Mambilla Journal of Science and the Environment Vol. 10, Issue 1, Page No. 41-46

MAMBILLA JOURNAL OF SCIENCE AND THE ENVIRONMENT

journal homepage: https://oer.tsuniversity.edu.ng/index.php/mjse

A MACHINE LEARNING-BASED EARLY WARNING SYSTEM FOR IDENTIFYING AT-RISK UNIVERSITY STUDENTS IN NIGERIAN HIGHER EDUCATION

[1*]Bashiru, S., [2]Malgwi, Y.M.

[1]Department of Computer Science, Taraba State University, Jalingo, Nigeria.
[2] Department of Computer Science, Modibbo Adama University, Yola, Nigeria.

ARTICLE INFO

Article history:
Received 02 December 2025
Received in revised form 19 January 2026
Accepted 20 January, 2026

Keywords:
Early Warning System, Academic Performance, At-Risk Students, Machine Learning, Nigerian Universities.

ABSTRACT

Early identification of academically at-risk students remains a persistent challenge in Nigerian universities, where large class sizes, limited academic support systems, and delayed interventions often contribute to poor academic outcomes and increased dropout rates. This study proposes a machine learning-based Early Warning System (EWS) for identifying at-risk undergraduate students using academic, behavioral, and socioeconomic indicators. Data were collected from 1,200 undergraduate students of Taraba State University between 2023 and 2025, of which 1,000 valid records were retained after preprocessing. The dataset comprised variables including continuous assessment scores, class attendance, study hours, learning management system activity, parental income, parental education, and demographic attributes. Two supervised learning algorithms, Logistic Regression and Random Forest, were implemented using Python, with a 70:30 train-test split. Model performance was evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve. Results indicate that the Random Forest model outperformed Logistic Regression, achieving an accuracy of 89% and an AUC of 0.92. Feature importance analysis revealed that attendance rate, study hours, LMS engagement, and parental income were the most influential predictors of academic risk. The proposed EWS demonstrates strong potential as a decision-support tool for academic administrators, enabling proactive interventions aimed at improving student retention and academic success in Nigerian higher education institutions.

1. Introduction

Academic performance remains a critical indicator of institutional effectiveness and national human capital development. In higher education, timely identification of students who are academically at risk is essential for improving retention rates, optimizing academic support services, and ensuring equitable educational outcomes. However, many universities in developing countries, including Nigeria, still rely on reactive and manual approaches to academic monitoring, which often fail to detect early signs of academic distress.

Traditional academic evaluation systems predominantly focus on examination results and cumulative grade point average (CGPA), offering limited insight into the underlying behavioral and socioeconomic factors that influence student success. As a result, interventions are often delayed until students have already experienced significant academic decline. This limitation has intensified interest in data-driven approaches capable of identifying risk patterns before failure occurs.

Machine learning (ML), a subfield of artificial intelligence, offers powerful tools for modeling complex, non-linear relationships within large educational datasets. In recent years, ML-based early warning systems have gained prominence in educational data mining due to their ability to process multidimensional student data and generate predictive insights with high accuracy (Romero & Ventura, 2020). These systems enable institutions to classify students into risk categories and implement targeted interventions at an early stage.

In the Nigerian higher education context, localized ML-driven early warning frameworks remain scarce, particularly in public universities located in underserved regions. Taraba State University faces challenges such as heterogeneous

* Corresponding author: +2348026434373
E-mail address: sanibasheer2013@gmail.com

student backgrounds, limited academic advisory capacity, and inconsistent monitoring of student engagement. The absence of an automated, predictive early warning mechanism constrains the institution's ability to proactively support vulnerable students.

This study addresses this gap by developing a machine learning-based Early Warning System designed to identify at-risk undergraduate students using a combination of academic, behavioral, and socioeconomic variables. Specifically, the objectives of the study are to:

(i) identify key predictors of academic risk among undergraduate students;
(ii) develop and evaluate machine learning models for early risk detection; and
(iii) propose a data-driven early warning framework suitable for Nigerian universities.

