Score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Where, the terms TP, FN, TN, and FP refer to the number of true-positive classifications [Kubara and Kopczewska (2024)], false-

negative classifications, true-negative classifications, and false-positive classifications, respectively [23].

Models Evaluation

In this study, various ML methods are employed to determine the most effective sentiment analysis techniques for online education. These models include NB, LR, RF, and SVM.

Naïve Bayes

Naive Bayes is a probabilistic classification algorithm grounded in Bayes' theorem and assumes independence among features [25]. This assumption makes NB computationally efficient and highly interpretable, particularly suitable for text classification tasks where each word is treated as an independent contributor to the overall sentiment. In this research, NB has been employed to calculate the probabilities of each sentiment class (e.g., positive, negative, neutral, or irrelevant) based on word occurrences within tweets. Its efficiency in handling high-dimensional, sparse data makes it a valuable traditional approach for SA. However, the independence assumption and its inability to account for complex relationships, such as word dependencies and negations, limit its effectiveness for nuanced text analysis.

The classification process in NB relies on Bayes' theorem:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)},$$

- P (C|X) is the posterior probability of class C given features X.
- P (X|C) is the likelihood of the features given the class.
- P I is the prior probability of the class.
- P (X) is the evidence or marginal likelihood of the features.

$$\hat{y} = \arg\max_{C} P(C) \prod_{i=1}^{n} P(x_i|C),$$

- y is the predicted class.
- P I is the prior probability of the class.
- P_n (x_i|C) is the likelihood of feature xi given the class.
- $\prod_{i=1}$ calculates the product of probabilities for all features.

Logistic Regression

Logistic Regression is a linear model widely employed for binary and multiclass classification tasks. It predicts the probability of an input belonging to a particular class (e.g., positive, negative, or neutral sentiment) using a sigmoid activation function. This simplicity makes LR highly interpretable and computationally efficient, making it a reliable baseline model for sentiment classification. In this research, LR has been utilized to classify tweet sentiments, leveraging its straightforward architecture that maps input features to sentiment labels. While effective in scenarios where relationships between features and output are linear, LR's performance is limited in handling complex, non-linear patterns, reducing its utility in more intricate classification tasks. The mathematical foundation of LR is based on the sigmoid function, defined as:

$$y = \frac{1}{1 + e^{-z}} \; where \; w^T x + b$$

- x is the feature vector.
- w represents the weights.
- b is the bias.

It calculates the probability y by applying the sigmoid function to the linear combination z of features. For example, x might represent the frequency of positive or negative words in a tweet. LR is effective as a baseline model for SA because it is interpretable and handles binary classification directly, extending naturally to multiclass classification tasks.

Random Forest

The RF is an ensemble of decision trees, where each tree is trained on a subset of data with randomly selected features [24]. Through majority voting, RF enhances prediction robustness, which is beneficial for capturing sentiment variations in Twitter data. In RF, each decision tree analyzes different tweet features (such as word frequency or sentiment score), classifying each tweet individually. By aggregating these classifications, the RF model produces a final, majority-based sentiment prediction, enhancing overall accuracy and reducing overfitting.

RF is an ensemble method that aggregates predictions from multiple decision trees:

$$y_k = f_k(x), \quad k = 1, 2, \ldots, K,$$

The final prediction is determined by majority voting:

$$\hat{y} = \text{majority vote}\{y_1, y_2, \ldots, y_K\}$$

Each tree learns to classify tweets based on features like word frequencies, hashtags, or emoticons. RF's strength lies in its ability to reduce overfitting by averaging results, making it particularly robust for noisy data like tweets.

Support Vector Machine

Support Vector Machines are a powerful supervised learning technique [23] widely utilized in Twitter SA due to their proficiency in handling high-dimensional, sparse data. SVMs work by finding an optimal hyperplane that separates sentiment classes (e.g., positive, negative, neutral, or irrelevant) with the maximum margin. This approach ensures robust classification performance, even in the presence of noisy, informal text often found in social media.

