



**STRATHMORE UNIVERSITY
SCHOOL OF COMPUTING AND ENGINEERING SCIENCES**

**A PREDICTIVE ANALYTICS PLATFORM FOR
STUDENT SUCCESS:**

**ML-BASED EARLY DETECTION OF AT-RISK STUDENTS IN
HIGHER EDUCATION**

BY

**KATHEMBO TSONGO DIEUDONNE
112721**

*A Thesis Submitted in Partial Fulfillment of the Requirements for the Award of the Degree of
Master of Science in Information Technology*

NAIROBI, KENYA

May, 2026

Declaration and Approval

I, the undersigned, declare that this thesis is my original work and has not been submitted for a degree in any other university.

Signature: _____ Date: _____

Kathembo Tsongo Dieudonne

This thesis has been submitted for examination with my approval as the University Supervisor.

Signature: _____ Date: _____

Dr. [Supervisor Name]

Strathmore University

Contents

1	Introduction	1
1.1	Background of the Study	1
1.2	Statement of the Problem	2
1.3	Purpose of the Study	3
1.4	Research Objectives	3
1.5	Research Questions	3
1.6	Significance of the Study	3

Chapter 1

Introduction

1.1 Background of the Study

Student retention and academic success represent critical challenges confronting higher education institutions worldwide. According to the Irwin et al. (2024), while the average adjusted cohort graduation rate for public high school students has reached 87%, undergraduate enrollment in degree-granting institutions decreased by 13% between 2012 and 2022. This persistent attrition represents not only significant personal and economic costs for individual students but also substantial challenges for institutional sustainability and mission fulfillment (Element451, 2024).

Traditional approaches to student support have been largely reactive, addressing academic difficulties only after students have already failed. The critical gap lies in the lack of integrated, automated tools to transform existing institutional data into actionable early warnings. The advent of artificial intelligence has created unprecedented opportunities to address this through predictive analytics. Contemporary platforms have evolved to become comprehensive ecosystem monitors that analyze hundreds of data points to forecast outcomes (Mapademics, 2025). Georgia State University's pioneering implementation exemplifies this, tracking over 800 risk factors for more than 40,000 students daily (Georgia State University, 2024).

Machine learning algorithms now process vast datasets, identifying patterns that precede disengagement with accuracy levels reaching 88–92% (Ahmed, 2024; Gandhi et al., 2025). In the Kenyan context, public universities face distinctive challenges, including dramatic enrollment expansion alongside institutional resource struggles (Musasia et al., 2025a). Recent studies reveal that while postgraduate retention improved to 87% by 2024, substantial attrition remains (Musasia et al., 2025b). These institutions confront funding gaps where government capitation covers only 57% of costs (Kenya Institute for Public Policy Research and Analysis, 2024). Financial pressures are acute, with public universities accumulating pending bills of Ksh 62

billion (Daily Nation, 2024), further complicated by severe infrastructure and staffing deficits (Visualdo Institute, 2024).

Strathmore University operates within this landscape, maintaining a strict 67% attendance requirement and a 2.0 GPA threshold. While the university collects data via Student Information Systems (SIS) and Learning Management Systems (LMS), these operate independently. Recent research demonstrates that ensemble methods, such as Random Forest and XGBoost, can achieve prediction accuracies exceeding 88% (Turkmenbayev et al., 2025; Wang et al., 2025). However, successful implementation requires explainable models that provide clear insights into why students are flagged (Ramos-Pulido et al., 2025). This study addresses this by developing a predictive platform tailored to the Strathmore context to transform reactive support into proactive intervention.

1.2 Statement of the Problem

Strathmore University faces significant challenges in identifying and supporting at-risk students before academic failure occurs. The university operates three comprehensive data systems—a Student Information System (SIS), Learning Management System (LMS), and attendance tracking system—that collectively capture student demographics, academic performance, engagement metrics, and class participation. However, these systems operate independently without integrated analytics capabilities, preventing the transformation of this data into actionable early warnings. Consequently, at-risk students are typically identified only after grades decline, attendance falls below the mandatory 67% threshold, or dropout occurs, at which point intervention opportunities have largely passed. Academic advisors and mentors currently lack data-driven tools to prioritize limited resources toward students most likely to benefit from support, resulting in reactive rather than proactive intervention strategies.