2. Literature Review

2.1 Early Warning Systems in Higher Education

Early Warning Systems (EWS) are analytical frameworks designed to identify students who are likely to experience academic difficulty before failure occurs. These systems support proactive academic advising by leveraging historical and real-time student data to detect early risk signals (Wolff *et al*., 2020). In higher education, EWS have been applied to predict dropout, course failure, and delayed graduation.

Studies conducted in developed educational contexts demonstrate that EWS significantly improve retention and completion rates when combined with timely interventions (Romero & Ventura, 2020). However, many existing systems rely heavily on academic records alone, limiting their predictive capacity in environments where non-academic factors strongly influence performance.

2.2 Behavioral Indicators of Academic Risk

Behavioral factors such as lecture attendance, study habits, assignment submission patterns, and engagement with learning management systems are strong predictors of academic outcomes. Sembiring *et al.* (2021) found that students with irregular attendance and poor study routines were significantly more likely to fail courses.

The integration of LMS activity logs into predictive models has further enhanced early detection accuracy. Metrics such as login frequency, access to course materials, and participation in online forums serve as proxies for student engagement and motivation (Jena & Nayak, 2020).

2.3 Socioeconomic Factors and Academic Vulnerability

Socioeconomic status (SES) plays a critical role in shaping academic outcomes, particularly in developing countries. Parental income, educational background, and living conditions influence access to learning resources, psychological well-being, and academic persistence (Okoye *et al*., 2020). Students from low-income households often face additional stressors such as financial responsibility and inadequate study environments, increasing their risk of academic failure. Empirical evidence from Sub-Saharan Africa consistently demonstrates that SES variables are among the most influential predictors of academic success (Aina & Ayodele, 2019; Mensah & Kiernan, 2019).

2.4 Machine Learning for Academic Risk Prediction

Machine learning techniques have been widely applied to academic risk prediction due to their ability to handle large, heterogeneous datasets. Algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines have demonstrated high predictive accuracy in educational contexts (Alzahrani & Seth, 2021).

Random Forest models, in particular, have shown superior performance due to their ensemble structure, robustness to noise, and ability to estimate feature importance (Goyal & Vohra, 2020). Logistic Regression, while simpler, remains valuable for baseline modeling and interpretability.

Despite these advances, few studies have developed ML-based early warning systems tailored to Nigerian universities, highlighting the need for localized, context-aware predictive frameworks.

3. Methods

3.1 Research Design

This study adopted a quantitative, predictive research design focused on the development and evaluation of a machine learning–based early warning system for academic risk detection.

3.2 Data Collection and Dataset Description

Data were collected from 1,200 undergraduate students across five departments at Taraba State University. After data cleaning and validation, 1,000 records were retained. The dataset included academic, behavioral, and socioeconomic variables, with academic performance categorized into "At-Risk" and "Not At-Risk" classes based on GPA thresholds.

3.3 Data Preprocessing

Preprocessing steps included missing value imputation, outlier removal, categorical encoding, and feature scaling. Redundant features were eliminated using correlation analysis and Recursive Feature Elimination to enhance model efficiency.

