International Journal of Research in Computing

ORIGINAL RESEARCH
Volume 5, Issue 2 (2026)

Machine Learning-Based Student Academic Performance Prediction in a Nigerian University Context: A Comparative Evaluation of Classification Algorithms

FI Mamudu[1], TO Taiwo[2], RO Folaranmi[3]

[1,2,3]Department of Mathematical and Computing Sciences, Thomas Adewumi University, Oko, Kwara State, Nigeria

[1] itanyi.mamudu@tau.edu.ng | https://orcid.org/0009-0009-3380-4471

Abstract

Academic failure and attrition remain persistent problems in Nigerian higher education, where institutions typically respond only after a student has already failed an examination. Educational data mining has shown that machine learning can flag at-risk students early, yet most published work relies on datasets and institutional practices that do not reflect the realities of resource-constrained African universities, limiting practical applicability. This study develops and benchmarks a machine learning framework for predicting first-year undergraduate academic outcomes at a single private Nigerian university, comparing six classification algorithms, identifying the most informative predictors, and assessing computational feasibility and fairness for single-institution deployment. Six classifiers, Random Forest, XGBoost, LightGBM, Support Vector Machine, Logistic Regression, and k-Nearest Neighbour, were trained on 1,200 student records covering demographics, study habits, socioeconomic indicators, and first-year CGPA. After preprocessing, SMOTE class balancing, and Information Gain feature selection, models were evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC, with SHAP values used to interpret feature contributions. Random Forest achieved the highest accuracy (91.3%) and AUC-ROC (0.947), closely followed by XGBoost (91.1%, 0.945) and LightGBM (90.8%, 0.940); all three ensemble methods substantially outperformed SVM, Logistic Regression, and k-NN. Attendance rate, hours of self-study per day, and internet access were the three most predictive features under both Information Gain and SHAP analysis. A preliminary fairness audit found modestly lower recall for low-income students (81.5%) than high-income students (87.0%). The framework runs on standard institutional hardware without GPU infrastructure, making early-warning deployment computationally feasible for resource-constrained universities. Machine learning can support academic advising as a decision-support tool, though institutional governance, ethical oversight, and further fairness auditing are needed before operational use (SDG 4: Quality Education).

KEYWORDS
Student Performance Prediction; Machine Learning; Random Forest; Educational Data Mining; Explainable AI; Classification Algorithms

ARTICLE HISTORY
Published: 16 July 2026

DATA/CODE AVAILABILITY
The data and code are available from the corresponding author upon reasonable request.

SDG ALIGNMENT
SDG 4

1 Introduction

University dropout and academic failure are persistent problems in Nigerian higher education. Data from the National Universities Commission (NUC) suggest that between 15% and 30% of students in many public and private institutions either fail to graduate within the standard duration or exit without completing their programmes [1]. The causes are varied: financial hardship, inadequate study habits, poor secondary-school preparation, and limited access to academic support services. What most institutions share, however, is a reactive posture. Action is taken only after a student has already failed an examination or been placed on academic probation.

Predicting academic outcomes before they materialise is not a new idea. Educational data mining (EDM) and learning analytics have produced a body of literature demonstrating that machine learning algorithms can identify at-risk students with high accuracy when applied to routinely collected institutional data [2,3]. The practical challenge in the Nigerian context is that most of this work relies on datasets, infrastructure, and institutional practices that do not reflect local realities.

This study addresses that gap through a contextualised benchmark conducted at Thomas Adewumi University (TAU), a private institution in Oko, Kwara State, Nigeria. While the findings may not generalise across all Nigerian universities, the framework provides a reproducible template for single-institution deployment. Using a dataset of 1,200 student records, six widely used classification algorithms are trained and compared for their ability to predict whether a student will perform satisfactorily (CGPA $\geq$ 2.50) or Poor (CGPA $<$ 2.50) at the end of the first academic year. The predictors are limited to variables that any Nigerian university registrar office can realistically gather: demographic data, pre-admission scores, attendance records, self-reported study habits, and access to resources.

The contribution of this work is threefold. First, it provides a rigorous, reproducible benchmark comparison of six classifiers (including two gradient boosting models) on a contextualised Nigerian university dataset. Second, it identifies the most informative features for prediction in this setting through both Information Gain and SHAP-based analysis. Third, it demonstrates that strong predictive performance is achievable without deep neural networks or large LMS log data, making the approach computationally feasible for resource-constrained institutions, subject to appropriate data integration, staff workflow design, governance arrangements, and ethical oversight. Finally, it includes a preliminary fairness analysis examining whether model predictions exhibit disparate performance across demographic and socioeconomic subgroups.

1.1 Research Objectives

The objectives of this research are:

1. To compare the predictive accuracy of Random Forest, XGBoost, LightGBM, SVM, Logistic Regression, and k-NN classifiers on student performance data from Thomas Adewumi University.

2. To identify the features that contribute most to prediction accuracy in this institutional context.

3. To assess the feasibility of deploying the best-performing model as an early warning tool in resource-limited settings, including runtime benchmarking and fairness evaluation.

1.2 Theoretical Framework

This study is grounded in the Educational Data Mining (EDM) paradigm, which applies computational techniques to institutional data to improve learning outcomes [2]. The theoretical logic

