



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

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.

Technology is changing that. Machine learning algorithms can now identify at-risk students weeks or even months before traditional warning signs appear, achieving prediction accuracies of 88–92% (Ahmed, 2024; Gathani et al., 2026). Georgia State University’s GPS Advising system illustrates what becomes possible: by tracking over 800 risk factors for 40,000 students daily and generating 90,000 targeted interventions each year, the university raised its graduation rate by 7.8 percentage points and virtually eliminated achievement gaps between demographic groups (Georgia State University, 2024). The technology, in short, is mature and proven—but its benefits have not yet reached most of the world.

Africa is a striking example of that gap. The African Union’s Continental Education Strategy (CESA 2016–2025) acknowledges retention as a priority, yet most African universities still rely on end-of-semester transcripts as their primary diagnostic tool. Some

sub-regions have begun to experiment. Mohammed V University in Morocco achieved an 81.66% dropout-prediction accuracy in a pilot study that utilized L2-regularized logistic regression (Bouihi et al., 2024). *****

East Africa lags further still. In Kenya—where government funding covers only 57% of students against an 80% target and public universities carry Ksh 62 billion in unpaid bills (Daily Nation, 2024; Kenya Institute for Public Policy Research and Analysis, 2024)—Kenyatta University offers the region’s clearest case study. Its early-warning system, which merged attendance and grade data, reduced first-year dropout rates by 23% over two years (Wachira & Kimani, 2024). That result is encouraging, but the system collapsed when its sole faculty champion left the institution, revealing a fragility common to initiatives built around individuals rather than institutions.

Strathmore University sits squarely within this unresolved challenge. It enforces a 67% minimum attendance threshold for examination eligibility and a 2.0 GPA floor for good standing—policies that make early identification of struggling students especially consequential. The university already collects rich data across its Academic Management System, Learning Management System, and attendance platform, yet these systems operate in isolation, with no mechanism to translate their data into timely warnings. Students surface for intervention only after breaching a threshold or withdrawing formally, long after a proactive response could have made a difference. This study addresses that gap directly: by developing a machine learning-based predictive analytics platform calibrated to Strathmore’s policies, data infrastructure, and resource context, it aims to shift the institution from reactive monitoring to proactive, evidence-driven student support—and to provide a replicable model for comparable universities across East Africa.

1.2 Statement of the Problem

Strathmore University currently faces a critical gap between data collection and actionable student support. Although the institution maintains robust data streams through its Academic Management System (AMS), Learning Management System (LMS), and

attendance tracking platforms, these systems operate as isolated administrative repositories rather than integrated predictive tools. This fragmentation prevents the university from transforming raw institutional data into early warnings for at-risk undergraduate students—a limitation mirrored across comparable East African institutions, where similar pilots at Kenyatta University collapsed due to unsustainable, individually-driven implementations rather than embedded institutional systems (Wachira & Kimani, 2024).

The primary consequence of this gap is a reactive support paradigm. Students are typically identified for intervention only after manifesting lag-indicators of failure—cumulative low grades, breach of the mandatory 67% attendance threshold for examination eligibility, or formal withdrawal. By this stage, the window for effective academic recovery has often closed. This reactive approach is not only inefficient but inconsistent with the Republic of Kenya Ministry of Education’s (Republic of Kenya Ministry of Education, 2024) Sessional Paper No. 1, which explicitly identifies ‘inaccurate, unreliable data and inadequate utilisation of technology’ as a core bottleneck in the national education sector.

Despite possessing the necessary digital infrastructure and longitudinal student datasets, Strathmore University lacks a specialised machine learning architecture capable of integrating these data streams into a coherent, real-time risk stratification system. Without an automated platform to predict dropout risk, course failure risk, and program delay risk simultaneously, the university remains unable to allocate its limited counselling and mentoring resources to the students who need them most, and advisors are left to act on intuition rather than evidence.

This study therefore develops, implements, and evaluates an integrated predictive analytics platform tailored to Strathmore University’s academic policies, data infrastructure, and resource constraints. By producing explainable, actionable risk predictions across three distinct outcomes, the platform is designed to transition the institution from reactive monitoring to proactive, data-driven intervention—and to provide a replicable model that addresses the sustainability failures observed in prior East African imple-

mentations.

1.3 Justification

This research is justified on both practical and scholarly grounds. Practically, Strathmore University has made substantial investments in digital infrastructure—its Academic Management System, Learning Management System, and attendance tracking platforms collectively generate rich, longitudinal student data on a daily basis. Yet the institution currently lacks the analytical architecture to transform this data into actionable intelligence. The proposed platform leverages information that is already collected, requiring no additional burden on faculty or students, and therefore maximises the return on technology investments already made. Given that recent advances in machine learning demonstrate prediction accuracy levels of 85—92% for student success outcomes (**Gandhi2025**; Ahmed, 2024), the technology is sufficiently mature for reliable institutional deployment. The practical case is straightforward: the data exists, the technology works, and the cost of continuing without such a system is measured in preventable student failures every semester.

From a scholarly standpoint, this study addresses a precise and documented gap in the literature. While predictive analytics has been extensively studied in Western, well-resourced contexts—with Georgia State University and Purdue University generating decades of evidence—research examining implementation in African universities remains critically limited, accounting for fewer than 8% of published studies in the field (Olawoyin et al., 2024).

