

Scholarship Award Recommendation System Using Predictive Modeling and Ensemble Learning

Bunmi Janet Bambi

Department of Computer Science, Adeleke University, Ede, Nigeria
E-mail: bambi.bunmi@adelekeuniversity.edu.ng
ORCID iD: <https://orcid.org/0009-0002-3272-249X>

Olusegun Lala

Department of Information Technology, Adeleke University, Ede, Nigeria
E-mail: olusegun.lala@adelekeuniversity.edu.ng
ORCID iD: <https://orcid.org/0000-0002-5766-0072>

Akintoye Onamade

Department of Computer Science, Adeleke University, Ede, Nigeria
E-mail: onamadeakintoye@gmail.com
ORCID iD: <https://orcid.org/0000-0002-5307-4754>

Oludayo Oduwale

Department of Computer Science, Faculty of Computing, University of Ilesa, Ilesa, Nigeria
E-mail: dayooduus@yahoo.com
ORCID iD: <https://orcid.org/0000-0002-8298-3826>

Anozie Onyezewe*

Department of Computer Science, Adeleke University, Ede, Nigeria
E-mail: anonyezewe@outlook.com
ORCID iD: <https://orcid.org/0009-0004-0418-0157>

*Corresponding author

Received: 31 January 2026; Revised: 10 April 2026; Accepted: 09 May 2026; Published: 08 June 2026

Abstract: Education is crucial for personal and economic growth, but financial challenges in developing countries hinder equitable academic success. NGOs administer scholarship programs to empower underprivileged individuals, a crucial step towards the attainment of Sustainable Development Goal 4, which aims to provide inclusive and equitable quality education for all. This study proposes a novel Scholarship Award Recommendation System that leverages predictive modelling and ensemble learning to identify deserving students for scholarship awards. The system utilizes a robust ensemble model that combines the strengths of Quadratic Discriminant Analysis (QDA), Random Forest (RF), and Extra Trees (ET) to predict students' academic performance. Additionally, we incorporate answers from the General Mental Health Questionnaire (GHQ-12). The GHQ-12 responses are pre-processed using a binary scoring approach (0-0-1-1) and integrated as predictive variables alongside academic and demographic features. We apply this framework to a case study of Nigerian university students in partnership with Springtime Development Foundation. The results indicate that incorporating GHQ-12 features significantly enhances prediction accuracy, with QDA, RF, and ET achieving accuracy scores of 0.90, 0.86, and 0.89, respectively. Statistical analysis using a t-test confirms the relevance of GHQ-12 features, with a p-value of 0.0013 establishing a significant correlation between student performance and mental health status. The study showed the effectiveness of the ensemble model in accurately predicting students' academic performance. It highlights the significance of incorporating variables from the GHQ-12 into the predictive model, indicating mental health as a crucial factor for predicting academic performance which in turn enhances the performances of the Classification Models considered.

Index Terms: Machine Learning, Classification Models, Prediction, Students' Academic performance, GHQ-12, Quadratic Discriminant Analysis, Random Forest, Extra Trees, Recommender System.

1. Introduction

Education is a lifelong process that prepares individuals starting from birth and continuing throughout one's life to develop their potential and achieve success. It is a fundamental need for both developed and developing countries' growth and development [1]. Financial issues in developing countries significantly impact education, affecting the quality and accessibility of learning opportunities. These issues can lead to limited access to quality education, learning materials, textbooks, and other resources, such as computers or laboratory equipment, which can hinder the learning process. Therefore, it is crucial to effectively address these financial issues to improve the educational system in these countries [2,3].

NGOs are non-profit organizations that address social, environmental, and humanitarian issues. They are often established by individuals or groups to advocate for causes, provide services, or undertake development projects [2]. Examples include the Springtime Development Foundation (SDF), which sponsors students at Adeleke University Ede in Nigeria, covering tuition, food, and accommodation. The foundation's goal is to ensure students' success by maintaining high academic performance, justifying its investment. These scholarships act as a lifeline for underprivileged communities, giving people the chance to escape poverty cycles and reach their full potential. Nonetheless, overseeing scholarship programs is not without its difficulties. NGOs frequently get a high volume of applications, each of which reflects a distinct set of circumstances, credentials, and goals. A thorough and impartial screening procedure is necessary in light of the volume of applications received to guarantee that scholarships are given to students who have the best chance of succeeding academically and making a positive effect on society [3].

Scholarship selection has traditionally been a manual process, prone to human bias and time-consuming. With the exponential growth of applicants, it's becoming increasingly challenging for NGOs to evaluate each application thoroughly [4]. Advancements in machine learning and artificial intelligence (AI) have shown promise in revolutionizing decision-making processes [5]. AI can automate the initial screening of applications, provide valuable insights into applicant qualifications, and aid NGOs in making more informed and objective decisions. This technology has the potential to enhance the scholarship selection process and improve the overall decision-making process for NGOs [3].

With machine learning algorithms, it is possible to develop predictive models that can identify 'students-at-risk' early in their academic journey [6]. These models can analyze a multitude of variables, including demographic information, prior academic performance, socio-economic indicators, and engagement metrics, to provide timely interventions and tailored support. The integration of a machine learning approach to scholarship decision-making holds the potential to streamline the evaluation process, reduce administrative burdens, and increase the overall efficacy of scholarship programs. By leveraging data-driven algorithms, NGOs can identify patterns and indicators that correlate with academic success, socio-economic needs, and potential for positive community impact [3].

This study aims to develop a predictive model that leverages a heterogeneous ensemble of machine learning algorithms to forecast students' academic performance thus facilitating scholarship decisions by analysing student data from select Nigerian universities. By combining the strengths of Random Forest, Extra Trees, and Quadratic Discriminant Analysis, we create a robust model that predicts students' academic outcomes. The ultimate goal is to inform recommendation decisions for students, enabling timely interventions and support. Our approach utilizes a range of academic and demographic variables, as well as mental health indicators from the General Mental Health Questionnaire (GHQ-12), to identify the most significant predictors of academic success. The proposed model has the potential to revolutionize academic advising, scholarship allocation, and student support services, providing a data-driven framework for institutions to promote student success and academic excellence.

