STRATHMORE UNIVERSITY

SCHOOL OF COMPUTING AND ENGINEERING SCIENCES

A PREDICTIVE ANALYTICS PLATFORM FOR STUDENT SUCCESS:

MACHINE LEARNING-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

Student's Declaration

I declare that this thesis is my original work and has not been presented to any other institution for examination or any other award. Where other sources of information have been used, they have been properly acknowledged.

Name: Kathembo Tsongo Dieudonne

Registration Number: 112721

Signature: _____ **Date:** _____

Supervisor's Declaration

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

Name: Dr. [Supervisor Name]

Title: [Lecturer/Senior Lecturer]

Department: [Department Name]

Signature: _____ **Date:** _____

Contents

Chapter 1

INTRODUCTION

1.1 Background to the Study

Student retention and academic success represent critical challenges confronting higher education institutions worldwide. According to Irwin et al. (2024), undergraduate enrollment in degree-granting institutions decreased by 13% between 2012 and 2022, while 23.3% of undergraduate students leave universities annually. Six-year graduation rates reveal substantial variation: 63% at public universities, 68% at private nonprofit institutions, and only 29% at private for-profit institutions. This persistent attrition represents significant personal and economic costs for individual students, alongside substantial challenges for institutional sustainability and mission fulfillment.

Traditional student support approaches have been largely reactive, addressing academic difficulties only after failure occurs. The critical gap lies in the lack of integrated, automated tools to transform existing institutional data into actionable early warnings. However, the advent of artificial intelligence and machine learning has created unprecedented opportunities for proactive intervention through predictive analytics. Contemporary platforms now analyze hundreds of data points to forecast student outcomes with remarkable accuracy (Mapademics, 2025). Georgia State University's pioneering implementation exemplifies this transformation, tracking over 800 risk factors for

more than 40,000 students daily and generating 90,000 targeted interventions annually (Georgia State University, 2024). Machine learning algorithms can identify patterns that precede student disengagement with accuracy levels reaching 88–92% (Ahmed, 2024; Gandhi et al., 2025), enabling early identification weeks or months before traditional indicators would surface.

In the Kenyan context, higher education institutions face distinctive challenges that intensify retention concerns. Public universities have experienced dramatic enrollment expansion while institutional resources have struggled to keep pace (Musasia et al., 2025). Kenyan institutions confront severe funding constraints, with government capitation covering only 57% of students against a target of 80% (Kenya Institute for Public Policy Research and Analysis, 2024). Financial pressures have intensified dramatically, with public universities accumulating pending bills of Ksh 62 billion as of February 2022 (Daily Nation, 2024), compounded by infrastructure deficits and severe staffing shortages where some lecturers teach triple their contracted student numbers (Visualdo Institute, 2024). In this resource-constrained environment, automated systems to identify and support at-risk students become essential for institutional survival and mission fulfillment.

Strathmore University operates within this challenging landscape while maintaining distinct academic standards, including a 67% attendance requirement for examination eligibility and a 2.0 GPA threshold for good academic standing. The university collects extensive student data through Student Information Systems (SIS), Learning Management Systems (LMS), and attendance tracking mechanisms. However, these systems operate independently without integrated analytics capabilities, preventing transformation of existing data into actionable early warnings. Recent research demonstrates that ensemble methods such as Random Forest and XGBoost can achieve prediction accuracies exceeding 88% when properly trained on comprehensive educational datasets (Turkmenbayev et al., 2025; Wang et al., 2025). Yet successful implementation requires explainable models that provide clear insights into why particular students are flagged as at-risk and what specific interventions might prove most effective (Ramos-Pulido

et al., 2025). This study addresses this gap by developing a machine learning-based predictive analytics platform tailored to Strathmore University's context, transforming reactive student support into proactive, data-driven intervention.

1.2   Statement of the Problem

Strathmore University faces significant challenges in identifying and supporting at-risk undergraduate students before academic failure occurs. The university operates three comprehensive data systems that capture student demographics, academic performance, engagement metrics, and class participation. However, these systems operate independently without integrated analytics capabilities, preventing transformation of this data into actionable early warnings. Consequently, at-risk students are identified only after grades decline, attendance falls below the mandatory 67% threshold, or dropout occurs—when intervention opportunities have largely passed. Academic advisors and mentors lack data-driven tools to prioritize limited resources, resulting in reactive rather than proactive intervention strategies.

This reactive approach creates multiple institutional challenges. Students who might have succeeded with timely support instead face academic probation, course repetition, extended time-to-graduation, or institutional departure. Faculty and administrators cannot identify systemic patterns in student risk, limiting evidence-based improvements. In Kenya's resource-constrained environment where government capitation covers only 57% of students (Kenya Institute for Public Policy Research and Analysis, 2024), each preventable failure represents both a personal setback and significant institutional loss.

