Ritu Suryavanshi, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

A Robust Ensemble Machine Learning Framework for Accurate Sentiment Classification of Twitter Data

Ritu Suryavanshi[1], Sharad Morolia[2]
[1]Research Scholar, Department of Information Security Mahakal Institute of Technology, Ujjain, India.
[2]Assistant Professor, Computer Science and Engineering Mahakal Institute of Technology, Ujjain, India

Abstract- Social media platforms generate a massive amount of opinion-based data that reflects public attitudes toward various topics such as politics, products, and social events. Among these platforms, Twitter is widely used for expressing opinions in the form of short textual messages known as tweets. Analyzing these tweets can provide valuable insights into public sentiment. However, sentiment classification of Twitter data is challenging due to informal language, abbreviations, emojis, and sarcasm. This study proposes an ensemble learning framework to improve the accuracy of Twitter sentiment classification. The framework involves several stages, including data collection, preprocessing, feature extraction using techniques such as Bag-of-Words and TF-IDF, and training multiple machine learning classifiers. Ensemble methods combine the predictions of these classifiers to generate more reliable results. The performance of the proposed model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The proposed approach aims to enhance sentiment analysis performance and provide more accurate insights from social media data.

Keywords: Sentiment Analysis, Twitter Data, Machine Learning, Ensemble Learning, Natural Language Processing, Text Classification, TF-IDF.

I. INTRODUCTION

In recent years, social media platforms have become one of the most important sources for sharing opinions, experiences, and feedback about various real-world issues. Among these platforms, Twitter is widely used for expressing thoughts on topics such as politics, products, entertainment, public policies, and social events. Every day, millions of tweets are posted by users across the world, generating a massive amount of opinion-rich textual data. This data provides valuable insights into public attitudes, preferences, and emotions [1]. Analyzing such large volumes of textual information manually is difficult; therefore, automated techniques are required to extract meaningful knowledge from it. Sentiment analysis, also known as opinion mining, is a Natural Language Processing (NLP) technique used to identify and classify opinions expressed in textual data into categories such as positive, negative, or neutral. It has become an essential tool for organizations, governments, and researchers to understand public perception and support data-driven decision making [2].

Sentiment analysis on Twitter data presents several challenges due to the nature of tweets. Tweets are usually short and often contain informal language, abbreviations, hashtags, emojis, and spelling errors [3]. Moreover, users sometimes express opinions indirectly through sarcasm or mixed emotions, making accurate classification difficult. These characteristics make traditional text analysis approaches less effective when applied directly to Twitter data [4]. Therefore, advanced techniques based on machine learning and natural language processing have been widely adopted to improve sentiment classification performance [5].

Machine learning algorithms have shown significant potential in addressing sentiment classification problems. Traditional supervised learning models such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Naïve Bayes are commonly used for analyzing textual data and predicting sentiment categories [6]. These models are trained using labeled datasets where tweets are already classified according to sentiment polarity. However, relying on a single classifier may not always provide optimal performance, especially when dealing with complex textual data [7]. To overcome this limitation, ensemble learning techniques have gained attention in recent research. Ensemble learning combines multiple classifiers to produce a more accurate and stable prediction compared to individual models [8]. By integrating different algorithms through methods such as bagging, boosting, or voting mechanisms, ensemble approaches can reduce classification errors and improve overall performance [9].

Therefore, this study focuses on developing a robust ensemble learning framework for Twitter sentiment classification [10]. The proposed framework collects Twitter data, performs preprocessing to remove noise and irrelevant information, and applies feature extraction techniques such as Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) to convert textual information into numerical form [11]. Multiple machine learning classifiers are then trained and combined using an ensemble mechanism to generate the final sentiment prediction [12]. The performance of the proposed system is evaluated using standard metrics such as accuracy, precision, recall, and F1-score to ensure reliable sentiment classification results [13].

Background

Social media has become one of the most influential platforms for communication in today's digital world. Among various social media platforms, Twitter is widely used for sharing opinions, thoughts, and feedback on real-world topics such as politics, entertainment, products, and public events. Every day, millions of users post short messages called tweets, which reflect their emotions, attitudes, and viewpoints [1]. This large amount of opinion-based data provides an opportunity for researchers and organizations to analyze public sentiment and understand trends. The process of identifying and classifying people's emotions or opinions expressed in text is known as Sentiment Analysis [2]. It is an important part of Natural Language Processing (NLP) and is widely used for business decision-making, customer feedback analysis, brand monitoring, and prediction of public responses [3].

