ISRG Journal of Arts, Humanities and Social Sciences (ISRGJAHSS)

ISRG PUBLISHERS
Abbreviated Key Title: ISRG J Arts Humanit Soc Sci
ISSN: 2583-7672 (Online)
Journal homepage: https://isrgpublishers.com/isrgjahss
Volume – III Issue –IV (July-August) 2025
Frequency: Bimonthly

SYSTEMATIC REVIEW OF THE USE OF MACHINE LEARNING IN PREDICTING STUDENT ACADEMIC SUCCESS IN SOUTH WESTERN NIGERIAN UNIVERSITIES

MOSES ADEKUNLE OGUNRINDE[1*], DR. MICHAEL OLUSOLA METIBEMU[2]

[1] A Ph. D. Candidate, Institute of Education, University of Ibadan, Ibadan, Oyo State, Nigeria.

[2] Institute of Education, University of Ibadan, Ibadan, Oyo State, Nigeria.

| **Received:** 24.07.2025 | **Accepted:** 29.07.2025 | **Published:** 19.08.2025

***Corresponding author:** MOSES ADEKUNLE OGUNRINDE

A Ph. D. Candidate, Institute of Education, University of Ibadan, Ibadan, Oyo State, Nigeria.

Abstract

Machine learning (ML) has become an increasingly important tool in educational research, particularly in predicting student academic performance. While this field has matured in many developed countries, its application in Nigeria remains emergent and fragmented. This systematic review synthesizes existing literature on the use of ML models to predict academic success among students in Nigerian universities. To evaluate the types of ML models used, features employed for prediction, model performance metrics, and methodological quality of studies conducted within the Nigerian university context. Following the PRISMA 2020 guidelines, a systematic search was conducted across major academic databases including Google Scholar, AJOL, IEEE Xplore, Science Direct, and Scopus. Studies published between 2010 and 2024 that used ML algorithms to predict academic performance in Nigerian universities were included. Data were extracted and synthesized using a narrative approach. Twenty-seven studies met the inclusion criteria. Decision Trees, Naïve Bayes, Logistic Regression, and Support Vector Machines were the most commonly applied models. Frequently used features included CGPA, WAEC/UTME scores, gender, and attendance. Reported prediction accuracies ranged from 70% to 92%, with ensemble models and XGBoost showing the highest performance. However, most studies lacked rigorous validation, large datasets, or feature explain ability, and few explored deployment in real-world academic systems. Machine learning holds significant potential to enhance academic prediction and student support systems in Nigeria. However, the field is still developing, with gaps in data quality, model robustness, and institutional integration. It is recommended that future research should focus on adopting advanced techniques, improving dataset diversity, and promoting real-world application in policy and university administration.

Keywords: Review, Machine learning, Students, Academic success, Nigerian universities.

DOI: 10.5281/zenodo.16917658

Introduction

In recent years, the rapid advancement of technology has transformed various sectors, including education. Among the most promising innovations is the application of Machine Learning (ML), a subfield of artificial intelligence in predicting student academic performance. Academic success remains a critical concern for educators, policymakers, and institutions, as it directly impacts graduate employability, national development, and the quality of human capital (Romero & Ventura, 2020; Kotsiantis et al., 2010). ML techniques offer powerful data-driven insights by identifying at-risk students, uncovering hidden learning patterns, and informing proactive educational interventions.

Globally, ML algorithms such as decision trees, support vector machines, artificial neural networks, and ensemble methods have been widely adopted to predict academic performance using diverse features including demographics, prior grades, attendance, behavioral logs, and digital interactions (Baker & Inventado, 2014; Al-Barrak & Al-Razgan, 2016). These models have demonstrated higher accuracy and adaptability compared to traditional statistical approaches, making them valuable tools for educational decision-making and personalized learning (Romero & Ventura, 2020).

In the Nigerian context, however, the application of ML in education is still evolving. While researchers have examined factors affecting academic performance through traditional methods (Adepoju & Owoeye, 2015; Arul & Aladesote, 2017), the integration of ML for predictive purposes is relatively recent. Emerging studies have begun exploring the use of decision trees, logistic regression, and naïve Bayes classifiers to predict student outcomes based on variables such as CGPA, UTME/WAEC scores, gender, and parental background (Onanuga et al., 2021; Ojo et al., 2022; Olatunji & Aladejana, 2023). Despite promising results, these studies often suffer from limitations such as small sample sizes, poor validation methods, and lack of interpretability tools.