2. Related Works

In a study by Eyman and Dilek [7], academic achievement at the College of Computer Science and Information Technology (CCSIT) in Saudi Arabia was predicted using WEKA, or Waikato Environment for Knowledge Analysis. Three decision tree classifiers—J48, REPTree, and Random Tree (RT)—were applied to the student dataset in order to predict the students' achievement in their graduation year. The results showed that the J48 classifier outperformed both REPTree and RT in terms of predicting students' academic performance, with an accuracy of 69.3%. Important factors that were found to have an impact on academic performance included the Prep year CGPA, the Computer Skills course, Communication Skills course, and Math.

Adejo et al. [8] conducted a study on predicting students' academic progress using data from UWS. The study employed a quantitative research methodology with data samples from 141 University of the West of Scotland students. PCA was used to identify the significant factors required for prediction. The study showed a relationship between the variables that were found and enhancements to the performance prediction model. The results indicate that the ensemble classifier outperforms the base classifiers in terms of accuracy, precision, recall, and RMSE. According to the study, the hybrid model ensemble classifier outperforms other ensemble classifiers in terms of performance.

Adekitan and Salau [9] used DM techniques in a different study to forecast the graduation CGPA of engineering students at Covenant University in Nigeria. Consequently, six distinct DM algorithms were used. Logistic regression yielded the maximum accuracy of 89.15% for the classification strategy. According to the findings, engineering students'

ultimate CGPA for their fifth year of study may be fairly predicted based on their CGPA for the first three years of study.

In the study conducted by Zabriskie et al. [10] three independent variables were utilized: the performance on three midterm examinations, the cumulative grade point average, and the grades of students in four required courses. In order to forecast academic success using the LIME algorithm, the study illustrated the most important components needed for each student. This presented the use of classification algorithms to predict academic accomplishment in students. Random forests and logistic regression were used to forecast students' grades. In order to better understand the classification model, which was employed to identify students who were at risk of failing, the LIME technique highlights the characteristics that were critical in predicting a student's future performance. The provided approach gauges the model's correctness using a 10-fold cross-validation.

Zeineddine et al. [11] found the optimal hyperparameters and classification model using AutoML. The Auto-Weka search strategy was used in conjunction with the hyper-parameter optimization option. An automated machine learning procedure iterated over the list of prediction approaches and the accompanying hyper-parameter values to find the model with the best accuracy. The outcome of this process is the Ensemble Model, which integrates many approaches like Decision Trees, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors, K-Means Clustering, Artificial Neural Network, and Logistic Regression.

Aybek and Okur [12] created a model employing an Artificial Neural Network (ANN) in order to forecast how well students will perform in the course overall and on final exams. The criteria that were applied were the midterm exercise, educational history, and demographic information. The student data was gathered and assessed using Multilayer Perceptron (MLP) and Radial Basis Function (RBF). The results show that the factors pertaining to the students' midterm exam scores, university entrance exam scores, and year of secondary school graduation had the most importance for explaining their final exam scores and success rates. The result shows that predictions made using Multilayer Perceptron MLP provided more accurate results than predictions made using RBF, with an error rate difference of over 4%.

Xu et al.'s study [13] employed machine learning algorithms to forecast academic success based on patterns in internet usage. The association between internet usage and academic accomplishment was found, and decision trees, neural networks, and support vector machines were used to predict academic outcomes from internet usage data. The volume of internet traffic, the frequency of connections, and the amount of time spent online by 4,000 pupils were all analysed. The true difference between usage characteristics and performance level was found using the Mann-Whitney U test, and the association between students' internet usage characteristics and academic accomplishment was measured using Spearman's correlation coefficient. The results of the study showed that pupils performed well when it came to how frequently they connected to the internet, but poorly when it came to upload and download volumes. Hence, the amount and frequency of Internet connections had a big influence on how well pupils performed academically.

A deep artificial neural network with sophisticated features was created and employed by Waheed et al. [14] to analyse a collection of precisely crafted characteristics that were taken from clickstream data in virtual learning environments. The objective was to identify pupils who could be at risk so that timely assistance might be provided. The results showed that the recommended model had a classification accuracy of between 84% and 93%. The research demonstrated that the deep artificial neural network outperformed the baseline models of logistic regression and support vector machines. Specifically, the support vector machine's accuracy range was 79.95% to 89.14%, while the logistic regression's accuracy range was 79.82% to 85.60%. The findings highlighted the important impact of incorporating assessment-related data as well as historical data.

According to Asif et al. [15], academic success at graduation time for undergraduate students in a four-year program may be predicted. The pre-university component of the student as well as transcript data from the first and second years of the degree program were utilized with a reasonable degree of accuracy. Based on the graduating CPGPA objective criterion, they divided the data into five categories in their study: A (90–100%), B (80–89%), C (70–79%), D (60–69%), and E (50–59%). For 210 undergraduate students at the University of Engineering in Pakistan, they utilized X-means, Naive Bayes, K-Nearest Neighbor classifier, Neural Networks, Rule Induction, Decision Tree, Random Forest, and information technology department methodologies. The findings showed that the Naive Bayes was regarded as the best (83.65%).

Hussain et al. [16] carried out a study on the use of machine learning to predict secondary and intermediate students' academic performance. Regression modeling and classification methods were employed in the study. In the study, there was an improvement in classification accuracy and a decrease in root mean square error. Thirty carefully selected features from student academic records were tagged and used to construct a regression model and a decision tree-based classifier. The regression model oversaw predicting marks, whilst the classification technique focused on predicting grades. Ten-fold Cross Validation was used to evaluate the model's accuracy. After that, a careful analysis of the outcomes generated by these models was done. According to the findings, machine learning technology was efficient and pertinent in accurately predicting student performance.