Research demonstrates that machine learning systems achieve 85–92% accuracy in identifying at-risk students (Ahmed, 2024; Gandhi et al., 2025), with proven implementations like Georgia State University's generating 90,000 targeted interventions annually (Georgia State University, 2024). Strathmore University currently lacks such capabilities despite possessing the necessary data infrastructure. This study develops a machine learning-based predictive analytics platform integrating 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   Justification of the Study

This research is justified on both practical and scholarly grounds. Practically, Strathmore University has invested significantly in digital infrastructure—Student Information Systems, Learning Management Systems, and attendance tracking—yet lacks analytical capacity to transform this data into actionable intelligence. The platform leverages already-collected data without requiring additional burdens on faculty or students, maximizing return on existing technology investments. As already indicated in the problem statement, recent advances in machine learning demonstrate 85–92% prediction accuracy for student success outcomes (Ahmed, 2024; Gandhi et al., 2025), indicating the technology is sufficiently mature for reliable implementation.

Scholarly, this study addresses a critical gap in the literature. While predictive analytics has been extensively studied in Western contexts, limited research examines implementation in African universities where resource constraints, infrastructure challenges, and student demographics differ substantially. This research generates knowledge about adapting predictive analytics to resource-limited environments, providing a replicable model for similar institutions across Africa.

Given Kenya's national emphasis on improving higher education quality through the transition to Competency Based Education and Training (CBET) and addressing sector-wide pressures—including a 12% increase in university enrollment between 2015 and 2022 and significant funding gaps where the percentage of students receiving government loans dropped from 95.6% to 71.3% (Republic of Kenya Ministry of Education, 2024) —this research positions Strathmore University at the forefront of data-driven educational innovation. In monitoring student outcomes, this study aligns with the national goal of 'leveraging technology in the provision of quality education'.

1.4 Research Objectives

1.4.1 Main Objective

To design, develop, and evaluate a machine learning-based predictive analytics platform for early identification of at-risk undergraduate students at Strathmore University.

1.4.2 Specific Objectives

1. To integrate student data from institutional systems (SIS, LMS, and attendance tracking) into a unified analytics repository.

2. To develop machine learning models that predict dropout risk, course failure risk, and program delay risk.

3. To design role-based dashboards for university administrators, school administrators, and academic mentors with actionable risk predictions.

4. To evaluate model accuracy and predictive performance using historical student data and standard evaluation metrics.

5. To assess system usability and institutional impact through stakeholder evaluation.

1.5 Research Questions

1.5.1 Main Research Question

How can machine learning be effectively applied to institutional student data to enable early identification of at-risk undergraduate students at Strathmore University?

1.5.2 Specific Research Questions

1. What student data features from SIS, LMS, and attendance systems are most predictive of dropout risk, course failure risk, and program delay risk?

2. Which machine learning algorithms provide the most accurate predictions for each of the three risk categories?

3. How can risk predictions be effectively presented to different stakeholder groups to support decision-making?

4. What level of prediction accuracy can be achieved using historical student data from Strathmore University?

5. How do university stakeholders perceive the usability and potential institutional impact of the predictive analytics platform?

1.6 Scope and Delimitations of the Study

1.6.1 Scope

This research focuses specifically on developing and evaluating a predictive analytics platform for undergraduate student success at Strathmore University, Kenya. The study encompasses undergraduate student data from all five schools: the School of Computing and Engineering Sciences (SCES), Strathmore Business School (SBS), the School of Tourism and Hospitality (STH), Strathmore Law School (SLS), and the School of Humanities and Social Sciences (SHSS). The research utilizes data from approximately 5000 undergraduate students, providing sufficient statistical power for reliable machine learning model development.

The technical scope includes integration of data from three primary institutional systems: the Student Information System (academic records), the Learning Management System (engagement metrics), and the attendance tracking system (class participation). The study develops and evaluates three distinct machine learning prediction models tar-

geting dropout risk, course failure risk, and program delay risk, employing ensemble algorithms including Random Forest, Gradient Boosting, and XGBoost. The platform design encompasses role-based dashboards serving university administrators, school administrators, and academic mentors. System effectiveness is evaluated through quantitative metrics (accuracy, precision, recall, F1-score, ROC-AUC) and qualitative stakeholder assessment.

1.6.2   Delimitations

Several deliberate delimitations define the boundaries of this research.

First, the study focuses exclusively on academic risk factors measurable through institutional data systems, excluding predictions of mental health concerns, family circumstances, or socioeconomic factors that fall outside institutional data collection.

Second, the research limits its focus to undergraduate students, excluding postgraduate and continuing education populations whose risk patterns differ substantially from traditional undergraduate trajectories.

Third, the study does not develop course-specific recommendation systems or personalized learning pathways; the platform identifies at-risk students and suggests broad intervention categories without prescribing specific courses or detailed academic plans.

Fourth, the platform is designed as a web-based application accessible through standard browsers, deliberately excluding mobile application development to prioritize cross-platform accessibility and reduce development complexity.

Fifth, the research does not integrate financial aid prediction or scholarship recommendation capabilities, despite the documented importance of financial factors in Kenyan student retention.

