Edupredict: A Comparative Data Analytics Framework for Student Academic Performance Prediction Using Machine Learning Algorithms

Olumide Abraham Temilade[1]*, Olajide Michael Segun[2], Lawal Olusegun Olayinka[1], Olu-Pereao Oluwaseyi Teminijesu[1], and Olumide Tabitha Funmilayo[3]

[1]*Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Nigeria*
[2]*Department of Computer Science, Adeyemi Federal University of Education, Ondo, Nigeria*
[3]*Department of Crop, Soil and Pest Management, The Federal University of Agriculture, Akure, Nigeria*
**Corresponding Author: atolumide@jabu.edu.ng*

ABSTRACT

The increased demand for data-based decision-making has led to a surge in the use of predictive analytics to identify students at risk of academic failure. However, many tertiary institutions still rely mostly on Cumulative Grade Point Average (CGPA) as an indicator of performance. In this study, we analyzed and compared the predictive capability of three supervised machine learning algorithms (Logistic Regression, Decision Tree, and Random Forest) to forecast student academic outcomes using demographic, behavioral, and academic attributes. We preprocessed the dataset through missing-value handling, feature encoding, and scaling, then split it into 70% training and 30% testing sets. We trained, hyperparameter-tuned, and evaluated the models using accuracy, precision, recall, F1-score, and AUC-ROC, along with Pearson correlation analysis. Results showed that Random Forest performed best (accuracy 85%, AUC-ROC 0.92), outperforming Logistic Regression (78% accuracy, AUC-ROC 0.88) and Decision Tree (74% accuracy, AUC-ROC 0.82). Correlation analysis showed that prior grade was the strongest predictor ($r = 0.94$), followed by study time, while absences and age showed negligible association; this association is descriptive, not causal. These findings informed the development of *Edupredict*, a web-based analytics tool that generates real-time probability-of-pass scores from student-level inputs. As a prototype trained entirely on synthetic data, *Edupredict* demonstrates that such a tool is technically feasible; the reported metrics should be read as evidence that the analytic pipeline works as intended, not as a validated real-world performance, pending future testing on real institutional data.

KEYWORDS: Machine Learning; Student Performance Prediction; Educational Data Mining; Random Forest; Logistic Regression; Decision Tree; Predictive Analytics

I. INTRODUCTION

The growing demand for tertiary education has placed sustained pressure on universities worldwide, many of which lack sufficient capacity to accommodate an expanding student population. Judging academic performance through CGPA or average grade remains common, but many newcomers still have first-year CGPAs below the minimum required level. This diagnosis comes too late once the student's academic career is already damaged. Schools that only realize a student is struggling after the student fails a semester have already missed the window to provide academic advice or adjust the student's course load. This is exactly why early-warning analytics will fill the gap. Recent research has shown that a student's academic difficulties can be highlighted even before a final grade is given, regardless of whether student data comes from online-learning platform interaction logs (Al-Azazi & Ghurab, 2023; Abdullah et al., 2023), secondary-school data (Hoyos & Caicedo-Castro, 2024;

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.22981408

Alghamdi & Rahman, 2023), or demographic & attendance records (Sideridis & Alamri, 2023). This shift reflects a transition from traditional reporting to more actionable, forward-looking data analysis. To structure a perspective that makes use of data analytics, the problem boils down to turning the different and varied data collected at the student level into a forecast of student performance through a well-devised set of phases including but not limited to data cleaning, exploration of correlations amongst the features, model development, and lastly, the transformation of these forecasts into instruments that would support decision-making. This paper compares the performance of three well-established machine learning models for classification problems, Logistic Regression, Decision Tree, and Random Forest models, in predicting student outcomes and is expected to identify the model that offers the best combination of accuracy and interpretability for use in institutional settings. We have several main goals in this article: (i) obtaining and pre-processing a dataset of student performance indicators that have been structured; (ii) training the three algorithms within a uniform evaluation setup; and (iii) comparing performance using typical classification metrics, while performing a correlation-based analysis of the predictors as a complement. This paper's novelty lies in adding another perspective on using correlation as a diagnostic method that, combined with different machine learning algorithms, can be more effective than simpler methods while also explaining why.

II. RELATED WORKS

A. Algorithmic Approaches in Educational Data Mining

Researchers use various machine learning models to predict student performance, usually balancing prediction accuracy with interpretability. Researchers prefer most tree-based methods. Decision Trees, one of the simplest such methods, clearly show how each prediction is made from a set of rules (Baradwaj & Pal, 2012). Random Forest combines several decision trees into an ensemble that, through diversification, becomes more reliable (Breiman, 2001). The literature often reports the benefits of diverse trees in a forest (Batool et al., 2023). In contrast, comparative studies show that no single algorithm performs best across all datasets (Kotsiantis et al., 2004; Chen & Zhai, 2023; Hoyos & Caicedo-Castro, 2024). New studies also show that this variety of algorithms appears at the deep learning level, i.e., models capable of handling sequences such as Long Short-Term Memory and interpretable networks using SHAP explanations (Al-Azazi & Ghurab, 2023; Kalita et al., 2025). A major obstacle is the imbalance between students who fail and those who pass; without a remedy, classifiers may prefer predicting passing over failure (Chawla et al., 2002; Masood et al., 2025).