3. Predictive Model

Machine learning, a subfield of artificial intelligence (AI), is concerned with creating statistical models and algorithms that let computers learn and make judgments on their own without needing to be explicitly programmed to do so [17]. Conventional programming involves human writers creating precise instructions for a computer to follow [18],

while, in contrast, a computer learns from data via machine learning. Based on the data it receives, it recognizes trends, decides, and improves its performance.

Predicting the category or class to which a given input belongs based on its properties is the aim of classification, a fundamental job in machine learning. This method of predictive modeling works best when the target variable is categorical—that is, when it represents several classes or categories. The goal of classification is to predict categorical target variables, which represent discrete classes or labels. For instance, it may be used to figure out if a patient is at high risk of heart disease or if an email is spam. Classification algorithms acquire the capability to link input information to one of the predefined classes. In essence, classification algorithms play a crucial role in automating decision-making processes, making predictions from input data in a diverse array of real-world applications.

This research examines a range of 17 machine learning classification algorithms, which include Decision Tree, Support Vector Classifier (SVC), Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Gradient Boosting, AdaBoost, Gaussian Naive Bayes (NB), Bernoulli Naive Bayes, Ridge Classifier, Multi-Layer Perceptron (MLP) Classifier, Quadratic Discriminant Analysis (QDA), Calibrated Classifier CV, Histogram Gradient Boosting, Passive Aggressive Classifier, Linear SVC, and Extra Trees. The three (3) best performing algorithms for our classification task are briefly explained in the following

3.1. Quadratic Discriminant Analysis

Quadratic Discriminant Analysis (QDA) is a supervised machine learning algorithm used for classification and discrimination. It is a generalization of Linear Discriminant Analysis (LDA) that allows for non-linear separation of classes. QDA estimates the parameters of the underlying Gaussian distributions and uses quadratic boundaries to separate the classes. It is particularly useful when the classes have different covariance matrices, and the data is not linearly separable. QDA has been widely used in various fields, including finance, marketing, and healthcare.

The prediction formula for QDA is shown as:

$$\hat{y} = \arg \max_k \left[-\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log \pi_k \right] \quad (1)$$

Where $\hat{y}$ is the predicted class label, k indexes the classes, x is the feature vector, μ_k is the mean vector for class k , Σ_k is the covariance matrix for class k , and π_k is the prior probability for class k .

3.2. Random Forest

Random Forest is a supervised ensemble learning algorithm used for classification and regression tasks. It combines multiple decision trees to produce a more accurate and stable prediction model. Random Forest works by constructing a collection of decision trees, each trained on a random subset of the training data and features. The algorithm then aggregates the predictions from individual trees to produce a final prediction. Random Forest is known for its ability to handle high-dimensional data, reduce overfitting, and provide feature importance scores.

The prediction formula for Random Forest is shown as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (2)$$

Where $\hat{y}$ is the predicted class label or regression value, B is the number of decision trees, $T_b(x)$ is the prediction from the b th decision tree, x is the feature vector.

3.3. Extra Trees

Extra Trees (ET) is a supervised ensemble learning algorithm used for classification and regression tasks. Similar to Random Forest, but with some key differences, Extra Trees combines multiple decision trees to produce a more accurate and stable prediction model. Unlike Random Forest, Extra Trees does not use bagging or feature randomness; instead, it uses all the training data and features to train each tree. Extra Trees is known for its ability to handle high-dimensional data, reduce overfitting, and provide feature importance scores.

The prediction formula for Extra Trees is similar to Random Forest in (2). Extra Trees has been used in various applications, including image classification, bioinformatics, and predictive modeling.

4. Ensemble Learning

An ensemble model combines the predictions of multiple individual models to improve overall performance. Rather than relying on a single model's prediction, ensemble methods base their predictions on the aggregate of several models to produce a final prediction that often outperforms any individual model. The heterogeneous voting ensemble method is a technique used to combine the predictions of different base models, in this case, QDA, RF, and ET, to produce a single, more accurate prediction. Each base model is trained on the same dataset and receives different weights

in the ensemble based on their individual performance. The final prediction is determined by a majority vote, where each model's prediction is treated as a vote.

Algorithm 1: Weighted Voting Heterogeneous Ensemble Model

Data: List of base models ($M1, M2, \dots, Mk$), weights ($w1, w2, \dots, wk$), test sample x

Result: Ensemble prediction for test sample x

1. Initialize:
 - base models ($M1, M2, \dots, Mk$)
 - weights for each model ($w1, w2, \dots, wk$) based on their performance
 - threshold for decision (optional)
2. **for each** test sample x **do**
3. **for each** model M_i in ($M1, M2, \dots, Mk$) **do**
4. $y_i = M_i.predict(x)$
5. **end**
6. $y_{ensemble_pred} = \sum(w_i * y_i)$ for $i = 1$ to k
7. **end**
8. **return** ($y_{ensemble_pred}$)

5. Recommender System

The predictive model's output is utilized by the recommender system to categorize students into three groups: Recommended for Scholarship, Waitlisted, or Not Recommended. This categorization is based on predicted academic performance and other relevant factors. The recommender system's primary goal is to provide a data-driven approach to scholarship award decisions, ensuring that deserving students receive timely support. The recommender system's workflow is presented in Figure 1.

