Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10, 1149-1180
2025
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v9i10.10606

Prediction of students' feedback on faculty performance using stacking ensemble method: Machine learning algorithm

Mayowa Samuel Alade[1], Samuel Olujimi Adejumo[2], Olufemi Deborah Ninan[3], Abidemi Emmanuel Adeniyi[4*], Emeka Ogbuju[5], Oluwasegun Julius Aroba[6], Manduth Ramchander[7], Timothy T. Adeliyi[8]

[1,2]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria.
[3]Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.
[4]Department of Computer Science, Bowen University, Iwo, Nigeria; abidemi.adeniyi@bowen.edu.ng (A.E.A.).
[4]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.
[5]Department of Computer Science, MIVA Open University, Abuja, Nigeria.
[6]Centre for Ecological Intelligence Department, Faculty of Engineering & Built Environment, University of Johannesburg, South Africa.
[6,7]Operations and Quality Department, Faculty of Management Science, Durban University of Technology, KwaZulu-Natal 4001, South Africa.
[8]Department of Informatics, University of Pretoria, Pretoria 0083, South Africa.

Abstract: Students' feedback is fundamental for the growth and development of higher education institutions. Feedback and comments from students are an extremely useful and valuable source of information that reflects the quality of education or educational services received by students. However, the effective management of qualitative opinions of students is a challenge. Undeniably, many organisations deal with quantitative feedback effectively, while qualitative feedback is either manually processed or ignored. This paper proposes an opinion mining or sentiment analysis system using a stacking ensemble-based method. Furthermore, four base models, comprising various base-level classifiers, including logistic regression (LR), support vector machine (SVM), multilayer perceptron (MLP), and Naïve Bayes (NB), predict the orientations as positive, negative, or neutral. The system has been evaluated using performance metrics such as accuracy, precision, recall and F1-measure; and compared with similar models. Experimental results show that the four base-independent algorithms yield the following classification accuracies: LR algorithm, 79.05%; SVM, 81.76%; MLP, 50.68%; and Multinomial Naïve Bayes, 50.68%. These forecasts can be used by Nigerian Public universities and higher education institutions to improve the educational system and assist students to receive a better and quality education.

Keywords: *Machine learning, Multilayer perceptron, Opinion mining, Orientation, Qualitative feedback, Student feedback.*

1. Introduction

Education is a key indicator of development and a key contributor to the well-being of humans. The United Nations Sustainable Development Goals (SDGs), established in 2015, outline specific guidelines for enhancing educational standards and protecting children's welfare. They emphasize that all individuals should have access to high-quality education and opportunities for lifelong learning [1]. Moreover, one of the essential components of the educational system (teaching and learning) is students' academic success as well as the quality of education, which indeed is the fourth of the UN's 17 Sustainable Development Goals (SDGs). Therefore, quality is without a doubt a great concern for the global community's governments, businesses, civil society, higher education institutions (HEIs), and government agencies [2-4].

History: Received: 4 August 2025; Revised: 1 October 2025; Accepted: 6 October 2025; Published: 16 October 2025
* Correspondence: abidemi.adeniyi@bowen.edu.ng

Over the last four decades, a much-debated and researched topic has been the role of student feedback and assessment in improving learning and teaching quality. The relevance of assessment and feedback has increased.

Significantly, these elements are now recognized as crucial to student happiness, which in turn promotes increased motivation for academic success and ultimately academic advancement [5].

Feedback is one of the most important interventions in learning and education [6]. Feedback can be conceptualised as the crucial link between teaching and learning [7]. In other words, Lawal et al. [8] defined feedback in education as a type of learning that is usually conducted through assessment. But Harvey [9] in his definition, describes feedback as "the expressed opinions of students about the service received by the students." Moreover, feedback is an essential component of assessment for learning, which, if used appropriately, can support students' learning and lead to substantial learning gains. Likewise, if done as required, teachers and learners work in partnership to aid improvement. This, in turn, implies that assessment in the teaching process has come a long way, and its effect reflects on both learners and teachers. Feedback is a type of embedded assessment.

In education, assessments are used to collect information about student learning and achievement. They can be formative or summative, quantitative or qualitative [10]. Feedback falls under the formative type of assessment. It can be a student giving feedback about a teacher or vice versa. The latter type of assessment (summative) is concerned with summarizing students' achievement status, while the former (formative) is referred to as a process rather than a test to continuously monitor, provide feedback, and respond to student learning progress [11]. In other words, formative assessment is concerned with how judgment is obtained from the quality of student responses, which can be used to sharpen their competence and generate results in performance to improve and accelerate learning.

Moreover, feedback is widely considered for several benefits. The integration of student feedback into higher educational institutions (HEI) and their quality assurance processes is becoming increasingly crucial. The quality of courses, as well as the practices, techniques, and resources used in the classroom and laboratory, can be improved with the help of student feedback. Additionally, student comments can boost an institution's reputation in the competitive global education market. Because of this, Nair et al. [12] emphasized the importance of monitoring and reflecting upon the full spectrum of student feedback to devise and implement the best quality assurance mechanism in engineering education. In the same vein, Alade and Nwankpa [13] stated that educational institutions have long worked to raise the standards of both student education and education in general. However, most existing feedback systems on learning, according to Foltz and Rosenstein [14], are focused on improving student writing skills rather than the knowledge development process.

In the same vein, student feedback allows teachers to understand students' learning behavior and improve education. It enables students to highlight issues that may differ from their lecturer's view. This occurs when students do not understand some of the lectures or instances, or when the instruction is too fast or too slow. Feedback is usually collected, and the end unit is more beneficial in real time [15]. This is an entity about past behaviour from the statement to analyze the future and current behaviour to achieve the desired result. It allows us to follow new knowledge and avoid repeating past mistakes. In education, feedback plays an important role as many people want to know whether their opinion must be serious or the most important in the decision-making process for higher education quality. Feedback is the process of helping and assessing an organization on a monitor and standardizing the overall working environment [16].

In addition, feedback is an essential part of communication. Feedback in education helps both students and teachers strengthen the learning process and can help students improve their chances of success [17]. Feedback is a key element in formative assessment. It is an essential part of education and training programmes. Feedback helps learners maximise their potential at different stages of training, raises their awareness of strengths and areas for improvement, and identifies actions to be taken to improve performance.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

Student feedback is essential for assessing and analyzing learning management systems, instruction, pedagogical practices, and courses in the field of education [18]. Thus, gathering comments after each semester, educational institutions use student feedback questionnaires to capture their thoughts on the courses they have taken that semester [19]. The feedback includes both quantitative and qualitative information, such as student demographics, course information, ratings, and comments. Quantitative data can offer statistical insights into student feedback on the courses, while qualitative data analysis can reveal students' intentions.

Technological advancements through the implementation of Artificial Intelligence (AI) and natural language processing (NLP) methodologies have revolutionized industries such as healthcare and education [20, 21]. Moreover, the improvement in technology has allowed students to explore new fields. Subsequently, it has also helped to keep track of performance and improve the abilities of faculty, with the opinions of students being helpful. Therefore, finding out the subjective meaning of opinions is the major task. Sentiment analysis can be one way to do it. Sentiment analysis is contextual mining of text that identifies and extracts subjective information in source material and helps to understand social sentiment or opinion when monitoring an organization. NLP has to do with the building of computational algorithms to automatically analyze and represent human language.

Students have not been taught to provide feedback on their learning over the years, particularly in developing countries, and the few that do so are unfamiliar with the process; some of the opinions provided have no bearing on their academic performance. However, efforts have been undertaken to understand and improve students' learning experiences in various ways. A useful method of evaluating instruction is to ask students about their experiences as students in higher educational institutions. There are numerous methods for gathering student input, such as small group instructional diagnosis (SGIDS), surveys, web questions, and open-ended feedback forms like FACETS [10]. For academic bodies to achieve the intended results, it is not enough to just gather student opinions; instead, the feedback must be carefully examined. We can better understand student feedback on their educational experiences by dissecting and analyzing the remarks made by students. Any institution can use online or offline feedback analysis methods to collect student feedback. As a result, participating in it can be an effective technique to enrich and enhance students' knowledge. One of the most significant influences on learning and success is feedback, yet this influence can also be very detrimental.

In today's digital world, the environment of the contemporary educational system is constantly enhanced by the vast amount of data produced and shared each day across different platforms, such as social media and learning management systems, much of which contains significant and useful information as well as comments [3, 22]. These vast amounts of data are what data analysts leverage to access views and opinions on various topics; hence, they predict business and social outcomes such as stock returns, product sales, and the political outcomes of elections [23-25]. Therefore, finding and extracting the subjective meaning from opinions user-generated content and opinions from the enormous volume of data is the major task that opinion mining and sentiment analysis can provide.

Sentiment analysis (SA), also referred to as opinion mining, subjectivity analysis, and appraisal extraction, is described as a process that automates the mining of opinions, views, attitudes, and emotions from text, tweets, speech, and database sources through NLP [26]. Sentiment analysis uses computer methods to examine how individuals feel and think about specific issues expressed in text data. The idea behind sentiment analysis is to analyze a collection of text data to understand the opinion or sentiment conveyed. This is typically achieved by determining the sentiment within the text and assigning a positive, negative, or neutral value, known as polarity. The overall sentiment can often be identified by the polarity's sign and classified accordingly. Therefore, sentiment analysis has a significant impact on texts containing emotions or dispositions of any kind.

Research has demonstrated that using student feedback to assess teaching and learning improves the provision of high-quality instruction. Moreover, a variety of manual techniques, including audience response, questionnaires, and polling, have been employed to evaluate the opinions that students have supplied. However, the existing approach, which involves manually assessing and handling qualitative

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

as well as formative remarks or opinions of thousands of students, is ineffective and presents a difficult issue in the education arena (HEI). This is because it could result in arbitrary and inconsistent interpretations, thereby decreasing the accuracy and consistency of the findings. Similarly, the manual approach is vulnerable to human error, primarily due to tiredness from performing repetitive tasks manually. Therefore, the need for a computational model that analyzes the opinions given by students and polarizes the results for decision-making emerges in the context of mining and evaluating university students' informal remarks.

There have been various approaches that have been employed for solving students' feedback problems using opinion mining [27, 28]. These include methodologies such as Naïve Bayes (NB). For example, Amusa et al. [29] and Alade and Nwankpa [13] addressed the challenge of obtaining feedback from education using sentiment analysis. However, other methodologies such as K-nearest neighbor (KNN), support vector machine (SVM), decision tree, regression, and others have been employed to predict student performance and their feedback. However, innovative and economical methods, according to Ngwira et al. [30], are required by researchers to effectively assess students' reviews and process them efficiently.

The main objective of the study is to develop a stacking ensemble model for students' feedback prediction with multiple predictors, to examine the existing research related to the performance of students' feedback models, and to test and validate the proposed model using various performance metrics. Specifically, four machine learning models, Linear Regression (LR), Support Vector Machine (SVM), Multilayer Perceptron neural network (MLP), and Naïve Bayes (NB) were used as base learners due to their high popularity and good performance in previous studies. A dataset from a higher educational institution was used, partitioned into training and test data, and the model was implemented using the Python programming language.

The rest of the paper is organized as follows: Section 2 presents the literature review and related work in the field of sentence-based sentiment analysis. In Section 3, the paper presents the proposed model and discusses a detailed methodology for accomplishing the task of mining students' feedback, the evaluation metrics, and sentiment prediction. Experiments, results, and discussion are presented in Section 4, including the comparison of model performances, the examination of the performance at different time scales, and the discussion of prediction results. Section 5 concludes the paper with future directions.

1.1. Background

1.1.1. Machine Learning

This is the field of study that analyzes or examines the use of computational algorithms and converts, changes, or transforms empirical data into usable models [31]. Therefore, it is considered a multidisciplinary and interdisciplinary field of artificial intelligence, which includes the following domains: mathematics, data mining, natural language processing, computer science, and deep learning [32]. ML techniques can be either supervised, unsupervised, or semi-supervised, and they use language features and well-known ML algorithms to categorize opinions into positive or negative sentiments. The lexicon-based method is a set of words or phrases that express information about positive or negative polarity. Compared to the ML methodology, the lexicon-based method is simpler to understand and implement. However, it is restricted by the requirement of engaging people in the text analysis process, as shown in Figure 1. The hybrid method includes both lexicon-based and machine-learning techniques.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

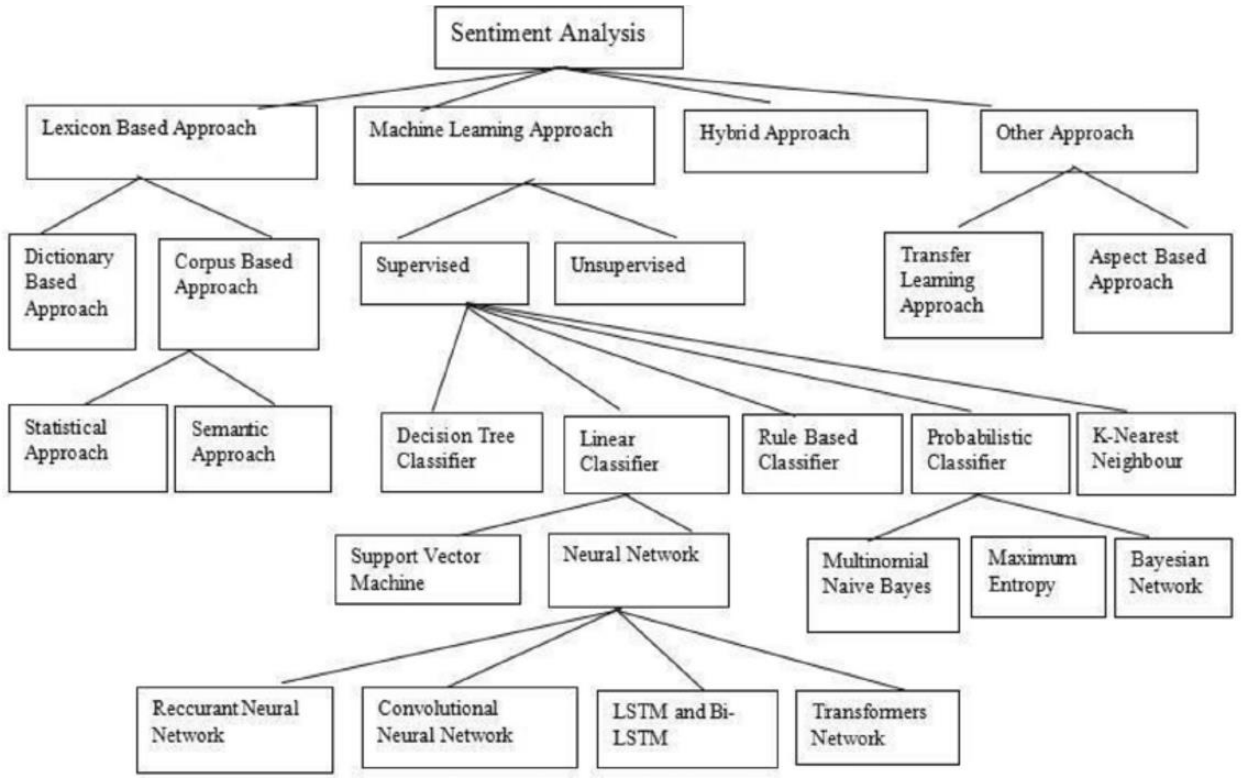


Figure 1.
Classification of Sentiment Analysis methods.
Source: Medhat, et al. [33] and Bhavitha, et al. [34].