SVMs leverage kernel functions, such as linear, polynomial, and radial basis function (RBF), to map input data into higher-dimensional spaces, enabling the model to capture. Non-linear relationships and intricate patterns within tweets. However, SVMs are heavily reliant on effective preprocessing and feature engineering to address linguistic challenges like sarcasm, irony, and negations. Careful parameter tuning, including the selection of kernel type and regularization parameters, is crucial for achieving optimal accuracy, making SVMs a strong and adaptable candidate for SA in diverse and complex Twitter datasets.

Mathematically, SVMs find the optimal hyperplane by solving the following optimization problem:

$$\min_{w} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i + b)).$$

- ||w||2: A term to maximize the margin between classes, ensuring better separation.
- C: A parameter that balances between maximizing the margin and minimizing classification errors.
- yi: Actual class label (+1 or −1) for the i-th sample.

SVM handles sparse tweet embeddings effectively, distinguishing sentiment categories based on the distance of tweet features from the hyperplane.

Results and Discussion

To evaluate our results, we split our data into a 70% training set and a 30% test set. We employed ML techniques, including NB, RF, LR and SVM, to classify the text based on sentiment polarity, and then assessed the outcomes., as illustrated in Figure 1.

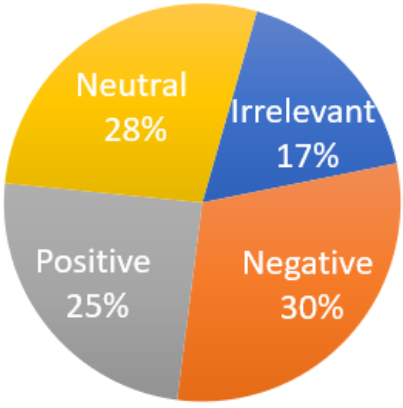

Figure 2. Sentiment Count

Table 1: Analysis through Precision, F-measure, and Recall

Model	Precision (%)	Recall (%)	F-Measure (%)
Logistic Regression	85.2	84.7	84.9
Naive Bayes	82.5	80.3	81.4
Random Forest	91	90.2	90.6
SVM	88.7	86.5	87.6
LR-RF	92.4	91.7	92

The findings demonstrate that LR-RF excels over alternative methods, achieving an accuracy rate of 0.92, an F1 score of 0.92, and a recall score of 0.92 RF excels over alternative methods, in contrast, RF achieve an accuracy rate of 0.92, an F1 score of 0.906, and a recall score of 0.902, the of RF results are notably underwhelming. Figure 3 depicts the total outcomes.

Table 2: Findings of LR-RF over alternative Methods

Model Comparison	Precision Difference (%)	Recall Difference (%)	F-Measure Difference (%)
LR vs. NB	2.7	4.4	3.5
LR vs. RF	-5.8	-5.5	-5.6
LR vs. SVM	-3.5	-2.1	-2.8
LR vs. LR-RF	-7.2	-7	-7.1
NB vs. RF	-8.5	-9.9	-9.1
NB vs. SVM	-6.2	-6.5	-6.3
NB vs. LR-RF	-9.9	-11.4	-10.6
RF vs. SVM	2.3	3.4	2.8
RF vs. LR-RF	-1.4	-1.5	-1.4
SVM vs. LR-RF	-3.7	-5.2	-4.4

Based on the results, Logistic Regression (LR) performs relatively worse than Naive Bayes (NB) but better than Random Forest (RF) and the combined Logistic Regression – Random Forest (LR-RF) model since it is a simple model with low interaction and complexity rate of the data. This demonstrates that NB is indeed inferior to the other models including the hybrid LR-RF and clearly elucidates its inability to handle complex feature extraction. Comparable but slightly less accurate results were obtained using Random Forest (RF) alone that performs much better than LR and NB and shows a good performance when working separately from the hybrid LR-RF. SVM performs better than NB and LR but slightly inferior to RF and close to comparable with LR-RF, thus indicating it as a competent but less efficient algorithm. Here, the hybrid LR-RF model proves to be optimal, surpassing traditional ML models while performing nearly as well as RF and exemplifying the utility of ensemble learning for maximizing performance.