Twitter sentiment classification involves categorizing tweets into different sentiment classes such as positive, negative, or neutral [4]. However, performing sentiment analysis on Twitter data is challenging because tweets are short and often include informal language, abbreviations, emojis, hashtags, and spelling errors. Additionally, users may express sentiment indirectly through sarcasm or mixed emotions, which makes accurate classification difficult [5]. Therefore, there is a need to develop effective and reliable methods for analyzing sentiment in Twitter data [6].

Machine learning techniques have shown strong potential in solving sentiment classification problems. Traditional supervised learning algorithms such as Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) are commonly used for text classification. These models can be trained using labeled datasets where tweets are already marked with sentiment categories [7]. Among these, ensemble learning methods have gained popularity because they combine the results of multiple models to improve overall performance [8]. The main idea behind ensemble learning is that instead of relying on a single classifier [9], a group of classifiers can work together to produce a more accurate and stable prediction [10]. This synopsis proposes a robust ensemble learning framework for Twitter sentiment classification [11].

The proposed framework aims to improve classification accuracy by using ensemble techniques that combine multiple machine learning algorithms. Ensemble learning can be implemented using methods such as bagging, boosting, or voting-based classifiers [12], where predictions from different models are merged [13]. These methods are particularly useful in reducing errors caused by overfitting or model bias and help in building a more reliable sentiment analysis system [14]. The framework begins with the collection of Twitter data, followed by preprocessing steps such as removal of unwanted characters, stop words, and noise. After preprocessing, feature extraction techniques like Bag-of-Words (BoW) or Term Frequency–Inverse Document Frequency (TF-IDF) are used to convert textual data into numerical form for machine learning. Then, multiple classifiers are trained, and the ensemble mechanism is applied to generate final sentiment predictions. The performance of the proposed model is evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

Figure 1: Social media for sentiment analysis [1]

II. CRITICAL ANALYSIS OF RELEVANT STUDIES

Bhattacharjee (2026) analyzed tourist reviews from northeastern Indian states to understand visitor perceptions using sentiment analysis. Data from TripAdvisor, Reddit, and travel blogs were preprocessed using tokenization, lemmatization, and stop-word removal. Sentiment polarity was determined using VADER and TextBlob, while emotions were extracted using the NRC Emotion Lexicon. K-means clustering identified thematic patterns. Results showed positive sentiment in Sikkim and Meghalaya, while Assam and Nagaland had mixed sentiment due to infrastructure issues [1]. Rifani et al. (2026) studied public opinion on the 2024 president-elect administration using Twitter data. A total of 1,074 tweets were collected and classified into positive, negative, and neutral sentiments. Text preprocessing and TF-IDF feature extraction were applied, while SMOTE handled class imbalance. Using the Naïve Bayes algorithm, the model achieved 71% accuracy. Results indicated that negative sentiment dominated online discussions regarding governmental performance [2].

Elmokhtar et al. (2025) proposed an enhanced sentiment analysis framework to classify consumer opinions on phone brands using Twitter data. The system applied preprocessing, TF-IDF, and singular value decomposition for feature extraction. A novel ensemble classifier combining neural networks, KNN, and random forest was optimized using a modified walrus optimizer. The approach achieved 99.84% accuracy and demonstrated strong potential for brand reputation monitoring and consumer feedback analysis [3].

Nawaz et al. (2024) introduced the FNACSPM framework for detecting fake news using sequential pattern mining. The framework extracts frequent textual patterns from multiple public datasets and uses them as features for classification. Eight machine learning classifiers were evaluated using different performance metrics. Experimental results demonstrated that sequential pattern mining effectively improves fake news classification accuracy and provides insights into the structural patterns of misinformation [4].

Arief and Samsudin (2023) conducted a comparative study on handling neutral sentiment in binary sentiment classification. Using supervised learning techniques such as Naïve Bayes and Support Vector Machine, the research evaluated three strategies for dealing with neutral customer reviews. Experimental results showed that SVM achieved better performance with an AUC value of 0.923. The study concluded that handling neutral sentiment carefully improves overall classification reliability [5].