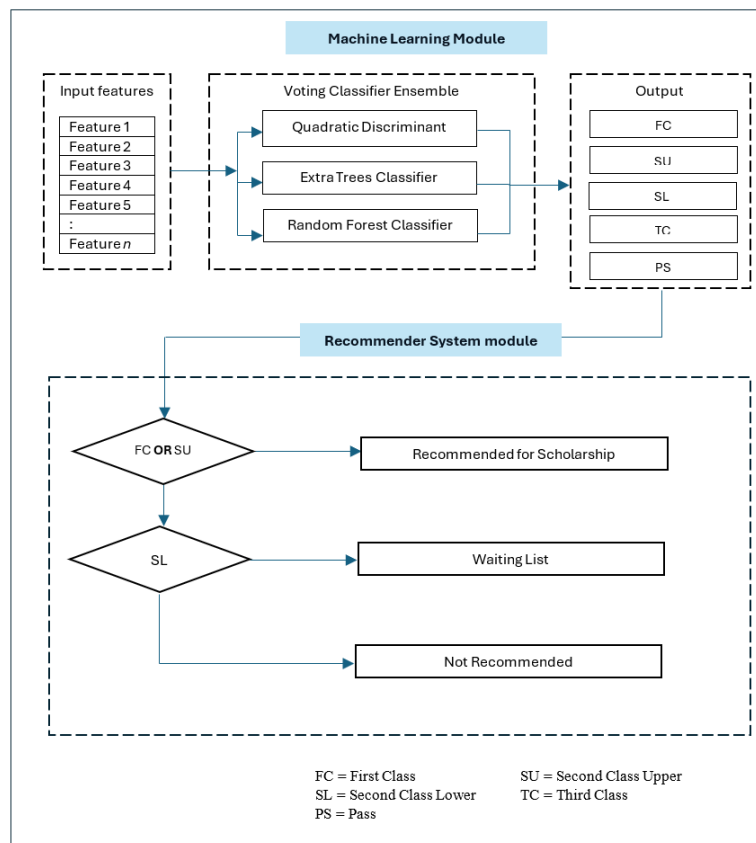


Fig. 1. Scholarship recommendation system architecture

6. Evaluation

6.1. Dataset Description

This study used a comprehensive dataset gathered from surveys of Adeleke University students that covered a wide range of student characteristics. The dataset consisting of 183,962 observations included demographic information,

academic history, and other relevant factors derived from the General Health Questionnaire (GHQ) that could impact academic performance. Our goal was to predict the final grade achieved by students, which was categorized into five levels: first class, second class upper, second class lower, third class, and pass.

To facilitate computational efficiency and manage memory usage during analysis, we partitioned the large dataset into smaller chunks, each containing 391 randomly selected observations, a number calculated using the sample size formula:

$$n = \frac{N}{1+e^2} \quad (3)$$

Where n is the sample size, N is the total population, and e is the margin of error.

This chunking was performed using a stratified random sampling approach, ensuring that each subset retained the overall distribution of key features, including academic performance and GHQ-12 scores. Preliminary experiments confirmed that the predictive model's performance remained consistent across different chunks, thereby preserving the integrity of the results.

Outlier detection and missing data handling were critical components of the data preprocessing stage. Outliers were identified using standard statistical methods (such as the interquartile range method) and were either removed or capped based on domain-specific thresholds. Missing values were addressed by applying imputation techniques (mean or median imputation for numerical variables and mode imputation for categorical variables) to ensure a complete dataset without introducing significant bias. These measures contributed to enhancing the overall reliability and robustness of the predictive model.

6.2. Experimental Setup

The predictive model was developed on a robust Python framework using Anaconda, which integrates essential libraries such as NumPy, Pandas, Scikit-learn, TensorFlow, and PyTorch. The dataset was first subjected to exploratory data analysis (EDA) to optimize memory usage. Column data types were adjusted to be more memory-efficient, reducing memory issues. Categorical features were converted into binary vectors using one-hot encoding for machine learning algorithms. The dataset was divided into training and testing sets, with 80% used for training and 20% for validation. This allowed for an unbiased assessment of model performance, ensuring accurate generalization to unseen data. This optimization was crucial for handling large datasets without memory issues.

To ensure reproducibility and clarity in our methodology, we employed stratified 10-fold cross-validation. Specifically, the entire dataset was randomly shuffled using a fixed random seed (e.g., seed=42) to guarantee consistency across experiments. The data was then divided into 10 equally sized folds, ensuring that each fold maintained the same class distribution as the overall dataset. In each of the 10 iterations, one (1) fold was used as the test set while the remaining 9 folds served as the training set. The performance metrics reported (accuracy, precision, recall, and F1-score) represent the average results over these 10 iterations, providing a robust estimate of the model's generalization capability.

Subsequently, we fed the same data to the 17 classification models under investigation: Decision Tree, Support Vector Classifier (SVC), Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Gradient Boosting, AdaBoost, Gaussian Naive Bayes (NB), Bernoulli Naive Bayes, Ridge Classifier, Multi-Layer Perceptron (MLP) Classifier, Quadratic Discriminant Analysis (QDA), Calibrated Classifier CV, Histogram Gradient Boosting, Passive Aggressive Classifier, Linear SVC, and Extra Trees.

6.3. Ensemble Model

Following the experimental results from the investigation of the 17 classification models, we selected the top-performing models to create an ensemble model aimed at optimizing predictive performance. The selected models (QDA, ET, and RF) demonstrated superior performance compared to the other models. The selection of these three models was driven not only by their superior performance metrics but also by their complementary methodological strengths. RF and ET, both tree-based algorithms, excel at capturing complex interactions in high-dimensional data and provide robust feature importance measures, while QDA leverages statistical assumptions to model non-linear decision boundaries, making it particularly effective in scenarios where class covariances differ. Together, these models offer a diverse perspective on the underlying data structure, which is critical for the multifaceted task of predicting academic performance.

To further optimize the ensemble model, we performed hyperparameter tuning for each of the selected models. The optimal hyperparameters were determined using a grid search approach, ensuring the best combination of parameters for each model. The voting classifier was employed as the metamodel to combine the predictions of the selected models. The weights assigned to each model were based on their individual performance, with QDA receiving a weight of 0.6, Extra Trees receiving a weight of 0.3, and Random Forest receiving a weight of 0.1. This ensemble approach allowed for the leveraging of the strengths of each model, leading to improved overall performance and generalization capabilities.