Moreover, the deployment of predictive ML models in real-world academic planning remains minimal in Nigerian universities due to challenges including limited access to quality data, infrastructural constraints, and ethical concerns around data use (Mhlanga & Moloi, 2020). Consequently, the current state of research in Nigeria lacks cohesion, scalability, and practical relevance, highlighting the need for a structured synthesis of existing studies. This systematic review aims to fill that gap by evaluating and synthesizing published studies on the application of machine learning models to predict academic performance in Nigerian universities. Specifically, it examines the types of models used, features selected, methodological quality, and predictive effectiveness. By aggregating these findings, this review seeks to uncover research trends, identify methodological gaps, and offer recommendations for future research and practical implementation.

Literature Review

The application of Machine Learning (ML) in educational research has gained substantial traction worldwide. ML models are capable of analyzing large datasets and identifying patterns that may not be visible through traditional statistical methods. In education, such models have been applied to predict student academic performance, classify risk levels, and inform personalized learning pathways (Baker & Inventado, 2014; Kotsiantis et al., 2010). Techniques such as decision trees, random forests, support vector machines, artificial neural networks, and more recently, XGBoost and deep learning, have shown promising accuracy in predicting student outcomes based on various features such as prior grades, attendance, socio-economic status, behavioral engagement, and digital interaction logs (Al-Barrak & Al-Razgan, 2016). Studies in countries like the United States, India, and China have demonstrated that ML can outperform traditional regression techniques in identifying at-risk students, thereby aiding early intervention (Romero & Ventura, 2020). Moreover, ML's adaptability to big data allows continuous learning and refinement, which is essential in dynamic educational environments.

In Africa, the adoption of educational data mining (EDM) and ML tools is still evolving. While countries like South Africa and Kenya have begun integrating ML into university systems, challenges such as limited infrastructure, data unavailability, and lack of technical expertise persist (Mhlanga & Moloi, 2020). Nonetheless, researchers have shown growing interest in leveraging ML models to solve local education challenges. Reviews conducted in broader African contexts highlight both the potential of predictive analytics and the urgent need for context-specific models that reflect African educational realities.

In the Nigerian context, a number of studies have explored academic performance predictors using conventional statistical tools. However, ML-based prediction studies are relatively new and fragmented. Recent studies such as those by Onanuga et al. (2021), Ojo et al. (2022), and Olatunji & Aladejana (2023) have experimented with algorithms such as decision trees, Naïve Bayes, and logistic regression on Nigerian university datasets. These studies commonly use features like students' WAEC scores, CGPAs, socio-demographic attributes, and attendance records to forecast final classifications or dropout risks. Notably, most of these studies: Are limited in sample size and generalizability, lack robust cross-validation and comparative model evaluations, rarely use feature importance analysis, explainability (e.g., SHAP), or advanced ensemble methods. There is also little evidence of real-world application of these models in academic planning, policy decisions, or adaptive learning systems. This reflects a disconnect between theoretical research and practical deployment.

Thus, from the limited but growing body of Nigerian ML research in education, several gaps emerge: No comprehensive synthesis exists that aggregates findings across Nigerian studies. Inconsistencies in model types, features selected, and evaluation metrics. Lack of attention to data privacy, ethical considerations, and deployment challenges in local institutions. Few comparative studies across federal, state, and private universities to explore institutional differences. These gaps underscore the importance of a systematic review that compiles, appraises, and contextualizes existing research efforts. Such a review would not only inform future studies but also guide educational stakeholders seeking to harness AI for academic improvement in Nigeria. The literature suggests that while machine learning has demonstrated great promise in educational prediction globally, its application in Nigeria remains under-explored and under-synthesized. By conducting a systematic review, this study addresses a critical gap, providing a structured understanding of how, where, and to what extent machine learning has been used to predict academic performance in Nigerian universities.

Methodology

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines to ensure methodological rigor, transparency, and reproducibility.

DOI: 10.5281/zenodo.16917658

The review aims to synthesize existing evidence on the use of machine learning techniques to predict student academic performance in Nigerian universities. A systematic review design was employed to collect, evaluate, and synthesize relevant peer-reviewed studies. The approach involved a structured literature search, application of inclusion and exclusion criteria, data extraction, and critical appraisal of the selected studies.

The inclusion and exclusion criteria were established using the PICOS framework (Population, Intervention, Comparison, Outcome, and Study design). Concerning population, we made use of undergraduate and postgraduate students in Nigerian universities. There was use of machine learning techniques for predicting academic performance. On the outcome, there was predictive accuracy, classification results, performance indicators (e.g., CGPA, dropout risk, final classification). This methodological approach ensures a transparent, reproducible, and evidence-based synthesis of literature on the application of machine learning in predicting academic outcomes in Nigerian universities.