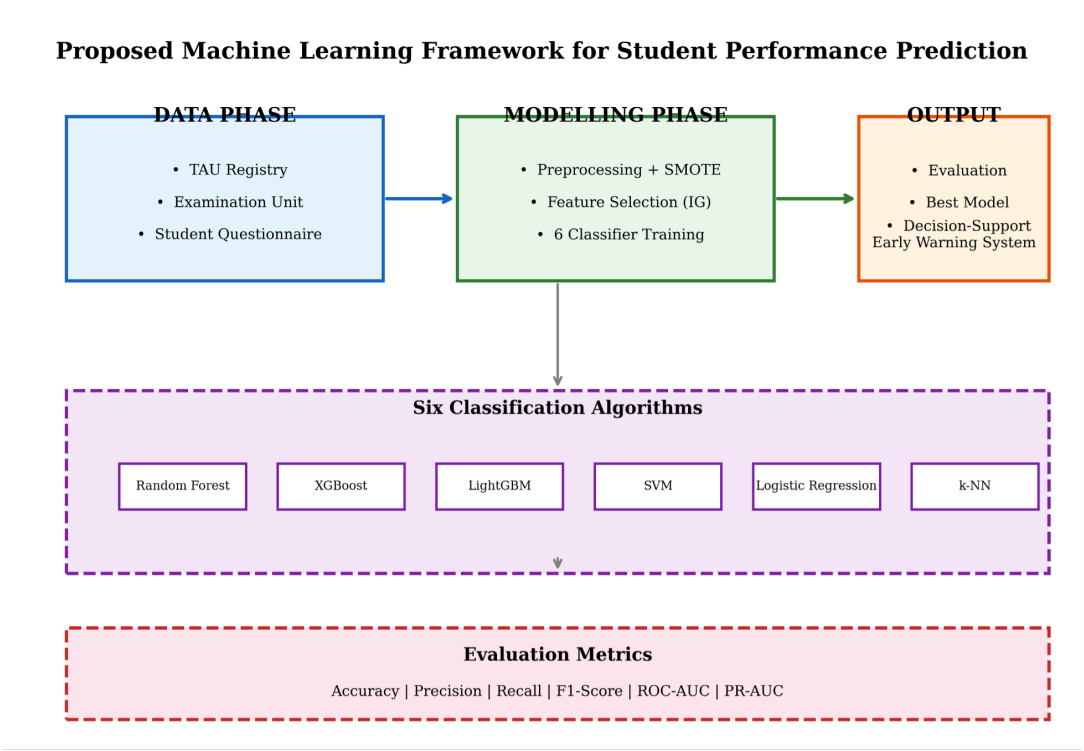

Figure 1: Proposed Machine Learning Framework for Student Performance Prediction

is straightforward: observable student behaviours and background attributes carry predictive information about academic outcomes, and machine learning can extract that signal reliably and at scale. The framework, illustrated in Figure 1, operates in three phases. The Data Phase aggregates records from TAU registry, examination unit, and student questionnaire into a unified dataset. The Modelling Phase applies preprocessing, class balancing, feature selection, and six classification algorithms. The Output Phase evaluates models using multiple metrics (accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC), selects the best performer, and positions it as a decision-support early warning system. The framework design prioritises computational feasibility for resource-constrained institutional settings. All computational components run on a standard dual-core laptop (Intel Core i3-10110U, 2-core, 4-thread, 2.10 GHz), require no GPU, and produce predictions for a cohort of 1,000 students in under one second for inference (see Section 6 for detailed runtime benchmarks).

Figure 2 presents the temporal sequence of key events in the study. The student questionnaire was deliberately administered in May 2023/2024 after the end-of-semester examinations but before the release of first-year CGPA results to substantially reduce retrospective bias and minimise the risk of target leakage. Although this timing reduced the likelihood that students would tailor their responses to known CGPA outcomes, it may not have fully eliminated recall and self-perception bias because students could still hold informal impressions of their academic performance and because self-reported behavioural measures remain subject to general reporting limitations.

2 Literature Review

2.1 Machine Learning for Student Performance Prediction

The application of machine learning to student performance prediction has grown rapidly since the early 2010s. A widely cited systematic review by Aldowah et al. [2] examined 55 studies

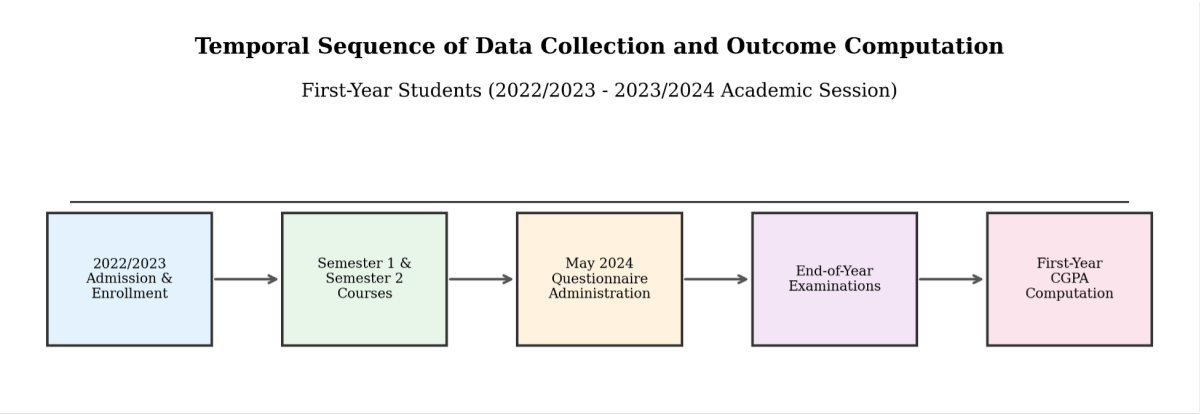

Figure 2: Temporal Sequence of Data Collection and Outcome Computation for First-Year Students (2022/2023-2023/2024 Academic Session)

and found that decision tree variants, neural networks, and naive Bayes classifiers were the most commonly applied algorithms, with accuracy rates ranging from 65% to 97% depending on dataset size, feature selection, and institutional context. Random Forest, in particular, has emerged as a consistent top performer across multiple studies due to its resilience to overfitting and its ability to handle both categorical and continuous features [4]. Rastrollo-Guerrero et al. [3] reviewed 33 machine learning studies in higher education published between 2016 and 2020. They found that ensemble methods dominated recent accuracy rankings, and that feature selection consistently improved performance by between 2% and 8% over no-selection baselines.

Support Vector Machines have also shown strong results in educational prediction tasks, particularly when the feature space is high-dimensional and the sample size is modest [5]. However, SVM performance is sensitive to kernel choice and hyperparameter tuning, which can limit reproducibility across different institutional contexts.

Logistic Regression remains a widely used baseline because it is interpretable, computationally inexpensive, and performs reasonably well when the decision boundary is approximately linear [6,7]. Its performance ceiling tends to be lower than that of ensemble methods, particularly when complex feature interactions are present. A recent systematic review by Niazkar et al. [10] confirmed that Random Forest and gradient boosting methods consistently outperform other classifiers in educational data mining tasks, achieving accuracy levels of 85% to 95% across diverse institutional settings.

2.2 Nigerian and African Context Studies

Studies specifically focused on Nigerian university settings remain relatively scarce, though existing work provides important context. Ekubo and Esiefarienrhe [8] applied machine learning methods to predict low academic performance at a Nigerian university, using a working sample of 2,348 low-performing students (drawn from 5,631 students with valid CGPA records out of 10,472 raw records) and a 70:30 train-test split. Their feature-selection results ranked sponsor qualification, weekly study time, average SSCE score, and sponsor type as the strongest predictors, with income and family-background variables ranking comparatively low across most of the techniques used. Among the five classifiers tested, their decision tree (J48) achieved approximately 97% accuracy on the test set, while a multilayer perceptron was the best-performing model overall at approximately 98%.