3.4 Early Warning System Architecture

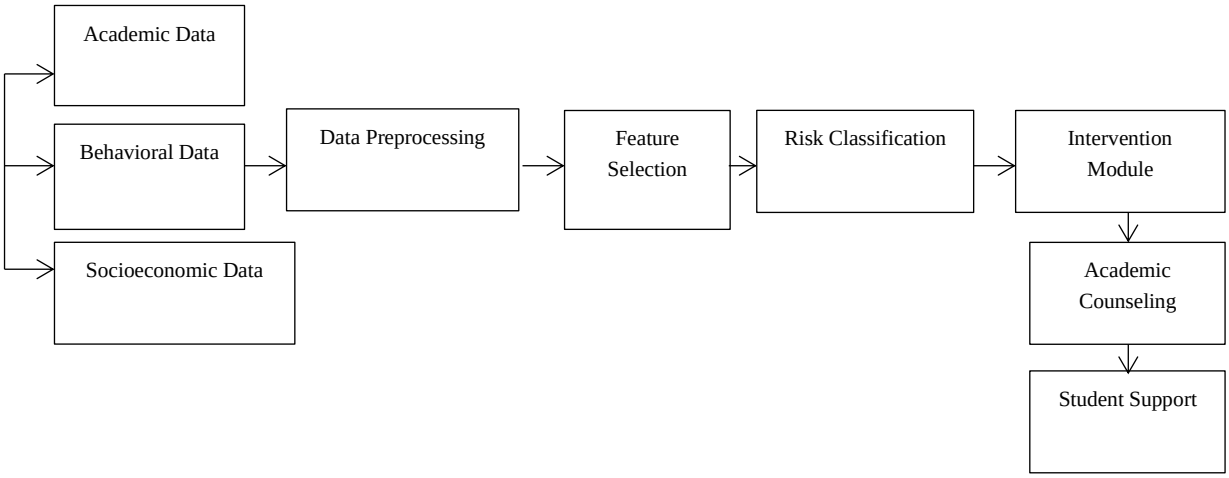

Figure 1: Early Warning System Architecture

3.5 Machine Learning Models

Two supervised learning algorithms were implemented:

3.5.1 Logistic Regression

Used as a baseline classifier for binary academic risk prediction due to its interpretability and efficiency.

3.5.2 Random Forest

An ensemble classifier employed to capture complex interactions among behavioral and socioeconomic variables.

3.6 Model Flowchart

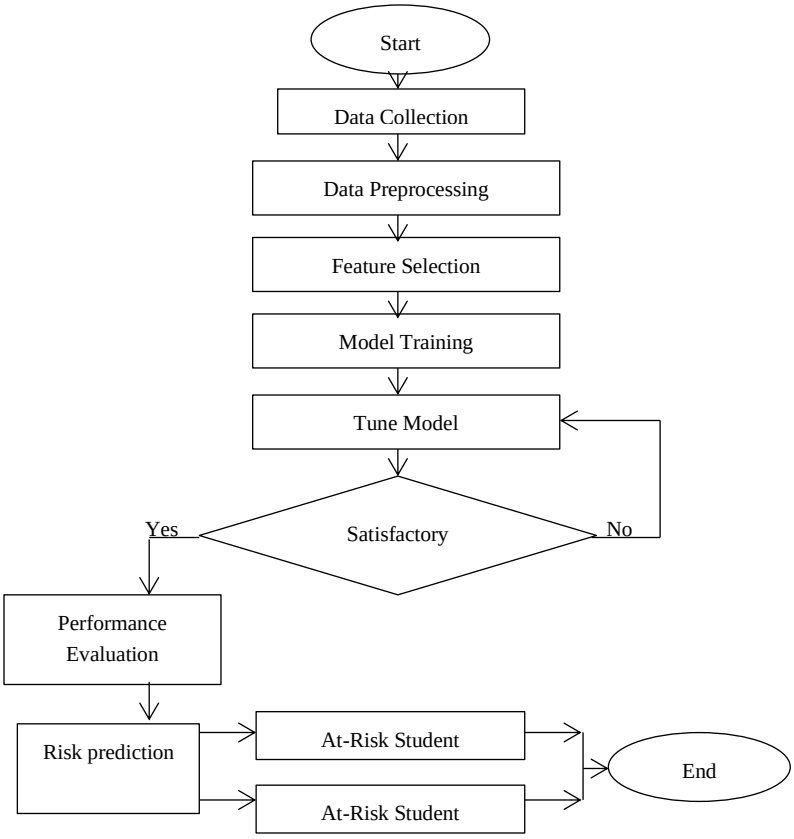

Figure 2: Early Warning Model Flowchart

3.7 Model Evaluation Metrics

Models were evaluated using accuracy, precision, recall, F1-score, and AUC-ROC to ensure robust performance assessment.