B. Educational Data Mining: Scope and Predictor Importance

Educational Data Mining (EDM) uses classification, clustering, and association-rule mining on educational data to increase student retention and support educators' decisions (Romero & Ventura, 2007). Previous academic records are generally regarded as an important factor for forecasting future academic attainment. Demographic factors have less influence than previous academic records (Khairy et al., 2024; Alghamdi & Rahman, 2023; Sideridis & Alamri, 2023; Abdullah et al., 2023). Engagement and attendance records contain useful signals, but their strength fluctuates. Motivation is a plausible but rarely recorded variable in most datasets, including this one, in terms of measurement. Another ongoing issue is the lack of correlation analysis alongside reported predictor importance; this project empirically demonstrates this shortcoming in the fourth main section (4.2).

C. Empirical Studies

Several empirical studies have greatly inspired the design of this work. Batool et al. (2023) synthesized approximately 260 research papers on predictive factors; Chen and Zhai (2023) compared seven

https://doi.org/10.5281/zenodo.22981408

algorithms across three task types and found that Random Forest outperformed the others. Khairy et al. (2024) and Hoyos and Caicedo-Castro (2024) tuned classifiers using secondary-school records (Hoyos' model reached almost 80% accuracy after five-fold cross-validation). Similarly, Bellaj et al. (2024), Ahmed (2024), and Hussain and Khan (2023) used multiple algorithms to benchmark secondary-level datasets. Al-Azazi and Ghurab (2023) and Abdullah et al. (2023) extended work on predicting behavioral logs, whereas Kalita et al. (2025) combined deep learning with SHAP-based interpretability. Masood et al. (2025) and Nguyen et al. (2024) also used SMOTE to address class imbalance. This study directly addresses the remaining gap: few articles have paired model comparison with explicit correlation analysis.

D. Research Gaps

At least three research gaps are still open. First, context-specific models: numerous studies are confined to a single institution, which restricts a study's applicability to other settings (Kotsiantis et al., 2004; Alghamdi & Rahman, 2023). The second issue was real-time prediction, as most existing systems take an initial batch of data for a single decision, rather than updating the model with new incoming data (and monitoring the performance). Also, for 'explainability' or 'interpretability', although simpler models tend to give better performance on predictions (Kalita et al., 2025), the output of the model does not show how the prediction is made based on the potentially relevant factors. The three issues were addressed in this work through (a) applying the same dataset to all three algorithms, (b) reporting the full multi-metric evaluation results, and (c) combining model comparison with explicit explanations translated into human-understandable terms (i.e., interpretability).

III. METHODOLOGY

A. Research Design and Data

This project uses quantitative methods, including machine learning, on a structured dataset of various student performance features. It comprises phases of data collection, pre-processing, model training, and evaluation; the trained models were operationalized as a browser-based prototype, *Edupredict*, built to demonstrate the pipeline end-to-end rather than as a validated institutional system. Figure 1 shows a snapshot of the research and analytics method developed here, depicting the pipeline of student records through pre-processing, model training, multi-metric assessment, and correlation-based diagnostics, up to its use as a live predictive tool. The dataset is an imitation of 100 student records created to cover the demographic, academic, and behavioral characteristics of the students – gender, age, level of education of the parents, hours devoted to studying each week, absences, access to the internet, extra-course enrollment, previous grades, and final grades (pass/fail dichotomized at 50). Instead of genuine institutional documents, we used synthetic data to build this prototype because confidentiality regulations prohibit disclosure of private student records. We constructed the sample to approximate realistic attribute distributions rather than draw it from an existing repository. Accordingly, all results reported in this study are preliminary: they demonstrate that the analytical pipeline functions as intended on data with these statistical properties, but they do not yet constitute evidence about actual university students, and should not be generalized to real institutional populations until validated on genuine data."

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.22981408

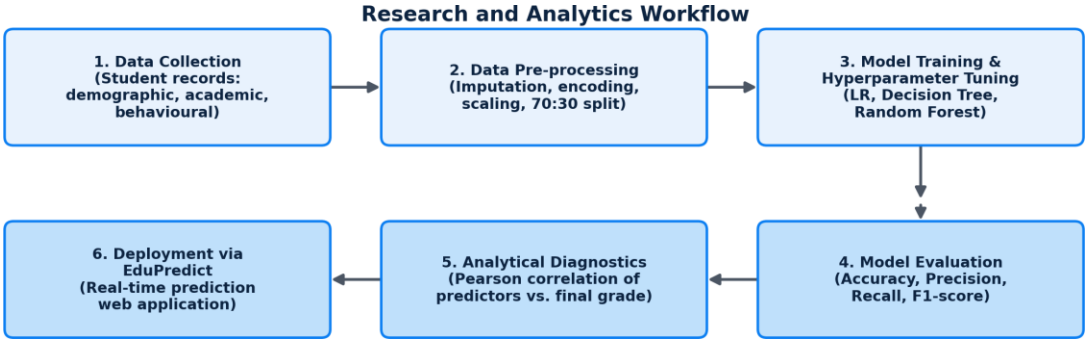

Figure 1: Research and analytics workflow adopted in this study

B. Data Pre-processing

Handling of missing values was performed by mean/mode imputation-a simple and highly effective approach (mainly effective for small datasets (samples), e.g., 100 records). We performed one-hot/label encoding (categorical attributes) and Min-Max scaling (numeric attributes), respectively. We split the dataset into train/test sets at a 70:30 ratio ($n \approx 30$) using stratified sampling. Because of the multi-metric predictive assumption, we used a relatively large test partition. We packaged the entire setup, including data processing, into *Edupredict*, a web-based data analytics application. See Fig. 2 for the platform's interface, which shows its main functionality: predicting student outcomes at the individual level and providing batch analysis for researchers and teachers.

