NIU Journal of Educational Research Copyright©2026
Nexus International University ISSN: *3007-1844;* 12(2): 199-211

Development of Machine Learning Models for Sentiment Analysis of University Students' Opinions on NELFUND

MUMINI OYETUNJI RAJI, OLUSOLA BAMIDELE AYOADE,
AMINAT ADEJOKE AKINDELE, KEMI JEMILAT YUSUF-MASHOPA,
MUINAT FOLAKE ABDULRAUFF, IBRAHIM ADEBAYO RAJI,
FATIMA BOLANLE MUSAH, ABIODUN TIMOTHY ADEGBIJI,
OLUWAFEMI MICHAEL AMUDA
Emmanuel Alayande University of Education, Oyo, Nigeria

Abstract. This study investigated students' perceptions of the Nigerian Education Loan Fund (NELFUND) at Emmanuel Alayande University of Education, concentrating on issues pertaining to the transparency of loan disbursement and perceived advantages, which influence students' confidence in the programme. A sentiment analysis methodology was utilised, employing opinion datasets classified into transparency and benefits attributes, with polarity levels predominantly indicating positive sentiments. Data preprocessing encompassed cleaning, tokenisation, part-of-speech tagging, stop-word elimination, stemming, and lemmatisation to enhance data quality. Feature extraction was performed utilising Term Frequency–Inverse Document Frequency (TF-IDF) and count vectorisation methods. Sentiment classification utilised the VADER model, and five machine learning algorithms, Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were assessed based on accuracy, precision, recall, and F1-score metrics. The findings indicated that in terms of transparency-related sentiments, SVM, RF, and XGBoost consistently surpassed NB and LR. Regarding benefits-related sentiments, Random Forest and Support Vector Machine demonstrated optimal performance for both negative and positive polarities. The findings indicate predominantly favourable student perceptions of NELFUND and illustrate the efficacy of advanced machine learning models in analysing policy-related opinions in educational finance.

Keywords: Educational Finance Policy, Extreme Gradient Boosting, Natural Language Processing, Nigerian Education Loan Fund, Sentiment Analysis

1. Introduction

In recent years, the proliferation of user-generated content on digital platforms has increased significantly, particularly on social networks, e-commerce sites, e-learning platforms, blogs, and online forums. These platforms contain extensive information that aids organisations and governments in making informed decisions and enhances the quality of products and services (Zhang et al., 2018). This content predominantly reflects users' opinions, emotions, experiences, and attitudes regarding the services or products they consume, which can significantly affect the behaviours and perceptions of other users (Coaquira & Villanueva, 2024). Social media platforms, including Twitter, WhatsApp, Facebook, Instagram, and YouTube, are increasingly popular as information disseminates more swiftly through these channels than through conventional news outlets (Ainapure et al., 2023). Consequently, these platforms have evolved into significant repositories of opinion-laden data, with user sentiments increasingly influencing societal perspectives and public discourse (Jung et al., 2020).

Sentiment analysis, a branch of Natural Language Processing (NLP), aims to create computational techniques for detecting and interpreting opinions conveyed in textual data, including online reviews, comments, and social media posts (Khan & Junejo, 2020). The primary aim of sentiment analysis is to assess individuals' attitudes, perceptions, and opinions concerning particular subjects or entities. This process is generally accomplished by attributing polarity values, such as positive, negative, or neutral, to textual expressions, allowing researchers to ascertain the overall sentiment conveyed in the data (Dahiya, 2019). Sentiment analysis seeks to identify opinions, emotions, and attitudes present in diverse textual sources, such as documents, brief messages, blogs, reviews, and news articles (Birjali et al., 2021). While predominantly employed in business contexts for analysing customer feedback and enhancing services, its application has broadened to various fields, including social media monitoring, electoral processes, stock market forecasting, disaster management, healthcare, and cyberbullying detection (Mika et al., 2018).

Sentiment analysis methodologies can derive insights from textual data via various analytical frameworks, including granular sentiment analysis (Xiao et al., 2022), intent-based sentiment analysis (Lavado et al., 2022), aspect-based sentiment analysis (Ismet et al., 2022), and emotion-oriented sentiment analysis (Nandwani & Verma, 2021). This study employs a detailed sentiment analysis method to assess the strength and orientation of opinions articulated in the dataset. Sentiments are typically classified into positive, negative, or neutral categories. This study utilises various machine learning algorithms for sentiment classification, including support vector machines (SVM), multinomial Naïve Bayes, logistic regression, random forest, and extreme gradient boosting. The use of individual algorithms has progressively diminished owing to their variable predictive efficacy across datasets (Sarker, 2021; Soori et al., 2023). As a result, researchers are progressively incorporating various algorithms or methodologies to improve model precision and dependability, such as ensemble methods and hybrid approaches that combine the strengths of multiple algorithms (Chatterjee & Byun, 2022; Van Fc et al., 2024). This study employs machine learning models to categorise students' sentiments regarding the Nigerian Education Loan Fund (NELFUND) into two polar classifications: positive and negative.

2. Related Studies

Oyekunle and Abdulkareem (2019) examined students' perceptions of computer-based testing (CBT) at the University of Ilorin by analysing Twitter posts featuring hashtags like Unilorin CBT and Unilorin. A dictionary-based sentiment analysis method was used to categorise tweets as positive, negative, or neutral. The findings revealed that 40.54% of the tweets conveyed negative sentiment, 34.17% were positive, and 25.29% were neutral. The study determined that a significant proportion of students reported dissatisfaction with CBT examinations, primarily due to the difficulties they faced during the testing process, such as unclear instructions, time constraints, and technical issues that hindered their performance.

