Engineering and Technology Journal e-ISSN: 2456-3358
Volume 11 Issue 07 July-2026, Page No.-10945-10956
DOI: 10.47191/etj/v11i07.18, I.F. – 8.595

Development of an Ensemble Machine Learning Model for Sentiment Analysis of University Students' Opinions on NELFUND

RAJI[1] Mumini Oyetunji, AYOADE[1*] Olusola Bamidele, AKINDELE[1] Aminat Adejoke, YUSUF-MASHOPA[1] Kemi Jemilat, ABDULRAUFF[1] Muinat Folake, RAJI[1] Ibrahim Adebayo, MUSAH[1] Fatima Bolanle, ADEGBIJI[1], Abiodun, Timothy, AMUDA[2] Oluwafemi Micheal,

[1]Department of Data Science, Informatics and Computer Science, Emmanuel Alayande University of Education, Oyo, Nigeria
[2]Directorate of Information Communication Technology, Emmanuel Alayande University of Education, Oyo, Nigeria

ABSTRACT: The Nigerian Education Loan Fund (NELFUND) was established to improve access to higher education by providing financial assistance to eligible students. Since its introduction, students have actively expressed their views on the programme through social media platforms and online discussion forums. Analysing these opinions is essential for understanding public perception and assessing the transparency and perceived benefits of the scheme. Traditional sentiment analysis methods often rely on individual machine learning algorithms whose predictive performance may be constrained by model-specific limitations. Ensemble learning provides an alternative by combining multiple classifiers to improve robustness and classification accuracy. This study develops and evaluates a voting-based ensemble framework for sentiment analysis of students' opinions on NELFUND. Opinions were collected from students of Emmanuel Alayande University of Education, Oyo, through WhatsApp discussion groups. The textual data were preprocessed using data cleaning, tokenisation, part-of-speech tagging, stop-word removal, stemming, and lemmatisation. Text features were generated using Term Frequency–Inverse Document Frequency (TF-IDF) and Count Vectorisation techniques. Five machine learning classifiers—Support Vector Machine, Multinomial Naïve Bayes, Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost)—were trained individually and subsequently combined using three ensemble strategies: Majority Voting, Weighted Average Voting, and Average Probability Voting. Experimental results indicate that the proposed models achieved strong sentiment classification performance for both the transparency and perceived benefits of NELFUND. Across the evaluated datasets, classification accuracies ranged from 92% to 94%, while the ensemble approaches demonstrated competitive performance relative to the individual classifiers. The findings further reveal that classifier effectiveness varies with the characteristics of the dataset, suggesting that no single model consistently outperforms others under all conditions. Overall, the study demonstrates that ensemble learning provides a reliable and practical approach for analysing students' opinions and offers valuable insights that can support the evaluation and continuous improvement of educational funding initiatives.

KEYWORDS: Nigerian Education Loan Fund, Voting Ensemble, Tokenisation, WhatsApp Platform, Term Frequency-Inverse Document Frequency.

1. INTRODUCTION

The rapid growth of social media and online communication platforms has transformed the way people express opinions, share experiences, and discuss public policies. Every day, millions of textual messages are generated on platforms such as WhatsApp, Facebook, Twitter, and online discussion forums, creating valuable sources of information for understanding public attitudes and behaviour. Analysing these opinions enables organisations and policymakers to identify public concerns, evaluate policy effectiveness, and make informed decisions based on citizens' experiences.

Sentiment analysis, also referred to as opinion mining, has become one of the most active research areas in natural language processing and machine learning. It focuses on automatically identifying the emotional orientation of textual data by classifying opinions as positive, negative, or neutral. In recent years, sentiment analysis has been successfully applied in diverse domains, including product review analysis, healthcare, finance, education, politics, tourism, and social media analytics. Its growing popularity is largely attributed to the increasing volume of user-generated content and the need to convert unstructured textual information into actionable knowledge.

In Nigeria, the establishment of the Nigerian Education Loan Fund (NELFUND) represents a significant policy initiative aimed at improving access to higher education through financial support for eligible students. Since the commencement of the scheme, students have actively shared their experiences and perceptions regarding the transparency of the loan disbursement process and the benefits derived from the programme. These opinions provide an important source of feedback that can assist policymakers in evaluating the effectiveness of the scheme and identifying areas requiring further improvement. However, the large volume of textual data generated across online platforms makes manual analysis both time-consuming and impractical, thereby creating the need for automated sentiment analysis techniques.

Machine learning algorithms have demonstrated considerable success in sentiment classification because of their ability to identify hidden patterns within textual data. Traditional classifiers such as Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) have been widely employed in previous studies and have achieved encouraging results across different sentiment analysis tasks. Nevertheless, each of these algorithms possesses inherent strengths and weaknesses. While some classifiers provide high predictive accuracy, others exhibit greater robustness or computational efficiency depending on the characteristics of the dataset. Consequently, relying on a single classifier may not always produce the most reliable classification performance.

To overcome these limitations, researchers have increasingly adopted ensemble learning techniques that combine the predictions of multiple classifiers. Ensemble models improve generalisation by exploiting the complementary strengths of different learning algorithms and reducing the influence of individual model errors. Voting-based ensemble methods, including majority voting, weighted average voting, and probability-based voting, have attracted considerable attention because they are straightforward to implement while often producing more stable and accurate predictions than individual classifiers. Although ensemble learning has been extensively investigated in areas such as financial forecasting, medical diagnosis, cybersecurity, and product review analysis, relatively few studies have examined its application to analysing students' opinions on educational funding programmes, particularly within the Nigerian context. Existing studies have largely focused on generic social media datasets or commercial product reviews, leaving limited evidence regarding public perceptions of NELFUND. This gap highlights the need for an intelligent sentiment analysis framework capable of accurately evaluating students' opinions on the transparency and perceived benefits of the scheme.