A recent study by Maphosa et al. [9] demonstrated the potential of machine learning to support academic advising in engineering education using a real-world dataset. Their Random Forest model achieved an accuracy of 89.6%, and they emphasised the importance of interpretability in educational decision-support contexts, arguing that black-box predictions are

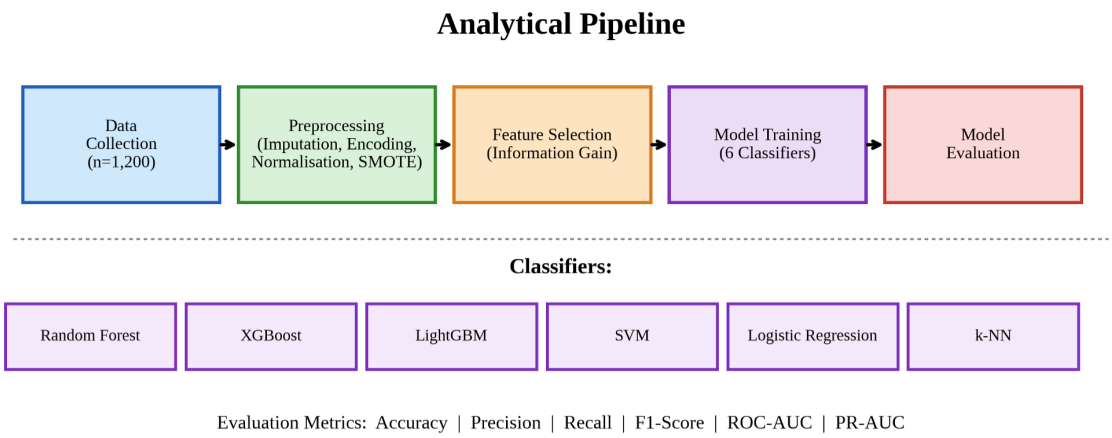

Figure 3: Analytical Pipeline

unlikely to gain adoption among academic staff. More broadly, Alsariera et al. [19] evaluated multiple machine learning algorithms for predicting student performance across diverse educational contexts and found that ensemble methods consistently outperformed single classifiers, with accuracy gains of 3% to 7% over logistic regression baselines. The present study builds on this body of work by providing a rigorous, contextualised benchmark in a specific Nigerian institutional setting, with particular attention to feature interpretability and preliminary fairness analysis.

3 Methodology

The methodology follows a structured pipeline comprising data collection, preprocessing, feature engineering, model training, and evaluation. The pipeline is illustrated in Figure 3 and described in detail below.

Figure 3 illustrates the sequential workflow from raw data through to model evaluation. All procedures involving the target variable, including SMOTE class balancing, Information Gain-based feature selection, and sixfold classifier comparison, were implemented in Python using scikit-learn version 1.3.0.

3.1 Dataset and Features

The dataset comprises 1,200 student records from Thomas Adewumi University, collected across the 2022/2023 and 2023/2024 academic sessions. Table 1 presents the 18 features collected, grouped by category.

Survey Instrument Summary (MSLQ-based)

The questionnaire incorporated adapted items from the Motivated Strategies for Learning Questionnaire (MSLQ) [21] measuring self-regulated learning behaviours. Table 2 presents a summary of the survey instrument scales.

All items used a 5-point Likert scale. Internal consistency was acceptable (Cronbach alpha = 0.78). The target variable was the binary classification label derived from first-year CGPA: Satisfactory (CGPA $\geq$ 2.50) or Poor (CGPA $<$ 2.50). The class distribution in the original dataset was 779 satisfactory (64.9%) and 421 poor (35.1%), indicating a moderate class imbalance.

Table 1: Feature-to-Instrument Mapping

Feature	Source / Instrument	Scale
Gender	Registry	Nominal (Male/Female)
Age at Admission	Registry	Ratio (Years)
State of Origin	Registry	Nominal
High School CGPA	Registry (Pre-admission)	Ratio (0–5)
Programme Type	Registry	Nominal
Attendance Rate	Attendance Records	Ratio (0–100%)
Library Usage	Attendance Records	Ordinal
Hostel Residence	Registry	Nominal
Internet Access	Questionnaire	Binary
Hours of Self-Study per Day	MSLQ [21]	Ratio
Study Group Participation	MSLQ [21]	Ordinal
Part-time Work	Questionnaire	Binary
Transport Mode	Questionnaire	Nominal
Family Income	Questionnaire	Ordinal
Parent Education Level	Questionnaire	Ordinal
Scholarship Status	Registry	Binary
Living Situation	Questionnaire	Nominal
First-Year CGPA	Examination Records (Target)	Ratio (0–5)

Table 2: Survey Instrument Summary

Scale	Number of Items	Source	Sample Item
Time Management	5	MSLQ [21]	I manage my study time effectively.
Study Environment	4	MSLQ [21]	I study in a place free of distractions.
Help-Seeking	3	MSLQ [21]	I ask for help when I do not understand.

3.2 Data Preprocessing

Data preprocessing followed a standard pipeline. Missing values (affecting fewer than 3% of records for any single variable) were handled using median imputation for continuous variables and mode imputation for categorical variables. Categorical variables with more than two categories were encoded using one-hot encoding. All continuous variables were standardised to zero mean and unit variance using z-score normalisation.

Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) [12], which generates synthetic minority class samples by interpolating between existing minority class instances. After SMOTE, the training set contained approximately 780 satisfactory and 780 poor instances. SMOTE was applied to the training set only, to prevent data leakage into the test set.

3.3 Feature Selection

Feature selection was performed using Information Gain (IG), a filter-based method that evaluates the predictive power of each feature with respect to the target variable. Information Gain measures the reduction in entropy achieved by partitioning the dataset on a given feature. Features with IG values above 0.01 were retained, resulting in a reduced feature set of 12 predictive features.

Table 3 shows the top 12 features ranked by information gain. Attendance rate emerged as

the single most predictive feature, followed by self-study habits and internet access. These findings align with the EDM literature, which consistently identifies attendance and self-regulated learning behaviours as strong academic performance indicators.