Krishna and Angeline (2022) introduced an extensive sentiment analysis framework utilising machine learning methodologies along with various preprocessing and feature selection techniques. Their methodology combines Part-of-Speech (POS) tagging with Natural Language Processing (NLP) techniques and WordNet to improve the understanding of semantic meaning and emotional nuances in text. Principal component analysis (PCA) was utilised to optimise the feature space, while an ensemble feature selection method incorporating elastic net regularisation was employed to identify the most pertinent features. The final classification phase employed a hybrid machine learning model that integrates the interpretability of Decision Trees with the resilience of Random Forest algorithms. Moreover, stacking and voting classifiers were employed to encapsulate both linear and nonlinear relationships within the dataset, resulting in enhanced sentiment classification efficacy.

Başarslan and Kayaalp (2023) performed sentiment analysis on user-generated feedback obtained from websites across two separate domains. The research employed two datasets: hotel evaluations from TripAdvisor and film critiques from Rotten Tomatoes. Sentiment classification was executed utilising two text representation methodologies: the Word2Vec word embedding model, which encapsulates semantic relationships among words, and the Term Frequency–Inverse Document Frequency (TF–IDF) model. The representations were assessed utilising four machine learning algorithms: Naïve Bayes (NB), Support Vector Machines (SVM), Logistic Regression (LR), and K-Nearest Neighbour (kNN), in addition to two ensemble methods, specifically stacking and majority voting. The evaluation of model performance was conducted utilising accuracy and F-measure metrics. The results indicated that ensemble techniques consistently surpassed individual machine learning algorithms, with the majority voting method yielding the most favourable outcomes.

Pandey et al. (2024) performed a comprehensive analysis of communication dynamics in WhatsApp group chats employing a hybrid sentiment analysis framework. Their methodology incorporates various elements, such as preprocessing protocols, lexicondriven analysis, generative AI methodologies, and ensemble scoring systems to deliver a more thorough comprehension of conversational sentiment.

Likewise, Sivri (2024) presented a stacking-based ensemble model aimed at analysing sentiment in financial news articles concerning BIST30 stocks traded on the Borsa Istanbul exchange. The model integrates predictions from multiple base classifiers, such as Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), Naïve Bayes, and Support Vector Machines (SVM). A logistic regression model functions as the meta-learner that consolidates the outputs of these foundational models. Experimental findings indicate that this ensemble approach markedly enhances accuracy and generalisation relative to standalone models.

Vinora et al. (2025) proposed a system for analysing WhatsApp conversations utilising Python-based data processing libraries, including pandas. The developed tool emphasises the identification of significant interactions and the application of sentiment analysis to derive insights into user behaviour. The system improves the comprehension and analysis of WhatsApp messages through the integration of machine learning and natural language processing techniques.

Kamath et al. (2025) presented a sentiment analysis methodology for restaurant reviews employing a majority voting ensemble classifier. The approach integrates multiple machine learning algorithms, namely Decision Tree, Support Vector Machine, Random Forest, Logistic Regression, K-Nearest Neighbours, and Multinomial Naïve Bayes. The models were assessed using standard performance metrics including accuracy, precision, recall, and F1-score. The findings demonstrated that the ensemble classifier, specifically the integration of Support Vector Machine, Logistic Regression, and Multinomial Naïve Bayes, attained enhanced accuracy and efficiency relative to standalone models.

Nandi et al. (2025) devised a fuzzy-based ensemble learning framework that amalgamates Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks with principles of fuzzy logic. Experimental evaluations demonstrated that this hybrid methodology substantially surpassed conventional machine learning and deep learning models. The proposed model attained an accuracy of 89.0%, a precision of 88.6%, a recall of 89.2%, and an F1-score of 88.9%. These results underscore the model's efficacy in managing the ambiguity and complexity typically found in social media text.

Anam et al. (2025) proposed an ensemble learning framework that amalgamates various techniques, including voting, bagging, boosting, and stacking, to enhance sentiment classification accuracy. Experiments performed on various datasets assessed the efficacy of individual classifiers against ensemble-based methods. The results indicated that the stacking-based ensemble model attained the highest accuracy of 92.45%, surpassing traditional classifiers like SVM (88.23%) and Naïve Bayes (85.67%). The stacking technique exhibited enhanced performance by efficiently integrating predictions from various base learners.

Elhadidi et al. (2025) constructed sentiment analysis models employing sophisticated deep learning architectures, including XLNet, RoBERTa, and BERT. The research used ensemble learning methods, such as stacking, bagging, and boosting, to identify varied linguistic patterns and improve classification efficacy. The researchers assessed individual model performance and the efficacy of the ensemble framework using the cardiffnlp/tweetsentimentmultilingual dataset. The findings indicated that the ensemble method consistently attained superior accuracy and an F1-score in comparison to individual models.

Devi and Vinod (2025) investigated sentiment analysis in e-commerce product reviews, emphasising polarity detection. The study employed a thorough preprocessing pipeline that encompassed addressing missing values, eliminating symbols and punctuation, converting text to lowercase, removing stop words, and executing stemming and tokenisation. The transformed text was subsequently converted into numerical features utilising the Count Vectorizer technique. Two machine learning algorithms—Naïve Bayes and Support Vector Machine—were employed to classify sentiment in reviews. The researchers proposed an ensemble model that integrates Random Forest and XGBoost algorithms to improve polarity detection accuracy.

Azim et al. (2025) presented the RRLS ensemble model, which amalgamates Random Forest, Recurrent Neural Network (RNN), Logistic Regression (LR), and Support Vector Machine (SVM) for sentiment analysis in Urdu. The research examined film critiques from the Internet Movie Database (IMDB) alongside Urdu-language tweets. Data preprocessing was executed utilising the UrduHack library, which undertook tasks including character normalisation, spacing corrections, and punctuation management. The individual SVM and LR models attained an accuracy of approximately 87%, whereas the LSTM and bidirectional LSTM (BiLSTM) models exhibited an accuracy of 84%. The proposed RRLS ensemble model achieved a performance accuracy of 90%, which rose to 92.77% upon the application of the Synthetic Minority Oversampling Technique (SMOTE) to mitigate class imbalance.