This study addresses this gap by developing a voting-based ensemble framework for sentiment analysis of students' opinions on NELFUND. Five supervised machine learning classifiers—Support Vector Machine, Multinomial Naïve Bayes, Logistic Regression, Random Forest, and Extreme Gradient Boosting—are trained individually and subsequently combined using three ensemble strategies: Majority Voting, Weighted Average Voting, and Average Probability Voting. Before classification, the textual data undergo comprehensive preprocessing involving data cleaning, tokenisation, part-of-speech tagging, stop-word removal, stemming, and lemmatisation, while TF-IDF and Count Vectorisation techniques are employed to generate representative text features. The effectiveness of the proposed framework is evaluated using students' opinions collected from WhatsApp discussion groups at Emmanuel Alayande University of Education, Oyo. Performance is assessed using accuracy, precision, recall, and F1-score, with comparative analysis conducted between the individual classifiers and the ensemble models. The findings are expected to contribute to the growing body of knowledge on sentiment analysis while providing useful evidence for policymakers seeking to improve the implementation, transparency, and public acceptance of the Nigerian Education Loan Fund.

2. LITERATURE REVIEW

The increasing availability of user-generated content on social media platforms has accelerated research on sentiment analysis, making it one of the most active application areas of natural language processing and machine learning. Sentiment analysis seeks to determine the emotional orientation of textual data by automatically classifying opinions as positive, negative, or neutral. Its applications extend across numerous domains, including product recommendation systems, healthcare, financial forecasting, political opinion mining, educational research, and public policy evaluation. As the volume and diversity of online textual data continue to expand, improving the accuracy and robustness of sentiment classification models has become a major research focus.

Early sentiment analysis studies relied heavily on lexicon-based techniques, where sentiment polarity was inferred from predefined dictionaries of positive and negative words. These methods are straightforward to implement and require little or no labelled training data. However, they often struggle with contextual expressions, sarcasm, negation, and domain-specific vocabulary, which limits their effectiveness when analysing informal social media conversations. To overcome these shortcomings, researchers have increasingly adopted supervised machine learning algorithms capable of learning complex relationships directly from labelled datasets.

Traditional machine learning classifiers, including Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), have consistently demonstrated strong performance in text classification tasks. Support Vector Machine is particularly recognised for its ability to handle high-dimensional sparse text features and its strong generalisation capability. Multinomial Naïve Bayes remains attractive because of its computational efficiency and effectiveness on bag-of-words representations, whereas Logistic Regression provides interpretable decision boundaries for binary and multiclass classification problems. Tree-based approaches such as Random Forest and XGBoost offer improved modelling of nonlinear relationships and interactions among textual features while providing robustness against noise and overfitting.

Several recent studies have reported encouraging results using these algorithms for sentiment classification. Ainapure et al. (2023) applied deep learning and lexicon-based methods to analyse sentiments expressed in COVID-19 tweets and demonstrated that combining different analytical techniques improved prediction accuracy. Similarly, Coaquira and Villanueva (2024) reviewed recent developments in sentiment analysis and concluded that supervised machine learning continues to provide competitive performance despite the rapid emergence of deep learning approaches. Wang and Jaber (2025) compared BERT and Long Short-Term Memory (LSTM) models for hotel review analysis and reported that transformer-based architectures achieved superior contextual understanding, although at considerably higher computational cost.

Despite their effectiveness, individual classifiers rarely perform optimally across all datasets because each learning algorithm possesses unique strengths and limitations. Model performance is strongly influenced by data characteristics such as vocabulary distribution, class imbalance, feature sparsity, and document length. Consequently, selecting a single classifier may not consistently provide the highest predictive accuracy or the best generalisation performance across different application domains. To address this limitation, researchers have increasingly adopted ensemble learning techniques that integrate the predictions of multiple classifiers. Ensemble methods improve predictive performance by exploiting the diversity of individual learning algorithms while reducing the effect of classification errors produced by any single model. Existing ensemble strategies are generally categorised into bagging, boosting, stacking, and voting methods, each employing a different mechanism for combining model outputs.

Among these approaches, voting-based ensembles have gained considerable attention because of their simplicity, computational efficiency, and ability to improve classification stability. In majority voting, the final prediction is determined by the class receiving the highest number of votes from the participating classifiers. Weighted voting extends this concept by assigning greater influence to classifiers with stronger predictive performance, while probability-based voting aggregates posterior probabilities to produce more refined classification decisions. Previous studies have shown that these strategies often outperform individual classifiers in a variety of application domains. Abellán and Castellano (2017) demonstrated that ensemble methods produced more reliable classification results than individual base learners in credit scoring applications. Similarly, Ala'raj and Abbod (2016) proposed a hybrid ensemble framework that significantly improved prediction accuracy by combining complementary classifiers. More recently, Krishna and Angeline (2022), Eyvazi-AbdolJabbar et al. (2024), and Muniappan and Subramanian (2025) reported that voting-based ensemble models consistently enhanced sentiment classification performance across financial, social media, and industrial datasets. These studies collectively indicate that classifier diversity plays an important role in improving predictive robustness.

Nevertheless, the superiority of ensemble learning is not universal. Several investigations have shown that well-tuned individual classifiers can occasionally outperform ensemble models depending on the characteristics of the dataset. Saleena (2018) observed that certain standalone classifiers achieved better sentiment prediction than ensemble approaches for specific Twitter datasets, while Varshney et al. (2020) similarly reported that classifier fusion does not always guarantee superior performance. These findings suggest that the effectiveness of ensemble learning depends not only on the combination strategy but also on the diversity and complementary strengths of the constituent classifiers.