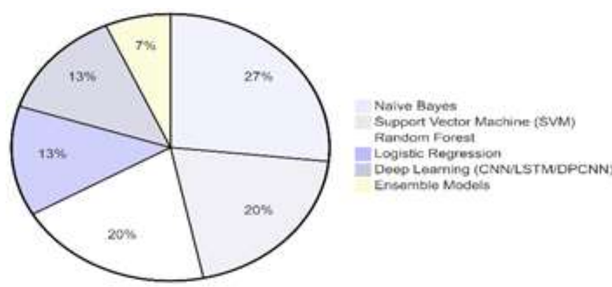

Figure 2: Distribution of Machine Learning Models Used in Sentiment Analysis Studies

Figure 2 illustrates the distribution of machine learning algorithms applied in sentiment analysis research based on the reviewed literature. Traditional machine learning models such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest are widely used for classifying sentiments in textual data. Among these, Naïve Bayes appears frequently due to its simplicity and efficiency in text classification tasks. Deep learning models such as

CNN and LSTM are also explored in some studies to capture contextual features from tweets. However, only a limited number of studies use ensemble learning approaches, highlighting the potential for improving sentiment classification accuracy by combining multiple classifiers.

Tsai and Shi (2022) proposed a sentiment classification model for Chinese text using FastText embeddings and a multi-scale Deep Pyramid Convolutional Neural Network (DPCNN). The approach first converts text into vector representations using FastText and then extracts features through multi-scale convolution filters. Experiments on the ChnSentiCorp dataset showed that the proposed hybrid model improves sentiment classification accuracy compared with traditional methods [6].

Talat et al. (2021) performed sentiment analysis on Roman Urdu product reviews collected from the Daraz e-commerce platform. The dataset was processed using natural language processing techniques before applying multiple machine learning classifiers. The study compared algorithms such as logistic regression and other classifiers for sentiment prediction. Results showed that logistic regression achieved the highest accuracy of 75%, demonstrating its effectiveness for Roman Urdu sentiment analysis [7].

Zahoor et al. (2020) analyzed restaurant reviews from a Facebook community in Karachi to classify customer sentiment regarding food, service, ambiance, and value for money. A dataset of approximately 4000 annotated reviews was used to train several machine learning models. Among Naïve Bayes, SVM, logistic regression, and random forest, the random forest classifier achieved the best performance with 95% accuracy [8]..

Figure 2 shows the distribution of datasets used in sentiment analysis studies. Twitter is the most commonly used data source in the reviewed literature because it provides a large volume of real-time user opinions on various topics such as politics, products, and social events. Other sources include tourism review platforms such as TripAdvisor and Reddit, e-commerce product reviews, and Facebook community discussions. Some studies also utilize mixed public datasets collected from multiple platforms. The dominance of Twitter datasets indicates its importance as a primary source for analyzing public sentiment.

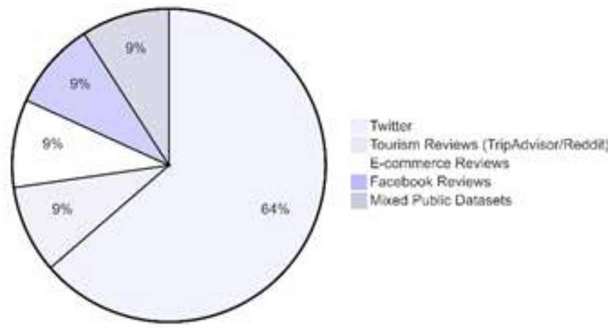

Figure 2: Data Sources Used in Sentiment Analysis Research

El_Rahman et al. (2019) proposed a sentiment analysis model that works with real-time Twitter data using both supervised and unsupervised machine learning approaches. The study analyzed tweets related to McDonald's and KFC to evaluate their popularity. From 7000 tweets for McDonald's, 2184 were positive, 1589 negative, and 3227 neutral. For KFC, 2076 were positive, 1311 negative, and 3613 neutral, indicating McDonald's was more popular. [9]
Rekha et al. (2019) analyzed Twitter data to understand public opinion about government schemes using hashtags. The dataset was cleaned and processed using machine learning techniques such as Naïve Bayes, Random Forest, and customized Random Forest. The results demonstrated that Random Forest achieved higher accuracy compared to Naïve Bayes in sentiment classification tasks. [10]