6.4. Statistical Justification for Inclusion of Mental health Information

In this study, the selection of mental health indicators from the GHQ-12 was driven by extensive literature suggesting that psychological wellbeing significantly influences academic performance. The GHQ-12 provides a concise yet powerful measure of mental health, capturing fluctuations in emotional and psychological states that may directly affect

student engagement and learning outcomes. Its inclusion was therefore essential to develop a more holistic predictive model that accounts for both academic and non-academic determinants of performance. To integrate the GHQ-12 responses into the predictive framework, we processed the questionnaire data using a binary scoring method (0-0-1-1). This approach converts qualitative responses into quantitative features, enabling their seamless incorporation into the ensemble model.

Statistical tests are crucial in validating hypotheses and determining the significance of relationships within data. Thus, the inclusion of mental health information in our study, as measured by the General Health Questionnaire (GHQ-12), necessitated statistical examination to ascertain its impact on student grades. To achieve this, a t-test was utilized, a statistical method designed to compare the means of two independent groups. The t-test helped in establishing whether the observed differences between these groups are statistically significant or likely to have occurred by chance.

Hypotheses

To conduct the t-test, the following hypotheses were formulated:

- *Null Hypothesis (H_0):* There is no significant difference in GHQ-12 scores between students with high grades and those with low grades.
- *Alternative Hypothesis (H_1):* There is a significant difference in GHQ-12 scores between students with high grades and those with low grades.

The null hypothesis assumes that any observed difference in the means of GHQ-12 scores between the two groups is due to random chance, while the alternative hypothesis suggests that the observed difference is statistically significant and not due to random chance.

Procedure

The data used for this t-test includes two columns: 'grade' and 'ghq_score'. The 'grade' column represents the academic performance score of students (1 to 5) categorized as first class (5), second class upper (4), second class lower (3), third class (2), and pass (1). The 'ghq' column represents the GHQ-12 scores of the students using a Binary Scoring (0-0-1-1) approach where "Not at all" and "Not more than usual" are scored as 0, while "Rather more than usual" and "Much more than usual" are scored as 1. Furthermore, the data were split into four (4) groups based on a threshold: Students with grades classified as high (threshold: grade ≥ 3), Students with grades classified as low (threshold: grade < 3), Students with GHQ-12 score classified as high (threshold: score ≥ 6), Students with GHQ-12 score classified as low (threshold: score < 6). An independent samples t-test is then performed to compare the mean GHQ-12 scores of these groups, with a p-value determined at alpha = 0.05 as statistically significant. By comparing GHQ-12 scores between students with varying academic performances, justification for the inclusion of GHQ-12 features in our predictive models is provided, ensuring that any identified relationships are both meaningful and reliable.

Table 1. Comparison of individual model performance (Accuracy, precision, recall, f1 score)

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.84	0.83	0.84	0.84
SVC	0.81	0.81	0.81	0.80
Logistic Regression	0.75	0.74	0.76	0.75
KNN	0.70	0.68	0.70	0.68
Random Forest*	0.86	0.86	0.86	0.86
Gradient Boosting	0.82	0.81	0.82	0.81
AdaBoost	0.27	0.42	0.27	0.29
Gaussian NB	0.52	0.56	0.52	0.44
Bernoulli NB	0.62	0.66	0.63	0.62
Ridge Classifier	0.62	0.62	0.63	0.60
MLP Classifier	0.85	0.84	0.84	0.84
Quadratic Discriminant Analysis*	0.90	0.91	0.92	0.91
Calibrated Classifier CV	0.71	0.68	0.71	0.68
Histogram Gradient Boost	0.80	0.78	0.80	0.78
Passive Aggressive Class	0.67	0.65	0.67	0.65
Linear SVC	0.71	0.69	0.71	0.70
Extra Trees*	0.89	0.89	0.89	0.89

6.5. Experimental Results

This section presents the results of the investigation into the performance of the 17 classification models, as well as

the ensemble model. The evaluation metrics used to assess the performance of the models include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (ROC/AUC). The results of the individual models are presented in Table 1, which shows that the top-performing models* were QDA, Extra Trees, and Random Forest, with accuracy values of 0.90, 0.89, and 0.86, respectively. The corresponding ROC/AUC curves for these models are shown in Figures 1-3, which illustrate their robust performance.

The results of the ensemble model* are presented in Table 2, which shows an accuracy value of 0.938, outperforming all individual models.

Table 2. Performance metrics for ensemble model vs base models (Accuracy, precision, recall, f1 score)

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.86	0.86	0.86	0.86
Quadratic Discriminant Analysis	0.90	0.91	0.92	0.91
Extra Trees	0.89	0.89	0.89	0.89
Ensemble Model*	0.91	0.93	0.91	0.91

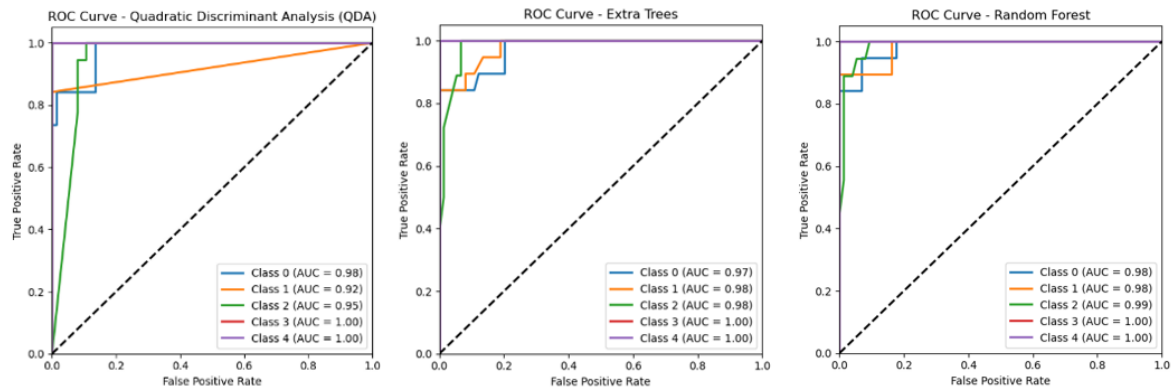


Fig. 1. ROC Curve for the best performing models

This study incorporated the General Health Questionnaire (GHQ-12) features, as predictive variables to evaluate their impact on the performance of the ensemble models in predicting students' academic performance. To assess the significance of these features, we compared the ensemble model's performance with and without the GHQ-12 variables. The results, presented in Table 3, reveal a substantial improvement in the model's predictive capabilities when the GHQ-12 variables are included.