Although substantial progress has been achieved in sentiment analysis, relatively limited attention has been devoted to analysing students' opinions on educational funding initiatives, particularly within developing countries. Most existing studies focus on product reviews, political discussions, healthcare, financial markets, or general social media conversations. Consequently, there remains limited empirical evidence regarding students' perceptions of the Nigerian Education Loan Fund (NELFUND), especially concerning its transparency and perceived benefits.

The present study addresses this research gap by developing a voting-based ensemble framework that combines Support Vector Machine, Multinomial Naïve Bayes, Logistic Regression, Random Forest, and Extreme Gradient Boosting classifiers. The proposed framework evaluates three ensemble strategies—Majority Voting, Weighted Average Voting, and Average Probability Voting—to determine their effectiveness in analysing students' opinions on NELFUND. By providing a comparative evaluation of individual and ensemble classifiers on real-world educational data, the study contributes to the growing body of knowledge on sentiment analysis. It demonstrates the applicability of ensemble learning to evidence-based evaluation of higher education financing policies.

3. MATERIALS AND METHODS

This section describes the overall research methodology adopted in the study, including the study design, data collection procedure, ethical considerations, text preprocessing, feature extraction, machine learning models, ensemble learning framework, and performance evaluation metrics used to assess the proposed sentiment analysis system.

3.1 Research Design

This study adopted an experimental research design to develop and evaluate a machine learning-based framework for sentiment analysis of students' opinions on the Nigerian Education Loan Fund (NELFUND). The framework combines conventional supervised machine learning algorithms with voting-based ensemble techniques to classify students' opinions regarding the transparency and perceived benefits of the NELFUND programme. The overall workflow of the proposed framework consists of six major stages: data collection, text preprocessing, feature extraction, individual classifier training, ensemble classification, and performance evaluation, as shown in Figure 1.

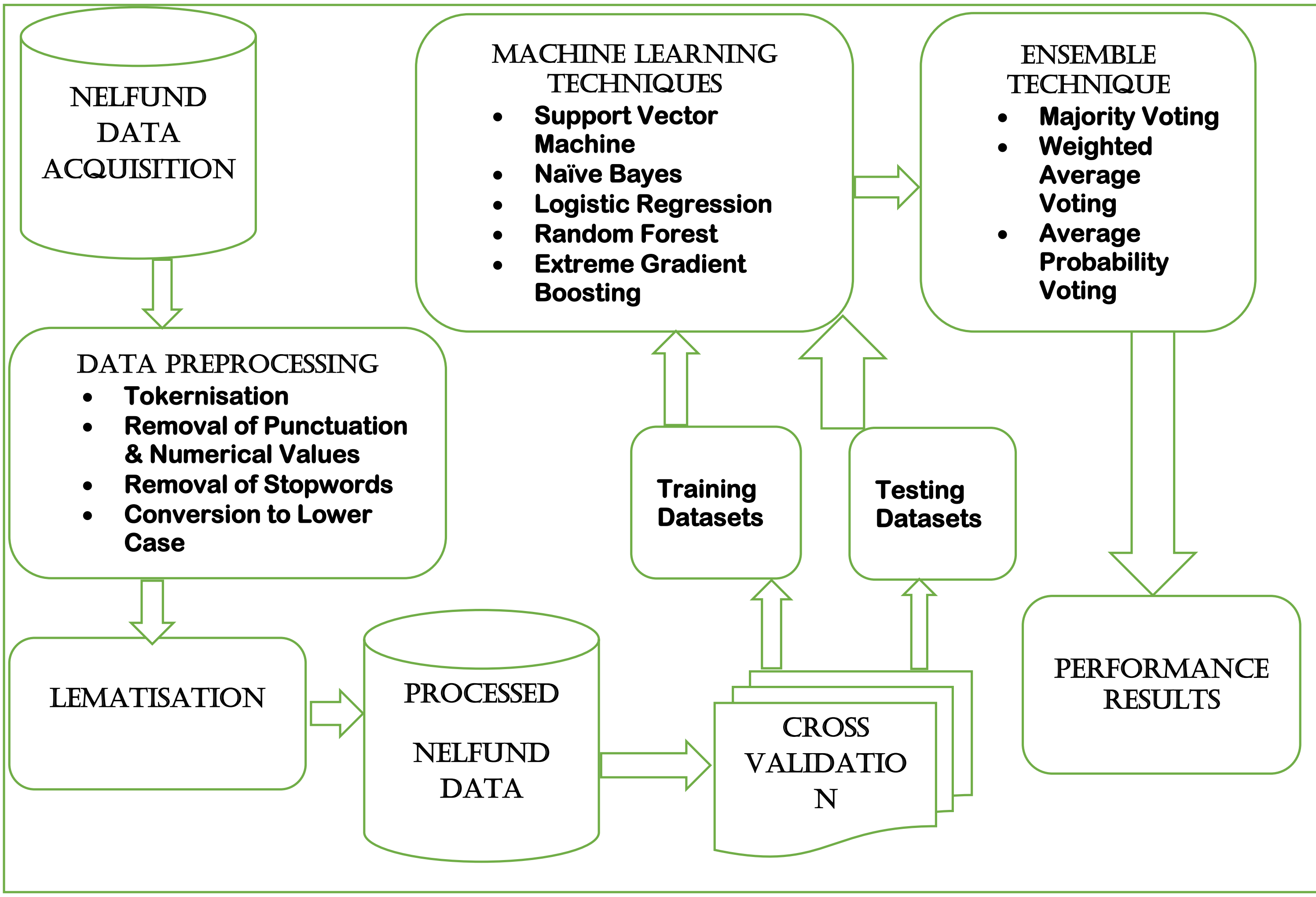

Figure 1: The Architectural Design of Ensemble Classification System for Students' Opinions on NELFUND

3.2 Data Collection

The dataset used in this study was obtained from WhatsApp discussion groups comprising undergraduate students of Emmanuel Alayande University of Education, Oyo, Nigeria, during the First Semester of the 2024/2025 academic session. The discussions focused on students' experiences, perceptions, and opinions regarding the implementation, transparency, and perceived benefits of the Nigerian Education Loan Fund (NELFUND). To ensure that the dataset reflected genuine opinions, only messages directly related to NELFUND were extracted for analysis. Messages unrelated to the objectives of the study, including greetings, advertisements, duplicated posts, emojis without textual content, multimedia-only messages, and unrelated conversations, were excluded during data preparation.