Saini et al. (2019) conducted sentiment analysis using the R programming language to preprocess, analyze, and visualize Twitter data. They also proposed an open-source framework for sentiment analysis using R tools. Approximately 3000 tweets were analyzed, and the results identified two main moods and eight different sentiment categories based on emotional expressions. [11]

Khun and Thant (2019) presented a visual sentiment analysis framework to examine public opinions about Huawei technology during the U.S.–China trade conflict. The framework combined sentiment analysis with spatial visualization techniques. The results provided insights that could help economists and policymakers understand and predict public opinion during international economic disputes. [12]

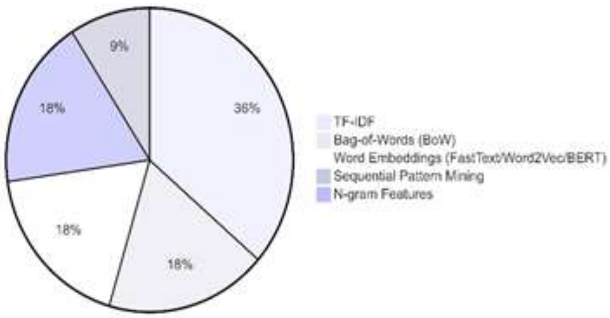

Figure 3: Feature Extraction Techniques Used in Sentiment Analysis

Figure 3 presents the distribution of feature extraction techniques used in sentiment analysis studies. Traditional text representation techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW) are widely adopted due to their effectiveness in converting textual data into numerical form for machine learning algorithms. In recent studies, advanced techniques such as word embeddings, including FastText, Word2Vec, and BERT, have been introduced to capture semantic relationships between words. Additionally, some research incorporates n-gram features and sequential pattern mining to improve feature representation and sentiment classification performance.

Kamış and Goularas (2019) compared deep learning models for Twitter sentiment analysis, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with LSTM architecture. Their experimental results highlighted limitations of both approaches in practical applications and demonstrated the challenges of applying deep learning techniques to real-world Twitter sentiment analysis tasks. [13]

Wang et al. (2018) proposed the SentiDiff algorithm to improve sentiment analysis on Twitter by combining textual information with sentiment diffusion patterns. The model utilized both text features and sentiment propagation structures to enhance prediction performance. Experimental results showed improvements in PR-AUC ranging from 5.09% to 8.38%, demonstrating the effectiveness of integrating diffusion patterns in sentiment prediction. [14]

Choudhary and Choudhary (2018) performed sentiment analysis on mobile brand reviews collected from Twitter. More than 5000 tweets were analyzed using a lexicon-based sentiment analysis approach. The results showed that Motorola, Samsung, and Oppo were the most frequently discussed brands in the dataset, indicating their strong presence in consumer discussions. [15]

Čišija et al. (2018) collected Twitter data using the RapidMiner tool and applied the AYLIEN extension to perform sentiment analysis. The study analyzed public opinion regarding political leadership and categorized sentiments into positive, negative, and neutral classes to understand the overall political atmosphere reflected on social media. [16]

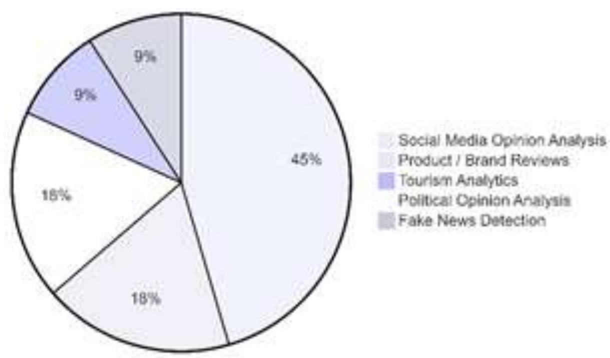

Figure 4: Application Domains of Sentiment Analysis in Literature

Naz et al. (2018) proposed a machine learning-based sentiment classification approach using textual patterns such as n-grams extracted from Twitter data. Experiments were conducted using four different n-gram feature sets and three weighting schemes. The results showed that unigram features produced better performance compared with other n-gram combinations. [17]

Figure 4 illustrates the different application domains where sentiment analysis techniques have been applied in the reviewed literature. Most studies focus on social media opinion analysis, particularly Twitter, to understand public attitudes toward various topics. Other application areas include product and brand review analysis, tourism experience evaluation, political opinion analysis, and fake news detection. These diverse application domains demonstrate the importance of sentiment analysis as a powerful tool for extracting insights from user-generated textual data across multiple fields.