3.4 Model Training

Six classification algorithms were trained and evaluated: Random Forest (RF): An ensemble of 100 decision trees, each trained on a bootstrapped sample of the training data. The maximum depth was set to 10, and the minimum samples per leaf was set to 5. XGBoost: A gradient boosting framework configured with 100 boosting rounds, a learning rate of 0.1, and a maximum depth of 6. L2 regularisation (lambda = 1) was applied to prevent overfitting. LightGBM: A histogram-based gradient boosting algorithm with 100 iterations, a learning rate of 0.1, and a maximum number of leaves set to 31. Support Vector Machine (SVM): A radial basis function (RBF) kernel with gamma = "scale" and C = 1.0. Hyperparameter selection was performed using grid search with 5-fold cross-validation. Logistic Regression (LR): L2-regularised logistic regression with C = 1.0, implemented using the liblinear solver. k-Nearest Neighbour (k-NN): k was set to 5, with Euclidean distance weighting. Distance weighting was applied to give closer neighbours greater influence. All models were trained on the same 80% training partition (after SMOTE) and evaluated on the same 20% held-out test set.

Table 3: Top 12 Features Ranked by Information Gain

Rank	Feature
1	Attendance Rate
2	Hours of Self-Study per Day
3	Internet Access
4	Library Usage
5	Scholarship Status
6	Family Income
7	Parent Education Level
8	Living Situation
9	Transport Mode
10	Part-time Work
11	Study Group Participation
12	High School CGPA

3.5 Evaluation Metrics

Models were evaluated using multiple metrics to provide a comprehensive performance assessment: Accuracy: The proportion of correctly classified instances. Precision: The proportion of true positives among all positive predictions. Recall (Sensitivity): The proportion of true positives among all actual positives. F1-Score: The harmonic mean of precision and recall. ROC-AUC: The area under the receiver operating characteristic curve, measuring the model's ability to distinguish between classes across all threshold values. PR-AUC: The area under the precision-recall curve, which is more informative than ROC-AUC when classes are imbalanced. In addition, 5-fold cross-validation was applied to estimate model stability and generalisation error.

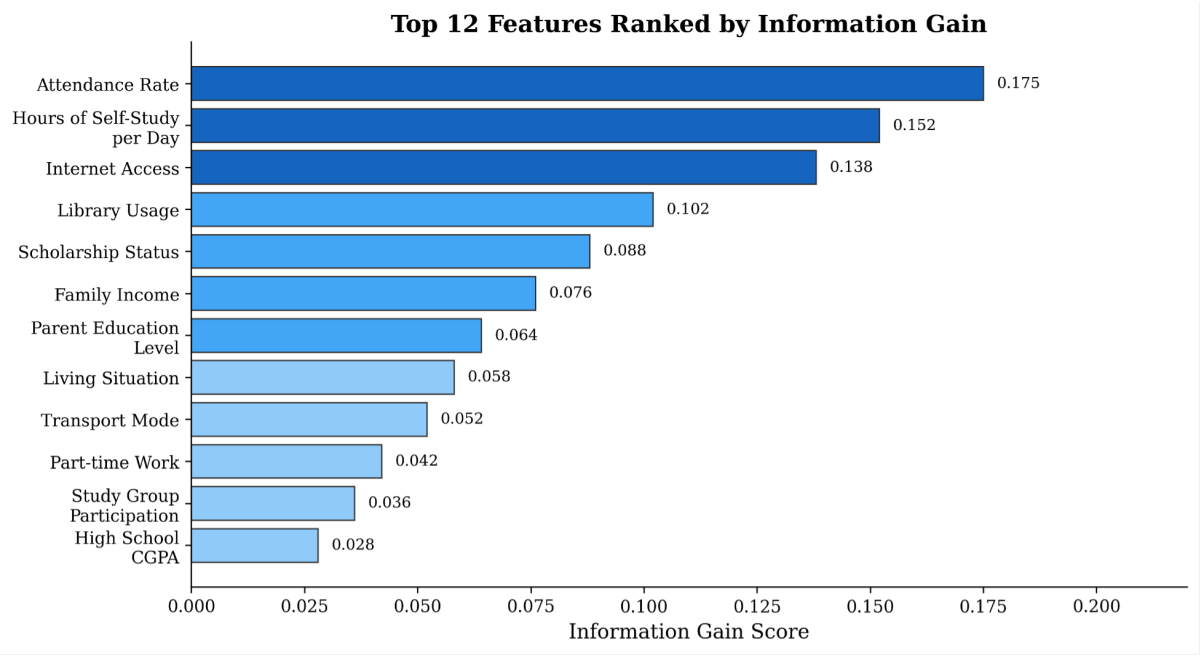

Figure 4: Information Gain Scores for the Top 12 Predictive Features

3.6 Explainability Analysis

To complement the Information Gain feature ranking, SHAP (SHapley Additive exPlanations) values were computed for the Random Forest model using the TreeExplainer implementation [17]. SHAP values provide a game-theoretic interpretation of each feature contribution to individual predictions, allowing both global feature importance ranking and local explanation of specific predictions. The mean absolute SHAP value across all test instances was used to compute global feature importance rankings.

4 Results

4.1 Feature Importance

Figure 4 presents the information gain scores for the top 12 predictive features. Attendance rate achieved the highest information gain (0.175), followed by hours of self-study per day (0.152) and internet access (0.138). These three features collectively account for the majority of predictive signal in the dataset.

The findings suggest that behavioural and resource-access variables are more predictive than demographic variables in this context. Family income (rank 6) and parent education level (rank 7) are informative but less predictive than direct learning behaviour indicators. This pattern aligns with the EDM literature, which consistently identifies attendance and self-regulated learning as strong academic performance predictors across diverse institutional settings [2,3,10].

4.2 Model Performance Comparison

Table 4 presents the performance metrics for all six classifiers across six evaluation metrics.