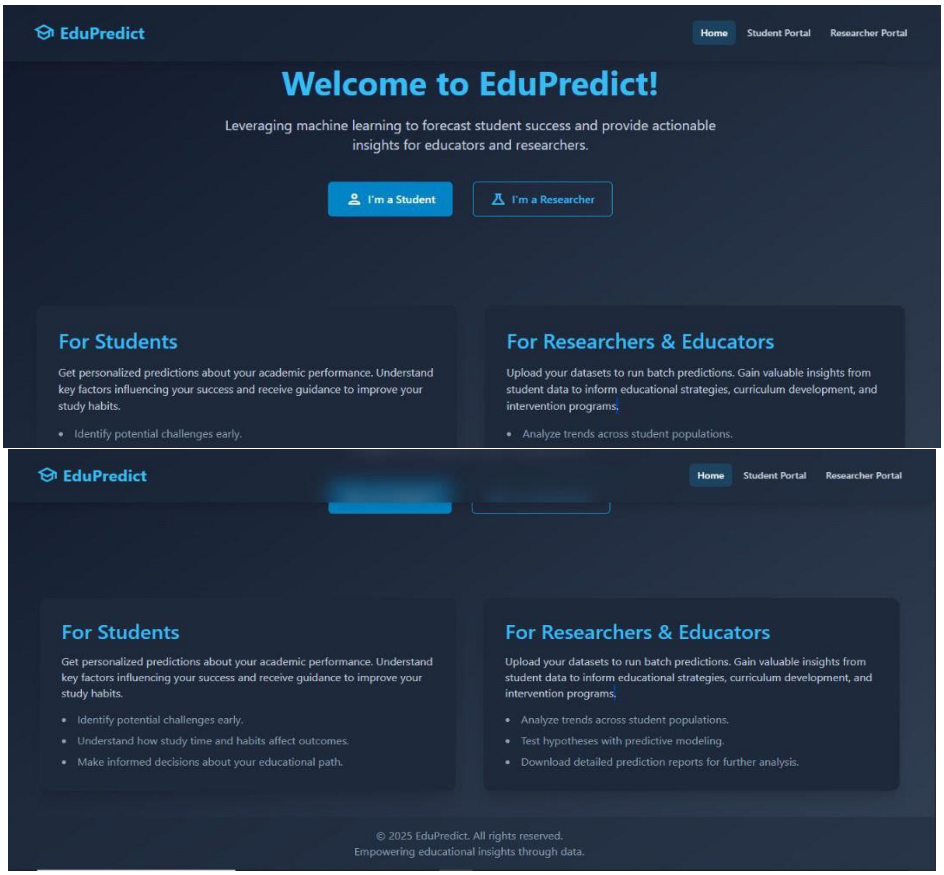

Figure 2: Landing Interface of the Edupredict Web Analytics Platform

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.22981408

C. 3.3 Machine Learning Algorithms

(i) Logistic Regression: It is a statistical regression model that uses a linear function and estimates the logistic probability of a binary outcome as a function of one or several explanatory variables in such a way that the coefficients can easily be interpreted as the marginal effects each variable exerts on the outcome variable (Hosmer & Lemeshow, 2000).

(ii) Decision Tree Classifier: This algorithm works automatically and intelligently so that, based on the information at hand, it decides on the best possible split for each level to predict the target variable (Kotsiantis et al., 2004).

(iii) Random Forest: This is an example of an ensemble technique where the classifier consists of a collection of classifiers (trees) that are trained independently on randomly selected samples of the dataset (the bootstrap samples), and their individual outputs are combined (aggregated) before finally giving the model output (class prediction). This reduces the model's tendency to overfit, improving performance in the presence of data noise (Breiman, 2001).

D. Model Training and the Edupredict Prototype

Each algorithm was first trained with an iterative search optimization method and then was fine-tuned using cross-validation and grid search. Three classification methods were embedded in a layered website. In Figure 3, the architecture of the deployed system is revealed, which is made up of a presentation layer that gathers the input through Student and Researcher Portals, an application layer that deals with data routing and validation, an analytics layer that handles data sanitization and the three machine learning models, respectively, and finally, a data layer that stores the data as well as the serialized models. Figure 4 shows the resulting data-entry interface. A user inputs their personal and educational details and immediately receives a computed probability of whether they will pass, based on the model.