Each opinion was manually reviewed and organised into two major analytical categories:

- opinions relating to the transparency of NELFUND loan disbursement; and
- opinions relating to the perceived benefits of the NELFUND scheme.

The resulting corpus formed the dataset used for training and evaluating the sentiment classification models.

3.3 Sentiment Tagging Module

The sentiment tagging stage was used to analyse the sentiment in the preprocessed data. Hutto and Gilbert (2014) developed the Valence Aware Dictionary and sEntiment Reasoner (VADER) technique, which was used in the sentiment tagging process. The VADER sentiment analysis tool uses a lexicon-based and rule-based approach to analyse text and classify it as positive, negative, or neutral sentiment. Furthermore, this produces a compound score that takes into account all three sentiments. This tool effectively analyses social media content, including emojis, acronyms, slang, and contractions (Hoti & Ajdari, 2023). Sentiment labelling was carried out at the word and sentence level. Positive words/sentences received a score of +1, negative words/sentences received a score of -1, and words/sentences with no opinion received a score of zero. Finally, the net score was calculated using the sum of positive and negative sentiments in a sentence.

3.4 Ethical Considerations

The study was conducted in accordance with accepted ethical principles governing research involving human-generated textual data. Before data collection commenced, permission was obtained from the administrators of the selected WhatsApp discussion groups to use anonymised discussion posts strictly for academic research purposes. Participants were informed that only publicly shared comments within the selected groups relating to NELFUND would be analysed and that no attempt would be made to identify individual contributors. To protect participants' privacy and confidentiality, all personally identifiable information, including names, telephone numbers, profile photographs, usernames, and any other identifying details, was removed before data analysis. Each record was assigned an anonymous identification code solely for data management purposes.

The dataset was stored securely and used exclusively for the objectives of this research. No individual participant is identifiable in the published results, and all findings are reported in aggregate form to ensure complete confidentiality. The study complied with the research ethics guidelines of Emmanuel Alayande University of Education for research involving publicly shared textual information and respected the principles of voluntary participation, confidentiality, privacy protection, and responsible data handling.

3.5 Text Preprocessing

Social media messages frequently contain spelling errors, abbreviations, punctuation inconsistencies, emojis, hyperlinks, and other forms of noisy textual information that can adversely affect machine learning performance. Consequently, several preprocessing operations were performed before feature extraction. The preprocessing pipeline consisted of:

- text cleaning;
- case normalisation;
- removal of URLs, punctuation marks, numbers, and special symbols;
- tokenisation;
- part-of-speech tagging;
- stop-word removal;
- stemming; and
- lemmatisation.

These operations standardised the textual data, reduced feature dimensionality, and improved the semantic consistency of the dataset before model training.

3.6 Feature Extraction

After preprocessing, numerical feature vectors were generated using two widely adopted text representation techniques:

- Count Vectorisation; and
- Term Frequency-Inverse Document Frequency (TF-IDF).

Count Vectorisation represents documents according to word occurrence frequencies, whereas TF-IDF assigns weights to terms based on both their frequency within individual documents and their rarity across the entire corpus. The combination of these representations enables the classifiers to capture both lexical occurrence patterns and discriminative textual information.

The resulting feature vectors served as inputs to all classification models.

3.7 Base Classification Models

Five supervised machine learning algorithms were selected because of their proven effectiveness in text classification tasks:

- Support Vector Machine (SVM)
- Multinomial Naïve Bayes (MNB)
- Logistic Regression (LR)
- Random Forest (RF)
- Extreme Gradient Boosting (XGBoost)

Each classifier was independently trained using identical training data to ensure a fair comparative evaluation. Their predictive outputs subsequently formed the basis for the ensemble learning framework.

3.8 Voting-Based Ensemble Classification

To improve classification robustness, three voting-based ensemble strategies were developed:

- Majority Voting;
- Weighted Average Voting; and
- Average Probability Voting.

Majority Voting determines the final sentiment by selecting the class predicted by the highest number of base classifiers.

Weighted Average Voting assigns different importance to each classifier according to its predictive capability before aggregating the individual predictions. Average Probability Voting computes the average posterior probability produced by the participating classifiers and selects the class with the highest aggregated probability.

The ensemble framework exploits the complementary strengths of individual classifiers while reducing the influence of isolated classification errors, thereby improving overall prediction stability.