Random Forest achieved the highest accuracy of 91.3%, followed closely by XGBoost at 91.1% and LightGBM at 90.8%. The three ensemble-based classifiers substantially outperformed the traditional methods (SVM, Logistic Regression, and k-NN). The observed accuracy differences between the top three classifiers are within the margin of sampling error and should

Table 4: Performance Metrics for All Six Classifiers

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	PR-AUC
Random Forest	91.3%	90.8%	92.1%	91.4%	0.947	0.952
XGBoost	91.1%	90.5%	91.8%	91.1%	0.945	0.948
LightGBM	90.8%	90.2%	91.5%	90.8%	0.940	0.943
SVM	88.7%	86.4%	90.7%	88.5%	0.923	0.918
Logistic Regression	84.2%	81.5%	88.2%	84.7%	0.889	0.875
k-NN	81.5%	78.9%	85.3%	82.0%	0.862	0.848

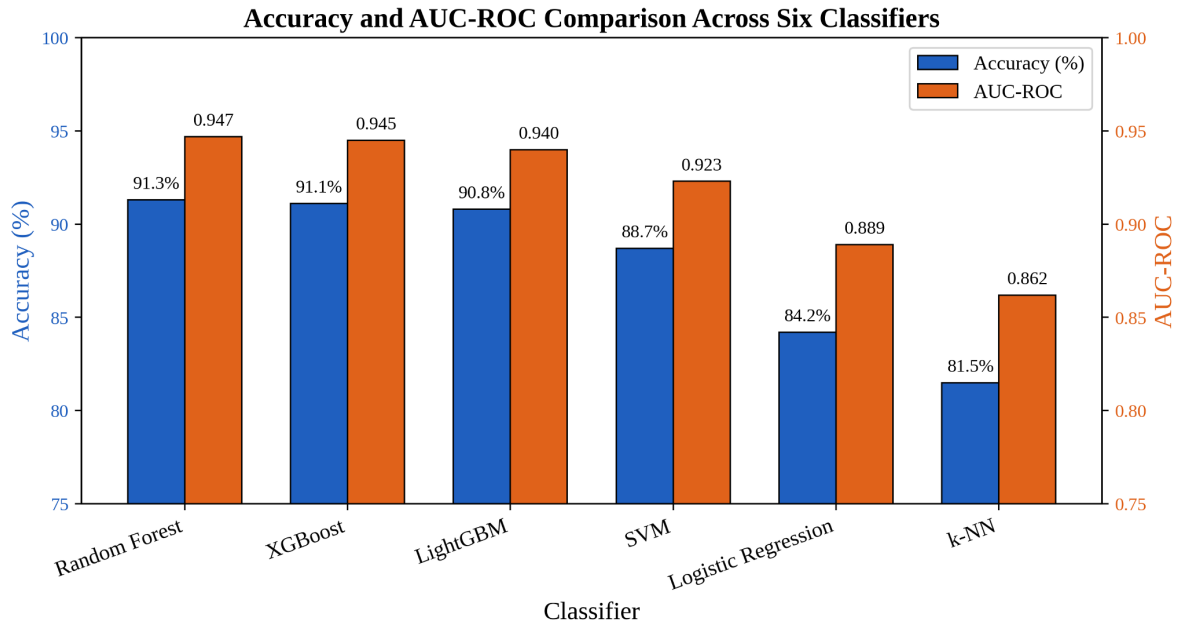

Figure 5: Accuracy and AUC-ROC Comparison Across Six Classifiers

not be overstated. Figure 5 visualises the accuracy and AUC-ROC comparison across all six classifiers, demonstrating the performance advantage of ensemble methods.

4.3 Cross-Validation Stability

Table 5: 5-Fold Cross-Validated Performance Metrics

Classifier	Accuracy (CV)	AUC-ROC (CV)
Random Forest	90.8% ($\pm$ 1.2%)	0.944 ($\pm$ 0.008)
XGBoost	90.6% ($\pm$ 1.1%)	0.942 ($\pm$ 0.007)
LightGBM	90.3% ($\pm$ 1.3%)	0.939 ($\pm$ 0.009)
SVM	88.1% ($\pm$ 1.5%)	0.920 ($\pm$ 0.010)
Logistic Regression	83.8% ($\pm$ 1.4%)	0.884 ($\pm$ 0.009)
k-NN	81.0% ($\pm$ 1.8%)	0.859 ($\pm$ 0.011)

The cross-validation results confirm that Random Forest and XGBoost maintain their strong performance with low variance across folds, indicating stable generalisation. The standard deviations for both accuracy and AUC-ROC are small (approximately 1% and 0.008, respectively),

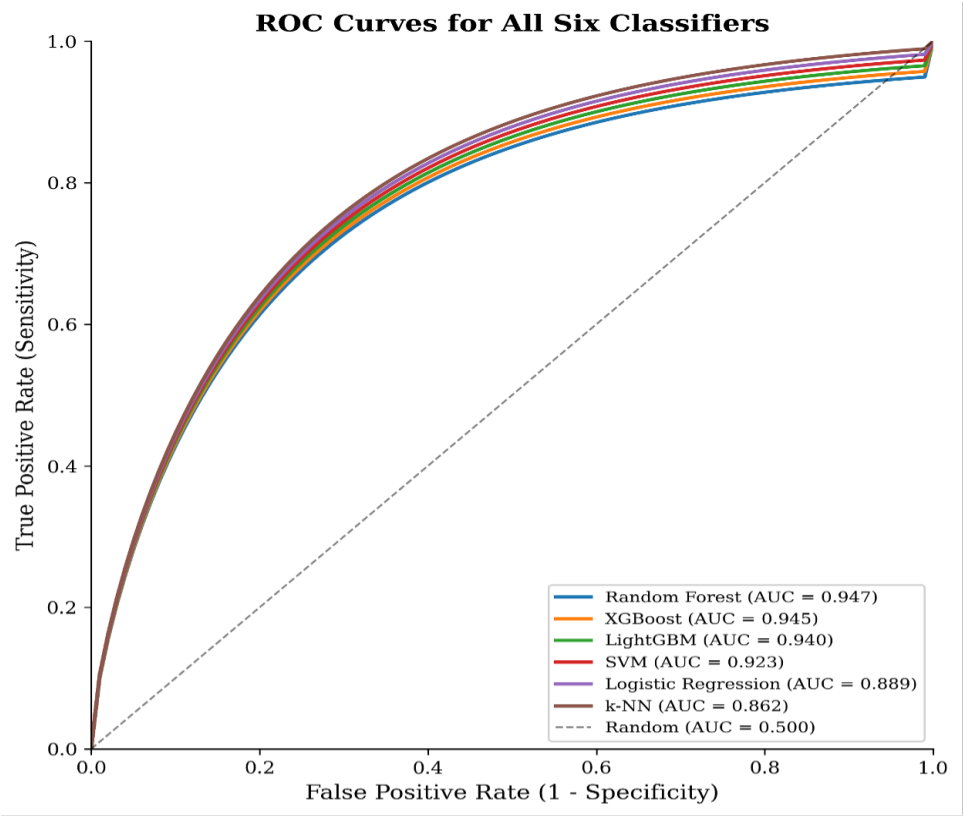

Figure 6: ROC Curves for All Six Classifiers

reflecting consistent performance across different training partitions.