Patil and Chavan (2018) introduced the SEG Analysis framework for real-time Twitter segmentation, sentiment analysis, and event detection. The framework combines Naïve Bayes classification with online clustering to detect emerging events and analyze associated sentiments from streaming Twitter data. This approach enables the detection of meaningful patterns in real-time social media discussions. [18]

Rani and Tondon (2018) explored sentiment classification using machine learning techniques such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). The study implemented the models using the Anaconda environment. Experimental results demonstrated that the KNN classifier achieved higher accuracy compared with the SVM classifier for sentiment classification tasks. [19]

Roy et al. (2018) conducted sentiment analysis on Twitter data related to the Goods and Services Tax (GST) over a period of 100 days. The analysis was performed using the R programming language and considered major political regions of the country. The results indicated that public opinion toward GST was evenly divided between positive and negative sentiments. [20]

Wagh and Punde (2018) presented a comprehensive survey on sentiment analysis techniques using Twitter datasets. The study reviewed various methods, algorithms, and research trends related to extracting sentiments from tweets and compared multiple approaches used in Twitter-based sentiment analysis systems. [21]

Table 1. Systematic Review

S. No.	Study (Author, Year)	Problem / Domain	Dataset Used	Feature / Representation	Model(s) Used	Optimization / Special Technique	Key Findings / Results
1	Rifani et al., 2026	Sentiment analysis of public opinion about the 2024 president-elect administration on Twitter	1,074 tweets collected via web scraping from Twitter	TF-IDF feature weighting	Multinomial Naïve Bayes	SMOTE used for class imbalance handling; preprocessing including tokenization, normalization, stopword removal, stemming	Model achieved 71.16% accuracy, 74.05% precision, 71.16% recall, 72.34% F1-score; negative sentiment dominated at 59.9% of tweets
2	Bhattacharjee, 2026	Sentiment analysis of tourism	Tourist reviews from	TF-IDF and word embeddin	Machine learning models	Emotion detection using NRC	Results identified positive

		experiences in North-East India for destination planning	TripAdvisor, Reddit, travel blogs, and GitHub datasets	gs (Word2Vec / BERT)	(Naïve Bayes, SVM, Random Forest) and deep learning (LSTM, BERT, FinBERT)	Emotion Lexicon; clustering using K-Means; visualization using t-SNE and PCA	sentiment trends for Sikkim and Meghalaya, while Assam and Nagaland showed mixed sentiment due to infrastructure challenges
3	Elmokhtar et al., 2025 Gomaa-2025	Twitter sentiment classification for phone brands/products	"Brands and Product Emotions" dataset (9094 tweets) Gomaa-2025	BoW → TF-IDF → SVD (hybrid categorization) Gomaa-2025	Novel 2-stage ensemble: KNN + Random Forest → Neural Network Gomaa-2025	Modified Walrus Optimizer for hyperparameter tuning Gomaa-2025	Achieved 99.84% accuracy, outperforming SoTA Gomaa-2025
4	Nawaz et al., 2024	Fake news detection/classification	6 public fake-news datasets + combined dataset	Sequential Pattern Mining (SPM) features	Evaluated using 8 classifiers	Proposed framework: FNACSPM	Outperformed state-of-art, high accuracy + faster classification
5	Arief & Samsudin, 2023	Handling neutral class in customer sentiment (binary setting)	Online customer reviews (3-star treated as neutral)	Standard text preprocessing	Naïve Bayes, SVM	Neutral handling scenarios tested	SVM best: AUC = 0.923, neutral may be treated as positive (AUC 0.904)
6	Tsai & Shi, 2022	Chinese sentiment classification	ChnSentiCorp dataset	FastText embeddings + multi-scale feature maps	Multi-scale DPCNN	Feature fusion + pyramid CNN	Improved accuracy vs baseline models
7	Talat et al., 2021	Roman Urdu product reviews sentiment	Kaggle Roman Urdu Daraz reviews	NLP preprocessing + feature extraction	Multiple ML classifiers	Comparative analysis	Logistic Regression achieved 75% accuracy (best)

8	Zahoor et al., 2020	Restaurant review sentiment + aspect classification	Facebook community data (~4000 samples)	Text categorization + sentiment polarity	NB, LR, SVM, RF	Also aspect classification: taste/service/ambience etc.	Random Forest achieved 95% accuracy (best)

III. Existing Methodology

Existing research on Twitter sentiment analysis mainly relies on machine learning and deep learning techniques to classify opinions expressed in tweets into sentiment categories such as positive, negative, and neutral. Most studies follow a general workflow consisting of data collection, preprocessing, feature extraction, model training, and evaluation.