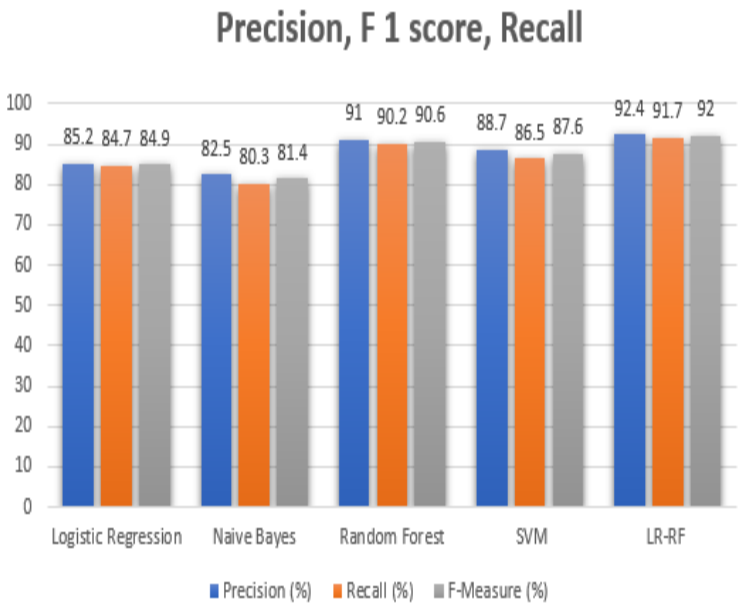

Figure 3: Analysis through Precision, F-measure, and Recall

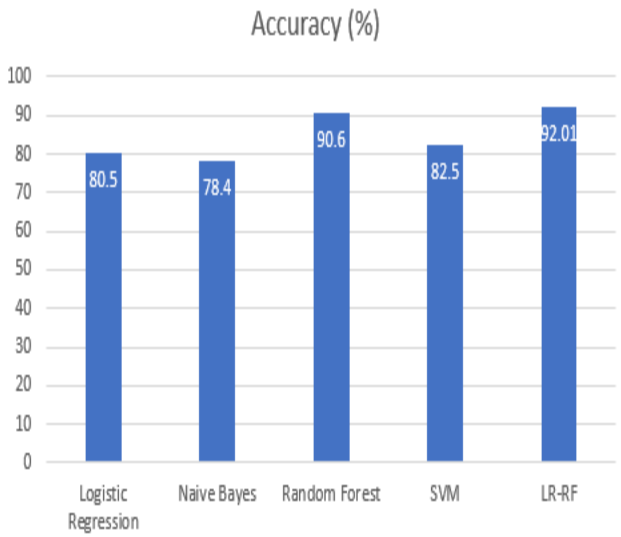

Figure 4: Accuracy Analysis of each Model

This accuracy analysis focuses on the results obtained for the traditional supervised learning models like Naïve Bayes, Logistic Regression, SVM and Random Forest for sentiment analysis in addition to the proposed LR-RF model. The performance Logistic Regression (LR) is assessed as 80.50% which is good in terms of computational efficacy and simplicity but poor in generalization capability of relational complexity in textual data. A conclusion derived from the research shows Naive Bayes (NB) has a lower accuracy of 78.40 percent because the method depends on word probabilities and rather makes independent assumptions about sentiments. RF used in our study achieves a classification rate of 90.60%, proving that RF is a stable single model due to a set of decision trees that help improve the predictiveness of the model. SVM obtains an accuracy of 82.50%, surpassing LR and NB and falling only behind RF, which shows a competitive nature of the model but somewhat smaller ability for generalization. Through the experiments, the proposed LR-RF model is determined as the best procedure as its accuracy is estimated higher than other models with 92.01%. Without a doubt, LR-RF outperforms both LR and RF due to synergy of LR's capability to define linear decision boundaries and RF's ensemble learning, proving that constructing a hybrid model using elements from multiple algorithms yields better results in comparison with using a single algorithm. The present study strengthens the case for a hybrid approach to improving the accuracy of sentiment analysis while revealing the drawbacks of solitary approaches.