4. Results

4.1 Model Performance

Table 1: Performance Comparison of Early Warning Models

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	81%	0.79	0.76	0.77	0.85
Random Forest	89%	0.88	0.86	0.87	0.92

4.2 ROC Curve Analysis

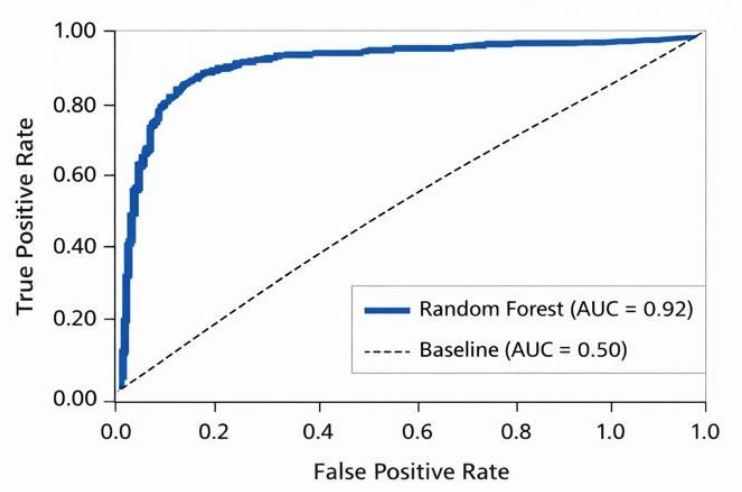

Figure 3: ROC Curve for Random Forest Early Warning Model

4.3 Feature Importance Results

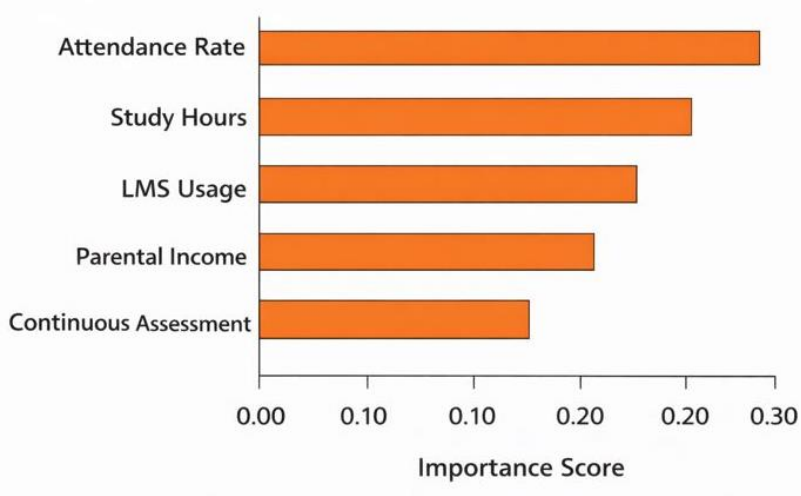

Figure 4: Feature Importance Ranking

5. Discussion

The results demonstrate that the proposed machine learning-based early warning system effectively identifies students at academic risk. The superior performance of the Random Forest model aligns with prior studies that emphasize the robustness of ensemble methods in educational prediction tasks (Goyal & Vohra, 2020).

Behavioral engagement indicators emerged as the most influential predictors, reinforcing the importance of continuous student monitoring beyond examination scores. The inclusion of socioeconomic variables further enhanced predictive accuracy, highlighting structural factors that contribute to academic vulnerability in Nigerian universities.

By aligning predictive outputs with early intervention strategies, the proposed EWS provides actionable insights for academic administrators and policymakers.

Conclusion

This study developed and evaluated a machine learning–based Early Warning System for identifying at-risk undergraduate students in a Nigerian university context. The findings confirm that integrating academic, behavioral, and socioeconomic data significantly improves early risk detection. The proposed framework offers a scalable, data-driven solution for proactive academic support and student retention. Future work should explore real-time data integration, deep learning models, and cross-institutional validation to enhance generalizability.

REFERENCES

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, *10*(1), 61-75. https://doi.org/10.1108/JARHE-09-2017-0106

Aina, J. K., & Ayodele, M. O. (2019). Socioeconomic status and students' academic performance: Evidence from Nigerian universities. *Journal of Education and Practice*, 10(4), 45-54.