1.1.2. Supervised Learning

The supervised ML method is used for making predictions of labelled data, according to Qureshi et al. [35], where the model is examined and trained depending on the necessary attributes or properties, and then tested using unlabeled data. As a result, the model learns during training and uses that information during the test stage with actual data [36]. The two methods of supervised learning that can be applied, depending on the available data, are classification (intended for discrete binary data) and regression (for continuous data).

In addition, the classification approach uses the category or class objective value to predict similar data. Hence, a classification algorithm balances input data so that the output is accurate, making it a necessary method for all types of data classification, including images and data mining [37]. There are numerous methods for classifying data in ML, such as KNN, Bayesian networks, decision trees, gradient boosting, neural networks, logistic regression, random forests, and many more for student performance and feedback predictions [38].

In this work, a decision-making process is presented where multiple opinions are weighed before making a final choice. Here, the results from various trained classifiers are integrated or combined to reach the ultimate conclusion. These approaches include creating several classifiers and then combining their outputs using joining criteria. This combination of classifiers' outputs greatly enhances the model's performance. In the experiment, four algorithms, namely Artificial Neural Network (ANN), Logistic Regression (LR), Support Vector Machine (SVM), and Naïve Bayes (NB), are selected as base-level classifiers.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

1.1.3. Artificial Neural Network: Multi-layer Perceptron (MLP)

ANN has emerged as one of the most popular ML techniques as a nonlinear fitting method [39] because of its benefits of simple training, flexible structure, and variable training parameters, extreme learning machines (ELM), backpropagation neural networks (BPNNs), general regression neural networks (GRNNs), and other ANN techniques, such as MLP, are now available due to algorithmic breakthroughs. MLP, being a supervised machine learning technique, is a typical ANN design used in this study. MLP is a feedforward network made up of an input layer, one or more hidden layers, and an output layer. External data is received by the input layer, and the output layer generates or produces the finished product. Between the input and output layers are neurons in the hidden layers, which offer nonlinearity functions. By using more hidden neurons or layers, more complicated or complex issues or problems can be solved as well, and adequate predictions can be achieved. This is because the hidden layer or layers in between are highly networked with neurons, which are computing units that process information linked together by weights. Moreover, the neural network, in terms of MLP, has been developed for the current issue because the target variable contains more than two classes and is a classification problem. Each neuron uses an activation function (the activation function for the hidden layer is hyperbolic tangent, and for the output layer, it is softmax) that processes a linear combination of inputs to yield outcomes in a non-linear transformation, with a displayed accuracy of about 81%. The mathematical description of this procedure is given in (1). MLP uses a set of attributes like x and a target y, which can learn a non-linear function estimator for classification.

$$f(,): R_m A \rightarrow R_0 \tag{1}$$

1.1.4. Support Vector Machine

SVM is a type of machine learning algorithm that is widely used for the collection of linear predictor functions that have been used to solve problems of function determination. It is a supervised machine-learning technique that can be used for both regression and classification applications, including multi-class classification. The key aspect of this method is explaining how independent and dependent variables relate to one another [40]. In this study, the classifier uses a different loss function from logistic regression, with the kernel mathematical machine function being utilized for data transformation in the SVM model. However, the optimal separating hyperplane is that which the SVM model seeks to identify. After the individual SVM datasets were converted to a high-dimensional feature space, a hyperplane was produced using the training datasets. The mathematical description is given in Equation 2.

$$f(x_i) = sign\,(w^T x_i + b) \tag{2}$$

While the functional margin is:

$$Yi = (w^T x_i + b) \tag{3}$$

Where w is the decision hyperplane normal vector, T is the data point, and i is the class of data (which is +1 or -1). Since the study deals with linear data, that is, it is not data that can be clustered into different groups, this leads to the use of a linear kernel in such an SVM learning algorithm. Hence, it provides high accuracy and good theoretical overfitting if the appropriate kernel is applied. However, the SVM is not sensitive to noise, where a small number of mislabelled examples can reduce the system's performance.

1.1.5. Logistic Regression (LR)

LR is one of the most used machine learning techniques for binary classification. It is also a multivariate data analytic model (linear and binary) that predicts the presence or absence of an attribute

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

or result based on the values of a collection of predictor variables. However, the LR model is robust to noise, and overfitting can be avoided in the feature selection process. Therefore, in this work, LR is regarded as a likelihood relevant to this investigation, as the variables do not have a normal distribution. In addition, the logistic function is given by the equation.

1.1.6. Naïve Bayes (NB)

This is a probabilistic classifier that works on Bayes' theorem of probability to predict the class of unknown datasets. In the Naive Bayes (NB) classifier, the presence of one feature in a class is unrelated to any other feature in the dataset. As a result, it is simple to implement with many counts. The NB algorithm has three different varieties, and it is worth noting that the variations of this algorithm yield different results. Although there are three variations of NB: (1) Multinomial Naive Bayes (MNB), (2) Bernoulli Naive Bayes (BNB), and (3) Gaussian Naive Bayes, the multinomial MNB is applied in this work because of the multiple occurrences of words or discrete frequency counts, which are essential in a classification problem. The model of NB is described as follows:

$$P(labels|features) = \frac{P(label) * P(features|label)}{P(features)} \tag{4}$$

$$P(Y|X_1, \dots X_n) = \frac{P(Y) * P(X_1, \dots , X_n|Y)}{P(X_1, \dots, X_n)} \tag{5}$$

Where $P(label)$ is the prior probability of the label occurring;
$P(features|label)$ is the prior probability of a given feature being classified as that label;
$P(features)$ is the prior probability of a given feature set occurring
$P(labels|features)$is the probability that the given features should have that label.

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

Where $P(c)$ is the probability of the class
N_c is the total count of a particular class in the training set
N is the total count of classes in the training set
Hence, Conditional independence is evaluated as follows

$$P(W|C)\frac{count(w,c) + 1}{count(c) + |V|} \tag{7}$$

Here, $P(W|C)$
W is the word attribute, and c is the class
$count\ w, c$ *is the total count of words attributes occuring in the class*
+1 is the Laplace function
Count c is the total number of word attributes in a particular class
|V| is the vocabulary

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

1.1.7. Ensemble Method

An ensemble is a classification technique that reuses one or more classification algorithms for robustness with multiple models or the same method for different parts of the data. In other words, an ensemble is a type of hybrid learning system in which multiple (base classifiers) analytics are intelligently coupled or combined to generate better results than single analytics can offer (more accurate, more robust, etc.). Even though certain ensemble classifiers may perform poorly, overall prediction quality is guaranteed. These methods have been an exceedingly powerful expansion of data mining and machine learning techniques, which synthesize manifold classifiers into a single, more accurate model. Furthermore, an ensemble model is built with two major goals while blending predictions from multiple models. The foremost intent is to augment the prediction accuracy of a model generated by hybridizing multiple classifiers over a single base classifier. Secondly, to minimize the overfitting problem in base classifiers and subsequently boost the classifier's stability and prediction accuracy. The fundamental standard is that an ensemble method can choose a set of instances from a wide spectrum of hypothesis sets and blend their predictions into a single prediction [41]. However, there are various ensemble techniques or meta-algorithms such as stacking, bagging, voting, and boosting. Rahman and Tasnim [42] described the base classifiers as individual or single classifiers, which are exploited or employed to construct the meta-classifiers. Yet, ensemble classifiers have been found more effective in the growth or development of educational data mining and machine learning models, particularly in evaluating student academic performance, thereby producing significant results than individual classifiers [43]. In this study, however, the stacking ensemble technique is used.

1.1.8. Stacking Ensemble

The stacking ensemble method is a variant of the voting approach in which various independent model types are built using the same training data. However, the independent models are all combined using a different machine learning model for three-group classification problems. It is a machine learning algorithm as a whole that is used to increase accuracy by combining predictions from multiple base-level learners into one general prediction using a higher-level base meta-learner algorithm [44]. It is well known that ensemble techniques can lower estimator variance, which enhances the accuracy of predictions. To increase the accuracy of averaging, various randomization or data augmentation techniques are frequently used to encourage a wide variety of predictions. The ensemble prediction is denoted in equation 8.

$$h^t : j^t . svm^t + nn^t + lr^t + nb^t \tag{8}$$

Where h^t is the ensemble predictor used to make a prediction of z^t based on the weighted results of SVM, NN, LR, and NB. The weight vector j^t is associated with the support vector machine result set, the weight vector w is associated with the neural network result, not^t and the weight vector not^t is associated with the naïve Bayes result r^t.

2. Literature Review

2.1. Sentiment Analysis in the Education Domain

SA is a task that concentrates on identifying polarity and feelings about a subject or an event. Finding people's opinions, identifying the sentiments they convey, and categorising them as positive, negative, or neutral are the general goals of sentiment analysis. Additionally, sentiment analysis systems search, retrieve, and synthesise knowledge and opinions from large amounts of textual data using natural language processing (NLP) and machine learning (ML) approaches [45]. Massive open online courses (MOOCs) in particular have drawn a lot of attention to research on sentiment analysis,

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

which is the process of identifying sentiment words and phrases that represent emotions [46, 47]. Sentiment analysis is a method that can be used to get the user's most crucial information.

The most essential information for the user can be extracted from plain text data using the approach of sentiment analysis. This has sparked an increase in research in the areas of opinion mining and sentiment analysis to create algorithms that can automatically analyze text passages or user evaluations and extract the data that is most pertinent to the user. Additionally, there has been a recent increase in studies in the fields of machine learning and NLP, particularly in education. These methods can be used to mine insights from evaluations. However, to guarantee consistency, it is important to create a roadmap for analysis that can be credible, reliable, and accurate.

There has recently been an increase in studies in the fields of machine learning and NLP, particularly in the area of education. These methods can be used to mine insights from evaluations. However, to guarantee consistency, it is important to create a roadmap for the analysis that is credible, reliable, and accurate. This research suggests a paradigm for examining students' comments. The framework utilizes existing machine learning, deep learning, and NLP methods. Among these approaches, ensembles have emerged as a successful paradigm for examining students' feedback [43]. However, based on earlier research, it is understandable that there have been only a few attempts to use ensembles in academic settings.

Sentiment analysis presents the ability to extract student opinions with their sentiment orientation at the document level, phrase level, entity level, and aspect level [48, 49]. However, Pathak and Warpade [50] asserted that opinions can be categorized into positive, negative, or neutral opinions. Sentiment analysis is performed for the entire text document at the document level. In addition, sentiment analysis at the document level is generic in that it does not provide the polarity score for individual reviews and evaluations contained in the text document. Hence, it is a shallow analysis that just provides an overview. Moreover, sentiment analysis at the sentence level is more in-depth compared to the sentiment analysis conducted at the document level. This is because sentiment polarity scores for each sentence in the document are computed in the text. In addition, entity-level sentiment extraction combines entity and sentiment analysis. Aspect-based sentiment analysis examines various data categories in a comment at a finer level and determines the sentiment orientation of each data category [51]. However, aspect-based sentiment analysis (ABSA) is a subfield of NLP that focuses specifically on sentiment targets known as aspects within a sentence [52]. It is more granular than document- and sentence-level sentiment analysis and, therefore, more complex to implement [2].

Dolianiti et al. [53] evaluated the effectiveness of five commercial sentiment analysis technologies at the document and sentence levels: IBM Watson Natural Language Understanding, Microsoft Azure Text Analytics API, Opinion Finder 2.0, Repustate, and SentiStrength. In the study, the authors used two educational datasets from the learning management system (LMS) containing student forum posts from two courses over a semester. Additionally, two alternative versions of the dataset's forum posts' sentiment orientation were manually annotated at the document and sentence levels. Consequently, SVM and k-fold cross-validation (CV) methods were used to construct four education domain tools, two for each course. According to the study, in one of the courses, educational domain tools outperformed commercial tools in terms of document and sentence quality.

Several studies have focused on predicting student performance while considering various variables such as income, family history, demographics, grades, and courses. Additionally, numerous research projects have been approved in this field to identify the variables necessary for modifying students' teaching and learning behaviors. The goal of the research study undertaken was to identify, detect, and estimate variables to understand students' learning behaviors by Kastrati et al. [22].

Previous studies on opinion mining have generally investigated an individual ML method with a single structure, demonstrating their respective superiority. Considering that feedback is affected by different factors, as well as that it shows different statistical characteristics, the individual ML model with a specific structure possesses limited ability to present the complex relationship between student feedback and diverse predictors in varying situations and circumstances. In recent years, ensemble

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

learning methods, which can combine multiple ML models, have shown their advantages. The stacking ensemble model is a popular one among them [54]. Stacking is a specific type of ensemble learning that can take advantage of different base model structures to generate theoretically more promising predictions [55]. However, Hutto and Gilbert [56] show that the majority of machine learning algorithms have limitations. To begin with, these methods frequently need large training datasets to represent different aspects. As a result of their extensive memory requirements and lengthy processing times, the approaches are frequently computationally expensive. Third, it is more difficult to modify, expand, or generalize the features that were retrieved from the text since they are difficult to interpret.

2.2. Related Work

A considerable amount of literature has been published on the use of machine learning techniques, particularly for predicting student opinions from their feedback in the academic world, including [57] presented an experiment using machine learning open-source data mining software tools, although no single tool consistently achieves the best results. The study aimed to improve SVM models on benchmark datasets from the Pang and Taboada corpora. SVM was selected for classification because it performs well in text classification and can handle large feature sets. The evaluation metrics included F1-measure, accuracy, and AUC (Area Under the ROC Curve). The study concluded that n-gram and bi-gram models have lower performance compared to the unigram model for both datasets. However, the experiment did not consider embedded feature extraction methods. Some models achieve better results more frequently than others.

Breiman [55] discussed faculty performance evaluation using a document-level sentiment analysis approach. In the study, a pre-processed dataset of about 5000 comments was used to train two machine learning classifiers, SVM and Naïve Bayes (NB), which had accuracy rates of 72.8% and 81%, respectively. Chatterjee and Chakma [58] compared the sentiment classification of student feedback questions at the sentence level and token level for different classifiers. The data obtained was classified using a supervised learning algorithm. The authors classified the questions in the form of natural language text using sentiment analysis methods. Additionally, feature extraction was performed with RapidMiner, and a standard POS tagger was used to tag all tokens. Different classifiers were employed at both the sentence and token levels. The results of the comparison showed that token-level sentiment analysis using a Decision Tree (DT)-based classifier yields improved results.