Wang and Jaber (2025) utilised BERT and LSTM deep learning models to categorise hotel reviews into positive, neutral, and negative sentiments. The dataset comprised 20,000 TripAdvisor reviews that underwent preprocessing, including stopword elimination, special character filtering, tokenisation, stemming, and lemmatisation. Sentiment labels were extracted from star ratings, with 4–5 stars denoting positive sentiment, 3 stars representing neutral sentiment, and 1–2 stars indicating negative sentiment. To rectify class imbalance, the predominant positive class was subjected to random undersampling. Both models were trained with an 80:20 training-validation split and assessed using accuracy, precision, recall, and F1-score through 5-fold cross-validation. The BERT model, which did not use undersampling, had the highest overall accuracy of 0.86. It also had strong F1-scores of 0.93 for the positive class and 0.79 for the negative class. Both models encountered difficulties in detecting neutral sentiment, achieving lower F1-scores (BERT: 0.43, LSTM: 0.25). Although undersampling enhanced recall for the neutral class (BERT: 0.79), it diminished overall accuracy (BERT: 0.73; LSTM: 0.67) and precision for the majority class as a result of information loss. BERT exhibited superior performance compared to LSTM in the analysis of sentiment in hotel reviews, especially in the context of imbalanced datasets.

3. Research Methodology

The methodology delineates the procedures employed in the development of machine learning models to analyse the sentiments articulated by university students regarding NELFUND. The procedure encompasses several critical phases, including dataset acquisition, data preprocessing, sentiment annotation, feature extraction, and sentiment classification employing five distinct machine learning algorithms. Figure 1 displays a block diagram depicting the architectural framework of the machine learning system employed to categorise students' perceptions of NELFUND. The following sections offer a comprehensive elucidation of each phase of the methodology.

3.1 Data Collection

This study gathered brief English texts expressing students' opinions on NELFUND at Emmanuel Alayande University of Education, Oyo, through a WhatsApp platform established for this research. Each opinion was classified into sentiment categories, either positive or negative, according to the student's overall perception of NELFUND. The dataset consisted of two primary attributes: transparency and perceived advantages. The transparency dataset comprised 747 entries, including 666 positive, 58 negative, and 23 neutral instances. The benefits dataset comprised 747 entries, including 685 positive, 52 negative, and 10 neutral responses.

3.2 Data Preprocessing

This study's preprocessing pipeline included data cleaning, tokenisation, part-of-speech tagging, stopword removal, stemming, and lemmatisation, all of which enhanced feature extraction and model performance.

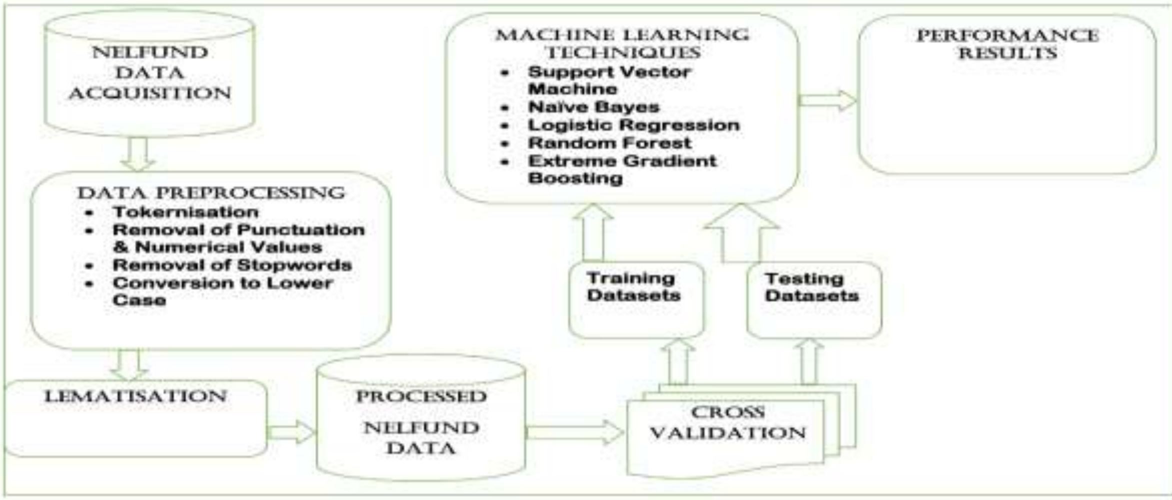

Fig. 1: The Architectural Design of Machine Learning Models for Classifying Students' Opinions on NELFUND

3.3 Data Cleaning

Initially, various preprocessing tasks were executed to clean the dataset, encompassing the elimination of slang, extraneous whitespace, numerals, special characters, punctuation, hashtags, emoticons, emojis, icons, reshared texts, duplicate entries, and orthographic errors. Furthermore, orthographic corrections were implemented, and all uppercase letters were converted to lowercase using Python's regular expressions (regex) library.

3.4 Tokenisation

This procedure entailed dividing the brief texts in the dataset into smaller components known as tokens. Tokenisation was executed at the sentence level utilising the word_tokenize() function from the Python Natural Language Toolkit (NLTK). The NELFUND loan is straightforward to obtain.The phrase is tokenised into the following individual tokens: ['NELFUND', 'loan', 'is', 'very', 'easy', 'to', 'secure'].