Aljohani, N. R. (2016). A comprehensive review of the major studies and theoretical models of student retention in higher education. *Higher Education Studies*, 6(2), 1-18. https://doi.org/10.5539/hes.v6n2p1

Almarabeh, H., Majdalawi, Y. K., & Mohammad, H. (2022). Predicting student academic performance using machine learning techniques. *Education and Information Technologies*, *27*, 543-561. https://doi.org/10.1007/s10639-021-10703-w

Almohammadi, K., Al-Sarem, M., Al-Harby, A., & Alshahrani, M. (2021). *Support vector machine-based predictive models for student academic performance*. Education and Information Technologies, 26(5), 6051-6072. https://doi.org/10.1007/s10639-021-10597-2

Alzahrani, A., & Seth, A. (2021). Predicting student academic performance: A systematic review of machine learning techniques. *Education and Information Technologies*, 26(6), 4007-4030. https://doi.org/10.1007/s10639-021-10558-y

Bhardwaj, R., & Pal, S. (2022). Application of machine learning in education: A review. *Procedia Computer Science*, 195, 218-225. https://doi.org/10.1016/j.procs.2021.12.073

Cortes, C., & Vapnik, V. (2021). *Support-vector networks*. Machine Learning, 20(3), 273-297. https://doi.org/10.1007/BF00994018

Fredricks, J. A., Wang, M. T., Schall Linn, J., Hofkens, T., & Allerton, J. (2019). Using qualitative methods to develop a measure of student engagement. *Educational Psychologist*, 54(1), 78-97. https://doi.org/10.1080/00461520.2018.1535782

Goyal, A., & Vohra, R. (2020). Machine learning techniques for student performance prediction: A comparative study. *Procedia Computer Science*, 172, 433-439. https://doi.org/10.1016/j.procs.2020.05.062

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.

Kumar, A., & Passi, K. K. (2019). A decision tree based model for academic performance prediction. *International Journal of Computer Applications*, 177(7), 1-6. https://doi.org/10.5120/ijca2019918633

Mensah, F. K., & Kiernan, K. E. (2019). Parental education, household resources and educational outcomes in Ghana. *International Journal of Educational Development*, 66, 61-71. https://doi.org/10.1016/j.ijedudev.2019.01.001

Okoye, K. R. E., Nwajiuba, C. A., & Ofoegbu, G. N. (2020). Socioeconomic status and students' academic performance in Nigeria: A machine learning approach. *African Journal of Education and Technology*, 10(1), 33-45.

Okoye, K. R. E., Nwajiuba, C. A., & Ofoegbu, G. N. (2021). Bridging digital inequality and academic performance in Nigerian universities: A policy perspective. *African Journal of Educational Management*, 19(2), 87-101.

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

Sembiring, M. G., Hasibuan, A., & Wahyuni, D. (2021). Behavioral analysis of students to predict academic success. *Journal of Educational Technology*, 20(2), 119-130. https://doi.org/10.21831/jtp.v20i2.39112

Umer, M., Ullah, I., & Alam, M. (2021). Predictive modeling for academic performance: A machine learning approach. *IEEE Access*, 9, 118322-118334. https://doi.org/10.1109/ACCESS.2021.3106971

UNESCO. (2020). *Global Education Monitoring Report: Inclusion and education All means all*. Paris: UNESCO Publishing.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2021). Data mining: *Practical machine learning tools and techniques* (5th ed.). Morgan Kaufmann.

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2020). Improving retention by identifying students at risk of dropout: A case study. *International Journal of Educational Technology in Higher Education*, 17(1), 1–20. https://doi.org/10.1186/s41239-020-00192-4

Yilmaz, R. M., Yilmaz, F. G. K., & Keser, H. (2020). The role of self-regulation, motivation, and learning strategies on academic performance of students. *Educational Sciences: Theory & Practice*, 20(3), 1–15. https://doi.org/10.14527/pegegog.2020.001

Zimmerman, B. J., & Schunk, D. H. (2011). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). New York: Routledge.