Furthermore, the cosine similarity method was used by Sivakumar and Reddy [59] to assess the semantic similarity of aspect terms and student opinion sentences. The data for the study was collected from the Twitter API, preprocessed, and the comments were categorized at the phrase level into seven elements. In an attempt to classify the texts into various features, three machine learning algorithms, namely decision trees, SVM, and NB, were applied. SentiWordNet, a lexicon-based technique, was used to attribute the sentiment orientation of subjective phrases after parts-of-speech (POS) tagging was used to extract them. To identify one or more features of the sentences and categorize their polarity as positive or negative sentiment, [43] performed a sentence-level analysis. For aspect extraction, the authors combined SVM, a cascade classifier, and rule-based approaches. The sentiment was also identified.

There are various methods to implement SA, and one of them is ML. However, many ML methods, including NB, SVM, neural networks, and k-nearest neighbor, have been employed to analyze student comments, Aung and Myo [60]. Dhanalakshmi et al. [61] emphasized that SVM is the best method for categorizing sparse text data, whereas neural networks use many layers of neurons to classify text, and KNN use Euclidean distances to evaluate the likelihood that a given text belongs to a specific feature. According to several studies, Zimbra et al. [23]; Altrabsheh et al. [62] and Balahadia et al. [63], neural networks are the ideal method for opinion mining.

El-Halees [64] investigated how opinion mining may offer an alternative way to improve course evaluation using students' attitudes posted on Internet forums, discussion groups, and/or blogs, which are collectively called user-generated content. The author proposed a model to mine knowledge from

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

students' opinions to improve teaching effectiveness in academic institutions, to achieve the purpose of the study. About 4957 data points were collected from discussion posts, pre-processed (data cleaning, removal of tags, non-textual contents, and stop words, tokenized, normalized, and stemmed, vectorized) and feature extraction was performed. Additionally, three machine learning methods, Naive Bayes (NB), KNN, and Support Vector Machine (SVM) were applied to classify opinions as positive or negative for each student's posts. The opinion classification was evaluated based on three performance metrics: precision, recall, and F-measure, and later compared with manually evaluated scores. With a precision of 77.58%, the study concluded that the NB method has better performance than the other two machine learning methods. Similarly, with a recall of 82.23%, SVM has better performance. However, overall, NB has the best F-measure with 77.83%.

Altrabsheh et al. [65] investigated different combinations of machine learning techniques, features, pre-processing levels, and the use of neutral classes for analyzing real-time students' feedback. About 1036 data points were collected from each student, with their distribution as 641 positive, 292 negative, and 103 neutral. Furthermore, the data collected was labeled by three experts, two of whom were linguistics experts. The reliability of the labels' inter-rater reliability was calculated, and the percentage agreement reached was 80.6%, the Fleiss kappa was 0.625, and Krippendorff's alpha was 0.626. Nearly all models performed better when pre-processing was applied, which was expected. However, some interesting exceptions include that unigrams gave high performance in several models; unigrams combined with bigrams performed well for Conditional Naïve Bayes (CNB), and trigrams performed relatively well with Maximum Entropy (ME). All methods except Naïve Bayes (NB) had relatively high accuracy, with the SVM linear kernel achieving the best performance at 95% and the SVM radial basis kernel the second best at 88%. Similarly, precision, recall, and F-score are high in both SVM and CNB models but low in NB and ME models. SVM and CNB also demonstrated good performance when the neutral class was considered.

In the study of predicting student performance in higher education, Jindal and Borah [66] used various decision tree categories, such as C5.0, C4.5-A1 and C4.5-A2 [10]. Implemented the framework as a prototype system, Student Feedback Mining Systems (SFMS), and tested it on selected courses using the topic extraction and sentiment extraction stage methodology. During the sentiment extraction stage, the labeled data was labeled manually for training and tested using the Lingpipe tool that adopts a logistic regression approach. Generally, the sentiment extraction stage achieved a precision of 80.1%, recall of 86.4%, and an F-score of 83.5%, which is significantly higher than the IMDb-trained classifier.

Menaha et al. [15] proposed a student feedback mining system using text analytics and sentiment analysis methods. In the study, the authors provided a deep analysis of qualitative feedback received from students to improve the student learning experience. In the experiment, feedback comments about each topic were collected and grouped into clusters, along with other data preparatory steps such as text processing to preprocess and clean the raw data, and features were extracted from the preprocessed documents. Additionally, the comments were classified using a sentiment classifier, and visualization techniques were applied to represent students' views. The results of the experiment showed that the frequency of each word was identified, and the topic with the highest frequency count was extracted. Similar comments within each topic were clustered, and the clustered words were classified into different orientations as negative and positive. Moreover, the classified comments were represented using charts for easy visualization. However, the feedback mining system built can also adopt semantic similarity to achieve the best results.

Kandhro et al. [67] proposed the SA model for enhancing the standard of instruction in HEI using a variety of ML techniques, including SVM, Multinomial Naïve Bayes, Random Forest, MLP Classifier, and Stochastic Gradient Descent. The study was successful in comparing various SA models to identify the best model for examining student feedback data in the classroom. Jena [68] conducted a study to examine sentiment polarity from students' views and model students' emotions (Anxiety, Amused, Confused, Enthused, Excited, Bored, Frustrated, etc.) using machine learning techniques such as

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

sentiment classifiers, Naive Bayes (NB), and Support Vector Machines (SVM) based on big data frameworks. Suppala and Rao [69] proposed and developed a sentiment analysis model with the Natural Language Toolkit (NLTK) on a dataset containing tweets, using the Naïve Bayes algorithm. The results showed the probability of each tweet being either positive or negative.

[70] proposed a method to predict student success/performance in an online learning environment. The method divides the mathematical material in an online math learning platform into activity scopes, which are then used to train classifiers such as Random Forest, Logistic Regression, Decision Tree with AdaBoost, Naïve Bayes, k-Nearest Neighbours, and Stochastic Gradient Descent. Finally, an ensemble of these classifiers was utilized to predict student performance, and the accuracy rate was 73.5%. Adejo and Connolly [71] Used decision trees, support vector machines, artificial neural networks, and a stacking ensemble approach to predict student performance. The accuracy, recall, precision, and root mean square error (RMSE) scores for the stacking ensemble are superior to those of the base classifiers.

Kumari et al. [72] used K-Nearest Neighbour (KNN), Iterative Dichotomiser 3 (ID3), Naive Bayes, and SVM as classifiers to assess the impact of student behaviour characteristics on students' performance. Similarly, the authors employed the group strategies of bagging, boosting, and voting. In comparison to the greatest accuracy value of 88.3% for the standalone classifier (ID3), the voting ensemble approach yielded a value of 89.0%. In a related study, Ajibade et al. [73] proposed a novel model for predicting student success based on data mining techniques and students' new behavioral characteristics. The experience of learners is connected to these behavioral characteristics. The proposed model in this instance made use of a variety of classifiers, including KNN, SVM, and Decision Tree.

Ensemble approaches are also applied to classifiers to enhance their performance. These methods include Random Forest, Bagging, and Boosting. The highest accuracy for ensemble approaches was 91.5%. The works being discussed use ensemble approaches to produce encouraging outcomes. Comparatively, these ensemble approaches yield better results than individual classifiers. According to our research, there is no information in the literature regarding how to choose and combine classifiers for ensemble modeling. It appears to be a random strategy for this combination because there is no clearly defined method for combining classifiers to improve outcomes. In this paper, we propose ensemble strategies that combine independent and supportive algorithms from the family of machine learning classification algorithms. In earlier investigations, this kind of algorithmic and ensemble technique integration was not observed.

Zounemat-Kermani et al. [74] asserted that using ensemble strategies is preferable to using individual machine learning models during their experiment in the hydrological domain. To anticipate mid-term streamflow, Li et al. [75] used SVR, RF, Elastic Net Regression (ENR), and Extreme Gradient Boosting (XGB). It was discovered that the stacking strategy enhanced the capability of individual models. When Wang et al. [76] compared the stacking model to individual models for predicting beach water quality, they discovered that the stacking model was the most reliable for predicting three beaches over five years. The promise of stacking ensemble models in students' feedback prediction has, however, received less attention.

Kesavaram et al. [77] proposed a customer feedback evaluation system for a particular product. The proposed approach provides a new Score Calculation Algorithm (SCA) method for score calculation based on various features from customer opinions. The methodology was implemented and tested in the Canon digital camera feedback dataset. In the feedback system, ontology-based feature extraction and a newly proposed score calculation method are used to evaluate customer feedback into positive and negative classes and rank their feature performances based on customer reviews. The sentiment analysis system developed is used to classify the comment data into optimistic and pessimistic categories to evaluate the overall performance of the product. Similarly, the SentiWordNet opinion word lexicon was adopted to facilitate identifying opinion-correlated words in the document. The performance estimation results show that the average accuracy of correctly classified features was found to be 81.11%.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

Nasim et al. [78] presented a combination of machine learning and lexicon-based approaches for sentiment analysis of students' feedback. In an attempt to grasp the sentiment polarity expressed in textual feedback by a student, the authors employed a hybrid approach to build the predictive model for sentiment analysis. The authors collected an unstructured textual dataset comprising 1230 comments extracted from the institution's educational portal. The dataset was preprocessed using Python NLTK libraries and manually labeled with sentiment polarity labels. The labeled data was partitioned into training and testing datasets on which feature extraction methods such as unigrams, bigrams, TF-IDF, and lexicon-based features were applied. The hybrid model was further trained using Random Forest and SVM algorithms, respectively. Results revealed that the best-performing model was achieved using TF-IDF and a domain-specific sentiment lexicon. However, the approach used in the study is limited to the computation of the overall sentiment of the student feedback.

Sultana et al. [79] presented a prediction of sentiment analysis on educational data based on the deep learning approach. In an attempt to accomplish the objective, a study on the comparative analysis of eight classifiers, namely SVM, MLP-Deep Learning, K-star, Bayes Net, Simple Logistics, Multi-class Classifier, Decision Tree, and Random Forest, was conducted on the dataset obtained from the Kiteboard 360 dataset repository to predict students' performance. In the same vein, ten-fold cross-validation was performed. The results indicated that SVM and MLP-Deep Learning were the best-performing learning methods, achieving 78.75% and 78.33% accuracy, respectively, as well as good performance in terms of their specificity, sensitivity, and ROC curve area. Ramadhani and Goo [80] compared the Multilayer Perceptron (MLP) and Deep Learning (DL) and achieved 52.60% and 75.03% test accuracy, respectively.

Altrabsheh et al. [65] employed an ensemble stacking approach and base classifiers, namely J48, Random Forest, and Random Tree to evaluate student performance with the aim of enhancing results. To corroborate the improvement, the SMOTE technique was applied. The findings revealed that the accuracy obtained from the base classifiers was 95.65%, while stacking Corollary achieved 95.96% and 96.11% using the SMOTE technique.

Olabode et al. [81] summarized research on the application of stacked ensemble learning approaches in developing a model for diagnosing head and neck cancer. The stacking ensemble method was selected, combining multiple classifiers, namely decision tree (C4.5), KNN, and Naive Bayes, via a meta-classifier, logistic regression, with cross-validation applied. When logistic regression was used at the meta-level on the reduced dataset, the results indicated that the chi-square method in a stacked ensemble model produced better predictions than the consistency method. The chi-square feature method on a stacked ensemble model can be used for the prediction of head and neck cancer. Iyanda and Abegunde [82] employed the ensemble approach (SVM, NB, and LR) to detect sentiment from Yorùbá sentences at the sentence level to extract users' opinions. The results show that Naïve Bayes outperforms other algorithms in sentiment analysis for Yorùbá sentences.

Kavitha and Kumar [83] discussed how sentiment analysis can be performed on the feedback collected in a learning management system to advance the teaching and learning process. This work presents the experimental results obtained after a comparison of various feature selection methods such as Chi-square, Information Gain, Mutual Information, and Symmetrical Uncertainty. Lalata et al. [84] proposed an ensemble learning approach of five machine learning algorithms, such as Naive Bayes, logistic regression, support vector machine, decision tree, and random forest classifiers, based on the majority voting principle. In the study, for each semester, student comments were compiled, and the sentiment of each comment was manually classified as positive, negative, or neutral. The authors performed individual classification and ensembled the classifier output. The authors used individual classification and combined the results of the classifiers. All model ensembles obtained accuracy, F1-score, and recall of 90.32%, 93.80%, and 90.86%, respectively, according to the results.

Sengar et al. [28] investigated the needs of teachers and students. In an attempt to predict student feedback coupled with the availability of a vast amount of educational text and speech data, the author applied the NLP approach, particularly the NLTK and Random Forest techniques. In the study, the

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

author collected student feedback about classrooms, exams, and laboratory facilities. The results revealed that the collected data was analyzed to be positive, negative, or neutral, which eventually helped in improving the performance of the institution and the institution's learning and teaching experience.

Wook et al. [85] investigated how data from students' feedback can be analyzed to give accurate results compared to the use of questionnaires. In an attempt to achieve this aim, an opinion mining (OM) feedback system, known as OM Feedback, was developed. However, only a small number of studies have utilized these features to enhance the ability of the opinion-mining technique to analyze students' feedback. Based on these reasons, this study has developed a new system to analyze students' feedback, known hereafter as the OM Feedback system. Katragadda et al. [86] explored opinion mining using supervised learning algorithms to find the polarity of student feedback based on predefined features of teaching and learning. In addition to providing a step-by-step explanation of the implementation process of opinion mining from student comments using the open-source data analytics tool known as RapidMiner, this paper also presents a comparative performance study of algorithms such as SVM, Naïve Bayes, KNN, and Neural Network classifiers. The opinion mining system was implemented using the Python programming language. Consequently, the results showed that neural networks achieved a performance accuracy of 88%, demonstrating better performance concerning various evaluation criteria for the different algorithms applied.

Ahamad and Ahmad [40] employed individual machine learning algorithms and ensemble methods, particularly stacking and voting ensembles, to predict feedback from students' assessments using the WEKA tool. The study indicated that the individual machine learning algorithm, notably the multilayer perceptron (MLP), achieved the highest accuracy of 92.30% compared to other algorithms. Additionally, these individual algorithms were combined to form a stacking ensemble (Fuzzy, Neural Network, Naïve Bayes, Random Forest, MLP). Qaiser et al. [87] employed sentiment analysis (SA) predict people's opinions on any topic of interest. The authors reviewed and compared several opinion mining techniques, including machine learning methods such as Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and deep learning. These ML techniques were applied to a single dataset to compare their performance in terms of accuracy. The study found that deep learning performed the best with 96.41% accuracy, followed by NB and SVM with 87.18% and 82.05%, respectively. Decision Tree performed the poorest with 68.21% accuracy.

Nidhi et al. [88] discussed the comparative analysis of heterogeneous ensemble learning using feature selection techniques for predicting student performance. The author employed a hybrid or heterogeneous approach of correlation attribute evaluation, ensemble learning, namely stacking, voting, and multischeme, in conjunction with seven other machine learning algorithms to improve accuracy. The results were validated using the K-fold cross-validation method to evaluate the performance of the classification algorithms.