Initially, data is collected from social media platforms such as Twitter, TripAdvisor, Reddit, Facebook, and online review websites. For example, Rifani et al. (2026) collected 1,074 tweets to analyze public opinion about the 2024 president-elect administration, while Elmokhtar et al. (2025) analyzed 9,094 tweets related to phone brands. After collecting, the data undergoes preprocessing steps including removal of stop words, tokenization, normalization, stemming, and noise removal to clean and prepare textual data for analysis.

Next, feature extraction techniques such as Bag-of-Words (BoW), TF-IDF, n-grams, word embeddings, and FastText representations are used to convert textual data into numerical vectors that machine learning algorithms can process. Several studies also use advanced representations such as Word2Vec, BERT embeddings, and sequential pattern mining features to capture semantic relationships between words.

Various machine learning classifiers are used in sentiment analysis research, including Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN). Some studies apply deep learning models such as CNN, LSTM, and DPCNN to improve classification performance. Additionally, ensemble learning techniques combine multiple models to increase prediction accuracy. Performance of these models is typically evaluated using accuracy, precision, recall, F1-score, and AUC metrics.

Although these methods have achieved good results, many existing systems rely on single classifiers or limited feature representations, which may reduce classification accuracy when dealing with complex Twitter data containing sarcasm, informal language, emojis, and abbreviations.

IV. RESEARCH GAP

Despite significant progress in sentiment analysis research, several limitations still exist in the current approaches. First, many studies rely on single machine learning models such as Naïve Bayes, SVM, or logistic regression, which may not fully capture the complexity of Twitter data. These models often struggle with informal language, sarcasm, emojis, and mixed sentiments that are commonly found in social media texts.

Second, although deep learning models such as CNN and LSTM have been explored, they often require large labeled datasets and high computational resources, which limits their practical implementation. In addition, some studies focus on specific domains such as tourism, politics, or product reviews, making it difficult to generalize the models to other domains.

Third, many existing works do not effectively address class imbalance problems in sentiment datasets, where neutral or negative tweets may dominate the dataset. This imbalance can lead to biased models that fail to correctly identify minority sentiment classes.

Furthermore, while several researchers have applied different machine learning algorithms individually, limited work has been done on developing robust ensemble learning frameworks that combine multiple classifiers to improve prediction performance and reduce classification errors.

Therefore, there is a need to develop a more robust ensemble-based sentiment classification framework that integrates multiple machine learning models, improves feature representation, and enhances classification accuracy for Twitter sentiment analysis. The proposed system aims to address these limitations by combining multiple classifiers through an ensemble learning approach, leading to more reliable and accurate sentiment prediction.

V. SYSTEM ARCHITECTURE

The proposed system architecture for Twitter sentiment classification is designed to analyze public opinions by applying machine learning and ensemble learning techniques. The framework consists of several stages, including data collection, preprocessing, feature extraction, model training, ensemble classification, and evaluation. Each stage plays an important role in transforming raw Twitter data into meaningful sentiment predictions.

Data Collection: The first stage of the framework involves collecting tweets from Twitter using APIs or publicly available datasets. The collected tweets represent users' opinions about specific topics such as political events, products, services, or public policies. The dataset may contain thousands of tweets which are then labeled into sentiment categories such as positive, negative, or neutral.

Data Preprocessing: Raw Twitter data often contains noise such as hashtags, URLs, emojis, punctuation marks, and abbreviations. Therefore, preprocessing is necessary to clean and prepare the data for analysis. The preprocessing stage includes several steps such as removal of URLs, stop words, hashtags, special characters, tokenization, case folding, normalization, and stemming or lemmatization. These processes improve the quality of textual data and make it suitable for machine learning models.