3.5 Parts of Speech (POS) Tagging

Part-of-Speech (POS) tagging designates each token with a tuple formatted as (word, tag), signifying its grammatical function. Part-of-speech tagging is crucial for maintaining the contextual significance of words, especially in lemmatisation. This can be executed in Python utilising the pos_tag() function from the Natural Language Toolkit (NLTK). For instance, the sentence "NELFUND loan is exceedingly facile to obtain" can be POS-tagged as follows:

('NELFUND', 'NN')
('loan', 'NN')
('is', 'VB')
('very', 'AVB')
('easy', 'ADJ')
('to', 'IN')
('secure', 'NN')

where, 'NN' = noun, 'VB' = verb, 'AVB" = adverb. 'ADJ' =adjective, 'IN' = preposition

3.6 Stop Word Removal

In English, stop words are prevalent terms that possess negligible semantic significance and frequently do not meaningfully enhance text analysis. During text preprocessing, these terms are generally eliminated to improve model efficiency. The Python Natural Language Toolkit (NLTK) offers a pre-established list of English stop words. In the sentence "NELFUND loan is very easy to secure," the stop words are ['is', 'very', 'to']." Eliminating them produces the sanitised sentence: ["NELFUND", "loan", "easy", "secure"].

3.7 Stemming and Lemmatisation

Stemming and lemmatisation are common methodologies for reducing words to their fundamental or root forms. The primary distinction resides in accuracy: stemming frequently yields truncated roots that may be devoid of meaning, whereas lemmatisation produces meaningful root words derived from dictionary forms. Here is an example:

Text: “He glanced at the dictionary for the meanings of loanee and guarantor”

Stem: [‘glanc’, ‘dictio’, ‘mean’, ‘loane, ‘guaran’]

Lemma: [‘glance’, ‘dictionary’, ‘meaning’, ‘loanee’, ‘guarantor’]

3.8 Sentiment Tagging Module

The sentiment tagging phase was utilised to analyse the sentiments within the preprocessed dataset. The Valence Aware Dictionary and sEntiment Reasoner (VADER), created by Hutto and Gilbert (2014), was employed for this purpose. VADER integrates lexicon-based and rule-based methodologies to categorise text as positive, negative, or neutral, yielding a compound score that encapsulates the overall sentiment. It proficiently manages social media content, encompassing emojis, acronyms, slang, and contractions (Hoti & Ajdari, 2023). Sentiment labelling was conducted at both the lexical and syntactic levels: positive words or sentences received a score of +1, negative ones -1, and neutral statements 0. The overall sentiment score of a sentence was calculated by aggregating the positive and negative sentiment values.

Feature Extraction

Lemmatisation reduces words to their base or canonical forms, which are subsequently utilised for feature extraction. This study utilised CountVectorizer and Term Frequency–Inverse Document Frequency (TF-IDF) methods to extract features from students' opinions regarding NELFUND. These techniques convert textual data into numerical vectors appropriate for training machine learning (ML) models.

TF-IDF and count vectorizer

CountVectorizer produces a document-term matrix that quantifies the frequency of each word within a document. This method is straightforward and efficient for text representation. Conversely, Term Frequency–Inverse Document Frequency (TF-IDF) allocates weights to words according to their frequency in a document and their scarcity throughout the entire corpus. Frequent words are assigned lower weights, whereas uncommon or unique words are allocated higher weights, thereby augmenting the efficacy of TF-IDF for text representation. Research indicates that TF-IDF, which considers both term frequency and uniqueness, offers a more effective representation than CountVectorizer (Azim et al., 2025). Equations 1 and 2 elucidate the computation of TF-IDF as articulated by Azim et al. (2025):

$$TF\ (t,d) = \frac{No.of\ times\ t\ appears\ in\ document\ D}{Total\ number\ of\ terms\ in\ document\ d} \quad (1)$$

$$IDF\ (t,D) = log\left(\frac{Total\ \ document\ in\ the\ corpus\ D}{Documents\ containing\ term\ t\ +1}\right) \quad (2)$$

The TfidfVectorizer from the scikit-learn feature extraction library transforms sentences into a TF-IDF matrix, indicating the relative significance of each word. A polarity function subsequently computes the sentiment score of each phrase by aggregating the TF-IDF values and multiplying them by their initial polarity. The output presents phrase-level polarity scores and TF-IDF values for individual words, indicating their importance and sentiment intensity within the corpus.

Conversely, CountVectorizer allocates weights exclusively according to word frequency, disregarding overall document significance. The resultant polarity scores reflect the frequency and sentiment of words or phrases. The statement "NELFUND loan is very easy to secure" may garner a positive score, while "NELFUND loan is difficult to secure" would yield a negative score. CountVectorizer provides a more straightforward alternative to TF-IDF, emphasising word frequency rather than relative significance within the corpus.

Sentiment Classification using Machine Learning Models

The opinions of students regarding NELFUND operations at Emmanuel Alayande University were evaluated utilising five machine learning classifiers: support vector machines (SVM), Naïve Bayes (NB), logistic regression (LR), random forests (RF), and extreme gradient boosting (XGBoost).

Support Vector Machine (SVM)

Support Vector Machine is another classification technique used. SVM excels at identifying optimal hyperplanes for dividing data into different classes. SVM determines the best hyperplane for differentiating between sentiments after training on the provided dataset. For a given training dataset, consider (xi, yi), where xi is the input feature vector and yi is the class label (- or + for binary classification) (Rakotomamonjy, 2003). Equation 3 from Rakotomamonjy (2003) describes the decision function for classifying a new data point x:

$$f\ (\boldsymbol{x}) = sign\ (w.x\ +\ b) \qquad (3)$$

where,
w: is the weight vector.
x: is the input feature vector.
b: is the bias term.
sign: is the sign function that returns -1if the argument is negative, 0 if it's zero, and +1if it's positive.