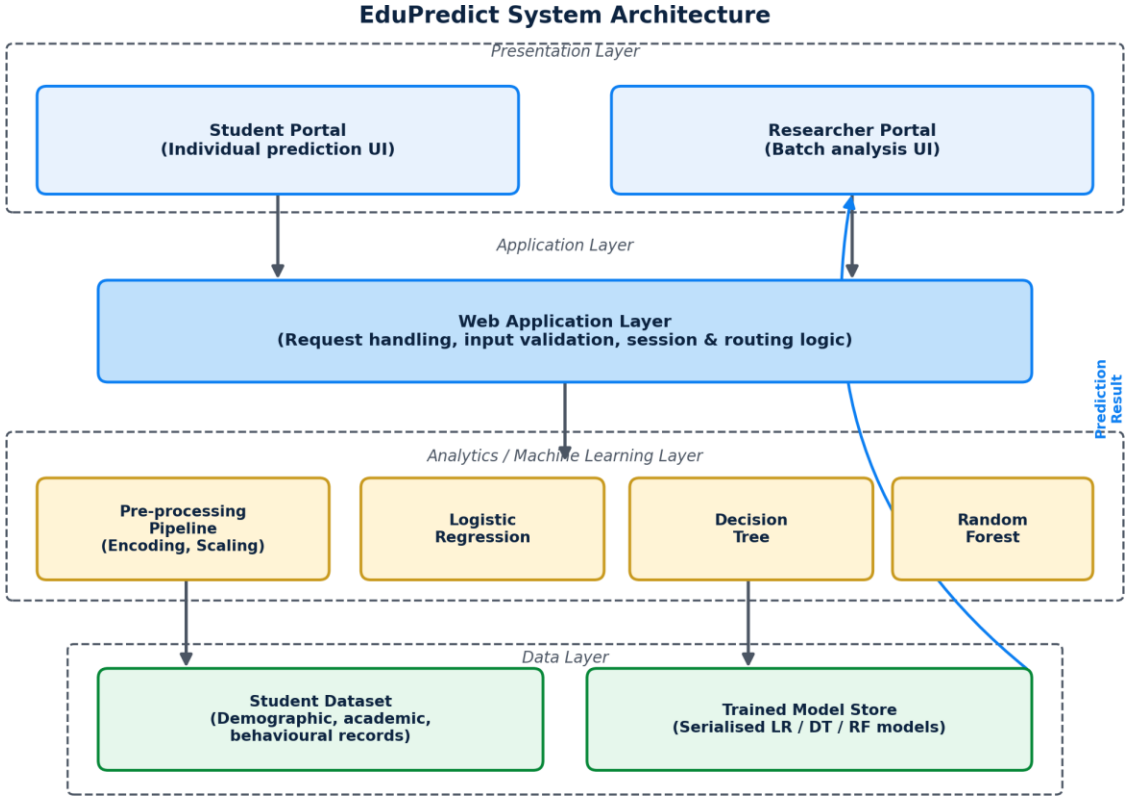

Figure 3: System Architecture of the Edupredict Web Analytics Platform

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.22981408

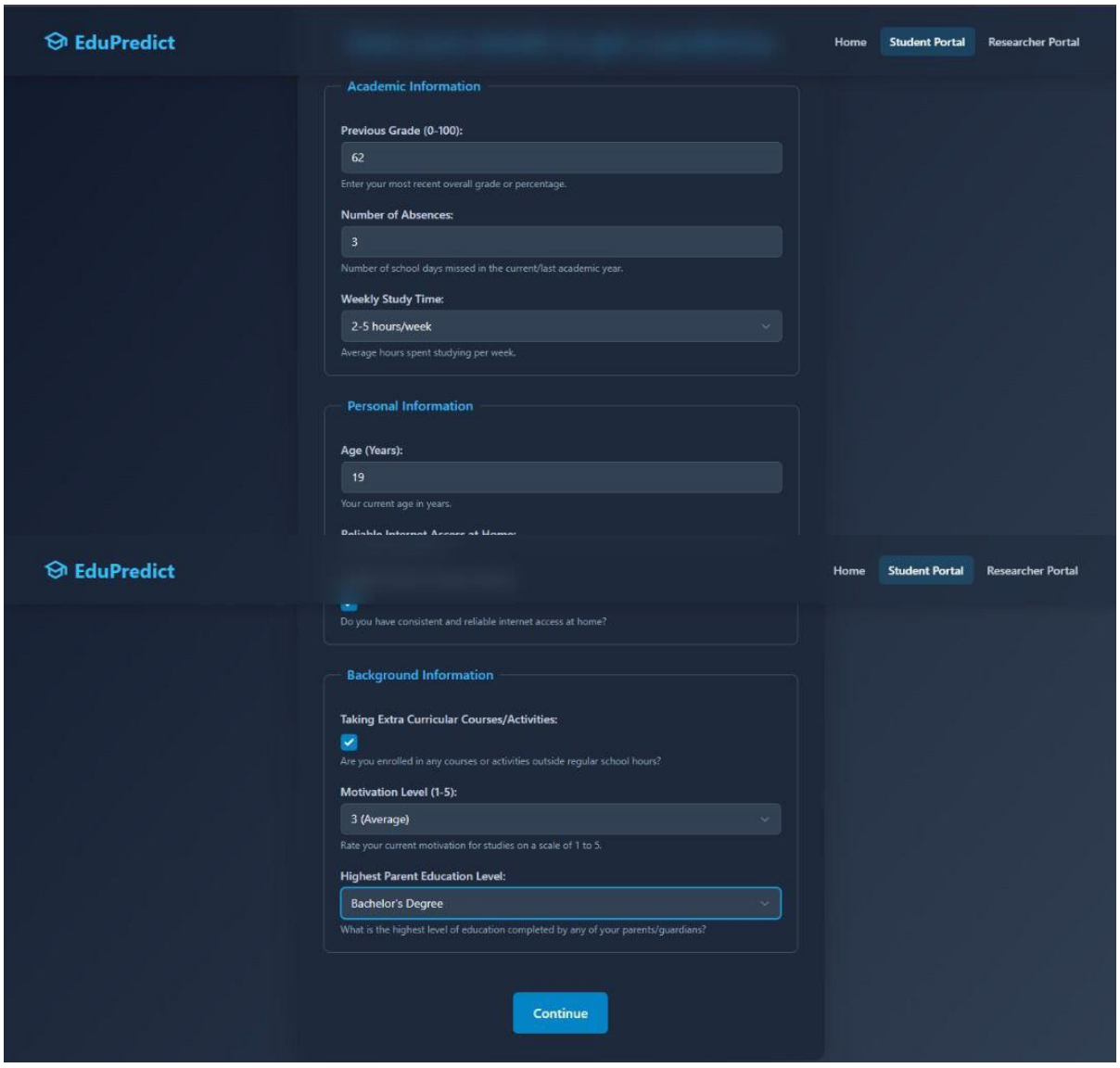

Figure 4: Edupredict Data-entry Interface for Academic and Personal Attributes

E. Model Evaluation

Model performance was assessed using:

(a). Accuracy = (TP+TN)/(TP+TN+FP+FN)

(b). Precision = TP/(TP+FP)

(c). Recall = TP/(TP+FN)

(d). *F1-Score = 2×(Precision × Recall)/(Precision + Recall)*

where TP, TN, FP, FN denote true/false positives/negatives, respectively. We supplemented these with AUC-ROC (Area Under the Receiver Operating Characteristic curve), a threshold-independent discrimination measure that reflects the data's favorable