Feature Extraction: After preprocessing, the textual data is converted into numerical form so that machine learning algorithms can process it. In this stage, feature extraction techniques such as Bag-of-Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), and n-gram representations are applied. These techniques capture important words and patterns within tweets that help in identifying the sentiment expressed in the text.

Model Training: Once the features are extracted, multiple machine learning classifiers are trained using the processed dataset. Common classifiers used in sentiment analysis include Naïve Bayes, Support Vector Machine (SVM), Random Forest, and Logistic Regression. Each classifier learns patterns from the training data and attempts to predict the sentiment of new tweets.

Ensemble Learning: To improve classification performance, the proposed framework applies ensemble learning techniques. Ensemble learning combines predictions from multiple classifiers using methods such as majority voting, bagging, or boosting. By combining multiple models, the system reduces prediction errors and improves overall accuracy and stability compared to using a single classifier.

Sentiment Classification: The ensemble model generates the final sentiment prediction for each tweet. Tweets are categorized into three main classes: positive, negative, and neutral. The classification results help identify overall public opinion regarding the topic being analyzed.

Performance Evaluation: The final stage evaluates the performance of the proposed system using standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics measure how effectively the ensemble model predicts sentiments and allow comparison with existing methods.

VI. CONCLUSION

In this study, a comprehensive review of sentiment analysis techniques for Twitter data has been presented. Social media platforms, particularly Twitter, generate large volumes of opinion-rich textual data that can be used to understand public sentiment on various topics such as politics, tourism, products, and social issues. However, the analysis of Twitter data is challenging due to the presence of informal language, abbreviations, emojis, and noisy textual structures. Therefore, effective computational approaches are required to extract meaningful insights from such data.

The literature review examined several machine learning and deep learning approaches used for sentiment classification, including Naïve Bayes, Support Vector Machine, Random Forest, Logistic Regression, Convolutional Neural Networks, and Long Short-Term Memory networks. Feature extraction techniques such as Bag-of-Words, TF-IDF, word embeddings, and n-gram representations have also been widely used to convert textual data into machine-readable formats. The review highlights that although these techniques have shown promising results, many existing studies rely on single classifiers or limited feature representations, which may reduce classification performance when dealing with complex Twitter data.

To address these challenges, the study proposes an ensemble learning framework for Twitter sentiment classification. The proposed framework integrates multiple machine learning models to improve prediction accuracy and stability. By combining the strengths of different classifiers through ensemble methods such as voting, bagging, or boosting, the system can reduce individual model bias and improve sentiment classification performance. The overall framework includes data collection, preprocessing, feature extraction, model training, ensemble classification, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. The proposed approach is expected to provide more reliable sentiment predictions and better insights into public opinion expressed on social media platforms.

Future Work

Although the proposed framework provides a promising approach for improving sentiment classification accuracy, several opportunities for future research remain. First, future studies can explore the integration of advanced deep learning models such as transformers, Bidirectional Encoder Representations from Transformers (BERT), and attention-based neural networks to capture contextual information more effectively in Twitter text. Second, incorporating multimodal data such as images, videos, and emojis along with textual information could further enhance sentiment detection performance. Social media posts often contain multimedia content that contributes to the overall sentiment expression.

Third, future research can focus on improving the detection of complex linguistic features such as sarcasm, irony, and mixed sentiments, which remain challenging for many existing sentiment analysis systems. Additionally, large-scale datasets from multiple social media platforms can be used to improve the generalization ability of the model across different domains. Real-time sentiment analysis systems can also be developed to monitor public opinion dynamically for applications such as brand monitoring, policy analysis, and crisis management. Finally, future work may investigate hybrid models that combine ensemble learning with deep learning architectures to further improve classification accuracy and scalability for large social media datasets.

REFERENCES

1. S. Bhattacharjee, "Machine Learning–Driven Sentiment Analysis for Tourism Insights: The Case of North-East Indian States," Applied Decision Analytics, vol. 2, no. 1, pp. 103–123, 2026.
2. P. Rifani, B. Irawan, A. Faqih, D. Pratama, and D. A. Kurnia, "Sentiment Analysis of Public Opinion on the 2024 President-Elect's Administration on Twitter using Naïve Bayes," Journal of Artificial Intelligence and Engineering Applications, vol. 5, no. 2, pp. 3247–3252, 2026.
3. G. Elmokhtar, W. M. Mohamed, E. El-Shafeiy, et al., "Enhanced sentiment classification of phone brands on Twitter with a modified walrus optimizer and novel ensemble method," Neural Computing and Applications, vol. 37, pp. 14299–14325, 2025.