Table 3: Comparisons of Models

Factor	**Random Forest (RF)**	**Naive Bayes (NB)**	**Logistic Regression (LR)**	**Support Vector Machines (SVM)**	**Hybrid Model (LR-RF)**
Model Structure	Ensemble of decision trees, aggregates results for better generalization	Probabilistic classifier assuming feature independence	Linear model that predicts sentiment probabilities through a decision boundary	Margin-based classifier, optimizes hyperplane for sentiment class separation	Combines LR's linear separability with RF's ensemble learning for improved predictions
Handling Feature Dependency	Captures complex patterns and dependencies	Assumes feature independence, limiting abilit	Handles linear relationships effectively but struggles	Captures nonlinear relationships well but limit	Leverages complementary strengths of LR and RF

	among features, improving sentiment classification	y to capture word dependencies	with nonlinear dependencies	ed by kernel functions	for better handling of dependencies
Data Imbalance Robustness	Manages class imbalance effectively, achieving high recall and F1 scores	Sensitive to class imbalance; may misclassify minority classes	Moderately robust to class imbalance; focus on precision can limit recall	Performs well with balanced data but less effective in highly imbalanced datasets	Excels in handling class imbalance, combining RF's robustness and LR's focus on classification
Handling High Dimensionality	Resilient to high-dimensional text data due to random feature selection at	Handles high-dimensional data but lacks flexibility to capture complex	Efficient in high-dimensional data but limited by linear assumptions	Handles high-dimensional data well, especially with appropriate kern	Robust in high-dimensional text data, combining LR's simplicity and RF's resilience
	each split	patterns		els	
Overall Effectiveness	Superior model for sentiment analysis due to a high balance between precision and recall	Moderate performance; effective but limited by the feature independence assumption	Computationally efficient but struggles with complex text patterns	Competitive performance, but slightly inferior to RF and hybrid models	Most effective, outperforming standalone models by combining linear and nonlinear feature-handling ability

Conclusion

For the purpose of this study, machine learning techniques such as Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), Support Vector Machines (SVM), and a combination of LR-RF were used for the analysis of sentiment on the Twitter platform. The performance of Random Forest was highly robust with an accuracy of 90.60% because of its ensemble learning ability to capture intricate patterns, while LR and NB, though efficient in computation, performed poorly in handling nonlinearity, thus giving moderate results. While doing that, SVM gave good accuracy rate but since RF has no kernel restriction it

outperforms SVM slightly. The best performance was noted for the LR-RF model, yielding an accuracy of 92.01 % by combining the efficacy of LR and RF algorithms. Consequently, with respect to the results presented in this paper, model selection should be dependent on data features and required application, yet it is worth mentioning that although traditional models such as the one presented here are accurate for basic sentiment classification NLI, the LR-RF classifier provides substantially better results, thus it should be investigated more in the future research.

Reference

P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 2601–2627, 2021.

P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," Social network analysis and mining, vol. 11, no. 1, p. 81, 2021.

M. Taboada, "Sentiment analysis: An overview from linguistics," Annual Review of Linguistics, vol. 2, pp. 325–347, 2016.

S. Kolkur, G. Dantal, and R. Mahe, "Study of different levels for sentiment analysis," International Journal of Current Engineering and Technology, vol. 5, no. 2, pp. 768770, 2015.