Verma et al. [89] proposed a scalable machine learning-based ensemble approach to predict student performance to enhance the accuracy of identifying students at risk. The author employed five single-supervised machine learning techniques: Decision Tree, Naïve Bayes, k-nearest neighbor, Support Vector Machine, and Logistic Regression (ensemble model) that integrates the most suitable data mining technique. Furthermore, the performance of the algorithms was evaluated with and without resampling methods such as Synthetic Minority Oversampling Technique (SMOTE), Borderline SMOTE, SVM-SMOTE, and Adaptive Synthetic (ADASYN). However, the Random Hold-Out method and GridSearchCV were used as model validation techniques and for hyper-parameter tuning, respectively. The results indicated that Logistic Regression was the best-performing classifier across all balanced datasets generated using the four resampling techniques, achieving the highest accuracy of 94.54% with SMOTE. Additionally, to improve prediction results and enhance scalability, the most suitable classifier was integrated using bagging, resulting in an accuracy of 95.45%. Roaring et al. [90] examined the connection between the numerical evaluation of teacher performance and the real opinions, sentiments, and observations.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

Gebashe et al. [91] proposed a two-part faculty assessment system built on machine learning and text analytics. Postgraduate students' qualitative and quantitative evaluations regarding their faculty, as well as information about student and teacher characteristics, were recorded using a standardized questionnaire. Sentiment analysis was used to analyze the qualitative input and convert the text feedback into polarity scores. Depending on the faculty and the length of the response, the polarity of the words and sentences in the qualitative feedback varied. Students were highly emotionally engaged by faculty members who employed case studies, practical experiments, and real-world analogies. Ten different machine learning algorithms were used to predict professor effectiveness based on the polarity ratings of the qualitative and quantitative evaluations. The random forest model was the best among all, outperforming others with high accuracy, precision, and area under the curve, achieving 98.87%, 97.71%, and 97.32%, respectively.

3. Methods and Materials

This section discusses the approach used to achieve the study's goals of developing a stacking ensemble model for students' feedback prediction with many predictors, as illustrated in Fig. 2. The methodology also comprises the following stages: dataset collection, cleaning, preparation, pre-processing, data balancing, partitioning (training and testing of data), cross-validation, and model development, implementation, and evaluation. Additionally, the ensemble methods and a mixture or combination of machine learning algorithms for creating an efficient ensemble approach are provided. Finally, a model was created to improve prediction accuracy by comparing the outcomes.

3.1. Proposed Model

In this paper, both stacking ensemble classifiers and independent machine learning classification techniques are employed across the educational dataset about HEI. The stacking ensemble in the current study comprises various base-level classifiers, including LR, SVM, MLP, and NB, whose outputs are combined to produce an improved prediction result. Moreover, the filtered dataset was obtained after performing a pre-processing step on the original dataset. The pre-processed data was then classified using several algorithms, from which the three best-performing classifiers were selected to enhance prediction accuracy on test data. Figure 2 illustrates the proposed model, where the dataset acquired from the university learning management system includes different features of students and the final class labels. The dataset primarily undergoes pre-processing to remove inconsistent and noisy data. Additionally, to achieve better forecasting results, balancing of instances is performed to ensure a uniform distribution of class labels. The selected base classifiers are trained and tested on the data, and their outputs are combined using the stacking ensemble method. Ultimately, the predictions are generated at the meta-level based on the combined predictions from different classifiers, aiming to reduce the generalization error. Furthermore, the pre-processed data is also evaluated on other classifiers without balancing the dataset, thereby providing a comparison between balanced and imbalanced data.

3.2. Data Collection and Description

The dataset used is collected from learning management systems (LMS) of a public university in Nigeria using a learner activity tracker tool called Experience API (xAPI) that monitors learning progress and learners' actions. The educational feedback dataset consists of textual reviews of about **1011** students with features such as student_id, student_name, review, and emotion. A snapshot of the sample feedback data collected from the LMS is illustrated in Fig. 3. These sentences are unstructured, which is not suitable for direct analysis. Therefore, preprocessing of the dataset is necessary. The dataset was manually labeled with sentiment polarity labels: positive, negative, and neutral, as shown in Figure 4. After collecting the data in its raw form with manual labeling, the comments are divided into sentences.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

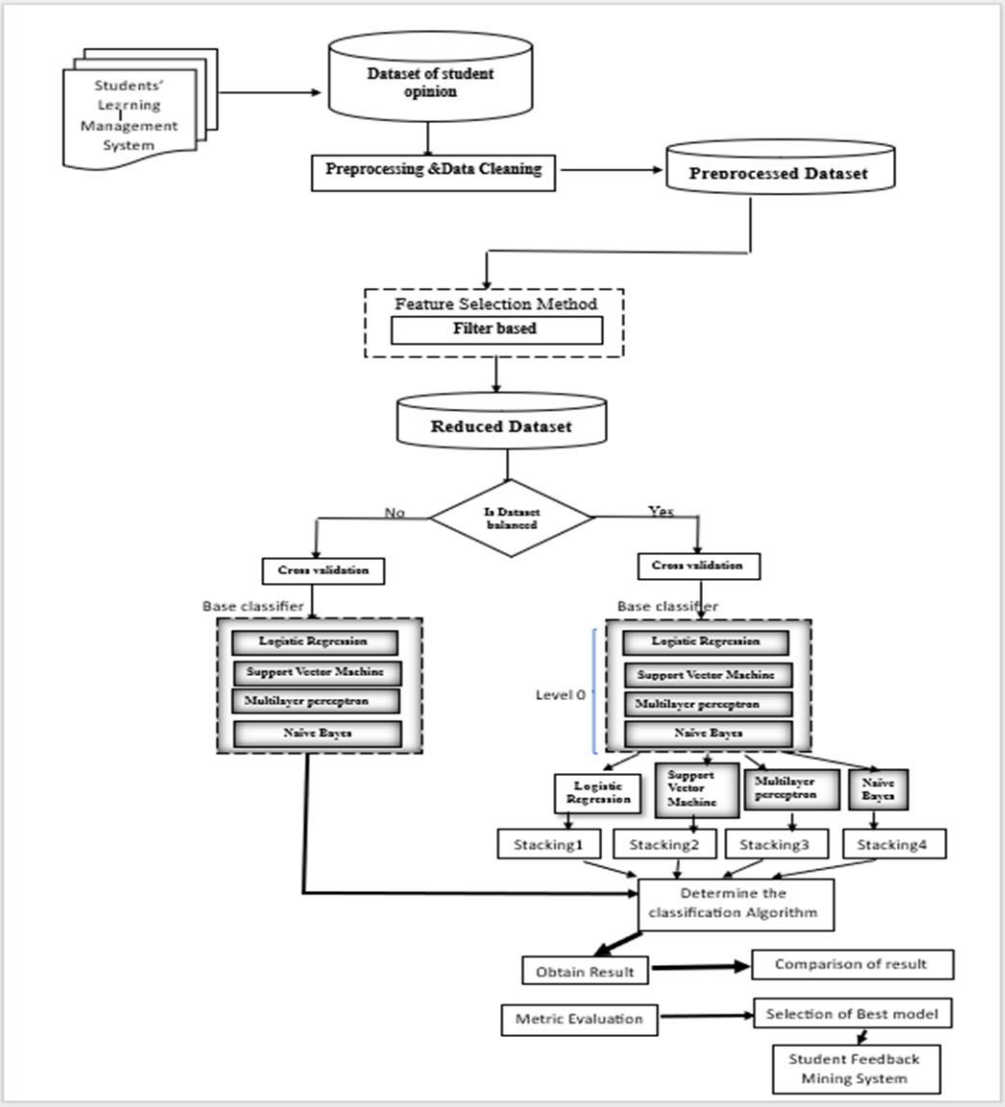
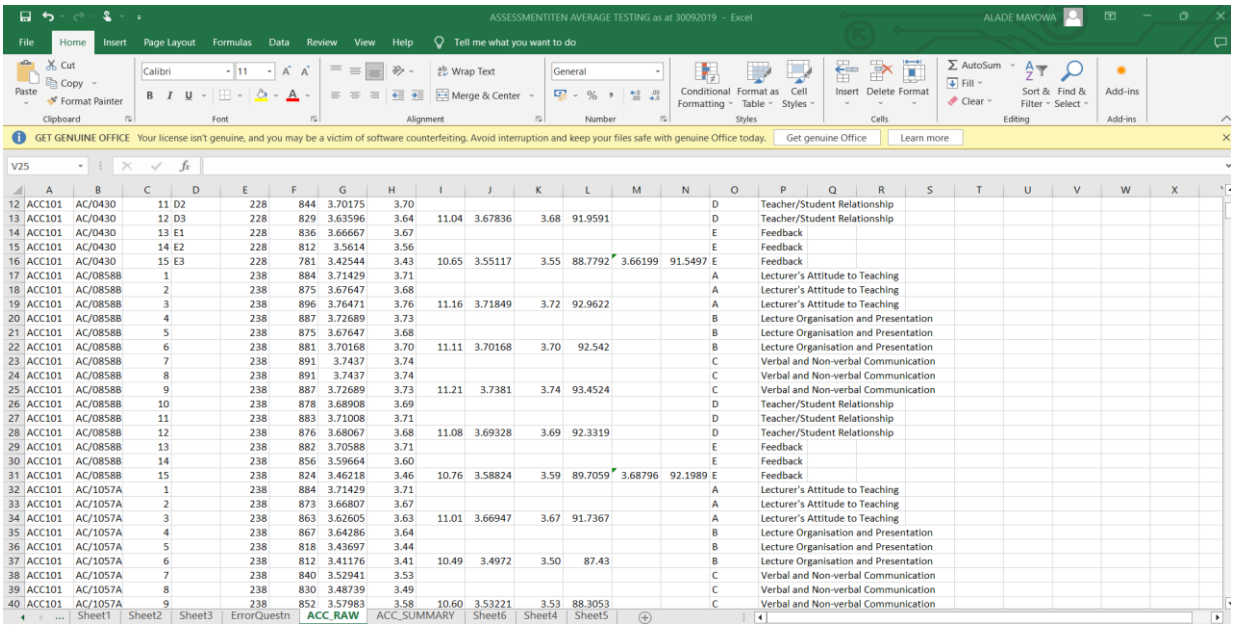


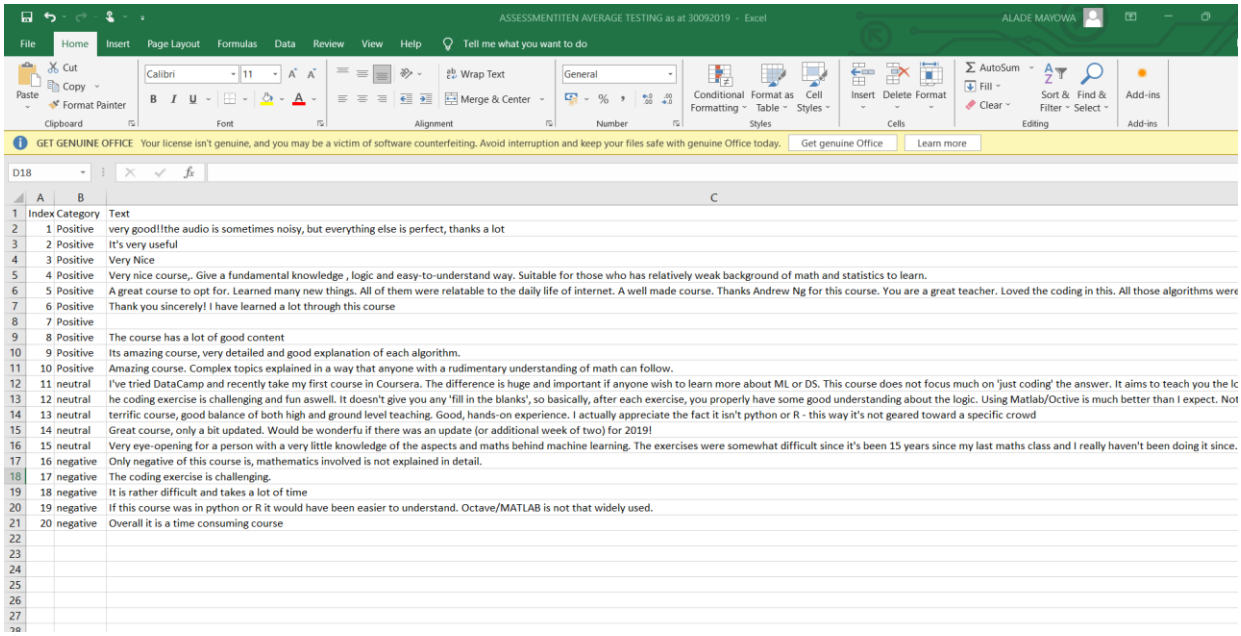
Figure 2.
Proposed model for student feedback opinion mining system.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
12	ACC101	AC/0430	11	D2	228	844	3.70175	3.70							D	Teacher/Student Relationship
13	ACC101	AC/0430	12	D3	228	829	3.63596	3.64	11.04	3.67836	3.68	91.9591			D	Teacher/Student Relationship
14	ACC101	AC/0430	13	E1	228	836	3.66667	3.67							E	Feedback
15	ACC101	AC/0430	14	E2	228	812	3.5614	3.56							E	Feedback
16	ACC101	AC/0430	15	E3	228	781	3.42544	3.43	10.65	3.55117	3.55	88.7792	3.66199	91.5497	E	Feedback
17	ACC101	AC/0858B	1		238	884	3.71429	3.71							A	Lecturer's Attitude to Teaching
18	ACC101	AC/0858B	2		238	875	3.67647	3.68							A	Lecturer's Attitude to Teaching
19	ACC101	AC/0858B	3		238	896	3.76471	3.76	11.16	3.71849	3.72	92.9622			A	Lecturer's Attitude to Teaching
20	ACC101	AC/0858B	4		238	887	3.72689	3.73							B	Lecture Organisation and Presentation
21	ACC101	AC/0858B	5		238	875	3.67647	3.68							B	Lecture Organisation and Presentation
22	ACC101	AC/0858B	6		238	881	3.70168	3.70	11.11	3.70168	3.70	92.542			B	Lecture Organisation and Presentation
23	ACC101	AC/0858B	7		238	891	3.7437	3.74							C	Verbal and Non-verbal Communication
24	ACC101	AC/0858B	8		238	891	3.7437	3.74							C	Verbal and Non-verbal Communication
25	ACC101	AC/0858B	9		238	887	3.72689	3.73	11.21	3.7381	3.74	93.4524			C	Verbal and Non-verbal Communication
26	ACC101	AC/0858B	10		238	878	3.68908	3.69							D	Teacher/Student Relationship
27	ACC101	AC/0858B	11		238	883	3.71008	3.71							D	Teacher/Student Relationship
28	ACC101	AC/0858B	12		238	876	3.68067	3.68	11.08	3.69328	3.69	92.3319			D	Teacher/Student Relationship
29	ACC101	AC/0858B	13		238	882	3.70588	3.71							E	Feedback
30	ACC101	AC/0858B	14		238	856	3.59664	3.60							E	Feedback
31	ACC101	AC/0858B	15		238	824	3.46218	3.46	10.76	3.58824	3.59	89.7059	3.68796	92.1989	E	Feedback
32	ACC101	AC/1057A	1		238	884	3.71429	3.71							A	Lecturer's Attitude to Teaching
33	ACC101	AC/1057A	2		238	873	3.66807	3.67							A	Lecturer's Attitude to Teaching
34	ACC101	AC/1057A	3		238	863	3.62605	3.63	11.01	3.66947	3.67	91.7367			A	Lecturer's Attitude to Teaching
35	ACC101	AC/1057A	4		238	867	3.64286	3.64							B	Lecture Organisation and Presentation
36	ACC101	AC/1057A	5		238	818	3.43697	3.44							B	Lecture Organisation and Presentation
37	ACC101	AC/1057A	6		238	812	3.41176	3.41	10.49	3.4972	3.50	87.43			B	Lecture Organisation and Presentation
38	ACC101	AC/1057A	7		238	840	3.52941	3.53							C	Verbal and Non-verbal Communication
39	ACC101	AC/1057A	8		238	830	3.48739	3.49							C	Verbal and Non-verbal Communication
40	ACC101	AC/1057A	9		238	852	3.57983	3.58	10.60	3.53221	3.53	88.3053			C	Verbal and Non-verbal Communication

Figure 3.
Sample of the raw dataset before pre-processing.