The hyperplane is determined by minimising the objective function represented in equation 4 as in Rakotomamonjy (2003):

$$\min_{w,b} \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N} max\ (0, 1\ - y_i\ (\boldsymbol{w}.\boldsymbol{x}_i\ + b)) \qquad (4)$$

where,

$||w||^2$ is the squared norm of the weight vector.
C is a regularisation parameter that controls the trade-off between maximising the margin and minimising the classification error.
The sum term represents the hinge loss, which penalises points that fall on the wrong side of the margin

Multinomial Naïve Bayes

The subsequent classifier is Multinomial Naïve Bayes (MNB). Training entails utilising the dataset to assess the likelihood that each word appears within various sentiment categories. This probabilistic framework is fundamental to Naïve Bayes classifiers and essential for sentiment prediction based on word occurrence. In text-based applications, words are generally transformed into quantifiable integer representations for analysis (Su et al., 2011). The probability computations for sentiment classification utilising MNB are articulated in Equation 5 (Su et al., 2011).

$$P\ (c|d) \propto P\ (c) \prod P(\ t_k\ |c) \qquad (5)$$

$P(t_k|c)$ is the conditional probability that the word appears in a document with class c. In equation (5), $P(t_k|c)$, the likelihood probability of class c is t_k and P(c) is the prior probability. The purpose of class determination is to compare posterior probability results obtained, and then the class with the largest posterior probability is chosen as the predicted result (Su *et al.,* 2011). The prior probability equation is represented in Equation 6 as in Su *et al.* (2011):

$$P\ (c) = \frac{N_c}{N} \qquad (6)$$

N_c is the sum of category c, while N is the sum of all categories. The likelihood probability Equation can be represented in Equation 7 as in Su *et al.* (2011):

$$P\ (t_k\ |c) = \frac{T_{kc}}{\sum t\prime \in V^{T}{}_{ct\prime\prime}} \qquad (7)$$

T_{kc} is the number of occurrences of the word t in the document having class c, and $Tkc\sum t' \in V^{T}{}_{ct''}$ is the total number of occurrences of all words in class c.

Logistic Regression

The logistic function, also known as the sigmoid function, serves as the foundation for probability estimation and provides information about the likelihood of a given sentiment (Peng *et al*., 2002). Consider the logistic regression model as f (X), where X is the input vector. The logistic function is defined using equation 8 as in Peng *et al*. (2002):

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad (8)$$

where,

z is the linear combination of input features and their associated weights and bias term given by equation 9 as in Peng *et al.* (2002):
$Z = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$ (9)

The logistic regression model can be represented using equation 10 as in Peng *ei al.* (2002):
$$f(X) = \sigma(z) = \frac{1}{1+e^{-(w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n)}} \quad (10)$$

where,

w: is the vector of weights
X: is the input vector
σ(.): is the logistic function.
The logistic regression model is trained by finding the weights w that maximise the likelihood of the observed data (Peng *et al.,* 2002).

Random Forest

This study employs the Random Forest (RF) classifier, which creates an ensemble of decision trees and consolidates their predictions to improve overall accuracy. The method's intrinsic diversity and ability to integrate multiple trees render it especially efficient for intricate classification tasks (Evans et al., 2010). Consider the Random Forest algorithm as RF. Given an input vector X, the prediction is made by RF(X), which is the sum of predictions from individual decision trees represented by equation 11 in Evans *et al*. (2010):
$RF(X) = \frac{1}{N}\sum_{i=1}^{N} T_i(X)$ (11)
where,
N is the number of decision trees in the forest.
T_i (X) is the prediction of the i-th decision tree.
For classification tasks, the final prediction is often the class that receives the majority of votes among the individual trees.
For regression tasks, the final prediction is often the average (mean) of the predictions from all trees.

Each tree T_i in the Random Forest is typically constructed using a subset of the training data and a subset of features, promoting diversity among the trees (Evans *et al.,* 2010).

Extreme Gradient Boosting

Finally, the XGBoost classification model was utilised in this study to analyse students' sentiments concerning NELFUND. XGBoost is an exceptionally efficient variant of gradient boosting, operating up to ten times faster than traditional implementations, thus being extensively utilised for classification and regression tasks, including seller prediction, customer behaviour analysis, ad prediction, and web text prediction (Chen & Guestrin, 2016).

Boosting is an ensemble method that incrementally incorporates models to rectify errors made by preceding models and persists until no additional enhancement is realised. XGBoost employs a tree ensemble model consisting of classification and regression trees, amalgamating predictions from various trees into a singular output (Sholahuddin & Abdullah, 2021). Each predictor is trained using the residual errors from the previous model. Following dataset input, a preliminary model is established, and the initial prediction values along with residual errors are calculated as outlined in Equations 12 and 13 (Kurniawanda & Tobing, 2022).
$h_0(X) = mean(Y)$ (12)

$\hat{Y} = Y - h_0(x)$ (13)

In this equation, $h_0(x)$ represents the first model's initial predictive value, and Y represents its residual error. The residual errors from the initial model are subsequently used to develop the second model, which produces its predictive values. The residuals from both the initial and second models are subsequently employed to construct the third model, which generates its respective predictions. This iterative process persists for the number of repetitions defined by the n_estimators parameter. Gradient Boosting, the foundational algorithm, is engineered to iteratively reduce prediction errors by learning from the residuals of prior models (Handayani et al., 2017).

Like conventional boosting, XGBoost builds an ensemble of decision trees, with each successive model reliant on the performance of its predecessor. The initial model generally exhibits limited predictive power, and the algorithm incrementally adjusts the weights across successive models to improve prediction accuracy. The forecasts from all individual models are subsequently consolidated and utilised in Equation 14 to minimise the objective function (Syukron et al., 2020).

4. Results and Discussion

This study assesses the efficacy of five machine learning models: Support Vector Machine, Naïve Bayes, Logistic Regression, Random Forest, and XGBoost, in conjunction with three ensemble fusion techniques: Majority Voting, Weighted Average Voting, and Average Probability Voting. Sentiment analysis was performed on a preprocessed dataset comprising two attributes: NELFUND transparency and benefits. Preprocessing encompassed stopword elimination, tokenisation, and lemmatisation. The evaluation of model performance was conducted utilising accuracy, precision, recall, and F1-score metrics for each sentiment category.