4. M. Z. Nawaz, M. S. Nawaz, P. Fournier-Viger, and Y. He, “Analysis and Classification of Fake News Using Sequential Pattern Mining,” Big Data Mining and Analytics, vol. 7, no. 3, pp. 942–963, Sept. 2024.
5. M. Arief and N. A. Samsudin, “Neutral Class Handling for Customer Sentiment Analysis in Binary Classification: A Comparative Study of Supervised Machine Learning Classification Algorithm,” in Proc. 8th Int. Conf. Informatics and Computing (ICIC), Manado, Indonesia, 2023, pp. 1–8.
6. S. Tsai and L. Shi, “Chinese Text Sentiment Classification Model Based on FastText and Multi-scale Deep Pyramid Convolutional Neural Network,” in Proc. Int. Conf. Computation, Big-Data and Engineering (ICCBE), Yunlin, Taiwan, 2022, pp. 75–77.
7. M. Talat, H. Asim, and A. Asmat, “Classification of Sentiments of the Roman Urdu Reviews of Daraz Products using Natural Language Processing Approach,” in Proc. Int. Conf. Innovative Computing (ICIC), Lahore, Pakistan, 2021, pp. 1–6.
8. K. Zahoor, N. Z. Bawany, and S. Hamid, “Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning,” in Proc. 21st Int. Arab Conf. Information Technology (ACIT), Giza, Egypt, 2020, pp. 1–6.
9. S. A. El_Rahman, F. A. AlOtaibi, and W. A. AlShehri, “Sentiment analysis of Twitter data,” in Proc. 2019 Int. Conf. Computer and Information Sciences (ICCIS), 2019, pp. 1–4.
10. V. Rekha, R. Raksha, P. Patil, N. Swaras, and G. L. Rajat, “Sentiment analysis on Indian government schemes using Twitter data,” 2019.
11. S. Saini, R. Punhani, R. Bathla, and V. K. Shukla, “Sentiment analysis on Twitter data using R,” in Proc. 2019 Int. Conf. Automation, Computational and Technology Management (ICACTM), Amity University, 2019.
12. N. H. Khun and H. A. Thant, “Visualization of Twitter sentiment during the period of US banned Huawei,” 2019.
13. S. Kamış and D. Goularas, “Evaluation of deep learning techniques in sentiment analysis from Twitter data,” in Proc. 2019 Int. Conf. Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019.
14. L. Wang, J. Niu, and S. Yu, “SentiDiff: Combining textual information and sentiment diffusion patterns for Twitter sentiment analysis,” Journal of Latex Class Files, vol. 14, no. 8, Aug. 2018.
15. M. Choudhary and P. K. Choudhary, “Sentiment analysis of text reviewing algorithm using data mining,” in Proc. 2018 Int. Conf. Smart Systems and Inventive Technology (ICSSIT), 2018.
16. M. Čišija, E. Žunić, and Dž. Đonko, “Collection and sentiment analysis of Twitter data on the political atmosphere,” in Proc. Symp. Neural Networks and Applications, 2018.
17. S. Naz, A. Sharan, and N. Malik, “Sentiment classification on Twitter data using support vector machine,” in Proc. 2018 IEEE/WIC/ACM Int. Conf. Web Intelligence (WI), 2018.
18. M. Patil and H. K. Chavan, “Event based sentiment analysis of Twitter data,” in Proc. 2nd Int. Conf. Computing Methodologies and Communication (ICCMC), 2018.
19. R. Rani and S. Tondon, “Chat summarization and sentiment analysis techniques in data mining,” in Proc. 2018 4th Int. Conf. Computing Science, 2018.
20. S. Roy, S. Sehgal, and S. Agrawal, “An approach to sentiment analysis of Twitter data on the goods and services tax,” in Proc. 2018 Int. Conf. Advances in Computing, Communication Control and Networking (ICACCCN), 2018.
21. R. Wagh and P. Punde, “Survey on sentiment analysis using Twitter dataset,” in Proc. 2018 2nd Int. Conf. Electronics, Communication and Aerospace Technology (ICECA), 2018.