	A	B	C
1	Index	Category	Text
2	1	Positive	very good!!the audio is sometimes noisy, but everything else is perfect, thanks a lot
3	2	Positive	It's very useful
4	3	Positive	Very Nice
5	4	Positive	Very nice course,. Give a fundamental knowledge , logic and easy-to-understand way. Suitable for those who has relatively weak background of math and statistics to learn.
6	5	Positive	A great course to opt for. Learned many new things. All of them were relatable to the daily life of internet. A well made course. Thanks Andrew Ng for this course. You are a great teacher. Loved the coding in this. All those algorithms were
7	6	Positive	Thank you sincerely! I have learned a lot through this course
8	7	Positive	
9	8	Positive	The course has a lot of good content
10	9	Positive	Its amazing course, very detailed and good explanation of each algorithm.
11	10	Positive	Amazing course. Complex topics explained in a way that anyone with a rudimentary understanding of math can follow.
12	11	neutral	I've tried DataCamp and recently take my first course in Coursera. The difference is huge and important if anyone wish to learn more about ML or DS. This course does not focus much on 'just coding' the answer. It aims to teach you the lo
13	12	neutral	he coding exercise is challenging and fun aswell. It doesn't give you any 'fill in the blanks', so basically, after each exercise, you properly have some good understanding about the logic. Using Matlab/Octive is much better than I expect. Not
14	13	neutral	terrific course, good balance of both high and ground level teaching. Good, hands-on experience. I actually appreciate the fact it isn't python or R - this way it's not geared toward a specific crowd
15	14	neutral	Great course, only a bit updated. Would be wonderfu if there was an update (or additional week of two) for 2019!
16	15	neutral	Very eye-opening for a person with a very little knowledge of the aspects and maths behind machine learning. The exercises were somewhat difficult since it's been 15 years since my last maths class and I really haven't been doing it since.
17	16	negative	Only negative of this course is, mathematics involved is not explained in detail.
18	17	negative	The coding exercise is challenging.
19	18	negative	It is rather difficult and takes a lot of time
20	19	negative	If this course was in python or R it would have been easier to understand. Octave/MATLAB is not that widely used.
21	20	negative	Overall it is a time consuming course

Figure 4.
Dataset after pre-processing.

3.3. Data Preparation

After the dataset from tertiary institutions was acquired or collected from the learning management system using the Experience API (xAPI), data preparation is an important step in the process of mining useful insights from educational datasets. In this stage, the following tasks were performed.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

3.3.1. Data Cleaning and Preprocessing

These are important steps involved in preparing data to make raw or original datasets suitable for data mining techniques and algorithms to be implemented. Owing to the large volume of datasets available in educational data repositories and institutional learning portals and platforms, there are challenges facing the database that affect the quality of data. To improve the quality of data available for the model, data cleaning was carried out to handle missing and inconsistent data, as well as noise removal. In the same way, removal of punctuations, numbers, special characters, hashtags, hyperlinks, stop words like verbs, and prepositions from the feedback text are necessary because they do not carry useful information. In an attempt to make the data cleaning process more efficient, a data cleaning pipeline (function) was created. The function removed punctuations, joined all the data as a string, converted the texts to lowercase, and removed stop words such as after, and in preparation for the tokenization process. The tokenization involves the splitting of sentences (text streams) into words, otherwise referred to as tokens. In the study, the token, which is a sequence of characters (words, emotions), is obtained using whitespace and punctuation as delimiters. In addition to tokenization, the tokenized text was normalized by applying stemming and lemmatization processes. Moreover, the dataset is reduced while preserving the most important information, owing to the large amount of redundant information contained in the dataset.

Consequently, the dimensionality reduction method is applied to reduce the number of dataset features using several techniques related to data compression, feature selection, and feature construction. Similarly, transformation of the dataset is necessary and is applied to convert the cleaned dataset to a suitable format using discretization and normalization methods. In this study, a total of 1001 feedback, i.e., comments that are in an unstructured form full of noise and unwanted information, were received using some Python programming language libraries and packages, all of which were used after data cleaning. Afterwards, string attributes were converted into a set of numeric attributes representing word occurrence information from the text contained in the strings. In this work, a string is split into an n-gram. In the same vein, the comments from the students' feedback dataset are then converted from numerical values into nominal values, which denote the three class labels (positive, negative, and neutral) for the classification problem using a discretization process.

3.3.2. Feature Selection

This is another important step carried out in data preprocessing in data mining to develop a students' feedback prediction model. The purpose of carrying out feature selection is to select an appropriate subset of features that can efficiently describe the input data, reduce the dimensionality of the feature space, and eliminate redundant or irrelevant data without losing reliability in classification. Thus, improving the quality of the data and, in turn, the performance of the learning algorithm. Notably, there are three methods of feature selection: manual selection based on pedagogical theories or expert experience, filter-based selection, and wrapper feature selection. In the present study, a filter-based technique for feature selection was applied. The filter-based method employed searches for the minimum set of relevant features while ignoring others. Moreover, the technique is a ranking method used to rank the features, where highly ranked features are selected while ignoring the rest, using an information gain-based selection algorithm to evaluate the feature ranks and determine which features are most important for developing a student feedback performance model.

3.3.3. Splitting

The dataset collected was partitioned into 90% for the training set and 10% for validation or testing. This resulted in a total of 589 data points assigned for training and 147 data points assigned to the validation set, to continue the experiments with the independent algorithms and without balancing the data. However, for experiment 2, the dataset was increased and balanced with a total of 1011 observations. Out of these, 90% (i.e., 809 data points) were allocated for training purposes, while the

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

remaining 10% (202 data points) were used for validation. Afterwards, the experiments were performed with the proposed model to improve the results obtained in the previous experiment.

3.4. Modelling and Evaluation

In this research, stacking ensemble methods are applied to provide an accurate evaluation of the features that may impact the performance of the student feedback system and to enhance the model's effectiveness. The stacking ensemble method is a learning approach that combines multiple models to solve a problem. Simply put, the process of stacking begins by obtaining the results predicted by a set of diverse base models or classifiers, and then optimally combining these outputs into a larger framework using a meta-learner or model to generate the final prediction. In other words, the outputs of various base learners (level-0) are merged by the meta-learner or algorithm in the stacking ensemble structure, as depicted in Fig. 2. LR, SVM, MLP, and NB were the basic level classifiers. During the modeling stage, independent models such as LR, SVM, MLP, and NB were employed. Additionally, four combined models that included stacking methods with various levels were proposed.

3.5. Experiment

The student feedback dataset used was collected from the academic information system at the computer division of Obafemi Awolowo University, a higher education institution in Nigeria. The dataset included information from students. The experiment was run on a PC containing 6GB of RAM, 4 Intel cores (2.67GHz each), with basic preprocessing and other exclusive hybrid techniques where term frequency and outlier removal were applied to the dataset. The cleaned dataset was fed into the existing machine learning techniques: LR, SVM, MLP, and Multinomial Naïve Bayes. The entire dataset was split into a ratio of 90% for training and 10% for testing. For each algorithm, the model was produced based on the training set. Using the model, the opinions in the testing dataset were predicted and compared. The results are displayed in tables, charts, and graphs, with MLP achieving the best accuracy among all applied machine learning algorithms. For the experiments, the Python programming language and its libraries such as scikit-learn, joblib, NumPy, mlxtend, pandas, NLTK, Django framework, and Langdetect were used to implement the model, render templates, and evaluate the proposed classification models and comparisons.

All the classification methods are trained using 10-fold cross-validation. Using this method, the dataset is divided into 10 equal-sized subsets, nine of which are utilized for training and one for testing. The process is iterated ten times, and the final result is estimated as the average error rate on test examples. Once the classification model has been trained, the validation process begins. The validation process is the last phase of building a predictive model, which is used to evaluate the performance of the prediction model by running the model over real data.

In the experiment conducted, the performance of machine learning models was evaluated for classification quality using commonly used measures: Accuracy, Precision, Recall, and F-Measure. Additionally, a confusion matrix was employed to assess the performance of the classification algorithm, which plots the number of correct predictions against incorrect predictions, accounting for dataset imbalances. The confusion matrix is based on actual and predicted values for positive, negative, and neutral classes. This is a multi-class (3-class) problem, with target values: positive (A), negative (B), and neutral (C). The performance measures precision, recall, and F-measure, were calculated for each class label to analyze individual class performance. These values were then averaged to determine overall precision, recall, and F-measure. The calculations follow equations 9, 10, 11, and 12, respectively.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F_{measure} = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \quad (12)$$

4. Results and Discussion

4.1. Results

This section discusses the findings from the individual models and proposed combined models with and without data balancing. The evaluation measures, which enable the assessment of the effectiveness of the suggested models, are also provided.

4.1.1. Algorithm confusion matrix

The confusion matrix (CM) helps us handle data imbalances; that is, if the number of observations varies significantly, we can also identify misclassified data. The CM is calculated from the top-left diagonal to the bottom-right diagonal, where the actual label and the predicted label intersect. It is a measurement tool created as a table to visualize the performance of an algorithm or classifier. The CM graph depicts the relationship between correctly predicted and incorrectly predicted reviews. Each row of the matrix represents a predicted value, while the column displays the actual value, or vice versa. The number of positive feedbacks correctly predicted by the classifier is represented by True Positives (TP), whereas the number of positive feedbacks incorrectly predicted is denoted by False Positives (FP). Similarly, True Negatives (TN) refer to the number of negative reviews or opinions correctly predicted, while False Negatives (FN) refer to negative reviews incorrectly classified as positive.

Figure 5 shows that from the first row, out of a total of 21 comments given by students, 10 have been correctly predicted as negative, and therefore incorrectly predicted as positive; 8 have been wrongly predicted as neutral; and 3 have been wrongly predicted as positive. From the second row, with a total of 9 comments, 1 has been correctly predicted as negative; 4 have been wrongly predicted as neutral; and 4 have been wrongly predicted as positive. In the third row, 0 comments have been correctly predicted as negative; 1 has been wrongly predicted as neutral; and 30 have been wrongly predicted as positive.

Figure 6 shows that from the first row, out of a total of 45 comments given by students, 37 have been correctly predicted as negative, and hence incorrectly predicted as negative, 5 have been wrongly predicted as neutral, and 3 have been wrongly predicted as positive. On the second row, out of a total of 52 comments, 4 have been correctly predicted as negative, and hence incorrectly predicted as negative, 32 have been wrongly predicted as neutral, and 16 have been wrongly predicted as positive. On the third row, out of a total of 106 comments, 3 have been correctly predicted as negative, and hence incorrectly predicted as negative, 14 have been wrongly predicted as neutral, and 89 have been wrongly predicted as positive.

Fig.7 shows that, from the first row, out of a total of 39 comments given by students, 38 have been correctly predicted as negative, and hence incorrectly predicted as positive; 0 have been wrongly predicted as neutral; and 1 has been wrongly predicted as positive. In the second row, out of 56

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

comments, 45 have been correctly predicted as negative; 3 have been wrongly predicted as neutral; and 8 have been wrongly predicted as positive. In the third row, out of 53 comments, 34 have been correctly predicted as negative; 12 have been wrongly predicted as neutral; and 7 have been wrongly predicted as positive.

4.1.2. Evaluation Metrics

4.1.2.1. Accuracy

The accuracy is the total number of correct predictions divided by the total number of predictions made for a dataset. It provides a positive indication of how well the model is performing in terms of generating correct predictions. The high accuracy value obtained indicates that the model is correctly identifying the underlying trends in the data and delivering reliable findings. It enables comparison between different algorithms or variations of the same model. Tables 1 and 2 show the results obtained from performance analysis concerning the accuracy of existing machine learning algorithms and the proposed stacking ensemble model. Similarly, based on the confusion matrix value, classification accuracy, and other metrics, namely precision, recall, and F-measure, have been evaluated, and the results are tabulated in Table 2.

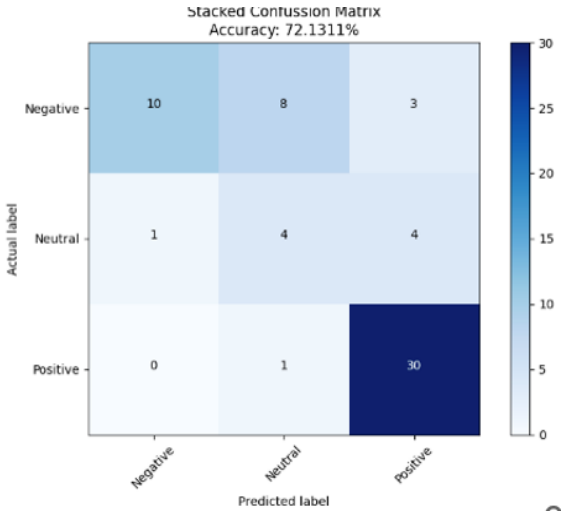


Figure 5.
Confusion matrix- Stacking 1 with data imbalance.