Algorithm 1: Training Model for Hard Voting and Soft Voting Ensemble Classifier

---

Input:

```
- Classifiers: C = {C₁, C₂, ..., Cₖ}
- Input instance: x
- Class set: Y = {y₁, y₂, ..., yₙ}
- Voting mode: MODE ∈ {"hard", "soft"}
```

Output:

```
- Final predicted class label: y_pred
```

Procedure:

```
1.    If MODE == "hard":
  a.    Initialize vote counter: votes = dict()
        For each class y in Y: votes[y] = 0
  b.    For each classifier Cᵢ in C:
        Try:
       yᵢ = Cᵢ.predict(x)
       votes[yᵢ] += 1
    Except:
       continue # Skip classifiers that fail
  c.    max_votes = max(votes.values())
        candidates = [y for y in Y if votes[y] == max_votes]
  d.    If len(candidates) == 1:
      y_pred = candidates [0]
    Else:
      y_pred = tie_break(candidates)
2.    Else if MODE == "soft":
  a.    Initialize probability accumulator: probs = dict()
        For each class y in Y: probs[y] = 0.0
  b.    valid_classifiers = 0
  c.    For each classifier Cᵢ in C:
         Try:
       class_probs = Cᵢ.predict_proba(x)
       For j in range(len(Y)):
          probs[Y[j]] += class_probs[j]
       valid_classifiers += 1
    Except:
       continue
  d.    For each class y in Y:
      probs[y] = probs[y] / valid_classifiers
  e. max_prob = max(probs.values())
    candidates = [y for y in Y if probs[y] == max_prob]
  f.    If len(candidates) == 1:
      y_pred = candidates [0]
    Else:
      y_pred = tie_break(candidates)
3. Return y_pred
```

3.9 Performance Evaluation

The developed models were evaluated using four standard classification metrics:

- Accuracy;
- Precision;
- Recall; and
- F1-score.

These metrics provide complementary information regarding the predictive capability of both the individual classifiers and the proposed ensemble models. Comparative analysis was conducted to determine the relative effectiveness of each classification strategy on the transparency and benefits datasets.

4. RESULTS AND DISCUSSION

This section presents the experimental findings obtained from the sentiment classification models developed in this study. The performance of the five individual machine learning classifiers and the three voting-based ensemble models was evaluated using Accuracy, Precision, Recall, and F1-score. Separate experiments were conducted on the transparency and perceived benefits datasets to examine the effectiveness of the proposed ensemble framework under different sentiment analysis scenarios.

4.1 Performance of Individual Machine Learning Classifiers

The classification results demonstrate that all five machine learning algorithms successfully identified students' sentiments regarding the Nigerian Education Loan Fund (NELFUND), although their predictive performance varied across the evaluation metrics. This variation reflects the different learning mechanisms employed by each classifier and the influence of dataset characteristics on model behaviour. Support Vector Machine and Logistic Regression consistently produced strong classification performance owing to their ability to model high-dimensional textual features effectively. Similarly, Multinomial Naïve Bayes achieved competitive results, particularly for opinions represented through term-frequency features, confirming its suitability for text classification tasks involving sparse document representations.

Random Forest and XGBoost also demonstrated satisfactory predictive capability by effectively modelling complex relationships within the textual features. However, slight differences in precision, recall, and F1-score indicate that no individual classifier consistently outperformed all others across every evaluation metric. These findings reinforce the widely accepted view that classifier performance depends on both the learning algorithm and the characteristics of the underlying dataset.

4.2 Performance of the Voting-Based Ensemble Models

The voting-based ensemble models were developed to exploit the complementary strengths of the individual classifiers while minimising their respective weaknesses. Three ensemble strategies—Majority Voting, Weighted Average Voting, and Average Probability Voting—were evaluated and compared with the standalone classifiers. The experimental results show that the ensemble models achieved consistently competitive performance across the transparency and perceived benefits datasets. By aggregating predictions from multiple classifiers, the ensemble framework reduced isolated classification errors and produced more stable sentiment predictions than individual learning algorithms. This improvement demonstrates the benefit of classifier diversity in sentiment analysis, where different models often capture complementary linguistic patterns within the same dataset. Among the evaluated ensemble techniques, the Weighted Average Voting model achieved the most balanced overall performance by assigning greater influence to classifiers with superior predictive capability. Majority Voting also produced reliable classification results, while Average Probability Voting provided competitive performance through probability aggregation, particularly in instances where the participating classifiers generated similar confidence scores.

Table 1: Performance Evaluation Metrics of the Developed Classification Models for Sentiment Analysis of Students' Opinion on NELFUND on Transparency NELFUND Datasets

Model	Performance Metrics						
	F1-Score (%)		Recall (%)		Accuracy (%)	Precision (%)	
	Negative	Positive	Negative	Positive		Negative	Positive
Support Vector Machine	24	96	14	100	93	89	93
Naïve Bayes	0	96	0	100	92	0	92
Logistic Regression	0	96	0	100	92	0	92
Random Forest	21	96	12	100	93	78	93
XGBoost	64	91	21	99	93	60	94

Majority Voting	10	96	5	100	92	75	92
Weighted Average Voting	21	96	12	100	93	78	93
Average Probability Voting	13	96	100	93	93	100	93

Table 2: Performance Evaluation Metrics of the Developed Classification Models for Sentiment Analysis of Students' Opinion on NELFUND on Advantages of NELFUND Datasets

Model	Performance Metrics						
	F1-Score (%)		Recall (%)		Accuracy (%)	Precision (%)	
	Negative	Positive	Negative	Positive		Negative	Positive
Support Vector Machine	33	97	20	100	94	100	94
Naïve Bayes	0	96	0	100	93	0	93
Logistic Regression	18	97	10	100	94	100	94
Random Forest	33	97	20	100	94	100	94
XGBoost	31	97	22	99	93	58	94
Majority Voting	24	97	44	100	94	100	94
Weighted Average Voting	16	100	27	97	94	100	94
Average Probability Voting	24	97	14	100	94	100	94

4.3 Comparative Analysis

Comparison of the individual classifiers and ensemble models reveals several important observations. Although some standalone classifiers achieved marginally higher scores for specific evaluation metrics, the ensemble approaches generally produced more consistent performance across all metrics. This behaviour reflects one of the principal advantages of ensemble learning, namely its ability to improve model robustness by combining diverse decision boundaries rather than relying on a single prediction mechanism. The findings further indicate that no single classifier universally dominates sentiment classification performance. Instead, the effectiveness of each model depends on the distribution of textual features, vocabulary characteristics, and sentiment patterns contained within the dataset. Consequently, integrating multiple classifiers provides a practical mechanism for reducing prediction variability and improving overall reliability.