class distribution; the 83%/17% pass/fail split (Section 4.1) reflects moderate class imbalance, under which accuracy alone can overstate performance by rewarding majority-class prediction. Figure 5 shows *Edupredict*'s model-selection interface, which reports live accuracy and AUC-ROC estimates for each candidate model; Table 3 reports the AUC-ROC values. The dataset used for system development was synthetic. No real student's personal

https://doi.org/10.5281/zenodo.22981408

information was collected or exposed. We reviewed model outputs for fairness across subgroups and noted the synthetic origin as a limitation in Section 4.4.

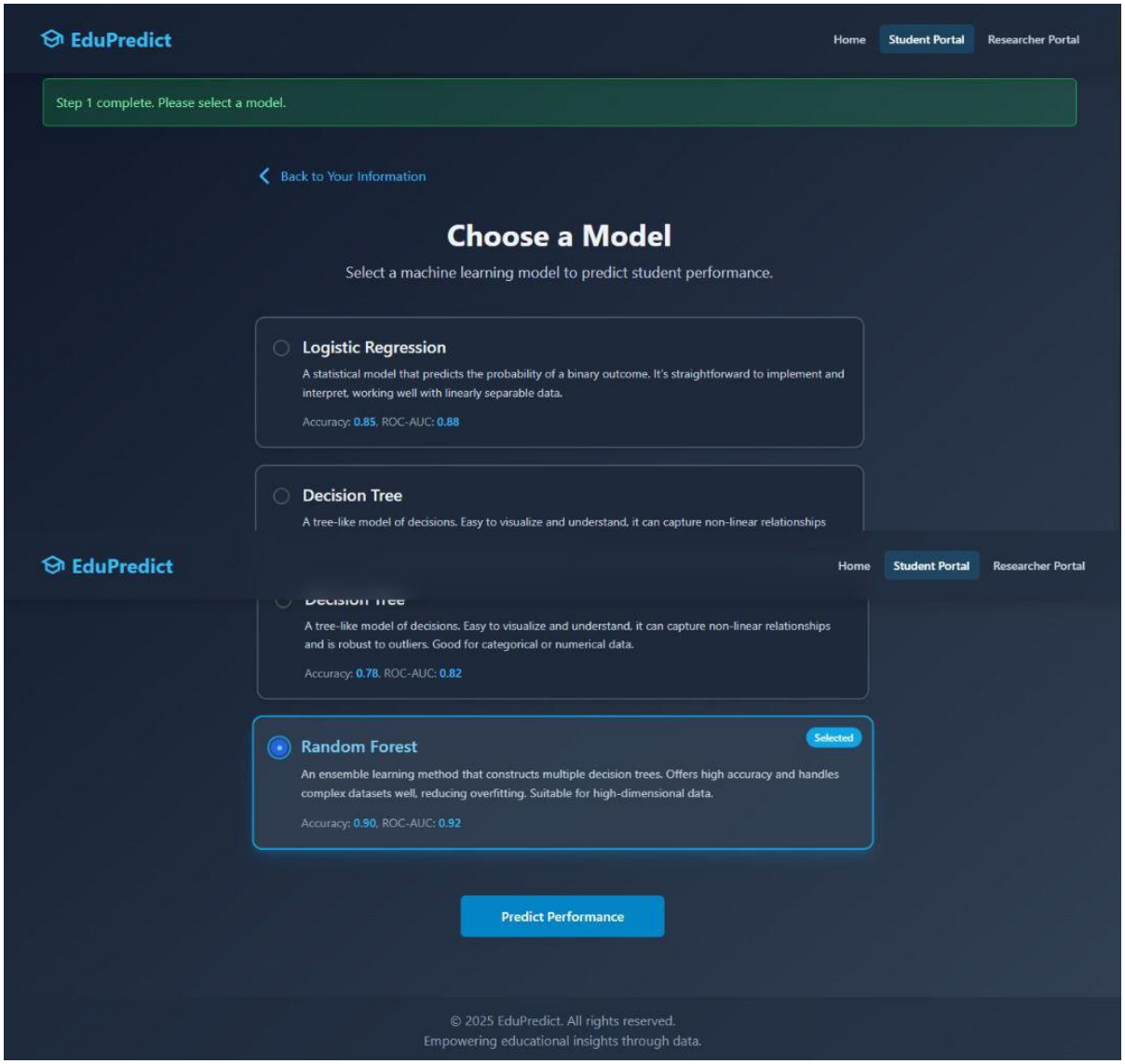

Figure 5: Edupredict Model-selection Interface with Live Accuracy and AUC-ROC Display