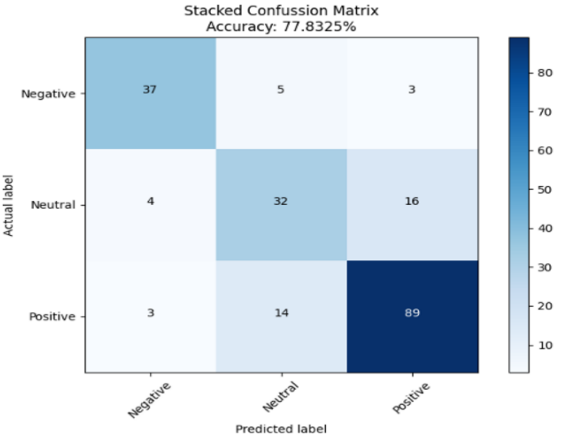


Figure 6.
Confusion matrix- Stacking 2 with data imbalance

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

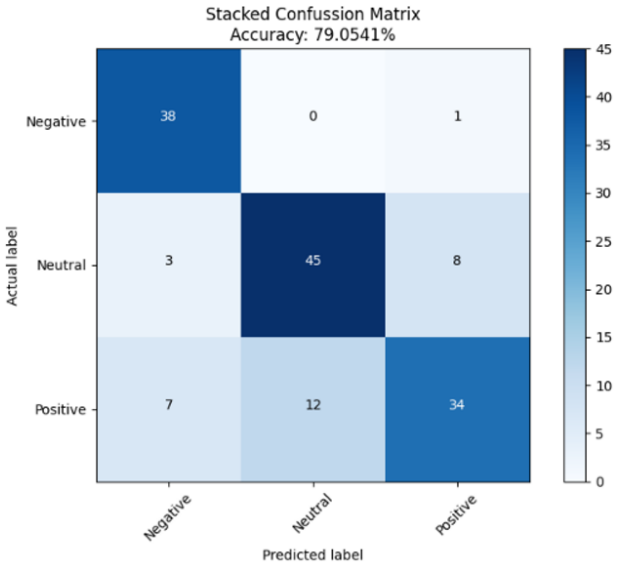


Figure 7.
Confusion matrix- Stacking 3 with data imbalance.

Table 1.
Summary of the classification accuracy produced by different machine learning algorithms and the proposed model.

Algorithms	Classification Accuracy (%)
Linear Regression	79.05
Support Vector Machine	79.05
Multilayer Perceptron	81.76
Naïve Bayes	50.68
Stacking	79.05

Table 2.
Performance measures along with their corresponding classification technique percentages (validation data).

	Accuracy	Precision	Recall	Fi-measure
Independent Algorithms				
Logistic Regression (LR)	79.05	79.52	79.78	79.64
Support Vector Machine (SVM)	79.05	79.07	80.20	79.36
Multilayer Perceptron (MLP)	81.76	81.71	83.20	82.11
Multinomial Naïve Bayes (MNB)	50.68	70.28	55.50	49.19
Combined Algorithms				
LR second-level Stacking 1 ensemble technique without data balancing	72.13	67.59	62.95	62.37
SVM Level Stacking 2 ensemble technique without balancing	77.83	76.41	75.91	76.15
MLP second-level Stacking 3 ensemble technique with data balancing	79.05	79.06	80.65	79.28
NB second-level stacking 4 ensemble technique without data balancing	50.68	70.28	55.50	49.19

The result shows that the LR algorithm yields a classification accuracy of 79.05%, while SVM, MLP, and Multinomial Naïve Bayes yield 79.05%, 81.76%, and 50.68%, respectively. Among all machine learning algorithms, MLP provides the best accuracy, as illustrated in Figures 8, 9, and 10. These figures clearly show the comparison of the accuracy metric of individual machine learning algorithms and the proposed stacking ensemble algorithm, as well as the accuracy of the combined algorithms.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

Other metrics, namely precision, recall, and F-measure, have been evaluated, and the results are presented in Table 2.

4.1.2.2. Precision

It is a metric that measures the number of correctly classified classes. It gives the ratio between true positives and the total number of cases categorized as positive (true positives and false positives). As a result, according to these criteria, the independent MLP scored 81.71%, followed by the combined algorithm utilizing ensemble stacking 3 with data balancing, which scored 79.06%, as shown in Table 2 and Fig.10.

4.1.2.3. Recall

Recall is a statistic that measures the number of correct positive predictions made out of all possible positive predictions, as determined by the formulas in equation 11. Table 2 and Fig. 10 present the results of the recall metric, where the values for this metric reached different percentages, which varied according to the models. It is observed that the independent MLP algorithm or model gave the highest recall percentage or value of 81.71%, followed by the logistic regression model and the stacking 3 ensemble method using MLP, with a percentage of 79.06%. The other models ranged from 67.59% to 76.41% without using balanced data, and from 70.28% to 81.71% with and without balanced data.

4.1.2.4. F1-Measure

The F1-score allows one to integrate precision and recall into a single metric that incorporates both attributes of the model, as given by the formula in equation 12. It is a metric that combines the values of the true positive rate and positive predictive value, namely, accuracy and recall. Table 2 and Fig. 10 show the evaluation metric "F1-Score," where the MLP with the Stacking 3 ensemble technique using the data balancing method obtained a value of 80.65%, considered the best model for predicting opinions from student feedback in a university. Meanwhile, Stacking 2 and Stacking 1 without data balancing achieved the lowest F1-score values (76.15% and 62.37%) compared to other individual models such as LR, MLP, and SVM, except for NB, which scored about 49.19%.

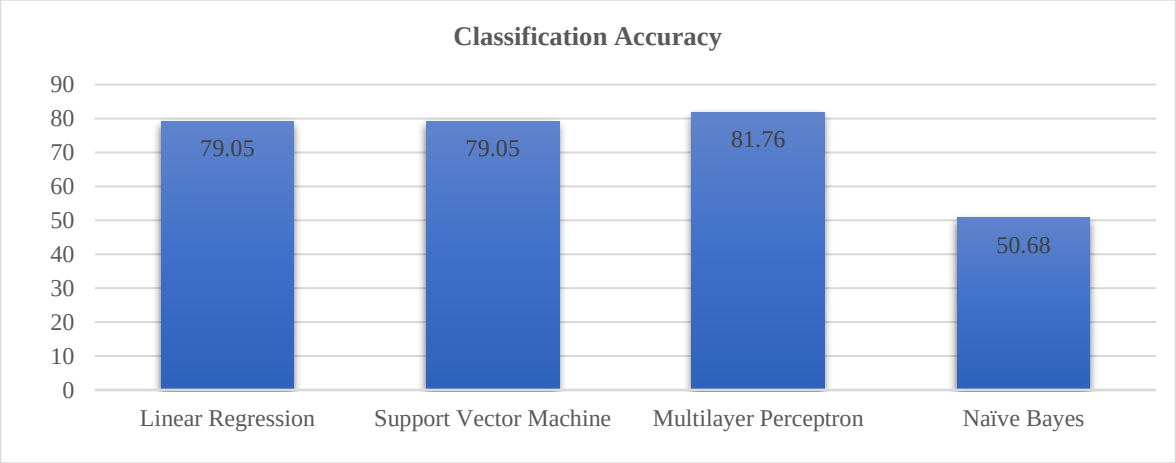


Figure 8.
Visualisation result classification accuracies of various supervised machine learning algorithms used in the experiment

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

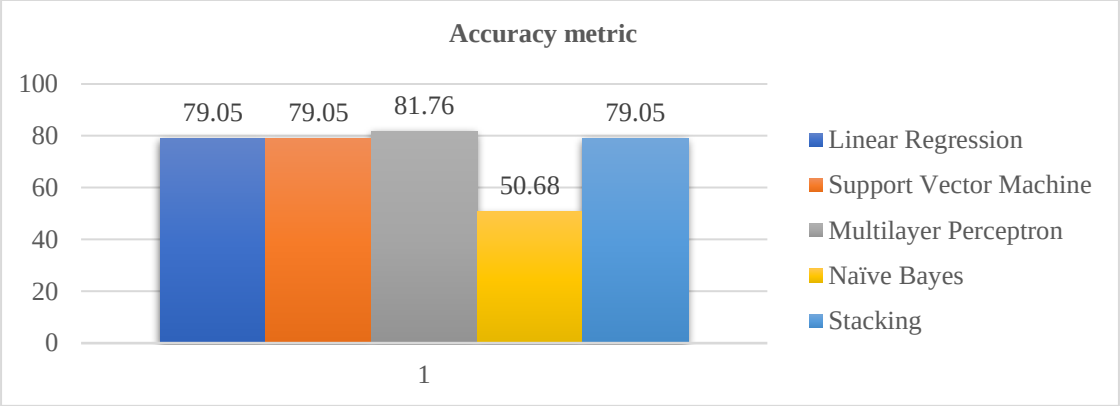


Figure 9.
Comparison of the Accuracy metric of various supervised machine learning algorithms with a stacking ensemble.

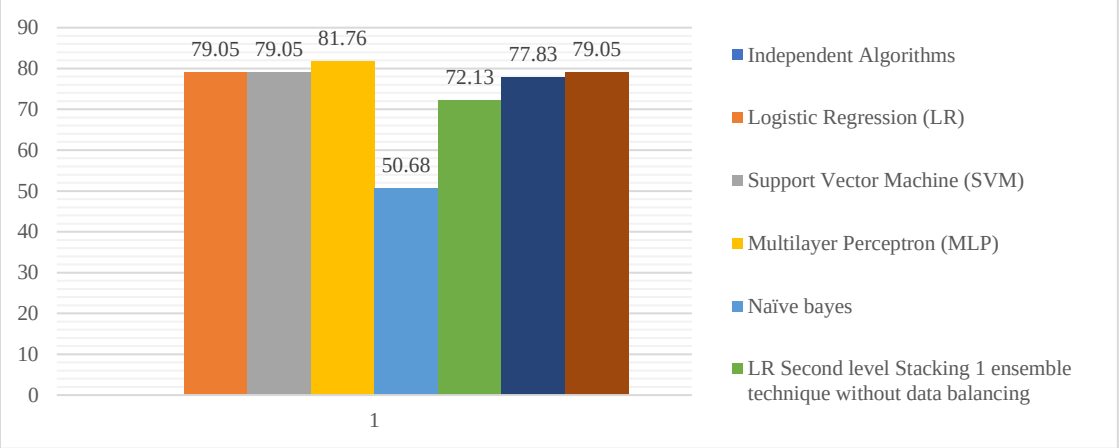


Figure 10.
Comparison of the percentages of the accuracy metric of the Stacking models to predict the opinion of university students, both with unbalanced data and with balanced data.

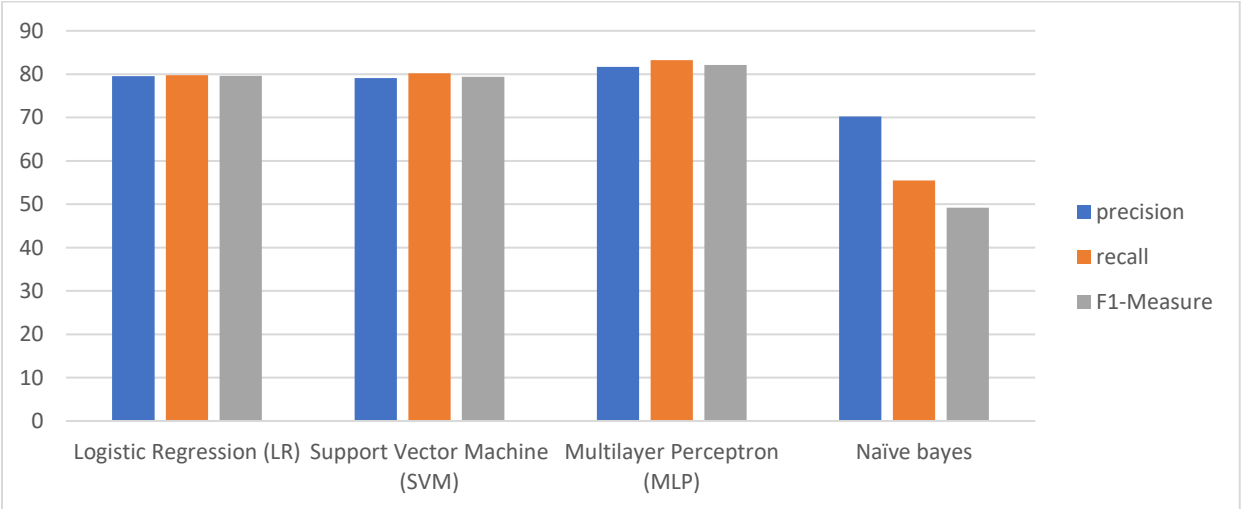


Figure 11.
Performance measure of various classification algorithms.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

In the same vein, from Table 2, the independent multilayer perceptron (MLP) has better metric performance, as shown in Fig. 11, than the rest of the algorithms. The naive Bayes algorithm performed poorer than the rest of the algorithms. It can be seen that this is not only due to having one higher class that can conclude the final results. The multilayer algorithm is a neural network algorithm with three layers and 500 hidden units per layer. From our results, the neural network algorithm is the best of all. Consequently, the authors of this paper proceeded to build the intelligent student opinion mining system interface as shown in Appendix A.

4.2. Discussion

Four base models and a stacking model are constructed in the experimental setup for classifying students' feedback performance. The best base machine learning methods used in classifying comments or observations collected from students' feedback obtained from the learning management system were the Multilayer Perceptron with an accuracy of 81.76% when using unigram and bigram features, which yielded the highest precision, recall, and F-measure. Logistic Regression (LR) achieved an accuracy of 79.05%, which is similar to the results obtained with Support Vector Machine (SVM) using an n-gram (1,2) approach to extract more features from the comments. However, there was a slight difference in their precision and recall values, with LR having 0.7952 precision and 0.7978 recall, while SVM had 0 and 0.802, respectively. Naive Bayes (NB) produced a lower accuracy of 0.50, with precision and recall values of 0.70 and 0.55, respectively, when using n-gram features. As shown in Table 2 and Figure 6, each algorithm's prediction on the same student comment indicates that reliance on a single algorithm is insufficient. NB is biased and predicted the comment incorrectly; thus, it cannot be used as the final classifier. The final result, highlighted with a green background, represents the outcome of the stacking model.

In addition, among the four models employed, the multilayer perceptron had the best classification accuracy, with the highest precision, recall, and F-measure. The Naïve Bayes model's accuracy was slightly lower across the four metrics used to measure performance, as illustrated in the figures. The multilayer perceptron demonstrated superior metrics, as shown in Fig. 13, compared to the other algorithms. The stacking ensemble algorithm, which forms the basis of our analysis, follows the multilayer perceptron. The Naïve Bayes algorithm performed worse than the other models. Findings indicate that no single class can solely determine the final results. The multilayer perceptron is a neural network with three layers and 500 hidden units per layer. Based on our results, the neural network algorithm is the most effective among all.

4.2.1. Comparison with Other Studies

Dhanalakshmi et al. [61] explored opinion mining utilizing supervised learning algorithms (SVM, NB, KNN, and MLP) to discover the polarity of student feedback using the RapidMiner tool. The findings from the experiment showed that the Naïve Bayes algorithm performed better than other learning algorithms in terms of accuracy and recall. The results obtained by the authors differ from the current findings in that the Naïve Bayes algorithm achieved the lowest accuracy, precision, recall, and F-score values. This is because the NB algorithm is biased and predicted the comments incorrectly. Therefore, it was not used as the final algorithm.