The classification accuracies obtained in this study, ranging from approximately 92% to 94%**,** demonstrate that voting-based ensemble learning provides an effective framework for analysing students' opinions on educational policy initiatives such as NELFUND. These results also confirm that machine learning techniques can successfully extract meaningful sentiment information from informal social media discussions despite the presence of abbreviations, spelling variations, and conversational language.

4.4 Discussion

The results obtained in this study demonstrate the effectiveness of ensemble learning for sentiment classification of students' opinions on the Nigerian Education Loan Fund. Unlike individual classifiers, which rely on a single learning strategy, the proposed ensemble framework combines multiple prediction models to produce more reliable and stable classification outcomes. This complementary learning behaviour reduces the influence of individual model errors and improves overall generalisation performance. The strong performance of the ensemble models can largely be attributed to the diversity of the participating classifiers. Support Vector Machine effectively separates high-dimensional feature spaces, Multinomial Naïve Bayes efficiently models word frequency distributions, Logistic Regression provides robust linear decision boundaries, while Random Forest and XGBoost capture nonlinear relationships among textual features. Combining these complementary learning mechanisms enables the ensemble framework to exploit the strengths of each classifier while compensating for their individual limitations.

The findings are consistent with previous studies reporting improved predictive performance with voting-based ensemble techniques. Similar improvements have been observed in sentiment analysis applications involving product reviews, financial opinions, healthcare discussions, and social media analytics, where ensemble learning frequently produces greater robustness than individual classifiers. The present study extends these findings to the educational domain by demonstrating that ensemble methods can effectively analyse students' perceptions of government-funded educational support programmes.

An important observation emerging from this study is that classifier effectiveness varies according to dataset characteristics. Although ensemble learning generally improved classification stability, certain individual classifiers performed competitively for specific evaluation metrics. This finding supports previous research suggesting that the superiority of ensemble learning depends on the diversity and complementary strengths of the constituent models rather than the number of classifiers alone. Beyond its predictive performance, the proposed framework has practical implications for educational policymakers and institutional administrators. Automated analysis of students' opinions provides timely evidence regarding public perceptions of policy implementation, programme transparency, and beneficiary satisfaction. Such information can support continuous policy evaluation, identify emerging concerns, and guide decision-making to improve educational funding initiatives. Consequently, the proposed sentiment analysis framework represents a valuable decision-support tool for assessing student feedback on large-scale educational intervention programmes such as NELFUND.

4.5 Practical Implications

The findings of this study have several practical implications for policymakers, educational administrators, researchers, and developers of intelligent decision-support systems.

For policymakers: The proposed sentiment analysis framework provides an evidence-based approach for monitoring public perceptions of the Nigerian Education Loan Fund (NELFUND). By continuously analysing students' opinions expressed on social media platforms, policymakers can identify emerging concerns, assess the effectiveness of policy implementation, and make timely adjustments to improve the accessibility, transparency, and impact of the loan programme. Such data-driven insights can complement traditional survey methods and support more responsive educational policy formulation.

For NELFUND administrators: The developed ensemble machine learning model can serve as an automated monitoring tool for analysing large volumes of student feedback in real time. The system enables administrators to detect recurring complaints, misinformation, delays in loan disbursement, application challenges, and positive experiences. This allows for faster intervention, improved stakeholder communication, enhanced customer support, and increased public confidence in the programme.

For higher education institutions: Universities, polytechnics, and colleges of education can adopt the proposed framework to evaluate students' perceptions of financial support initiatives and other institutional services. The insights obtained from sentiment analysis can assist institutional management in identifying areas requiring improvement, designing targeted student support programmes, and strengthening communication strategies between students and administrators.

For educational researchers: The annotated sentiment dataset and the ensemble machine learning framework developed in this study provide a valuable benchmark for future research on educational policy evaluation, natural language processing, and social media analytics. Researchers can extend the framework to other government intervention programmes, educational reforms, scholarship schemes, and student welfare initiatives, thereby contributing to the advancement of intelligent educational data analytics.

For software developers and data scientists: The proposed ensemble sentiment classification framework demonstrates the practical application of machine learning techniques for extracting actionable insights from unstructured textual data. The methodology can be integrated into web-based dashboards and decision-support systems for real-time monitoring of public opinion across diverse sectors, including education, healthcare, governance, and public service delivery.

Overall, this study demonstrates that automated sentiment analysis offers a scalable, efficient, and cost-effective mechanism for understanding students' perceptions of educational policies. The proposed framework supports evidence-based decision-making and provides stakeholders with timely information for improving policy implementation, service delivery, and student satisfaction.

5. CONCLUSION

This study developed and evaluated a voting-based ensemble machine learning framework for sentiment analysis of university students' opinions on the Nigerian Education Loan Fund (NELFUND). The framework combined five supervised machine learning classifiers—Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—with three ensemble strategies, namely Majority Voting, Weighted Average Voting, and Average Probability Voting. The objective was to investigate whether combining multiple classifiers could improve the reliability of sentiment classification compared with individual learning algorithms.