Results and Discussion

A total of 547 records were identified through database searches, with an additional 23 records from reference lists and institutional repositories. After removing duplicates, 492 unique records were screened based on titles and abstracts. Following full-text eligibility assessment of 72 articles, 27 studies met the inclusion criteria and were included in this review. The selection process is summarized in the PRISMA flow diagram (see Figure 1). Besides, the included studies were published between 2011 and 2024. Most were conducted in federal universities (59%), followed by state universities (30%) and private institutions (11%). Sample sizes ranged from 150 to over 3,000 students, and datasets were often obtained from university examination records, departmental archives, or institutional management systems. A variety of machine learning algorithms were used across the studies. The most commonly adopted models include:

ML Algorithm Number of Studies	Most Reported	Accuracy (%)
Decision Tree (e.g., C4.5, CART)	12	72–85%
Naïve Bayes	8	70-78%
Support Vector Machine (SVM)	8	75–83%
Logistic Regression (as ML).	7	65-80%
Random Forest / Ensemble Models	4	80–89%
XGBoost / Gradient Boosting	2	85–92%
Artificial Neural Networks (ANN)	3	78–88%

Note: Several studies compared multiple models in the same paper. The following were the most frequently used predictive features:

Cumulative Grade Point Average (CGPA) (all 27 studies)

WAEC/UTME scores (18 studies)

Gender and Age (15 studies)

Attendance and course registration data (12 studies)

Socioeconomic background (parents' occupation/education) (9 studies)

Engagement in e-learning platforms / LMS logs (4 studies)

Few studies explored psychological factors or real-time learning analytics, highlighting a gap in behavioral data usage. The most frequently used performance metrics were:

Accuracy (all studies)

Precision, Recall, and F1-Score (9 studies)

Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) (used in regression-based studies)

ROC-AUC (only in 3 studies)

Most studies reported moderate to high accuracy, with ensemble methods and XGBoost outperforming others in multi-class prediction tasks. Only 10 out of 27 studies used cross-validation (k-fold or holdout). Few studies employed hyper parameter tuning, feature selection techniques, or explainable AI tools (like SHAP or LIME), suggesting a need for stronger methodological rigor. South-West Nigeria was the most represented region (48%), followed by South-East (19%) and North-Central (15%). Studies from private universities reported higher data availability and digital record usage. Several federal institutions cited data access limitations, outdated systems, or ethical barriers in accessing student records. This review shows that machine learning is increasingly used to predict academic performance in Nigerian universities, with promising results. However, the field is still in its early stages, with limited use of advanced models, real-time data, or deployment in practice.

This systematic review synthesized evidence from 27 empirical studies on the use of machine learning (ML) for predicting student academic performance in Nigerian universities. The findings indicate a growing interest in ML-based educational prediction, yet also reveal important methodological, contextual, and implementation gaps that need to be addressed for broader impact. The increasing adoption of ML models in Nigerian education research reflects a positive shift toward data-driven academic management. The dominance of models like decision trees, Naïve Bayes, and logistic regression aligns with the ease of use and interpretability of these models. However, the limited application of ensemble methods, neural networks, and advanced models like XGBoost suggests that many studies are still at a foundational level of ML experimentation.

Moreover, despite some models reporting accuracies above 85%, model sophistication remains low, with minimal use of hyper parameter tuning, feature selection, or model explainability techniques such as SHAP or LIME. This limits the interpretability and real-world applicability of the findings in administrative or policy environments. The most frequently used features: CGPA, WAEC/UTME scores, gender, and attendance are consistent with earlier findings in educational psychology and performance modeling. However, the lack of real-time, behavioral, or psychological indicators (e.g., motivation, learning styles, LMS engagement logs) reduces the models' ability to adapt to modern, digitally enabled educational environments. Few studies incorporated contextual or environmental variables such as family background, socio-economic status, or access to school facilities; factors which, according to previous research, significantly influence student performance in Nigeria.

However, a major concern is the limited validation techniques applied across most studies. Less than half of the studies used cross-

DOI: 10.5281/zenodo.16917658

validation, and even fewer applied model evaluation beyond basic accuracy metrics. This raises questions about model generalizability and robustness when applied to unseen student populations across different institutions. Moreover, small dataset sizes, often below 1,000 students, restrict the scalability of the models. Most datasets were institution-specific, making it difficult to develop generalizable insights for the entire Nigerian university system. The lack of publicly available educational datasets further compounds this limitation. While these studies demonstrate academic promise, real-world deployment of ML models remains nearly nonexistent in Nigerian universities. There is little to no evidence that predictive models are being used for student advising, early warning systems, or adaptive teaching interventions. This signals a disconnect between research and policy or administrative practice. The contributing factors include: Low digital infrastructure in some institutions, privacy and ethical concerns around student data usage and lack of capacity or institutional will to integrate ML tools in student management systems.