Table 3. Performance metrics for ensemble model based on target class (Precision, recall, f1 score)

Class	With GHQ-12 Features			Without GHQ-12 Features		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
First Class	0.93	0.68	0.79	0.73	0.49	0.59
Second-Class Upper	1.00	0.94	0.97	0.80	0.74	0.77
Second-Class Lower	0.72	0.95	0.82	0.52	0.75	0.62
Third Class	1.00	1.00	1.00	0.80	0.80	0.80
Pass	1.00	1.00	1.00	0.80	0.80	0.80
Accuracy			0.91			0.70
Macro Average	0.93	0.92	0.92	0.73	0.72	0.72
Weighted Average	0.93	0.91	0.91	0.73	0.71	0.71

With the GHQ-12 variables, the ensemble model achieved an accuracy of 0.91, precision of 0.93, recall of 0.91, and an F1 score of 0.91. In contrast, without the GHQ-12 variables, the model's performance significantly dropped, with an accuracy of 0.70, precision of 0.73, recall of 0.71, and an F1 score of 0.71. This comparison highlights an average increase of approximately 20% in accuracy, precision, recall, and F1 score when the GHQ-12 variables are utilized.

Results of t-test for Grades Based on GHQ-12 scores

In this section, the results of the t-test conducted to evaluate the relationship between student grades and their GHQ-12 scores, justifying its inclusion as predictive variables, are presented. The t-test is performed to determine if there is a significant difference in GHQ-12 scores between students with high grades and those with low grades, yielding the following results:

$$t\text{-statistic: } 3.2676093895518545$$

$$p\text{-value: } 0.00126974069510269$$

From the results, the t-statistic of 3.27 indicates that the mean difference in GHQ-12 scores between students with high grades and those with low grades is more than three (3) standard deviations away from zero, suggesting a strong association. The p-value of 0.0013 which is significantly lower than the conventional alpha level of 0.05, further corroborates that this difference is not due to random chance, indicating that the difference observed is statistically significant. Therefore, the null hypothesis (H_0), which states that there is no significant difference in GHQ-12 scores between students with high grades and those with low grades, can be rejected based on the obtained p-value.

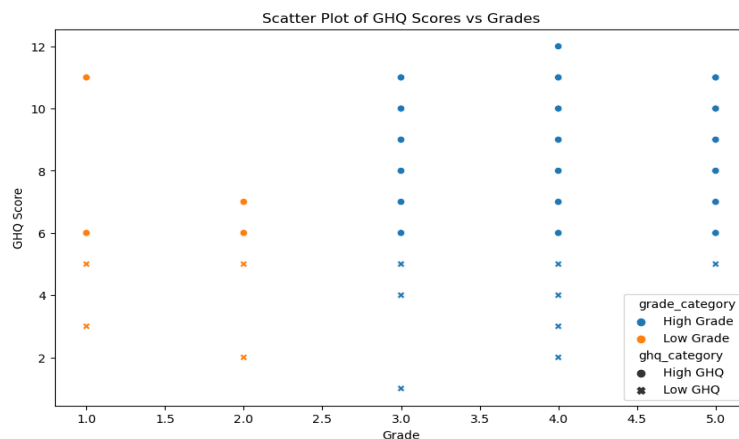


Fig. 2. Relationship between GHQ-12 and student's performance

Figure 3 illustrates the relationship between GHQ-12 scores and student grades using a scatter plot. In the plot: high GHQ scores are represented with a dot; low GHQ scores are represented with an 'x'; high grades (grade ≥ 3) have a blue hue, while low grades (grade < 3) have an orange hue. From the figure, it is evident that students with high GHQ-12 scores are more frequently found in the high grades (blue region).

7. Discussion

The experimental results demonstrate that the heterogeneous ensemble model, which integrates RF, ET, and QDA, outperformed the individual classifiers. While RF and ET are renowned for their predictive power, they are also known to be prone to overfitting, particularly when handling a large number of features. To counteract this, we implemented rigorous hyperparameter tuning using grid search and 10-fold cross-validation. Specific parameters such as maximum tree depth, number of trees, and the minimum samples required for a split were carefully optimized. Additionally, pruning techniques and constraints on feature splits were applied to both RF and ET to ensure that the models did not become overly complex and maintained good generalization on unseen data. The ensemble model's performance exceeded that of the individual models due to its ability to integrate and balance the strengths of each component. The weighted voting ensemble was designed to offset the limitations of individual classifiers by assigning weights based on their performance across multiple metrics. QDA's capability to discern complex statistical relationships combined with the robustness of RF and ET against noise creates a synergistic effect, resulting in a model that minimizes both bias and variance. This comprehensive integration leads to enhanced overall accuracy, precision, recall, and F1-score. Also, while accuracy, precision, recall, and F1-score are all important performance metrics, their relative importance can vary depending on the specific problem context. In the context of scholarship recommendations, where the cost of misclassification is significant—such as denying scholarships to deserving students—recall and F1-score are particularly critical. High recall ensures that most deserving candidates are identified, while a high F1-score indicates a balanced trade-off between precision and recall, reducing both false negatives and false positives.