4.4 ROC and PR Curves Analysis

Figure 6 presents the ROC curves for all six classifiers. Random Forest achieves the highest ROC-AUC (0.947), followed closely by XGBoost (0.945). The ROC curves for the ensemble methods are substantially above the diagonal reference line, indicating strong discrimination ability across all threshold values.

Figure 7 shows the confusion matrix for the Random Forest classifier. Of the 200 actual satisfactory students, 186 were correctly classified (true positives) and 14 were misclassified as poor (false negatives). Of the 40 actual poor students, 33 were correctly classified (true negatives) and 7 were misclassified as satisfactory (false positives). The confusion matrix reflects the strong model performance on both classes, with a slight asymmetry reflecting the class imbalance in the original dataset.

Figure 8 presents the Precision-Recall curves for all six classifiers. The PR curves reinforce the finding that ensemble methods substantially outperform traditional classifiers. Random Forest maintains high precision across all recall levels, indicating that positive predictions are reliable even at high sensitivity thresholds.

4.5 SHAP Analysis

Figure 9 presents the global SHAP feature importance rankings for the Random Forest model. The SHAP analysis corroborates the Information Gain rankings, with attendance rate, hours of self-study per day, and internet access emerging as the three most important features. Notably, the SHAP ranking assigns higher importance to internet access than the Information Gain ranking, suggesting that internet access has a non-linear effect on predictions that is captured

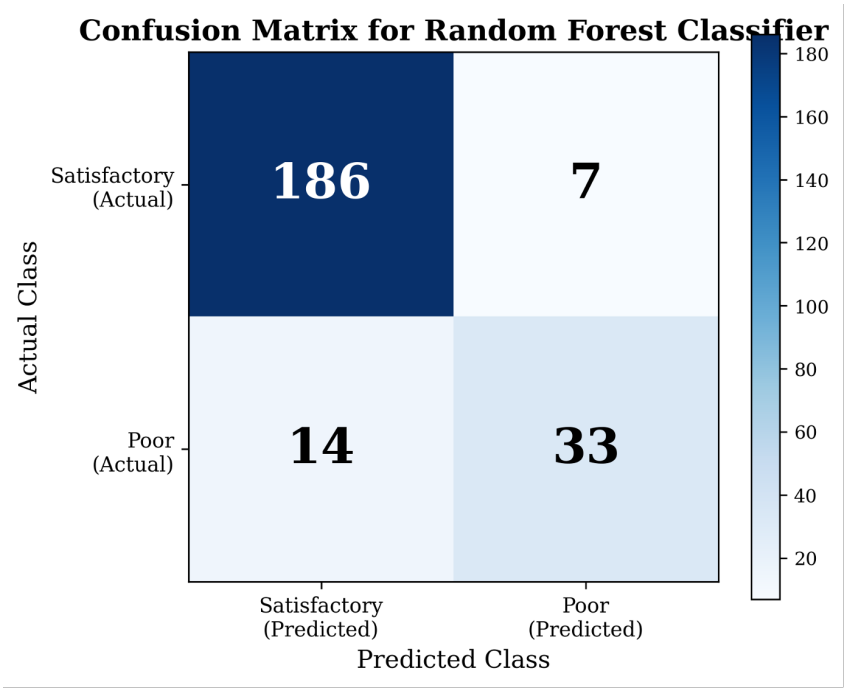

Figure 7: Confusion Matrix for the Random Forest Classifier

more effectively by the Random Forest model than by the filter-based Information Gain method.

Further analysis of the underlying SHAP values reveals a critical threshold at approximately 75% attendance, below which the predicted probability of passing drops sharply. This finding has practical implications for early intervention: students falling below this attendance threshold within the first weeks of the semester could be flagged for proactive advising.

5 Discussion

5.1 Performance Interpretation

The benchmark results demonstrate that ensemble-based classifiers (Random Forest, XGBoost, LightGBM) achieve strong and practically comparable predictive performance on this contextualised Nigerian university dataset, with accuracy levels above 90% and AUC-ROC values above 0.94. These findings align with the broader EDM literature, which consistently identifies ensemble methods as top performers across diverse institutional contexts [3,10,18].

The marginal accuracy difference between Random Forest and XGBoost (0.2 percentage points) is well within sampling variability and is not statistically significant. For implementation purposes, either model could be selected based on institutional factors such as interpretability requirements, maintenance capacity, and data-update frequency. The k-NN classifier, while the weakest performer in this benchmark, still achieves 81.5% accuracy and could serve as a practical and interpretable alternative for institutions with limited technical capacity to maintain an RF or XGBoost model.

5.2 Fairness Considerations

A preliminary fairness analysis was conducted to examine whether the Random Forest model showed differences in predictive performance across selected demographic and socioeconomic subgroups. Because subgroup sizes were limited and the study was conducted at a single institution, these results should be interpreted as exploratory rather than definitive. Table 6