4.1 Analysis of Student Opinion on NELFUND Transparency and Benefits through the use of Polarity Scores in Machine Learning Model Performance Evaluation

The findings demonstrate that all models exhibited superior performance with the preprocessed datasets for both attributes. In terms of NELFUND transparency, the Random Forest, XGBoost, and Support Vector Machine consistently surpassed other models across all evaluation metrics, with XGBoost attaining the highest F1-score and recall (refer to Table 1). Conversely, in the NELFUND benefits dataset, the Random Forest and Support Vector Machine outperformed all other metrics. Both Logistic Regression and XGBoost outperformed Naïve Bayes in most metrics, with the exception of XGBoost's accuracy. XGBoost surpassed Logistic Regression in F1-score and recall, while Logistic Regression attained superior accuracy and precision (refer to Table 2).

Table 1: Analysis of Student Opinion on NELFUND Transparency through the use of Polarity Scores in Machine Learning Model Performance Evaluation

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

Table 2: Analysis of Student Opinion on NELFUND Benefits through the use of Polarity Scores in Machine Learning Model Performance Evaluation

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

4.2 Graphical Analysis of Student Opinion on NELFUND Transparency and Benefits through the Use of Polarity Scores in Machine Learning Model Performance Evaluation

Figure 2 illustrates a graphical assessment of the effectiveness of machine learning models, indicating that the majority of students' perceptions regarding NELFUND's transparency are categorised as favourable. In terms of F1-score and recall, Extreme Gradient Boosting, followed by Support Vector Machine and Random Forest, detected a greater number of negative opinions compared to Naïve Bayes and Logistic Regression. Likewise, precision metrics demonstrate that Support Vector Machine, Random Forest, and Extreme Gradient Boosting identified a greater number of negative sentiments compared to Naïve Bayes and Logistic Regression.

Figure 3 depicts the categorisation of students' perceptions regarding the advantages of NELFUND. The majority of opinions were categorised as positive across all models. Regarding F1-score, Support Vector Machine and Random Forest, succeeded by Extreme Gradient Boosting and Logistic Regression, detected a greater number of negative sentiments than Naïve Bayes. Recall and Precision analyses indicate that these four models identified a greater number of negative opinions than Naïve Bayes.

5. Discussion

The main aim of this study was to create machine learning models for the sentiment analysis of students' perceptions regarding NELFUND. In the analysis of loan disbursement transparency and benefits datasets, Random Forest, Support Vector Machine, and Extreme Gradient Boosting demonstrated superior performance compared to Logistic Regression and Naïve Bayes classifiers, corroborating the findings of Nandi et al. (2025) and Muniappan and Subramanian (2025). Conventional models, including Naïve Bayes, Random Forest, and Logistic Regression, frequently encounter overfitting or restricted generalisability. The machine learning techniques utilised in this study exhibited enhanced performance relative to previous methods and further advanced research by affirming their efficacy in analysing students' sentiments

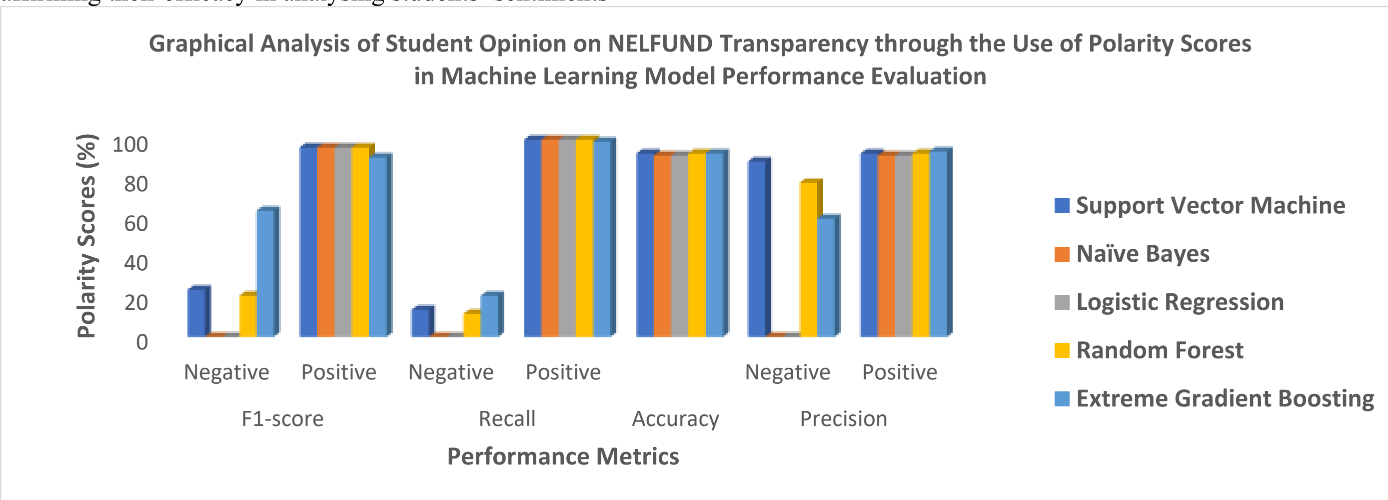

Fig. 2: *Graphical Analysis of Student Opinion on NELFUND Transparency through the use of Polarity Scores in Machine Learning Model Performance Evaluation*