Furthermore, we justified the inclusion of GHQ-12 mental health variables as important predictive features in predicting students' academic performance (Table 3), using a t-test to establish the relationship between GHQ-12 variables and student's performance (Figure 3), with the resulting p-value demonstrating the statistical significance of our findings.

8. Conclusions

In this study, we investigated the application of a heterogeneous voting ensemble approach to improve the predictive performance of classification models in predicting students' academic performance. Seventeen classification models were assessed, with the top-performing models being Quadratic Discriminant Analysis (QDA), Extra Trees, and Random Forest, achieving accuracy rates of 0.90, 0.89, and 0.86, respectively. These models were further integrated into an

ensemble model, which demonstrated superior accuracy compared to individual models. The ensemble model's output was instrumental in developing an algorithm for a scholarship recommender system. Notably, the research highlights the significant impact of incorporating mental health variables into machine learning models, which significantly enhances the accuracy in predicting academic performance. These findings underscore the potential of data-driven approaches to improve educational outcomes and support equitable scholarship allocation.

All the Declarations and Statements

Author Contributions Statement

B. J. Bambi: Conceptualization, Methodology, and Supervision: Proposed research ideas, constructed the overall framework, and supervised project execution.

O. G. Lala: Data Curation and Software Implementation: Handled data acquisition, dataset preprocessing, and implementing the research model.

A. A. Onamade: Model Training, Validation, and Performance Evaluation: Led the model training process, validated results using standard metrics, and benchmarked performance against existing methods.

O. A. Oludayo: Formal Analysis, Visualization, and Statistical Analysis: Performed in-depth analysis of experimental results, prepared performance charts, and ensured the statistical robustness of the evaluation.

A. Onyewe: Writing – Original Draft and Literature Survey: Drafted the initial manuscript, contributed to the literature survey, and documented the technical background of the study. Writing – Review and Editing, and Project Management: Reviewed and edited the manuscript, ensured clarity and coherence, and helped coordinate project milestones and deadlines.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this article.

At the time of conducting this research and preparing the manuscript, Dr. Bunmi Janet Bambi was a PhD student at Adeleke University and an employee of Springtime Development Foundation, a Non-Governmental Organization that awards scholarships to Adeleke University undergraduate students. Springtime Development Foundation did not provide funding or exert influence over the design, analysis, or conclusions of this research. The affiliation with Springtime Development Foundation is disclosed here to ensure transparency, and it does not alter the authors' adherence to the journal's policies on sharing data and materials.

Funding Declaration

This research received no external funding.

Data Availability Statement

The data used in this study are not publicly available due to privacy and ethical restrictions, as they comprise primary student records containing personally identifiable and sensitive academic information. The dataset was collected in partnership with Springtime Development Foundation and Adeleke University under conditions that preclude public disclosure. Access to the data may be requested from the corresponding author, subject to appropriate ethical approvals and institutional data-sharing agreements.

Ethical Declarations

This study involved the analysis of existing student records and General Health Questionnaire (GHQ-12) responses collected in partnership with Springtime Development Foundation and Adeleke University, Ede, Nigeria. All data were handled in strict accordance with applicable data protection regulations and institutional guidelines. Student identities were anonymized prior to analysis to protect privacy. The use of GHQ-12 data for research purposes was conducted in line with established ethical standards for the use of mental health screening instruments in academic research.

Acknowledgments

The authors sincerely thank the reviewers and domain experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiments and the reliability of the results. We also thank Springtime Development Foundation and Adeleke University, Ede, for their cooperation in facilitating access to the institutional context that motivated this research.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used Grammarly for grammar and spelling checks, and QuillBot to assist with paraphrasing during the writing and revision process. These tools were used solely to improve the language,

clarity, and readability of the manuscript. No AI tool was used to generate original content, ideas, data, analysis, or conclusions. All intellectual contributions, including conceptualization, methodology, experimentation, and interpretation of results, were carried out exclusively by the authors. The authors have reviewed and take full responsibility for the final content of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

AI - Artificial Intelligence
 AdaBoost - Adaptive Boosting
 AUC - Area Under the Curve
 CGPA - Cumulative Grade Point Average
 DL - Deep Learning
 EDA - Exploratory Data Analysis
 ET - Extra Trees
 GHQ-12 - General Health Questionnaire (12-item version)
 KNN - K-Nearest Neighbors
 LDA - Linear Discriminant Analysis
 ML - Machine Learning
 MLP - Multi-Layer Perceptron
 NB - Naive Bayes
 NGO - Non-Governmental Organization
 NLP - Natural Language Processing
 PCA - Principal Component Analysis
 QDA - Quadratic Discriminant Analysis
 RF - Random Forest
 ROC - Receiver Operating Characteristic
 SDF - Springtime Development Foundation
 SVC - Support Vector Classifier