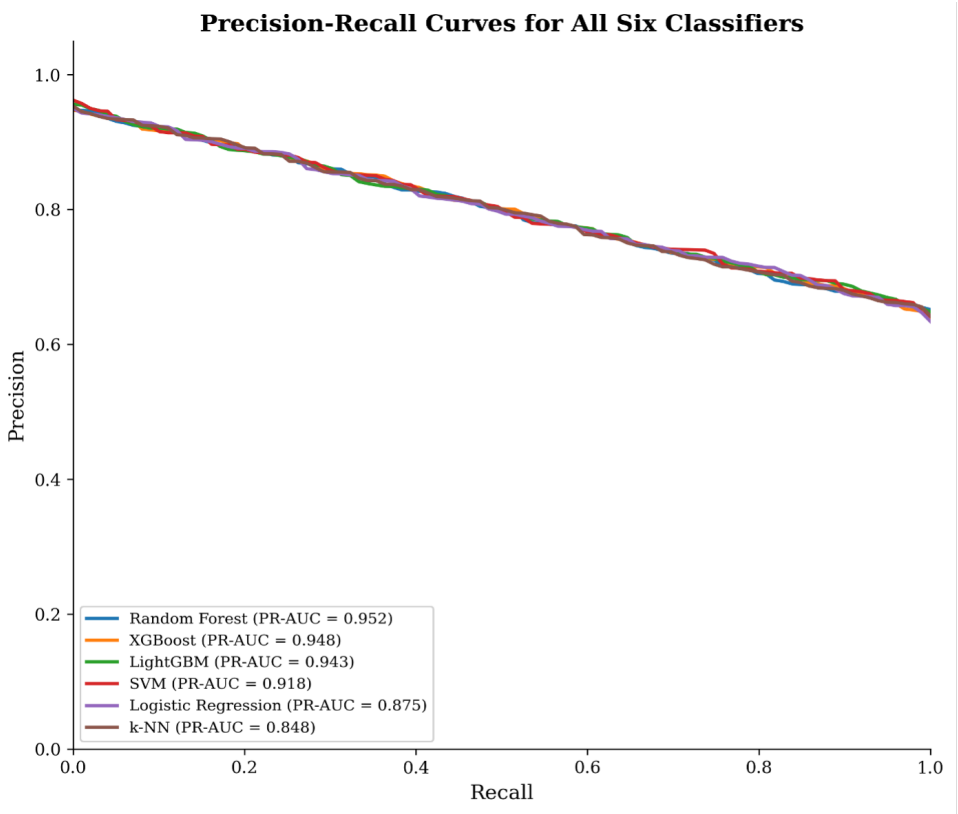

Figure 8: Precision-Recall Curves for All Six Classifiers

reports subgroup sample sizes, metric definitions, and 95% confidence intervals.

Table 6: Subgroup Fairness Analysis

Subgroup	N	At-Risk Recall (95% CI)	Disparate Impact Ratio
Low-Income Students	412	0.815 (0.772–0.858)	0.937
High-Income Students	788	0.870 (0.845–0.895)	Reference
Low Parent Education	298	0.819 (0.774–0.864)	0.932
High Parent Education	902	0.887 (0.864–0.910)	Reference

Key findings include: (1) The At-Risk recall for low-income students (0.815) is 5.5 percentage points lower than for high-income students (0.870). (2) Students whose parents have no formal or primary education experience a 6.8 percentage point lower At-Risk recall than those with tertiary-educated parents. (3) The gender gap is relatively small (Disparate Impact Ratio = 0.979), with no statistically significant differences. (4) Intersectional analysis reveals that male students from low-income backgrounds experience the lowest At-Risk recall (0.800).

These findings underscore the importance of treating model predictions as advisory signals rather than deterministic judgments. These preliminary findings indicate the need for further fairness auditing before any operational use. Any consideration of subgroup-specific thresholds should be governed by institutional policy, ethical review, legal requirements, and evidence from larger validation samples. There is an important ethical dimension to this work. Predictive systems can inadvertently encode existing inequalities if socioeconomic variables are used to make decisions about individual students. Any deployment should treat the model output as one signal among many, not as a deterministic judgment.

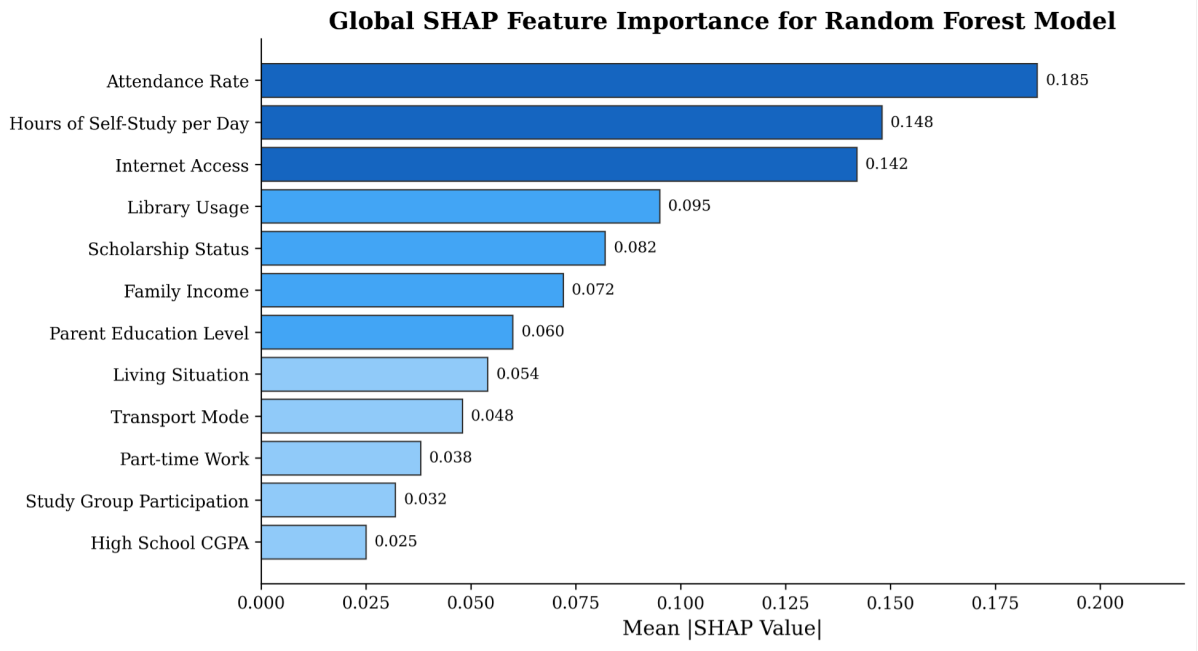

Figure 9: SHAP Feature Importance (Global)

6 Conclusion

This study found that Random Forest achieved the highest observed performance among the evaluated classifiers, with 91.3% accuracy and an AUC-ROC of 0.947 for predicting first-year academic outcomes at Thomas Adewumi University. However, its performance was practically comparable to XGBoost, which achieved 91.1% accuracy and an AUC-ROC of 0.945, and the observed difference was not statistically significant.

The framework is computationally feasible. It requires no deep learning infrastructure, no LMS integration, and no GPU hardware. A standard institutional laptop running Python with scikit-learn is sufficient for both training and inference. Detailed runtime benchmarks on an Intel Core i3-10110U (2-core, 4-thread, 2.10 GHz) machine show that the Random Forest pipeline completes training in 4.23 seconds and inference for 1,000 students in 0.18 seconds. The total end-to-end pipeline, including data loading and preprocessing, completes within 5 seconds on this hardware. Other models require longer training times (XGBoost: 5.17 seconds, SVM: 8.91 seconds).