Sultana et al. [79] presented a prediction of sentiment analysis on educational data based on the deep learning approach. In the study, results indicated that SVM and MLP deep learning were the best-performing learning methods, achieving 78.75% and 78.33% accuracy, respectively, as well as good performance in terms of their specificity (recall), sensitivity (precision), and ROC curve area. Similarly, Ramadhani and Goo [80] applied basic ML methods in SA, where they only achieved 78.33% and 75.03% accuracy. However, it can be observed that the results obtained from the present study achieved better performance compared to those obtained in Sultana et al. [79] and Ramadhani and Goo [80], respectively, with MLP and the stacking ensemble method employed, having higher accuracy, precision, recall, and F1-score of 81.76% respectively.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

Ashraf et al. [43] investigated the pedagogical dataset using a more effective ensemble classifier called stacking. Moreover, in the study by Ashraf et al., the researchers aimed to compare meta and base classifiers to determine which are more effective for making predictions in an educational context. It was observed that the meta classifier, through stacking, achieved an outstanding performance of 95.65%, while among the three base classifiers, random forest achieved a noteworthy prediction accuracy of 95.76%. Furthermore, when the dataset was further analyzed and subjected to undersampling and oversampling (SMOTE) techniques, the authors found no discrepancy in the results.

Olabode et al. [81] summarized, reviewed, and developed a model for diagnosing head and neck cancer through the effective application of stacked ensemble learning approaches. The classification methods, such as Decision Tree (C4.5), KNN, and Naive Bayes (the base-level classifiers), were combined via a meta-classifier, namely logistic regression, with cross-validation applied. The results showed that the stacked model outperformed individual machine learning models, achieving the best classification accuracy, precision, recall, and F1-score of 98.57%, 98.54%, 97.64%, and 98.09%, respectively, using the Chi-square method for feature selection. The methodology employed in this work aligns with the present study, as four independent classification algorithms, LR, SVM, MLP, and NB were applied together with a stacked ensemble through Decision Tree (C4.5), KNN, Naive Bayes, and the stacked ensemble method. However, the classification accuracy, precision, recall, and F1-score obtained in the present study are lower.

Iyanda and Abegunde [82] employed the ensemble approach (SVM, NB, and LR) to detect sentiment from *Yorùbá* sentences at the sentence level to extract users' opinions from the sentences. It was found that the application of Naïve Bayes outperforms the other machine learning algorithms for language text sentiment analysis, with an accuracy of 65.23% and average precision, recall, and F-score of 62.49%, 60.60%, and 60.69%, respectively. However, the results of the present study, which used individual machine learning algorithms and a stacking ensemble approach (LR, SVM, MLP, and NB) to predict sentiment from student feedback at the sentence level, are better than those of other machine learning algorithms, including the stacking ensemble with MLP, achieving an accuracy of 81.76% and an F-score of 79.05%, respectively.

Sengar et al. [28] addressed the needs of teachers and students to provide opinions to improve communication, education quality, and institutional performance using two (2) machine learning approaches, namely the NLTK toolkit and the random forest algorithm approach. The result of the prediction showed the students' opinion was classified into positive, negative, and neutral classes. This is similar to our findings in that the current study classified students' opinions into three (3) classes. However, the present study achieved an accuracy value of 81.76% using the MLP algorithm in opinion mining of student feedback.

Katragadda et al. [86] investigated opinion mining using SVM, NB, and ANN to search/examine the emotion of the student input, bolstered characterized choices of teaching and learning. The findings revealed that the accuracy of the representation is 88% by using the artificial neural network algorithm. This showed that the ANN algorithm has outperformed all other machine learning algorithms employed in the study. This aligns with the present study, where the ANN MLP technique also performed better than the other machine learning algorithms, including the stacking ensemble technique.

San Lwin and Xiangqian [92] a feedback analysis system using a dataset collected was partitioned into two sets with ratios of 90:10. The researchers developed a model using Naïve Bayes, trained and tested with 10-fold cross-validation and a 10% test dataset. The findings revealed that among the machine learning methods applied, Naïve Bayes provided optimal precision and recall values. In contrast, the present study employed four individual machine learning algorithms and a stacked ensemble method. The results, obtained through cross-validation, showed that the model trained on the split dataset achieved superior performance, with the Multilayer Perceptron (MLP) reaching an accuracy of 81.76%, outperforming other methods in San Lwin and Xiangqian [92] which shows that

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

when using KNN with the Ensemble Stacking 4A technique, you achieve better results, with an accuracy of 98.44%.

Osmanoğlu et al. [93] examined student feedback gathered from a university using six (6) machine learning techniques: multinomial logistic regression, decision tree, multi-layer perceptron, XGBoost, support vector classifier, Gaussian Naive Bayes, and KNN to classify the materials into positive, negative, or neutral sentiments. The results showed that logistic regression performed better than the other five classifiers. This contrasts with the present study, where MLP outperforms all the machine learning algorithms and the stacking ensemble method used in the opinion mining model for predicting student feedback in tertiary institutions.

Ahamad and Ahmad [40] employed individual machine learning algorithms and ensemble methods, particularly the stacking ensemble and voting ensemble, to predict feedback from students' assessments using the WEKA tool. The results of the study indicated that the individual machine learning algorithm, notably the multilayer perceptron (MLP), achieved the highest accuracy of 92.30% compared to other learning algorithms. Additionally, when these individual algorithms were combined, forming a stacking ensemble (Fuzzy, Neural Network, Naïve Bayes, Random Forest, MLP), an accuracy of 93.79% was achieved. The initial results obtained by the author align with the results obtained in the current study, where it is evident that MLP achieved the highest accuracy of 81.76%. Conversely, when the stacking ensemble method was applied (LR, SVM, MLP, and NB), the results differed from the previous work of Suppala and Rao [69] in that the current study achieved an accuracy of 79.05%.

Gebashe et al. [91] analyzed the qualitative input and turned the text feedback into polarity scores using sentiment analysis. Ten different machine learning algorithms were used to predict professor effectiveness based on the polarity ratings of the qualitative and quantitative evaluations. The random forest model outperformed all others with high accuracy, precision, and area under the curve, coming in at 98.87%, 97.71%, and 97.32%, respectively. This contrasts with the result achieved in the present study, where MLP achieved the highest accuracy performance compared to other ML algorithms, including the stacked ensemble method.

Gottipati et al. [10] predicted the student opinion from the Student Feedback Mining Systems (SFMS) developed using a logistic regression approach, where the sentiment extraction stage was identified to achieve a precision of 80.1%, recall of 86.4%, and F-score of 83.5%, which is significantly higher than the IMDB-trained classifier. Consequently, the results obtained from the current study showed lower performance using logistic regression, Naive Bayes, and SVM, but better results in terms of accuracy, precision, recall, and F1-score using MLP.

5. Conclusion

The primary goal of this study was to enhance the functionality of the students' feedback opinion mining system by implementing prediction methods using meta and base classifiers. To forecast university students' perspectives, the study proposed a model and four combined models based on stacking. A new approach was undertaken to analyze the feedback dataset using several fundamental machine learning methods and the stacking ensemble, a more powerful ensemble classifier. Additionally, the main aim was to compare meta and base classifiers to identify which classifiers are most effective at making predictions using the student feedback educational dataset.

All classifiers employed in this experiment or investigation were able to reasonably predict students' outcomes with an accuracy of greater than 70%. Among all the classifiers used, the multilayer perceptron (MLP) was the best-performing algorithm with both balanced and imbalanced datasets. The multilayer perceptron demonstrated better metrics, as illustrated in Fig. 13, than the other algorithms. The stacking method, which forms the basis of our argument, follows the multilayer perceptron algorithm. The naïve Bayes algorithm achieved the poorest performance compared to the other algorithms. Therefore, from Table 2, it can be seen that the classifier does not only have one higher class that can determine the final results.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

Additionally, when multiple Logistic Regression models were applied to the balanced dataset, the accuracy and prediction rate for detecting low performers as well as high performers increased. It was found that among the four base classifiers, MLP achieved a significant prediction accuracy of 81.76% when working with imbalanced data, while the meta-classifier stacking ensemble performed with an accuracy of 79.05% on the balanced data. These values may vary depending on the size and quality of the dataset. The evaluation indicators used in this study (Table 3 and Figs. 8–12) to forecast student feedback comments suggest that machine learning techniques combined within a stacking model enable efficient classification of students' feedback comments. It should also be noted that different classifiers may yield the highest prediction accuracy with different datasets; therefore, the system must be scalable for various circumstances. The primary benefit of this approach is its adaptability to different datasets. To improve prediction accuracy, this methodology could be applied to various fields of data mining and machine learning. The current study is limited by the small sample size and mildly unbalanced data in the dataset; future research should employ larger sample sizes and severely unbalanced data for predicting student comments. Additionally, applying other ensemble techniques such as bagging, boosting, layered generalization, mixtures of experts, and subspace methods could uncover further hidden patterns in educational datasets. Furthermore, large datasets with diverse features may benefit from incremental learning algorithms to address scalability issues.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgements:

The authors would like to acknowledge to Department of Computer Sciences, Nnamdi Azikiwe University, Nigeria, for allowing us to conduct the studies.

References

[1] R. P. S. Kaurav, K. Suresh, S. Narula, and R. Baber, "New education policy: Qualitative (contents) analysis and Twitter mining (sentiment analysis)," *Journal of Content, Community and Communication*, vol. 12, no. 1, pp. 4-13, 2020.

[2] Z. Lazić, A. Đorđević, and A. Gazizulina, "Improvement of quality of higher education institutions as a basis for improvement of quality of life," *Sustainability*, vol. 13, no. 8, p. 4149, 2021. https://doi.org/10.3390/su13084149

[3] E. Oghu, E. Ogbuju, T. Abiodun, and F. Oladipo, "Baseline study of different sentiment analysis computing methods to enhance quality assurance in teaching and learning," *Journal, Advances in Mathematical & Computational Sciences*, vol. 11, no. 3, pp. 1-20, 2023.

[4] Pooja and R. Bhalla, "A review paper on the role of sentiment analysis in quality education," *SN Computer Science*, vol. 3, p. 469, 2022. https://doi.org/10.1007/s42979-022-01366-9

[5] S. Bloxham and P. Boyd, *Developing effective assessment in higher education: A practical guide: A practical guide*. UK: McGraw-Hill Education, 2007.

[6] J. Hattie and H. Timperley, "The power of feedback," *Review of Educational Research*, vol. 77, no. 1, pp. 81-112, 2007. https://doi.org/10.3102/003465430298487

[7] C. Plank, H. Dixon, and G. Ward, "Student voices about the role feedback plays in the enhancement of their learning," *Australian Journal of Teacher Education (Online)*, vol. 39, no. 9, pp. 98-110, 2014.

[8] A. R. Lawal, S. Aina, S. Ayeni, S. D. Okegbile, and A. I. Oluwaranti, "Systems. Theory approach to engineering education: A review of feedback structures," in *Proceedings of 12th International Multi-Conference on ICT Applications, Application of Information Communication and Technologies in Teaching, Research and Administration Conference*, 2018.

[9] L. Harvey, "Student feedback: A report to the Higher Education Funding Council for England," Research Report, Centre for Research into Quality, The University of Central England, Birmingham, United Kingdom, 2001.

[10] S. Gottipati, V. Shankararaman, and S. Gan, "A conceptual framework for analyzing students' feedback," in *2017 IEEE Frontiers in Education Conference (FIE) (pp. 1-8). IEEE*, 2017.
[11] W. Harlen and M. James, "Assessment and learning: Differences and relationships between formative and summative assessment," *Assessment in education: Principles, Policy & Practice*, vol. 4, no. 3, pp. 365-379, 1997. https://doi.org/10.1080/0969594970040304
[12] C. S. Nair, A. Patil, and P. Mertova, "Re-engineering graduate skills–a case study," *European Journal of Engineering Education*, vol. 34, no. 2, pp. 131-139, 2009. https://doi.org/10.1080/03043790902829281
[13] M. S. Alade and J. M. Nwankpa, "Sentiment analysis of nigerian students' tweets on education: A data mining approach'," *International Journal of Computer*, vol. 45, no. 1, pp. 1-27, 2022.
[14] P. W. Foltz and M. Rosenstein, "Analysis of a large-scale formative writing assessment system with automated feedback," in *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, 2015.
[15] R. Menaha, R. Dhanaranjani, T. Rajalakshmi, and R. Yogarubini, "Student feedback mining system using sentiment analysis," *International Journal of Computer Application Technology and Research*, vol. 6, no. 1, pp. 1–69, 2017.
[16] D. D. Dsouza, D. P. N. Deepika, E. J. Machado, and N. Adesh, "Sentimental analysis of student feedback using machine learning techniques," *International Journal of Recent Technology and Engineering*, vol. 8, no. 14, pp. 986-991, 2019.
[17] R. Sara, F. Fabbro, and M. Eleonora, "Collective feedback as a formative assessment practice in an e-learning platform for teachers' professional development," *Q-Times Webmagazine*, vol. 2, no. 1, pp. 563-576, 2023.
[18] A. I. M. Elfeky, T. S. Y. Masadeh, and M. Y. H. Elbyaly, "Advance organizers in flipped classroom via e-learning management system and the promotion of integrated science process skills," *Thinking Skills and Creativity*, vol. 35, p. 100622, 2020. https://doi.org/10.1016/j.tsc.2019.100622
[19] L. McKinney, A. B. Burridge, M. M. Lee, G. V. Bourdeau, and M. Miller-Waters, "Incentivizing full-time enrollment at community colleges: What influences students' decision to take more courses?," *Community College Review*, vol. 50, no. 2, pp. 144-170, 2022. https://doi.org/10.1177/00915521211061416
[20] X. Chen, G. Cheng, F. L. Wang, X. Tao, H. Xie, and L. Xu, "Machine and cognitive intelligence for human health: Systematic review," *Brain Informatics*, vol. 9, p. 5, 2022. https://doi.org/10.1186/s40708-022-00153-9
[21] T. Shaik *et al.*, "A review of the trends and challenges in adopting natural language processing methods for education feedback analysis," *Ieee Access*, vol. 10, pp. 56720-56739, 2022. https://doi.org/10.1109/ACCESS.2022.3177752
[22] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799–106810, 2020. https://doi.org/10.1109/ACCESS.2020.3000739
[23] D. Zimbra, H. Chen, and R. F. Lusch, "Stakeholder analyses of firm-related Web forums: Applications in stock return prediction," *ACM Transactions on Management Information Systems*, vol. 6, no. 1, pp. 1-38, 2015. https://doi.org/10.1145/2675693
[24] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information Systems Research*, vol. 19, no. 3, pp. 291-313, 2008. https://doi.org/10.1287/isre.1080.0193
[25] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2010.
[26] M. Bansal, S. Verma, K. Vig, and K. Kakran, "Opinion mining from student feedback data using supervised learning algorithms," in *International Conference on Image Processing and Capsule Networks (pp. 411-418). Cham: Springer International Publishing*, 2022.
[27] I. Sindhu, S. M. Daudpota, K. Badar, M. Bakhtyar, J. Baber, and M. Nurunnabi, "Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation," *IEEE Access*, vol. 7, pp. 108729-108741, 2019. https://doi.org/10.1109/ACCESS.2019.2928872
[28] N. S. Sengar, K. Chourey, S. Bajaj, and S. Abinayaa, "Student feedback prediction using machine learning," *International Journal of Advanced Research, Ideas and Innovations in Technology*, vol. 2, no. 5, pp. 818–821, 2019.
[29] L. Amusa, W. Yahya, and A. O. Balogun, "Data mining of Nigerian's sentiments on the administration of federal government of Nigeria," *Annals. Computer Science Series*, vol. 14, no. 2, pp. 69-75, 2016.
[30] B. Ngwira, B. Gobin-Rahimbux, and N. G. Sahib, "A Deep-learning framework for analysing students' review in higher education," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 8462575, 2023. https://doi.org/10.1155/2023/8462575
[31] T. W. Edgar and D. O. Manz, *Machine learning. In T. W. Edgar & D. O. Manz (Eds.), Research methods for cybersecurity*. Amsterdam, Netherlands: Elsevier, 2017.
[32] M. Hammad, A. M. Iliyasu, A. Subasi, E. S. Ho, and A. A. Abd El-Latif, "A multitier deep learning model for arrhythmia detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-9, 2020.
[33] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014. https://doi.org/10.1016/j.asej.2014.04.011
[34] S. Bhavitha, K. G. Srinivasa, and S. Krishnamurthy, "Sentiment analysis approaches for student feedback," *International Journal of Computer Applications*, vol. 168, no. 12, pp. 31–36, 2017.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1149-1180, 2025
DOI: 10.55214/2576-8484.v9i10.10606