APPENDIX A/B/C..., with appendix title

None

References

- [1] J. Cantiello, M. D. Fottler, D. Oetjen and N. J. Zhang, "The impact of demographic and perceptual variables on a young adult's decision to be covered by private health insurance," *BMC Health Services Research*, vol. 15, no. 1, pp. 1-15, 2015.
- [2] S. C. Paik, "NGOs' Partnership for Social Development: A Case Study on an NGO," *Dinkum Journal of Social Innovations*, vol. 3, no. 2, pp. 115-123, 2024.
- [3] W. Rodgers, G. Udo, Y. Zenad and D. Adekemi, "Ai Ethical Algorithmic Pathways Influencing Ngos Decisions," 10.2139/ssrn.4579225., 2023.
- [4] D. Lewis, N. Kanji and N. S. Themudo, "Non-governmental organizations and development," Routledge, 2020.
- [5] Y. Cai and L. Tang, "Correlation Analysis between Higher Education Level and College Students' Public Mental Health Driven by AI," *Computational Intelligence and Neuroscience*, 2022.
- [6] R. Soobramoney and A. Singh, "Identifying Students At-Risk with an Ensemble of Machine Learning Algorithms," *Conference on Information Communications Technology and Society (ICTAS)*, 2019.
- [7] E. Alyahyan and D. Dusteaur, "Decision Trees for Very Early Prediction of Student's Achievement," *IEEE Xplore: International Conference on Computer and Information Sciences (ICCIS)*, 2020.
- [8] 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, p. 61-75, 2018.
- [9] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, Art. no. e01250, 2019. DOI: 10.1016/j.heliyon.2019.e01250.
- [10] C. Zabriskie, J. Yang, S. DeVore and J. Stewart, "Using machine learning to predict physics course outcomes," *Physical Review Physics Education Research*, vol. 15, no. 2, 2019.
- [11] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," *Computers and Electrical Engineering*, vol. 89, Art. no. 106903, 2021. DOI: 10.1016/j.compeleceng.2020.106903.
- [12] H. S. Y. Aybek and M. R. Okur, "Predicting achievement with artificial neural networks: The case of Anadolu University open education system," *International Journal of Assessment Tools in Education*, vol. 5, no. 3, pp. 474-490, 2018.
- [13] X. Xu, J. Wang, H. Peng and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Computers in Human Behavior*, vol. 98, pp. 166-173, 2019.
- [14] H. Waheed, S. U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human behavior*, 2020.
- [15] R. Asif, A. Merceron, S. Abbas and N. Ghani, "Analyzing undergraduate students' performance using educational data mining,"

Comput. Educ., vol. 113, p. 177–194, 2017.

- [16] S. Hussain and M. Q. Khan, "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," *Annals of data science*, vol. 10, no. 3, pp. 637-655, 2023.
- [17] B. Albreiki, N. Zaki and H. Alashwal, "A systematic literature review of student'performance prediction using machine learning techniques," *Education Sciences*, vol. 11, no. 9, p. 552, 2021.
- [18] S. G. Anjara, C. Bonetto, T. Van Bortel and C. Brayne, "Using the GHQ-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations," *International journal of mental health systems*, vol. 14, pp. 1-13, 2020.

Authors' Profiles



Dr. Bunmi J. Bambi is a Lecturer in the department of Computer Science, Adeleke University, Ede. She obtained a B.Sc. degree in Computer Science with Economics from Obafemi Awolowo University, Ile-Ife in 2012 and M.Sc. in Computer Science from Babcock University, Ilishan-Remo in 2020. She bagged her Ph.D. in Computer Science in 2024 from Adeleke University, Ede.

Dr. Bambi worked as an ICT staff for over 10 years at Springtime Development Foundation, Ede, Nigeria, where young potentials are being empowered through access to quality education via scholarship. Her research interests include Information Systems, AI, Data Science, and Mobile Computing. She is particularly focused on exploring innovative solutions in these fields to address real-world problems and contribute to the advancement of technology and sustainable development.



Prof. Olusegun G. Lala is a Professor of Computer Science at the Department of Information Technology, Adeleke University, Ede. He received B.Sc. degree in Computer Science in 2000 from University of Ilorin. He obtained M.Sc. degree in Computer Science in 2004 from University of Ibadan. He bagged his Ph.D. degree in Computer Science in 2015 from Ladoké Akintola University of Technology, Ogbomoshó.

Dr. Lala's research interest includes Software Engineering, Information Systems, Artificial Intelligence, Cloud Computing, Distributed Computing and Mobile computing. He is a full member of Nigeria Computer Society (NCS) and a full member of Computer Professionals Registration Council of Nigeria (CPN).



Dr. Akintoye Abraham Onamade is a dedicated academic, researcher, and educator in the field of Computer Science. He currently serves as a Head of the Department of Computer Science at Adeleke University, Ede, Osun State, Nigeria.

He obtained his Bachelor of Science (B.Sc.) degree in Computer Science from Federal University of Agriculture, Abeokuta, Ogun State, Nigeria, Master of Science (M.Sc.) degree in Computer Science from University of Ibadan, Nigeria, and Doctor of Philosophy (Ph.D.) in Computer Science with specialization in Computer Science and Health Informatics from University of South Africa, Pretoria.

Dr. Onamade's research interests include Machine Learning, Cybersecurity, Computer Science, Software Engineering, Mobile Application, Artificial Intelligence, Statistical Analysis and Information Systems.



Dr. Oduwale A. Oludayo is a Senior Lecturer in the department of Computer Science, University of Ilesa, Ilesa, Nigeria. He holds a B. Tech. degree in Mathematics/Computer Science in 2000 from Federal University of Technology, Minna Nigeria. He obtained M.Sc. degree in Computer Science in 2006 from University of Ibadan. He bagged his Ph.D. degree in Computer Science in 2022 from Adeleke University, Ede, Nigeria.

Dr. Oduwale's research interest includes Data Communication, Software Engineering, Information Systems, Cloud Computing, Data science, and Mobile computing. He is a full member of Nigeria Computer Society (NCS)



Anozie Onyewe received the B.Sc. degree in Computer Science from Michael Okpara University of Agriculture Umudike in 2015 and M.Sc. in Computer Science from Ahmadu Bello University Zaria in 2020. He served as a Data Analyst at the Ministry of Budget and Planning, Internal Revenue Service, and as a Data Scientist at the Ministry of Health of the Abia State Government of Nigeria between 2018 and 2023.

Onyewe is a Ph.D. student and an Assistant Lecturer in the Department of Computer Science, Adeleke University, Ede. His research interest includes Machine Learning, Computational Theory, and Knowledge Discovery. He is a member of the Nigeria Computer Society (NCS).

How to cite this paper

Bunmi Janet Bambi, Olusegun Lala, Akintoye Onamade, Oludayo Oduwale, Anozie Onyewe, "Scholarship Award Recommendation System Using Predictive Modeling and Ensemble Learning", *International Journal of Information Technology and Computer Science(IJITCS)*, Vol.18, No.3, pp.45-56, 2026. DOI:10.5815/ijitcs.2026.03.04