T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in Proceedings of the 2nd international conference on Knowledge capture, pp. 70–77, 2003.

K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE transactions on knowledge and data engineering, vol. 28, no. 3, pp. 813–830, 2015.

B. J. Paulich and V. Kumar, "Relating entertainment features in screenplays to movie performance: An empirical investigation," Journal of the Academy of Marketing Science, vol. 49, no. 6, pp. 1222–1242, 2021.

S. Behdenna, F. Barigou, and G. Belalem, "Document level sentiment analysis: a survey," EAI endorsed transactions on context-aware systems and applications, vol. 4, no. 13, pp. e2–e2, 2018.

V. Jagtap and K. Pawar, "Analysis of different approaches to sentence-level sentiment classification," International Journal of Scientific Engineering and Technology, vol. 2, no. 3, pp. 164–170, 2013.

J. Brooke, "A semantic approach to automated text sentiment analysis," 2009.

L.-T. Bei and Y.-C. Chiao, "An integrated model for the effects of perceived product, perceived service quality, and perceived price fairness on consumer satisfaction and 42 loyalty," Journal of consumer satisfaction, dissatisfaction and complaining behavior, vol. 14, pp. 125–140, 2001.

E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," Information retrieval, vol. 12, pp. 526–558, 2009.

P. Charoenkwan, C. Nantasenamat, M. M. Hasan, B. Manavalan, and W. Shoom buatong, "Bert4bitter: a bidirectional encoder representation from transformers (bert)-based model for improving the prediction of bitter peptides," Bioinformatics, vol. 37, no. 17, pp. 2556–2562, 2021.

Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," IEEE access, vol. 5, pp. 2870–2879, 2017.

A. P. Rodrigues, R. Fernandes, A. Shetty, K. Lakshmanna, R. M. Shafi, et al., "Real-time twitter spam detection and sentiment analysis using machine

learning and deep learning techniques," Computational Intelligence and Neuroscience, vol. 2022, 2022.

F. Safari and A. Chalechale, "Emotion and personality analysis and detection using natural language processing, advances, challenges and future scope," Artificial Intelligence Review, vol. 56, no. Suppl 3, pp. 3273–3297, 2023.

A. Alslaity and R. Orji, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," Behaviour & Information Technology, vol. 43, no. 1, pp. 139–164, 2024.

B. Abimbola, E. D. L. C. Marin, and Q. Tan, "Enhancing legal sentiment analysis: A cnn-lstm document-level model," 2024.

T. Singh and M. Kumari, "Role of text pre-processing in twitter sentiment analysis," Procedia Computer Science, vol. 89, pp. 549–554, 2016.

A. Radaideh and F. Dweiri, "Sentiment analysis predictions in digital media content using nlp techniques.," International Journal of Advanced Computer Science & Applications, vol. 14, no. 11, 2023.

V. Sahayak, V. Shete, and A. Pathan, "Sentiment analysis on twitter data," International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 2, no. 1, pp. 178–183, 2015.

B. Kabra and C. Nagar, "Attention-emotion-embedding bilstm-gru network based sentiment analysis," Journal of Integrated Science and Technology, vol. 11, no. 4, pp. 563–563, 2023.

J. Su, C. Jiang, X. Jin, Y. Qiao, T. Xiao, H. Ma, R. Wei, Z. Jing, J. Xu, and J. Lin, "Large language models for forecasting and anomaly detection: A systematic literature review," arXiv preprint arXiv:2402.10350, 2024.

I. Fadhli, L. Hlaoua, and M. N. Omri, "Sentiment analysis csam model to discover pertinent conversations in twitter microblogs," International Journal of Computer Network and Information Security, vol. 10, no. 5, p. 28, 2022.

T. Rusko, "Lexical Features of Scientific Discourse," International Conference on Brain Informatics, vol. 20, no. 1, pp. 82–88, 2014.