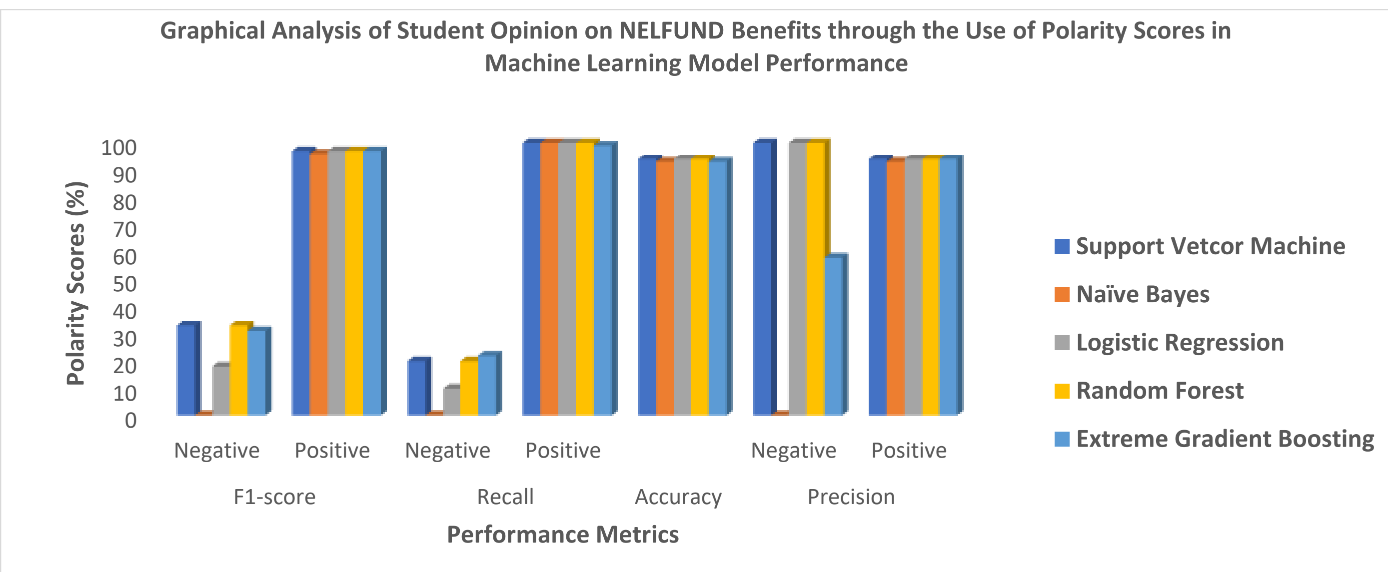

Fig. 3: *Graphical Analysis of Student Opinion on NELFUND Benefits through the use of Polarity Scores in Machine Learning Model Performance Evaluation regarding NELFUND disbursement.*

These findings underscore the capability of sophisticated machine learning models to improve accuracy and robustness, providing significant insights for both scholarly research and practical applications in assessing students' views on NELFUND.

6. Conclusion and Future Work

This study examined sentiment analysis regarding students' perspectives on NELFUND disbursement. Five machine learning models were created utilising a dataset of brief texts gathered from the WhatsApp group of students at Emmanuel Alayande University of Education, Oyo, during the initial semester of the 2024/2025 academic year. The dataset included two attributes: the transparency of NELFUND disbursement and its advantages. A thorough preprocessing pipeline, encompassing data cleaning, tokenisation, part-of-speech tagging, stopword elimination, and stemming and lemmatisation, was implemented. Feature extraction was conducted utilising TF-IDF and CountVectorizer.

The models were executed utilising Python and the Pandas library, facilitating the efficient management of extensive datasets. The models demonstrated efficacy in classifying positive and negative sentiments, attaining an accuracy of 92–94%. Subsequent research will augment datasets to encompass students from additional universities, integrate multilingual and cross-cultural data, investigate sophisticated ensemble methodologies such as boosting and stacking, deploy deep learning frameworks including LSTM and Transformers, employ advanced hyperparameter optimisation techniques such as Bayesian Optimisation and Aquila Optimiser, and leverage embedding methods like Word2Vec, GloVe, and BERT for enhanced contextual text representation.

References

Abellán, J. and Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications*, 73, 1–10.

Ainapure, B. S., Pise, R. N., Reddy, P., Appasani, B., Srinivasulu, A., Khan, M. S. and Bizon, N. (2023). Sentiment analysis of COVID-19 tweets using deep learning and lexicon-based approaches. *Sustainability*, 15, 1-16.

Ala'raj, M., and Abbod, M. F. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications*, 64, 36–55.

Anam, M. K., Lestari, T. P., Yenni, H., Nasution, T. and 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.

Azim, K., Tahir, A., Shahroz, M., Karamti, H., Vazquez, A. A., Vistorte, A. R. and Ashraf, I. (2025). Ensemble stacked model for enhanced identification of sentiments from IMDB Reviews. *Scientific Reports*, 15(13405), 1-18.

Başarslan, M. S. and Kayaalp, F. (2023). Sentiment analysis with ensemble and machine learning methods in multi-domain datasets, *Turkish Journal of Engineering*, 7(2), 141-148.

Birjali, M., Kasri, M. and Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-based Systems*, 226, 1-20.

Chatterjee, S. and Byun, Y. C. (2022). Voting ensemble approach for enhancing alzheimer's disease classification. *Sensors*, 22(19), 1-14.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 16, 785-794.

Coaquira, Y. M. and Villanueva, E. (2024). A review on text sentiment analysis with machine learning and deep learning techniques. *IEEE Access,* 12, 193115-193130.

Dahiya, A. (2019). A study on data science with python in comparison to other languages. *Indian Journal of Applied Research*, 9, 60-61.

Elhadidi, E. A., Salah, A., Abdellah, M. and 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.

Evans, J. S., Murphy, M. A., Holden, Z. A. and Cushman, S. A. (2010). Modeling species distribution and change using random forest. *Cham: Springe*r; 2010.

Eyvazi-AbdolJabbar, S., Kim, S., Feizi-Derakhshi, M., Farhadi, Z. and 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.

Hambali, M. A., Oladele, T. O., Adewole, K. S., Sangaiah, A. K. and Gao, W. (2022).