Compared to global studies, particularly from developed countries, Nigerian studies are relatively nascent in terms of complexity, scale, and implementation. International reviews show more integration of real-time data, use of deep learning, and actual application of models in online learning platforms. Nigeria still lags in this regard, partly due to infrastructural, technical, and data governance challenges. This review highlights the need for:

Larger, collaborative datasets across multiple institutions

Advanced ML techniques with proper validation and interpretation

Increased focus on deploying predictive tools in real institutional settings

Training programs to build ML capacity among educational researchers and administrators

The integration of explainable AI, ethical frameworks, and responsible data practices is also crucial to ensure trust and fairness in predictive academic systems.

In sum, while ML-based prediction of academic performance is gaining ground in Nigerian universities, much work remains in scaling, validating, and translating these models into usable tools for improving student success. A collaborative, data-informed approach backed by institutional commitment and ethical practices is essential to unlock the full potential of machine learning in Nigerian higher education.

Conclusion and Recommendations

This systematic review examined 27 studies focused on the application of machine learning (ML) techniques for predicting student academic performance in Nigerian universities. The findings reveal a growing body of evidence supporting the feasibility of ML models—particularly decision trees, logistic regression, and ensemble models—in identifying factors that influence academic outcomes. While many studies reported high prediction accuracies, the field remains in an early stage of development, characterized by:

Small, institution-specific datasets

Basic ML models with limited validation

Minimal integration of real-time or behavioral data

Lack of deployment or policy integration in academic environments

Despite these limitations, machine learning holds strong potential to improve early intervention strategies, guide academic advising, and foster data-informed decisions in Nigerian universities.

To unlock this potential, future research and institutional policies must bridge the gap between theoretical model building and practical implementation, with careful attention to data ethics, quality, and interpretability. Based on the findings of this review, the following recommendations are proposed for researchers, policymakers, and university administrators:

1. Adopt advanced ML techniques such as XGBoost, Random Forests, and deep learning models to improve predictive power.
2. Incorporate explainable AI tools (e.g., SHAP, LIME) to interpret predictions and support human decision-making.
3. Use larger and more diverse datasets that include multi-institutional samples across Nigeria for better generalizability.
4. Apply robust validation techniques (e.g., cross-validation, bootstrap methods) and report multiple evaluation metrics.
5. Expand the scope of predictive features to include psychosocial, behavioral, and digital interaction data from LMS platforms.
6. Invest in data infrastructure and digital record systems to enable ML-ready data collection and management.
7. Collaborate with researchers and data scientists to develop student success prediction dashboards and early warning systems.
8. Pilot ML models within student support units to identify at-risk students and provide timely interventions.
9. Support the development of open-access educational datasets to facilitate collaborative research across institutions.
10. Develop national guidelines for the ethical use of AI and data in education, ensuring data privacy, fairness, and transparency.

Machine learning is not a replacement for educational judgment, but rather a powerful complement when used responsibly. By embracing its potential and addressing existing limitations, Nigeria's higher education system can move toward a more predictive, personalized, and inclusive model of student success.

References

1. Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting Students' Performance through Classification: A Case Study. Journal of Theoretical and Applied Information Technology, 86(1), 27–34.
2. Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In J. A. Larusson & B. White (Eds.), Learning Analytics (Pp. 61–75). Springer.
3. Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2010). Predicting Students' Performance in Distance Learning using Machine Learning Techniques. Applied Artificial Intelligence,18(5), 411–426. https://doi.org/10.1080/08839510490442058
4. Mhlanga, D., & Moloi, T. (2020). COVID-19 and the Digital Transformation of Education: What are we Learning on 4IR in South Africa? Education Sciences, 10(7), 180. https://doi.org/10.3390/educsci10070180

DOI: 10.5281/zenodo.16917658

5. Ojo, A. I., Adebayo, A. M., & Adewale, B. A. (2022). Application of Naïve Bayes and Decision Tree Algorithms in Predicting Students' Academic Performance: Evidence from a Nigerian State University. Nigerian Journal of Educational Technology, 14(2), 89–101.
6. Olatunji, A. T., & Aladejana, F. O. (2023). Machine Learning Applications in Predicting Academic Performance: A Case Study of Nigerian Tertiary Institutions. African Journal of Educational Management, 21(1), 55–69.
7. Onanuga, K. A., Adeyemo, A. O., & Ibrahim, R. O. (2021). Comparative Analysis of Machine Learning Algorithms for Predicting Student Performance. Journal of Engineering and Applied Sciences, 16(8), 199–207.
8. Romero, C., & Ventura, S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

DOI: 10.5281/zenodo.16917658