F. Analytical Framework and Tools

Beyond model evaluation, the study builds on descriptive analytics, correlation diagnostics, and comparative performance diagnostics to move from model accuracy to a fact-based explanation of why models and features perform as they do. Implementation tools include Python with Scikit-learn, Pandas, and Matplotlib.

IV. RESULTS AND DISCUSSION

G. Descriptive Analytics of the Study Sample

Table 1 summarizes the sample: fairly balanced by gender (56%/44%), aged 16–21 (mean 18.5), with a mean previous grade of 68.8 and a mean final grade of 69.2. The overall pass rate of 83% corresponds

https://doi.org/10.5281/zenodo.22981408

to 83 passing and 17 failing cases out of 100, a moderate class imbalance. We applied no class weighting, oversampling (e.g., SMOTE), or other explicit balancing techniques when training the three models; all three were trained directly on this imbalanced distribution, which we revisit as a limitation in Section 4.4.

Table 1: Demographics and Attributes Overview

Attribute	Value(s)	Mean/Rate
Gender	Male/Female	56%/44%
Age range	16–21 yrs	18.5
Study time	1–6 hrs/day	3.1
Absences	0–29 days	14.2
Parental education	Bachelor's/Master's/HS	37%/33%/30%
Internet/Extra courses	Yes/No	43% / 58%
Previous/Final grade	40–99 / 30–100	68.8 / 69.2
Pass rate	Final grade ≥ 50	83%

H. Correlation Analysis

Previous academic grade exhibited a strong positive correlation with final grade ($r = 0.94$). This should be interpreted with caution: a strong association between past and future academic performance is expected to some degree, given that the two are measuring closely related, temporally adjacent outcomes rather than independent phenomena, and correlation alone does not establish that previous grade causally drives final grade or that it is intrinsically more 'important' than other factors — only that it is the strongest linear signal available in this dataset. By contrast, study time ($r = -0.04$), absences ($r = -0.005$), and age ($r = -0.07$) each showed negligible correlation — a finding discussed further in Section 4.4.

Table 2: Correlation Analysis

Feature	r	Interpretation
Previous grade	+0.94	Strong positive
Study time	−0.04	Weak negative
Absences	−0.005	Negligible
Age	−0.07	Negligible

Figures 6 and 7 illustrate this dynamic through two representative *Edupredict* outputs. Figure 6 shows the prediction output for a profile with a previous grade of 62%, three absences, and age 19, assigned a probability-of-pass of only 32.0% — attributable primarily to the below-average previous grade rather than absences or age, which are both negligible predictors. In contrast, Figure 7 shows the Logistic Regression prediction output for a different comparison profile, which returned a substantially higher probability-of-pass score of 94.1%.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.22981408

Figure 6: Prediction Output for a Below-average Previous-grade Profile (32.0% pass probability)

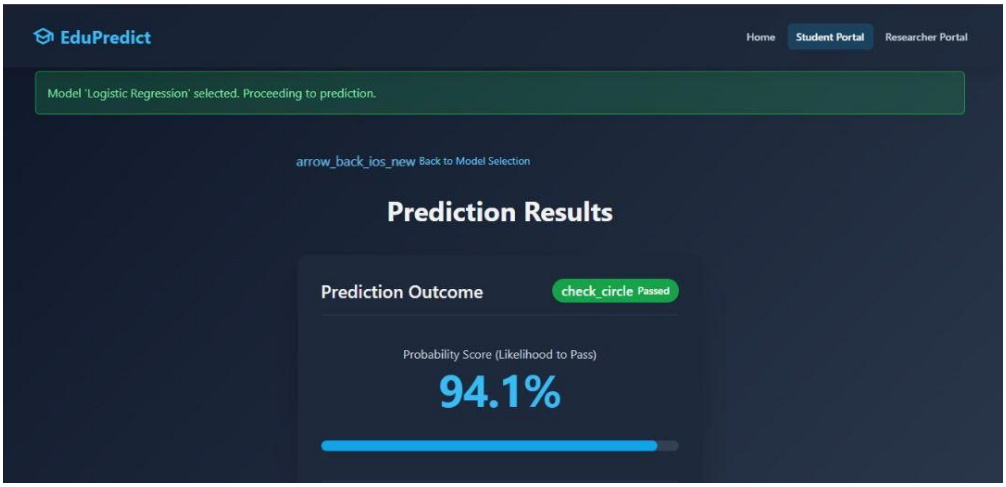

Figure 7: Logistic Regression Prediction Output for a Comparison Profile (94.1% pass probability)

I. Model Performance Comparison

Table 3 shows that Random Forest achieved the highest performance across all five metrics.