Feature selection and computational optimization in high-dimensional microarray cancer datasets via InfoGain-modified bat algorithm. *Multimedia Tools and Applications*, 81(12), 36505–36549

Harimoorthy, K. and Thangavelu, M. (2021). Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. *Journal of Ambient Intelligence and Humanized Computing,* 12(5), 3715–3723.

Hoti, M. T. and 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.

Handayani, A., Jamal, A. and Septiandri, A. A. (2017). EvaluasiTiga Jeis Algoritme Berbasis Pembelajaran Mesin untuk Klasifikasi Jenis Tumor Payudara. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*

Hutto, C.J. and 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).

Ismet, H. T., Mustaqim, T. and Purwitasari, D. (2022). Aspect based sentiment analysis of product review using memory network. *Scientific Journal of Informatics, 9(1),* 73-75.

Jung, J., Petkanic, P., Nan, D. and Kim, J. H. (2020). When a girl awakened the world: A user and social message analysis of Greta Thunberg. *Sustainability*, 12, 1.15.

Kamath, R. S., Kaushik, M. R. and Ramakrishna, M. (2025). Opinion mining using ensemble model for restaurant feedback analysis. *Discover Computing*, 28125), 1-27.

Khan, M. Y. and 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.

Krishna, S. and Angeline P. G. (2022). Sentiment analysis using ensemble machine learning techniques. *International Journal of Communication Networks and Information Security*, 14(3), 490-506.

Kumar, V. (2020). Evaluation of computationally intelligent techniques for breast cancer diagnosis. *Neural Computing and Applications,* 33 (4), 1-15.

Kurniawanda, M. R. and Tobing, F. A. T. (2022). Analysis sentiment cyberbullying in instagram comments with XGBoost method. International *Journal of New Media Technology (IJNMT)*, 9(1), 28-34.

Lavado, J. L., Cantador, I., Cort´esCediel, M. E.and Fern´andez, M. (2022). Automatic intent-based classification of citizen-to-government tweets. In: Proceedings of Ongoing Research, 2021, pp. 47.

Mika, V., Mäntylä, D. and 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.

Muniappan, M. and 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. Available online at:https://www.mdpi.com/1911-8074/18/4/197/pdf?version=1743772184. Accessed on May 20, 2025

Nandi, S., Subrahmanyan, P. and Singh, S. (2025). Fuzzy-based ensemble learning for sentiment analysis in social media data. *Global Journal of Engineering Innovations & Interdisciplinary Research*, 5(1): 1-6.

Nandwani, P. and Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(8): 1-21.

Nazareth, N. and Reddy, Y. V. R. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, 219, 1-15.

Oyekunle, R. A. and 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.

Pandey, G., Gupta, N. K., Paul, B. and 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.

Patil, P., Kale, G., Bivalkar, N. and Kolhatkar, A. (2023). Comparative analysis of weighted ensemble and majority voting algorithms for intrusion detection in open stack cloud environments. *International Journal of Advanced Computer Science and Applications (IJACSA),*14 (12), 741-747.

Peng, C. Y. J., So, T. S. H., Stage, F. K. and St. John, E. P. (2002). The use and interpretation of logistic regression in higher education journals 1988–1999. *Research in Higher Education*, 43, 259–293.

Rakotomamonjy, A. (2003). Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, 3,1357–1370.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2(3),* 1-14.

Saleena, A. N. (2018). An ensemble classification system for twitter sentiment analysis. *Procedia Computer Science*, 132, 937-946.

Saputra, M. F. A., Widiyaningtyas, T. and Wibawa, A. P. (2018). Illiteracy classification using k means naïve bayes algorithm. *International Journal on Informatics Visualization,* 2(3), 153-158.

Sholahuddin, A. and Abdullah, A. S. (2021). Extreme gradient boosting (XGBoost) method in making forecasting application and analysis of USD exchange rates against rupiah. *Journal of Physics: Conference Series,* 1722(1), 1-16.

Sivri, M. S. (2024). Combining sentiment analysis models using stacking ensemble learning techniques on BIST30 stocks. *Journal of Hendese Technical Sciences and Engineering*, 1(2), 91-97.

Soori, M., Arezoo, B. and Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 3, 54.

Su, J., Shirab, J. S. and Matwin, S. (2011). Large scale text classification using semi-supervised multinomial naive bayes. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, 97–104.

Syukron, M., Santoso, R. and Widiharih, T. (2020). Perbandingan metode smote random forest dan smote xgboost untukklasifikasi tingkat penyakit hepatitis c pada imbalance class data. *Jurnal Gaussian,* 9(3), 227-36

Tan, K. L., Lee, C. P. and Lim, K M. (2016). Sentiment analysis with ensemble hybrid deep learning model. *IEEE Access*, 4, 1-12.

Van Fc, L. L., Anam, M. K., Firdaus, M. B., Yunefri, Y. and Rahmi, N. A. (2024). Enhancing machine learning model performance in addressing class imbalance. *COGITO Smart Journal,* 10(1), 1-13.

Varshney, C. J., Sharma, A. and Yadav, P. (2020). Sentiment analysis using ensemble classification technique. IEEE Students' Conference on Engineering & Systems (SCES), July 10-12, 2020, Prayagraj, India

Vinora, A., Suryakumar. L. S. S., Harivishnu, T. S., Jaspin, K., Selvaraj, A. K. and 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.

Xiao, Y., Li, C., Th¨urer, M., Liu, Y. and Qu, T. (2022). Towards lean automation: Fine-grained sentiment analysis for customer value identification. *Computers & Industrial Engineering*, 169(108186), 1-24.

Wang, G. and 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.

Zhang, L., Wang, S. and Liu, B. (2018). Deep learning for sentiment analysis: A Survey. WIREs *Data Mining Knowledge. Discovery*, 8(4), 1-15.