A dataset comprising 747 student opinions collected from WhatsApp discussion groups at Emmanuel Alayande University of Education was used for the experimental evaluation. Before model development, the textual data underwent a comprehensive preprocessing pipeline that included text cleaning, tokenisation, part-of-speech tagging, stop-word removal, stemming, and lemmatisation. Feature extraction was performed using Count Vectorisation and Term Frequency–Inverse Document Frequency (TF-IDF), enabling the classifiers to capture meaningful linguistic patterns from the students' comments.

The experimental findings demonstrate that both the individual classifiers and the ensemble models achieved strong sentiment classification performance. Across the transparency and perceived benefits datasets, the models recorded classification accuracies ranging from approximately 92% to 94%**,** with consistently high precision, recall, and F1-scores. Although some individual classifiers performed competitively for specific evaluation metrics, the ensemble approaches provided more stable and balanced performance by combining the complementary strengths of the constituent classifiers. In particular, the voting-based strategies reduced the influence of isolated prediction errors and improved the robustness of the overall classification process.

The study also shows that ensemble learning offers a practical and reliable approach for analysing students' perceptions of educational funding policies. Automated sentiment analysis enables large volumes of textual feedback to be processed efficiently, providing policymakers and educational administrators with timely evidence on programme transparency, implementation challenges, and perceived benefits. Such insights can support evidence-based decision-making, continuous policy evaluation, and improvements in future educational intervention programmes. Despite these encouraging results, certain limitations should be acknowledged. The dataset was obtained from a single institution, which may limit the generalisability of the findings to students in other universities or regions. In addition, the study relied on traditional machine learning techniques using handcrafted text features rather than contextual language representations generated by modern transformer-based models. The performance of the proposed framework may therefore vary when applied to larger, more linguistically diverse datasets. Overall, this research demonstrates that voting-based ensemble learning provides an effective framework for sentiment analysis of students' opinions on NELFUND. The findings contribute to the growing application of machine learning in educational policy evaluation and illustrate the potential of ensemble methods as reliable decision-support tools for analysing public opinion on government-funded educational initiatives.

6. FUTURE WORK

Several opportunities exist for extending the present study. Future research should evaluate the proposed framework using datasets collected from multiple universities across different geopolitical zones in Nigeria. A larger and more diverse dataset would improve the generalisability of the findings and provide a broader understanding of students' perceptions of the Nigerian Education Loan Fund. Another promising direction is the integration of advanced deep learning and transformer-based language models such as Bidirectional Encoder Representations from Transformers (BERT), RoBERTa, DistilBERT, and other domain-adapted language models. These approaches can capture contextual and semantic relationships more effectively than traditional feature engineering techniques and may further improve sentiment classification performance, particularly for complex expressions, sarcasm, and context-dependent opinions.

Future studies should also investigate hybrid ensemble frameworks that combine conventional machine learning algorithms with deep learning architectures. Such hybrid approaches may benefit from both the interpretability of traditional classifiers and the contextual learning capabilities of neural language models. Comparative evaluations involving stacking, boosting, and bagging techniques would provide further insight into the relative strengths of different ensemble strategies for educational sentiment analysis. In addition, incorporating multilingual sentiment analysis would enable the framework to process opinions expressed in English as well as widely spoken Nigerian languages such as Yoruba, Hausa, Igbo, and Nigerian Pidgin. This enhancement would improve the inclusiveness of the analysis and better reflect the linguistic diversity of Nigerian students.

Future research may also extend sentiment classification beyond polarity detection by incorporating aspect-based sentiment analysis and emotion recognition. Such an approach would enable researchers to identify specific issues discussed by students, including loan application procedures, disbursement timelines, eligibility requirements, repayment conditions, communication, and customer support. These finer-grained insights would provide policymakers with more actionable information for improving the design and implementation of NELFUND. Finally, future investigations should incorporate explainable artificial intelligence (XAI) techniques to improve model transparency and interpretability. Methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) can help identify the textual features that influence classification decisions, thereby increasing stakeholders' confidence in automated sentiment analysis systems and supporting their adoption in educational policy evaluation.

REFERENCES

1. Abellán, J. & Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications*, 73, 1–10.
2. Ainapure, B. S., Pise, R. N., Reddy, P., Appasani, B., Srinivasulu, A., Khan, M. S. & Bizon, N. (2023). Sentiment analysis of COVID-19 tweets using deep

learning and lexicon-based approaches. *Sustainability*, 15, 1-16.