Table 3: Model Performance Comparison

Algorithm	Accuracy	Precision	Recall	F1	AUC-ROC
Logistic Regression	0.78	0.76	0.79	0.77	0.88
Decision Tree	0.74	0.72	0.75	0.73	0.82
Random Forest	0.85	0.83	0.86	0.84	0.92

J. Discussion

Random Forest's advantage is consistent with, though not fully explained by, its ensemble design (Breiman, 2001): averaging across many bootstrapped trees reduces the variance that likely hurt the single Decision Tree on this small sample, and this pattern is echoed elsewhere in the literature (Chen & Zhai, 2023; Kalita et al., 2025). Beyond architecture, the correlation results in Section 4.2 offer circumstantial support for why Random Forest performed well: with previous grade carrying almost all of the linear signal in the dataset, a model built to repeatedly re-split on the most informative feature across many trees is well positioned to exploit it. For instance, different methods do show different

https://doi.org/10.5281/zenodo.22981408

performance numbers (Batool et al., 2023). The AUC-ROC score of 0.92 here supports this, independent of the performance threshold, showing that the model is more likely to support use as a risk-scoring tool than simple classification. Logistic Regression offered a reasonable balance of accuracy and interpretability, while Decision Tree's weaker performance is most plausibly due to overfitting on a small (n=100) sample. Analytically, the correlation results show nearly all predictive signal concentrated in one feature — previous grade, with the behavioral/demographic ones bringing only a very low contribution. This is an indication that the modest accuracy ceiling (74–85%) is more likely a result of the data limitations than of the algorithms.

These results also carry implications specific to the Nigerian tertiary context in which this study is situated. Nigerian universities and colleges of education typically serve large, growing student populations against a comparatively small base of academic-support staff, making automated, low-cost early-warning tools such as *Edupredict* potentially valuable precisely where individual monitoring is least feasible. At the same time, the dataset's internet-access feature (43% of the synthetic sample) reflects a genuine constraint in many Nigerian institutions, where uneven connectivity could limit which students a web-based tool such as *Edupredict* can reach in practice, and where parental education and socioeconomic background often vary sharply within a single institution and even more so between institutions. Any future validation on real data should therefore stratify by institution type (e.g., federal, state, and private universities) and connectivity level, rather than assuming the correlation and performance patterns found here transfer uniformly across Nigeria's highly heterogeneous tertiary landscape. These findings have an important limitation: the 100 records analyzed are synthetic. The reported 85% Random Forest accuracy, and every other metric in Table 3, should therefore be read strictly as evidence that the pipeline behaves as intended on data with these statistical properties, not as an estimate of how the models would perform on real students. Generalization to actual institutional populations requires validation on genuine data and cannot be assumed from these results. The small sample and the single 70:30 train–test split, rather than repeated k-fold cross-validation, are a particular concern: with only 30 test-set records, the reported 85% Random Forest accuracy is sensitive to which records happened to fall into the test partition, and a different split could plausibly yield a noticeably different figure. Repeated k-fold cross-validation is therefore a necessary step for any future reporting of these metrics, not an optional refinement, and the absence of psychosocial factors further limits generalizability. However, the alignment between accuracy and AUC-ROC-based rankings, together with the case-level correspondence in Section 4.2, confirms that it is quite feasible to evaluate both predictive accuracy and explainability together - in other words, closing the gap in explainability mentioned in Section 2.4. As a prototype trained entirely on synthetic data, *Edupredict* demonstrates that such a tool is technically feasible and illustrates how it would present results to a student adviser, it is not yet a validated decision-support system, and reliability for actual early-warning use is bounded both by data richness and by the real-data validation step described above.

K. Benchmark Comparison with Related Studies

Table 4 benchmarks *Edupredict* against four closely related recent studies from Section 2, each applying machine learning to student-performance prediction but differing in explainability treatment and deployment.

Table 4: Benchmark Comparison with Recent Related Studies

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.22981408

Study	Algorithms	Explainability	Deployed Tool	Key Difference from *Edupredict*
Chen & Zhai (2023)	7 algorithms, 3 tasks	Not addressed	No	Accuracy ranking only; no correlation diagnostics or prototype
Hoyos & Caicedo-Castro (2024).	Multiple tuned classifiers	Not addressed	No	Accuracy-focused tuning (>80% after 5-fold CV); no interpretability layer
Kalita et al. (2025).	Bi-LSTM (deep learning)	Post-hoc SHAP	No	Explainability added after the fact to a black-box model
Ahmed (2024)	Multiple ML algorithms	Not addressed	No	Benchmarks algorithms only; no deployment or decision support
This study (*Edupredict*)	LR, Decision Tree, RF	Built-in correlation + AUC-ROC	Yes – prototype web app	Multi-metric benchmarking, quantified explainability, and live deployment in one framework

Most comparative studies merely report a ranked order of accuracy (Chen & Zhai, 2023; Hoyos & Caicedo-Castro, 2024; Ahmed, 2024), or they attach an explainability method to a highly complicated model only after making predictions (Kalita et al., 2025). None combine an explanatory, statistically based diagnostic layer with a system-in-production, explainable application like Edupredict. *Edupredict's* main competitive edge stems from this feature: explanation is an essential part of the pipeline, where the correlation diagnostic process identifies which predictors drive model outputs, rather than being added later as an afterthought. Furthermore, the models are quickly turned into a live, user-facing tool accessible to non-technical decision-makers, rather than left as a static leaderboard. By taking a comprehensive, continuous-cycle approach that connects benchmarking, interpretation, and deployment, *Edupredict* surpasses other methods and provides a holistic solution to the explainability challenge outlined in Section 2.4.