This gap is particularly critical given the proven effectiveness of predictive analytics in comparable institutional contexts. Research demonstrates that machine learning systems achieve 85–92% accuracy in identifying at-risk students when trained on comprehensive multi-source data (Ahmed, 2024; Gandhi et al., 2025), with institutions like Georgia State University generating 90,000 targeted interventions annually through automated risk tracking (Georgia State University, 2024). Strathmore University currently lacks such capabilities despite possessing the necessary raw data infrastructure. This study addresses this gap by developing a machine learning-based predictive analytics platform that integrates existing institutional data to predict dropout risk, course failure risk, and program delay risk, enabling proactive intervention through role-based dashboards for administrators and mentors.

1.3 Purpose of the Study

The purpose of this research is to develop a predictive analytics platform that leverages student demographic and engagement data to improve retention rates.

1.4 Research Objectives

- To identify the key indicators of academic risk in higher education.
- To develop a machine learning model for predicting student performance.
- To evaluate the accuracy and reliability of the predictive platform.

1.5 Research Questions

1. What features in student data are most predictive of academic failure?
2. How accurately can a Random Forest or Neural Network model detect at-risk students?

1.6 Significance of the Study

This study is significant for university administrators, educators, and students, as it provides a data-driven approach to academic support.

References

- Ahmed, E. (2024). Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024(1), 4067721. <https://doi.org/10.1155/2024/4067721>
- Daily Nation. (2024, May). *University dropout risk looms large. here's why* [Accessed: February 8, 2026]. <https://nation.africa/kenya/news/education/budget-cuts-set-to-cripple-learning-in-four-varsities--4625512>
- Element451. (2024, September). *2024 guide: What is student retention in higher education?* [Accessed: February 8, 2026]. <https://element451.com/blog/what-is-student-retention>
- Gandhi, M., et al. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*, 15(1), 12353. <https://doi.org/10.1038/s41598-025-12353-4>
- Georgia State University. (2024, December). *Approaching student success with predictive analytics* [Accessed: February 8, 2026]. <https://success.gsu.edu/approach/>
- Irwin, V., Wang, K., Jung, J., Kessler, E., Tezil, T., Alhassani, S., Filbey, A., Dilig, R., & Bullock Mann, F. (2024). *Report on the condition of education 2024 (nces 2024-144)*. National Center for Education Statistics, U.S. Department of Education. Washington, DC. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2024144>
- Kenya Institute for Public Policy Research and Analysis. (2024). *Improving the performance of public universities in delivering higher education in kenya* [Accessed: February 8, 2026]. <https://kippra.or.ke/improving-the-performance-of-public-universities-in-delivering-higher-education-in-kenya/>
- Mapademics. (2025, September). *Transforming student success: How predictive retention analytics is revolutionizing higher education in 2025* [Accessed: February 8, 2026]. <https://mapademics.com/whitepapers/predicting-student-success-with-smart-analytics>
- Musasia, C. M., Achoka, J. S., Wamocha, L. M., & Odebero, S. O. (2025a). Postgraduate students' retention and survival rates between 2015 and 2024 in public universities, kenya. *African Journal of Empirical Research*, 6(3), 679–687. <https://doi.org/10.51867/ajernet.6.3.52>
- Musasia, C. M., Achoka, J. S., Wamocha, L. M., & Odebero, S. O. (2025b). Postgraduate students' retention and survival rates between 2015 and 2024 in public universities,

kenya. *African Journal of Empirical Research*, 6(3), 679–687. <https://doi.org/10.51867/ajernet.6.3.52>

Ramos-Pulido, A., et al. (2025). Machine learning models for academic performance prediction: Interpretability and application in educational decision-making. *Frontiers in Education*, 10, 1632315. <https://doi.org/10.3389/feduc.2025.1632315>

Turkmenbayev, A., Abdykerimova, E., Nurgozhayev, S., Karabassova, G., & Baigozhanova, D. (2025). The application of machine learning in predicting student performance in university engineering programs: A rapid review. *Frontiers in Education*, 10, 1562586. <https://doi.org/10.3389/feduc.2025.1562586>

Visualdo Institute. (2024, October). *What are challenges facing colleges in kenya* [Accessed: February 8, 2026]. <https://visualdo.ac.ke/what-are-challenges-facing-colleges-in-kenya>

Wang, Y., et al. (2025). Artificial intelligence in student management systems to enhance academic performance monitoring and intervention. *Scientific Reports*, 15(1), 19159. <https://doi.org/10.1038/s41598-025-19159-4>