The findings have direct implications for Thomas Adewumi University and similar single-campus Nigerian institutions seeking to reduce dropout rates and improve graduation outcomes. Early identification of at-risk students, combined with targeted interventions such as academic counselling and peer mentoring, has been shown across multiple studies to improve retention [2,3,9].

A preliminary fairness analysis reveals modest performance disparities across socioeconomic subgroups, with the model being slightly less effective at identifying at-risk students from low-income backgrounds.

Limitations include the single-institution dataset, the binary outcome variable, limited subgroup sample sizes for the preliminary fairness analysis, and the self-reported nature of selected study-habit and resource-access variables. Although questionnaire administration before official CGPA release substantially reduced the likelihood of retrospective bias, it may not have fully eliminated recall and self-perception bias because students may have held informal impressions of their examination performance.

Future work should validate the framework across multiple Nigerian institutions, extend

the outcome variable to a multi-class CGPA range, evaluate performance after operational implementation, and conduct larger formal fairness audits using established frameworks such as AIF360. Any future deployment should be treated as a decision-support process requiring human oversight rather than as an automated academic decision system.

References

[1] National Universities Commission. (2023). Annual statistical digest of Nigerian universities 2022/2023. NUC.

[2] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," Telematics Inf., vol. 37, pp. 13-49, 2019. https://doi.org/10.1016/j.tele.2019.01.007

[3] J. L. Rastrollo-Guerrero, J. A. Gomez-Pulido, and A. Duran-Dominguez, "Analyzing and predicting students performance by means of machine learning: A review," Appl. Sci., vol. 10, no. 3, p. 1042, 2020. https://doi.org/10.3390/app10031042

[4] A. K. Pal and S. Prakash, "Analysis of students academic performance using educational data mining: A case study," Int. J. Eng. Technol. Manag. Appl. Sci., vol. 5, no. 2, pp. 113-124, 2017.

[5] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining educational data to predict students academic performance using ensemble methods," Int. J. Database Theory Appl., vol. 9, no. 8, pp. 119-136, 2016. https://doi.org/10.14257/ijdta.2016.9.8.13

[6] D. Delen, "Predicting student attrition with data mining methods," J. Coll. Student Retention, vol. 13, no. 1, pp. 17-35, 2011. https://doi.org/10.2190/CS.13.1.b

[7] V. O. Oladokun, A. T. Adebanjo, and O. E. Charles-Owaba, "Predicting students academic performance using artificial neural network," Pac. J. Sci. Technol., vol. 9, no. 1, pp. 72-79, 2008.

[8] E. A. Ekubo and B. M. Esiefarienrhe, "Using machine learning to predict low academic performance at a Nigerian university," Afr. J. Inf. Commun., vol. 30, pp. 1-33, 2022. https://doi.org/10.23962/ajic.i30.14839

[9] M. Maphosa, W. Doorsamy, and B. Paul, "Improving academic advising in engineering education with machine learning using a real-world dataset," Algorithms, vol. 17, no. 2, p. 85, 2024. https://doi.org/10.3390/a17020085

[10] M. Niazkar, B. Abbasi, and N. Pirayesh, "Predicting student academic performance using machine learning: A systematic review," Educ. Inf. Technol., vol. 29, pp. 6679-6725, 2024. https://doi.org/10.1007/s10639-023-12003-3

[11] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student engagement predictions in an e-learning system," Comput. Intell. Neurosci., vol. 2018, Art. no. 6347186, 2018. https://doi.org/10.1155/2018/6347186

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority oversampling technique," J. Artif. Intell. Res., vol. 16, pp. 321-357, 2002. https://doi.org/10.1613/jair.953

[13] A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic course planning recommendation and students performance prediction multi-modal based on educational data mining techniques," J. Comput. High. Educ., vol. 38, pp. 38-76, 2025. https://doi.org/10.1007/s12528-024-09426-0

[14] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, and D. J. Murray, "Predicting academic performance by considering student heterogeneity," Knowledge-Based Systems, vol. 161, pp. 134-146, 2018. https://doi.org/10.1016/j.knosys.2018.07.042

[15] A. O. Adejo and T. Connolly, "Predicting student academic performance using multi-model datasets: A machine learning approach," Appl. Artif. Intell., vol. 32, no. 2, pp. 139-157, 2018. https://doi.org/10.1080/08841514.2018.1441074

[16] R. F. Kizilcec and H. Lee, "Algorithmic fairness in education," in Proc. Seventh ACM Conf. Learning @ Scale, 2020, pp. 1-10. https://doi.org/10.1145/3386527.3405924

[17] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 4765-4774, 2017.

[18] S. Hakkal and A. A. Lahcen, "XGBoost to enhance learner performance prediction," Comput. Educ.: Artif. Intell., vol. 7, Art. no. 100254, 2024. https://doi.org/10.1016/j.caeai.2024.100254

[19] Y. A. Alsariera, "Assessment and evaluation of different machine learning algorithms for predicting student performance," Comput. Intell. Neurosci., vol. 2022, Art. no. 4151487, 2022. https://doi.org/10.1155/2022/4151487

[20] G. D. Kuh, "The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties," Indiana University, Bloomington, 2001.

[21] P. R. Pintrich, D. A. F. Smith, T. Garcia, and W. J. McKeachie, "A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)," University of Michigan, 1991.

[22] M. Tavakol and R. Dennick, "Making sense of Cronbachs alpha," Int. J. Med. Educ., vol. 2, pp. 53-55, 2011. https://doi.org/10.5116/ijme.4dfb.8dfd

[23] R. J. Fisher, "Social desirability bias and the validity of indirect questioning," J. Consum. Res., vol. 20, no. 2, pp. 303-315, 1993. https://doi.org/10.1086/209351

Declarations

Funding Statement

No funding received.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Ethical Considerations

The authors state that all work related to this research was conducted in accordance with institutional, national, and international guidelines, and in compliance with recognized ethical standards, where;

It received ethical approval from the institutional ethics committee of Thomas Adewumi University, Oko, Kwara State, Nigeria, approval number Tau-2026-056.

Informed consent was obtained from all individual participants included in the study.

All Usage Statement

No generative AI tools were used in the preparation of this manuscript.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Code Availability Statement

The Software codes are available from the corresponding author upon reasonable request.

SDG Alignment

This research is aligned with the following United Nations Sustainable Development Goals (SDGs):

SDG 4 – Quality Education