V. CONCLUSION

Universities have long used CGPA to indicate students' past performance, with few opportunities to act before underperformance sets in. Because of these limitations, this research evaluated and compared Logistic Regression, Decision Tree, and Random Forest for predicting student academic performance; it also conducted a correlation analysis and identified the predictors that most influence model outputs. On all measures, Random Forest was clearly the most accurate algorithm. Logistic Regression proved to be a reasonable compromise, providing high accuracy while remaining interpretable. Decision Tree, however, performed markedly worse on every measure, consistent with its well-documented sensitivity to small training samples. From the data analysis, the most important point is not the ranking of the algorithms. The real point is the source of the predictive power: previous academic grade alone accounted for the overwhelming share of the model's predictive signal ($r = 0.94$, corresponding to $r^2 \approx 0.88$, i.e., roughly 88% of the variance in final grade). The data imbalance should help institutions decide what to collect next. In conclusion, this study makes a methodological contribution. First, the authors developed and used a unified evaluation system. Second, the analysis combines group comparisons (pairwise comparisons) with quantified correlation diagnostics. Finally, the authors propose a working decision-support prototype, Edupredict, as a practical contribution. Therefore,

https://doi.org/10.5281/zenodo.22981408

institutions that want to apply predictive analytics effectively will need ensemble methods that enable early detection of problems that can lead to failure. Second, more behavioral and psychosocial data need to be collected, and lastly, the system should be developed in a user-friendly way that allows easy interpretation. Researchers are advised to take the analysis to the next level by using larger datasets from different universities and real scenarios. In addition, consider time-series methods at the semester level. This would extend the analysis and allow the results to reflect different situations.

REFERENCES

Abdullah, M., Al-Ayyoub, M., Shatnawi, F., Rawashdeh, S., & Abbott, R. (2023). Predicting students' academic performance using e-learning logs. IAES International Journal of Artificial Intelligence, 12(2), 831–839. https://doi.org/10.11591/ijai.v12.i2.pp831-839

Ahmed, E. (2024). Student performance prediction using machine learning algorithms. Applied Computational Intelligence and Soft Computing, 2024, 4067721. https://doi.org/10.1155/2024/4067721

Al-Azazi, F. A., & Ghurab, M. (2023). ANN-LSTM: A deep learning model for early student performance prediction in MOOCs. Heliyon, 9(4), e15382. https://doi.org/10.1016/j.heliyon.2023.e15382

Alghamdi, A. S., & Rahman, A. (2023). Data mining approach to predict success of secondary school students: A Saudi Arabian case study. Education Sciences, 13(3), 293. https://doi.org/10.3390/educsci13030293

Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. arXiv. https://doi.org/10.48550/arXiv.1201.3417

Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H.-Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies, 28(1), 905–971. https://doi.org/10.1007/s10639-022-11152-y

Bellaj, M., Ben Dahmane, A., Boudra, S., & Lamarti Sefian, M. (2024). Educational data mining: Employing machine learning techniques and hyperparameter optimization to improve students' academic performance. International Journal of Online and Biomedical Engineering, 20(3), 55–74. https://doi.org/10.3991/ijoe.v20i03.46287

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. Education and Information Technologies, 28(9), 12039–12057. https://doi.org/10.1007/s10639-023-11672-1

Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression (2nd ed.). Wiley. https://doi.org/10.1002/0471722146

Hoyos, W., & Caicedo-Castro, I. (2024). Tuning data mining models to predict secondary school academic performance. Data, 9(7), 86. https://doi.org/10.3390/data9070086

Hussain, S., & Khan, M. Q. (2023). Student-Performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of Data Science, 10(3), 637–655. https://doi.org/10.1007/s40745-021-00341-0

Kalita, E., Alfarwan, A. M., El Aouifi, H., Kukkar, A., Hussain, S., Ali, T., & Gaftandzhieva, S. (2025). Predicting student academic performance using Bi-LSTM: A deep learning framework with SHAP-

based interpretability and statistical validation. Frontiers in Education, 10, 1581247. https://doi.org/10.3389/feduc.2025.1581247

Khairy, D., Alharbi, N., & Amasha, M. A. (2024). Prediction of student exam performance using data mining classification algorithms. Education and Information Technologies, 29, 21621–21645. https://doi.org/10.1007/s10639-024-12619-w

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting students' performance in distance learning using machine learning techniques. Applied Artificial Intelligence, 18(5), 411–426. https://doi.org/10.1080/08839510490442058

Masood, S. W., Gogoi, M., & Begum, S. A. (2025). Optimized SMOTE-based imbalanced learning for student dropout prediction. Arabian Journal for Science and Engineering, 50, 7165–7179. https://doi.org/10.1007/s13369-024-09287-w

Nguyen, G. C., Nghiem, T. L., Ngo, T. H., Nguyen, M. T. L., & Phung, H. Q. (2024). Improvement in academic performance prediction of at-risk students by addressing the data imbalance problem. Applied Computational Intelligence and Soft Computing, 2024, 4795606. https://doi.org/10.1155/2024/4795606

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Sideridis, G., & Alamri, A. A. (2023). Predicting academic achievement and student absences in high school: The roles of student and school attributes. Frontiers in Psychology, 14, 987127. https://doi.org/10.3389/fpsyg.2023.987127