Finally, while the research develops a fully functional prototype system, full production deployment across the entire Strathmore student population represents future work beyond the thesis scope, requiring additional institutional approvals, infrastructure pro-

visioning, and ongoing technical support arrangements.

1.7   Definition of Terms

At-Risk Student: A student exhibiting data patterns that statistically correlate with increased probability of dropout, course failure, or program delay. Risk status is determined through predictive model outputs, enabling proactive identification before failure occurs.

Dropout Risk: The predicted probability that a currently enrolled student will discontinue university enrollment before completing degree requirements, encompassing both formal withdrawal and informal cessation of attendance.

Course Failure Risk: The predicted probability that a student will receive a failing grade (below 40%) or become ineligible for examination due to attendance below the 67% requirement.

Program Delay Risk: The predicted probability that a student will require additional semesters beyond the standard program duration to complete degree requirements due to course repetition, reduced course loads, or academic probation.

Learning Management System (LMS): A software platform (such as Moodle or Canvas) that delivers course content, facilitates communication, manages assignments, and tracks student engagement through behavioral data.

Student Information System (SIS): An administrative database maintaining comprehensive student records including demographics, enrollment history, grades, and academic standing—the institutional source of truth for official student data.

Predictive Analytics: Statistical and machine learning techniques that analyze historical and current data to generate probabilistic forecasts of future events, identifying students likely to experience academic difficulties.

Ensemble Learning: A machine learning approach combining predictions from multi-

ple algorithms (Random Forest, Gradient Boosting, XGBoost) to achieve higher accuracy and reduce overfitting.

Feature Engineering: The process of selecting, transforming, and creating variables from raw data that effectively capture patterns relevant to prediction, translating institutional data into optimized machine learning inputs.

Early Warning System: An integrated platform combining data collection, predictive modeling, and alert generation to identify students requiring intervention before academic problems become severe.

Engagement Metrics: Quantifiable measures of student participation derived from LMS data, including login frequency, time on platform, discussion posts, and assignment submission timeliness.

Model Accuracy: The proportion of correct predictions among all predictions made by a machine learning model, evaluated alongside precision, recall, and F1-score.

Intervention: Targeted support actions designed to address identified student risk factors, including academic advising, tutoring, workshops, counseling, or course load adjustments.

1.8 Significance of the Study

1.8.1 Theoretical Contribution

This study advances scholarly understanding of educational data mining in resource-constrained environments. While existing literature predominantly focuses on well-resourced Western institutions (Gandhi et al., 2025; Turkmenbayev et al., 2025), this research demonstrates the applicability of advanced predictive analytics where data infrastructure may be limited. The study operationalizes Tinto's (1975, 1993) Student Integration Model—originally developed for undergraduate populations—by mapping theoretical constructs of academic and social integration to quantifiable data elements

extracted from institutional systems, providing empirical validation of retention frameworks within an African context.

1.8.2 Practical Contribution

For Institutions

The platform provides Strathmore University and similar institutions with tools to transform reactive support into proactive intervention. Data-driven resource allocation directs limited counseling and mentoring resources toward students most likely to benefit. The platform generates actionable insights about program-level risk patterns, informing curriculum development and policy adjustments (Modern Campus, 2025).

For Educators

The platform equips academic advisors with evidence-based insights into student risk factors, moving beyond intuition to data-informed strategies. Role-based dashboards highlight specific concerns—attendance patterns, engagement deficits, assignment completion— enabling targeted interventions. Early identification creates opportunities for proactive outreach before problems become insurmountable.

For Students

While not direct users, undergraduate students are ultimate beneficiaries. Timely interventions can prevent academic failures, reduce program delays, and increase graduation likelihood. The system ensures that institutional policies like the 67% attendance requirement are supported by compliance mechanisms rather than serving merely as punitive measures.

1.8.3 Policy Contribution

This research demonstrates a cost-effective approach to addressing retention challenges in African universities, providing a replicable model. Given Kenya's higher education

challenges (Daily Nation, 2024; Kenya Institute for Public Policy Research and Analysis, 2024), the study illustrates how strategic application of artificial intelligence can amplify limited resources without proportional expenditure increases. By documenting implementation challenges and lessons learned, the study creates knowledge that can accelerate adoption across Kenyan and African institutions (Republic of Kenya Ministry of Education, 2024).

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

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

Modern Campus. (2025, November). *Student retention strategies: 8 proven methods to boost higher ed success* [Accessed: February 8, 2026]. https://moderncampus.com/blog/strategies-for-improving-retention-in-higher-education.html

Musasia, C. M., Achoka, J. S., Wamocha, L. M., & Odebero, S. O. (2025). 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

Republic of Kenya Ministry of Education. (2024). *Sessional paper no. x of 2024*. Ministry of Education. Nairobi, Kenya. https://www.education.go.ke/sites/default/files/BILLS/Sessional%5C%20Paper_2802024.pdf

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