[35] M. A. Qureshi *et al.*, "Sentiment analysis of reviews in natural language: Roman Urdu as a case study," *IEEE Access*, vol. 10, pp. 24945-24954, 2022. https://doi.org/10.1109/ACCESS.2022.3150172

[36] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Systems with Applications*, vol. 97, pp. 205-227, 2018. https://doi.org/10.1016/j.eswa.2017.12.020

[37] Y. Reddy, P. Viswanath, and B. E. Reddy, "Semi-supervised learning: A brief review," *International Journal of Engineering Technology*, vol. 7, no. 1.8, pp. 81-85, 2018.

[38] A. D. Vergaray, J. C. H. Miranda, J. B. Cornelio, A. R. L. Carranza, and C. F. P. Sánchez, "Predicting the depression in university students using stacking ensemble techniques over oversampling method," *Informatics in medicine unlocked*, vol. 41, p. 101295, 2023. https://doi.org/10.1016/j.imu.2023.101295

[39] A. Barham, M. S. Ismail, M. Hermana, E. Padmanabhan, Y. Baashar, and O. Sabir, "Predicting the maturity and organic richness using artificial neural networks (ANNs): A case study of Montney Formation, NE British Columbia, Canada," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3253-3264, 2021.

[40] M. Ahamad and N. Ahmad, "Students' knowledge assessment using the ensemble methods," *International Journal of Information Technology*, vol. 13, pp. 1025-1032, 2021. https://doi.org/10.1007/s41870-020-00593-8

[41] S. Sagayaraj and M. Santhoshkumar, "Heterogeneous ensemble learning method for personalized semantic web service recommendation," *International Journal of Information Technology*, vol. 12, pp. 983-994, 2020. https://doi.org/10.1007/s41870-020-00479-9

[42] A. Rahman and S. Tasnim, "Ensemble classifiers and their applications: a review. arXiv preprint arXiv: 14044088," 2014.

[43] M. Ashraf, M. Zaman, and M. Ahmed, "Using ensemble StackingC method and base classifiers to ameliorate prediction accuracy of pedagogical data," *Procedia Computer Science*, vol. 132, pp. 1021-1040, 2018. https://doi.org/10.1016/j.procs.2018.05.018

[44] N. M. Nhleko, O. J. Aroba, and C. T. Chisita, "A systematic review of information and communication technologies (ICTs) on student motivation: Researchers' reflections on a selected higher education institution (HEIs)," *Global Knowledge, Memory and Communication*, vol. 74, no. 11, pp. 77-100, 2025.

[45] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15-21, 2013. https://doi.org/10.1109/MIS.2013.30

[46] K. Lundqvist, T. Liyanagunawardena, and L. Starkey, "Evaluation of student feedback within a MOOC using sentiment analysis and target groups," *The International Review of Research in Open and Distributed Learning*, vol. 21, no. 3, pp. 140-156, 2020. https://doi.org/10.19173/irrodl.v21i3.4783

[47] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 23, p. e5909, 2021. https://doi.org/10.1002/cpe.5909

[48] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: A tertiary study," *Artificial Intelligence Review*, vol. 54, pp. 4997-5053, 2021. https://doi.org/10.1007/s10462-021-09973-3

[49] M. Anandarajan, C. Hill, and T. Nolan, *Text preprocessing. Practical text analytics: Maximizing the value of text data.* Cham: Springer, 2018.

[50] G. Pathak and S. Warpade, "Impact of lockdown due to COVID-19 on consumer behaviour while selecting retailer for essential goods," *Pathak, GP, & Warpade, S.(2020, July 31). Impact of Lockdown due to COVID*, vol. 19, pp. 282-289, 2020.

[51] O. D. Olanloye, P. A. Idowu, A. E. Adeniyi, A. A. Badmus, and O. J. Aroba, "Virtual learning environment on satisfaction and academic performance of students in institutions of higher learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 39, no. 1, pp. 258–271, 2025.

[52] N. Nikolić, O. Grljević, and A. Kovačević, "Aspect-based sentiment analysis of reviews in the domain of higher education," *The Electronic Library*, vol. 38, no. 1, pp. 44-64, 2020. https://doi.org/10.1108/EL-06-2019-0140

[53] F. S. Dolianiti *et al.*, "Sentiment analysis on educational datasets: A comparative evaluation of commercial tools," *Educational Journal of the University of Patras UNESCO Chair*, 2019. https://doi.org/10.26220/une.2987

[54] W. Sun and B. Trevor, "A stacking ensemble learning framework for annual river ice breakup dates," *Journal of Hydrology*, vol. 561, pp. 636-650, 2018. https://doi.org/10.1016/j.jhydrol.2018.04.008

[55] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996. https://doi.org/10.1007/BF00058655

[56] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2014.

[57] N. Zainuddin and A. Selamat, "Sentiment analysis using support vector machine," in *2014 International Conference on Computer, Communications, and Control Technology (I4CT) (pp. 333-337). IEEE*, 2014.

[58] C. Chatterjee and K. Chakma, "A comparison between sentiment analysis of student feedback at sentence level and at token level," *International Journal of Computer Science Network*, vol. 4, no. 3, pp. 1-7, 2015.

[59] M. Sivakumar and U. S. Reddy, "Aspect based sentiment analysis of students opinion using machine learning techniques," in *2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 726-731). IEEE*, 2017.

[60] K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using lexicon based approach," in *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) (pp. 149-154). IEEE*, 2017.

[61] V. Dhanalakshmi, D. Bino, and A. M. Saravanan, "Opinion mining from student feedback data using supervised learning algorithms," in *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-5). IEEE*, 2016.

[62] N. Altrabsheh, M. M. Gaber, and M. Cocea, "SA-E: Sentiment analysis for education," in *Proceedings of the International Conference on Intelligent Decision Technologies (KES-IDT) (Vol. 255, pp. 353–362). Amsterdam, Netherlands: IOS Press*, 2013.

[63] F. F. Balahadia, M. C. G. Fernando, and I. C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis," in *2016 IEEE Region 10 Symposium (TENSYMP) (pp. 95-98). IEEE*, 2016.

[64] A. El-Halees, "Mining opinions in user-generated contents to improve course evaluation," in *International Conference on Software Engineering and Computer Systems (pp. 107-115). Berlin, Heidelberg: Springer Berlin Heidelberg*, 2011.

[65] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Sentiment analysis: Towards a tool for analysing real-time students feedback," in *2014 IEEE 26th International Conference on Tools with Artificial Intelligence (pp. 419-423). IEEE*, 2014.

[66] R. Jindal and M. D. Borah, "A survey on educational data mining and research trends," *International Journal of Database Management Systems*, vol. 5, no. 3, p. 53, 2013.

[67] I. A. Kandhro, M. A. Chhajro, K. Kumar, H. N. Lashari, and U. Khan, "Student feedback sentiment analysis model using various machine learning schemes: A review," *Indian Journal of Science and Technology*, vol. 12, no. 14, pp. 1-9, 2019. https://doi:10.17485/ijst/2019/v12i14/143243

[68] R. Jena, "Understanding academic achievement emotions towards business analytics course: A case study among business management students from India," *Computers in Human Behavior*, vol. 92, pp. 716-723, 2019.

[69] K. Suppala and N. Rao, "Sentiment analysis using Naïve Bayes classifier," *International Journal of Innovation Technology Explorative Engineering*, vol. 8, no. 8, pp. 264–269, 2019.

[70] M. Stapel, Z. Zheng, and N. Pinkwart, "An ensemble method to predict student performance in an online math learning environment," *Proceedings of the International Educational Data Mining Societ*, 2016.

[71] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *Journal of Applied Research in Higher Education*, vol. 10, no. 1, pp. 61-75, 2018. https://doi.org/10.1108/JARHE-09-2017-0113

[72] P. Kumari, P. K. Jain, and R. Pamula, "An efficient use of ensemble methods to predict students academic performance," in *2018 4th International Conference on Recent Advances in Information Technology (RAIT) (pp. 1-6). IEEE*, 2018.

[73] S.-S. M. Ajibade, N. Bahiah Binti Ahmad, and S. Mariyam Shamsuddin, "Educational data mining: Enhancement of student performance model using ensemble methods," in *IOP Conference Series: Materials Science and Engineering (Vol. 551, No. 1, p. 012061). IOP Publishing*, 2019.

[74] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," *Journal of Hydrology*, vol. 598, p. 126266, 2021. https://doi.org/10.1016/j.jhydrol.2021.126266

[75] Y. Li, Z. Liang, Y. Hu, B. Li, B. Xu, and D. Wang, "A multi-model integration method for monthly streamflow prediction: Modified stacking ensemble strategy," *Journal of Hydroinformatics*, vol. 22, no. 2, pp. 310-326, 2020.

[76] L. Wang *et al.*, "Improving the robustness of beach water quality modeling using an ensemble machine learning approach," *Science of The Total Environment*, vol. 765, p. 142760, 2021. https://doi.org/10.1016/j.scitotenv.2020.142760

[77] K. A. Kesavaram, N. M. Periya, and D. Indra, "Customer feedback evaluation system using feature-based opinion mining," Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India, 2016.

[78] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in *2017 International Conference on Research and Innovation in Information Systems (ICRIIS) (pp. 1-6). IEEE*, 2017.

[79] J. Sultana, N. Sultana, K. Yadav, and F. Alvarez, "Prediction of sentiment analysis on educational data on a deep learning approach," in *Proceedings of the 21st Saudi Computer Society National Computer Conference (NCC 2018) (pp. 1–5). Riyadh, Saudi Arabia*, 2018.

[80] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," in *2017 7th International Annual Engineering Seminar (InAES) (pp. 1-4). IEEE*, 2017.

[81] O. O. Olabode, A. O. Adetunmbi, F. Akinbohun, and A. Ambrose, "Stacked ensemble model for diagnosis of head and neck cancer in primary healthcare system," in *Proceedings of the 13th International Multi-Conference on ICT Applications: Application of Information Communication and Technologies in Teaching, Research and Administration Conference (pp. 5–10). Lagos, Nigeria*, 2019.

[82] A. R. Iyanda and O. Abegunde, "Predicting sentiment in yorùbá written texts: A comparison of machine learning models," in *Proceedings of SAI Intelligent Systems Conference (pp. 416-431). Cham: Springer International Publishing*, 2020.

[83] M. Kavitha and A. J. Kumar, "Contextual-cognitive attentive LSTM approach for sentiment analysis of Tamil review," *Turkish Journal of Computer and Mathematics Education (TuRCOMAT)*, vol. 11, no. 3, pp. 2493–2504, 2020.

[84] J.-a. P. Lalata, B. Gerardo, and R. Medina, "A sentiment analysis model for faculty comment evaluation using ensemble machine learning algorithms," in *Proceedings of the 2019 International Conference on big Data Engineering*, 2019.

[85] M. Wook *et al.*, "Opinion mining technique for developing student feedback analysis system using lexicon-based approach (OMFeedback)," *Education and Information Technologies*, vol. 25, no. 4, pp. 2549-2560, 2020.

[86] S. Katragadda, V. Ravi, P. Kumar, and G. J. Lakshmi, "Performance analysis on student feedback using machine learning algorithms," in *2020 6th international conference on advanced computing and communication systems (ICACCS) (pp. 1161-1163). IEEE*, 2020.

[87] S. Qaiser, N. Yusoff, R. Ali, M. A. Remli, and H. K. Adli, "A comparison of machine learning techniques for sentiment analysis," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 3, pp. 1738-1744, 2021.

[88] N. Nidhi, M. Kumar, and S. Agarwal, "Comparative analysis of heterogeneous ensemble learning using feature selection techniques for predicting academic performance of students," in *2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST) (pp. 212-217). IEEE*, 2021.

[89] S. Verma, R. K. Yadav, and K. Kholiya, "A scalable machine learning-based ensemble approach to enhance the prediction accuracy for identifying students at-risk," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, pp. 185-192, 2022.

[90] B. F. Roaring, F. F. Patacsil, and J. M. Parrone, "Analyzing Pangasinan state university student's faculty teaching performance rating using text mining technique," *WSEAS Transactions on Information Science and Applications*, vol. 19, pp. 161-170, 2022.

[91] K. Gebashe, Q. Mthethwa, M. Nzimande, F. Mlima, and O. J. Aroba, "The integration of virtual reality chemistry laboratory: A case study Durban university of technology," in *International Conference on Innovations in Bio-Inspired Computing and Applications (pp. 479-492). Cham: Springer Nature Switzerland*, 2023.

[92] C. San Lwin and W. Xiangqian, "Myanmar handwritten character recognition from similar character groups using K-means and convolutional neural network," in *2020 IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE) (pp. 172-176). IEEE*, 2020.

[93] U. Ö. Osmanoğlu, O. N. Atak, K. Çağlar, H. Kayhan, and T. Can, "Sentiment analysis for distance education course materials: A machine learning approach," *Journal of Educational Technology and Online Learning*, vol. 3, no. 1, pp. 31-48, 2020. https://doi.org/10.31681/jetol.663733