3. Ala'raj, M., & Abbod, M. F. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications*, 64, 36–55.
4. Anam, M. K., Lestari, T. P., Yenni, H., Nasution, T. & Firdaus, M. B. (2025). Enhancement of machine learning algorithm in fine-grained sentiment analysis using the ensemble. *ECTI Transactions on Computer and Information Technology*, 19(2), 159-167.
5. Başarslan, M. S. & Kayaalp, F. (2023). Sentiment analysis with ensemble and machine learning methods in multi-domain datasets, *Turkish Journal of Engineering*, 7(2), 141-148.
6. Birjali, M., Kasri, M. & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-based Systems*, 226, 1-20.
7. Chatterjee, S. & Byun, Y. C. (2022). Voting ensemble approach for enhancing alzheimer's disease classification. *Sensors*, 22(19), 1-14.
8. Coaquira, Y. M. & Villanueva, E. (2024). A review on text sentiment analysis with machine learning and deep learning techniques. *IEEE Access*, 12, 193115-193130.
9. Dahiya, A. (2019). A Study on Data Science with Python in Comparison to other Languages. *Indian Journal of Applied Research*, 9, 60-61.
10. Devi, S. S. & Vinod, V. (2025). An ensemble method for sentiment analysis on textile dataset using machine learning algorithms. *SEEJPH*, 26, 1700-1722.
11. Elhadidi, E. A., Salah, A., Abdellah, M. & Darwish, S. M. (2025). Sentiment analysis: a comparison of deep learning neural network algorithms with ensemble learning algorithms. *Journal of Information Systems Engineering and Management*, 10(34), 210-220.
12. Eyvazi-AbdolJabbar, S., Kim, S., Feizi-Derakhshi, M., Farhadi, Z. & Mohammed, D. A. (2024). An ensemble-based model for sentiment analysis of persian comments on Instagram using deep learning algorithms. *IEEE Access*, 12, 1-13.
13. Hoti, M. T. & Ajdari, J. (2023). Sentiment analysis using the vader model for assessing company services based on posts on social media. *SEEU Review*, 18(2), 19-33.
14. Hutto, C.J. & Gilbert, E. (2014). *VADER:* A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the Eighth International Conference on Weblogs and social media (ICWSM-14).
15. Ismet, H. T., Mustaqim, T. & Purwitasari, D. (2022). Aspect based sentiment analysis of product review using memory network. *Scientific Journal of Informatics*, 9(1), 73-75.
16. Jabbar, G. H. (2024). Advanced threat detection using soft and hard voting techniques in ensemble learning. *Journal of Robotics and Control (JRC)*, 5(4), 1-15.
17. Jung, J., Petkanic, P., Nan, D. & Kim, J. H. (2020). When a girl awakened the world: A user and social message analysis of Greta Thunberg. *Sustainability*, 12, 1.15.
18. Kamath, R. S., Kaushik, M. R. & Ramakrishna, M. (2025). Opinion mining using ensemble model for restaurant feedback analysis. *Discover Computing*, 28(125), 1-27.
19. Khan, M. Y. & Junejo, K. N. (2020). Exerting 2D-space of sentiment lexicons with machine learning techniques: A hybrid approach for sentiment analysis. *International Journal* of *Advanced Science and Computer Applications (IJASCA)*, 11, 599–608.
20. Krishna, S. & Angeline P. G. (2022). Sentiment analysis using ensemble machine learning techniques. *International Journal of Communication Networks and Information Security*, 14(3), 490-506.
21. Lavado, J. L., Cantador, I., Cort´esCediel, M. E.& Fern´andez, M. (2022). Automatic intent-based classification of citizen-to-government tweets. In: *Proceedings of Ongoing Research*, 2021, pp. 47.
22. Mika, V., Mäntylä, D. & Graziotin, M. K. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16–32.
23. Muniappan, M. & Subramanian, N. D. P. (2025). A majority voting mechanism-based ensemble learning approach for financial distress prediction in Indian automobile industry. *Journal of Risk and Financial Management*, 18(4), 1-31. https://www.mdpi.com/1911-8074/18/4/197/pdf?version=1743772184
24. Nandwani, P. & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(8), 1-21.
25. Oyekunle, R. A. & Abdulkareem, R. O. (2019). Sentiment analysis of students tweets on Unilorin CBT examinations. *International Journal of Applied Business and Information Systems*, 3(2), 66-75.
26. Pandey, G., Gupta, N. K., Paul, B. & Paul, M. (2024). Decoding digital conversations: A hybrid

sentiment analysis framework for whatsapp chat behavioural intelligence. *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, 12(21s), 4953–4960.
27. Rai, N., Kaushik, N., Kumar, D., Raj, C. & Ali, A. (2022). Mortality prediction of COVID-19 patients using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 3, 1-10.
28. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 1-14.
29. Saleena, A. N. (2018). An ensemble classification system for twitter sentiment analysis. *Procedia Computer Science*, 132, 937-946.
30. Song, Y. Q., Yao, X., Liu, Z., Shen, X. & Mao, J. (2020). An improved C4.5 algorithm in bagging integration model. *IEEE Access*, 8(1), 1-12.
31. Soori, M., Arezoo, B. & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 3, 54.
32. Supriatna, D. J. I., Saputra, H. & Hasan, K. (2023). Enhancing the red wine quality classification using ensemble voting classifiers. *Infolitika Journal of Data Science*, 1(2), 1-16.
33. Syafei, R. M. & Efrilianda, D. A. (2023). Machine learning model using extreme gradient boosting (XGBoost) feature importance and light gradient boosting machine (LightGBM) to improve accurate prediction of bankruptcy. *Recursive Journal of Informatics*, 1(2), 1-10.
34. Van Fc, L. L., Anam, M. K., Firdaus, M. B., Yunefri, Y. & Rahmi, N. A. (2024). Enhancing machine learning model performance in addressing class imbalance. *COGITO Smart Journal*, 10(1), 1-13.
35. Varshney, C. J., Sharma, A. & Yadav, P. (2020). Sentiment analysis using ensemble classification technique. *IEEE Students' Conference on Engineering & Systems (SCES)*, July 10-12, 2020, Prayagraj, India
36. Vinora, A., Suryakumar. L. S. S., Harivishnu,. T. S., Jaspin, K., Selvaraj, A. K. & Nancy, D. M. (2025). Visualization of hidden patterns in whatsapp chat using sentimental analysis. *3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, Bengaluru, India, 05-07 February 2025
37. Xiao, Y., Li, C., Th¨urer, M., Liu, Y. & Qu, T. (2022). Towards lean automation: Fine-grained sentiment analysis for customer value identification. *Computers & Industrial Engineering*, 169(108186), 1-24.
38. Wang, G. & Jaber, M. M. (2025). A deep learning approach to sentiment analysis of hotel reviews: Comparing BERT and LSTM models. *International Journal of Advances in Artificial Intelligence and Machine Learning*, 2(2), 67-75.
39. Zhang, L., Wang, S. & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining Knowledge. Discovery*, 8(4), 1-15.