The few African initiatives that have been documented reveal a consistent pattern: technically feasible models undermined by unsustainable implementation. The University of Pretoria achieved 81% prediction accuracy at minimal cost but without real-time capability or intervention infrastructure (Moloi et al., 2024); Kenyatta University reduced dropout rates by 23% but saw its system collapse when the sole faculty champion departed (Wachira & Kimani, 2024). This study contributes original knowledge

by demonstrating how a predictive analytics platform can be designed for institutional sustainability—not just technical performance—in a resource-constrained African private university context, providing a replicable model for comparable institutions across the region.

The study is further justified by Kenya’s national policy context. The Republic of Kenya Ministry of Education’s (Republic of Kenya Ministry of Education, 2024) Sessional Paper No. 1 of 2024 explicitly identifies the inadequate utilisation of technology and unreliable institutional data as core bottlenecks in higher education, and national enrolment pressures—a 12% increase in university intake between 2015 and 2022 alongside a decline in student loan coverage from 95.6% to 71.3% (Republic of Kenya Ministry of Education, 2024)—make efficient, data-driven allocation of limited student support resources not merely desirable but necessary. Strathmore University, as a leading private institution committed to academic excellence, is uniquely positioned to demonstrate that proactive, evidence-based student support is achievable within the Kenyan higher education landscape. In doing so, this research directly advances the national goal of leveraging technology to improve educational quality and positions the institution at the forefront of data-driven innovation in East African higher 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 consolidate and harmonise student data from institutional systems (AMS, LMS, and attendance tracking) into an integrated analytics pipeline capable of supporting real-time risk computation.

2. To develop and validate explainable machine learning models that predict three distinct risk outcomes: dropout risk, course failure risk, and program delay risk.
3. To design role-based dashboards for university administrators, school administrators, and academic mentors that present actionable risk predictions with supporting intervention guidance.
4. To evaluate model accuracy and predictive performance using historical student cohort data (\$2021-2024\$) and standard machine learning evaluation metrics including Accuracy, Precision, Recall, F1-score, and ROC-AUC.
5. To assess the platform's usability and perceived institutional impact through structured evaluation with academic advisors, school administrators, and mentors, with specific attention to intervention feasibility and long-term sustainability.

1.5 Research Questions

1. What student data features from AMS, 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 and explainable predictions for each of the three risk categories, and what features drive each prediction?
3. How can risk predictions be effectively presented to different stakeholder groups—university administrators, school administrators, and academic mentors—to support timely and informed decision-making?
4. To what extent does training predictive models on historical student cohort data (\$2021-2024\$) with known outcomes produce valid and generalisable predictions for current enrolled students?
5. How do academic advisors, school administrators, and mentors perceive the platform's usability, intervention feasibility, and potential for long-term institutional

adoption?

1.6 Scope and Delimitations of the Study

1.6.1 Scope

This research focuses on the development and evaluation of a predictive analytics platform for approximately 5,000 undergraduate students across all five schools at Strathmore University, Kenya. The technical scope involves integrating data from the Academic Management System (AMS), Learning Management System (LMS), and attendance tracking records into a unified analytics pipeline. The study develops three prediction models—targeting dropout risk, course failure risk, and program delay risk—using ensemble algorithms including Random Forest, Gradient Boosting, and XGBoost, trained on historical student cohort data from \$2021-2024\$ with known academic outcomes. Model predictions are accompanied by SHAP-based explainability outputs to ensure advisors and administrators can understand and act on the reasons behind each risk classification (Lundberg & Lee, 2017). The resulting platform includes role-based dashboards for university administrators, school administrators, and academic mentors, with model performance measured through standard machine learning evaluation metrics (Accuracy, Precision, Recall, F1-score, and ROC-AUC) and platform effectiveness assessed through structured stakeholder feedback.

1.6.2 Delimitations

The research boundaries are defined by the following deliberate constraints. Regarding data factors, the analysis is restricted to academic risk factors measurable through institutional systems, such as attendance rates, GPA, LMS engagement, and course grades. External variables such as socioeconomic background, mental health status, and family circumstances are excluded, as they are not systematically captured in Strathmore's existing data infrastructure and their collection would raise ethical concerns beyond the current scope. In terms of student population, the study is limited to undergradu-

ate students across Strathmore University’s five schools; postgraduate and continuing education cohorts are excluded due to their unique assessment patterns.

The system capabilities are delimited to identifying risk levels and suggesting general intervention categories, such as academic counselling or tutoring referrals, rather than providing course-specific remediation. Within the algorithmic scope, the study employs ensemble tree-based methods like Random Forest and XGBoost (Chen & Guestrin, 2016) rather than deep learning architectures. This decision is justified by the dataset size of approximately 5,000 students and the critical need for feature interpretability. Furthermore, the technical platform is developed as a web-based application, with mobile-native development remaining outside the scope. Financial domains, including fee payment risk and scholarship management, are also excluded to focus purely on academic and engagement indicators. Finally, the implementation phase concludes with a functional prototype and structured evaluation; full-scale production deployment and long-term longitudinal impact assessment are reserved for future research.

1.7 Definition of Terms

At-Risk Student A student whose data patterns statistically indicate a high probability of dropout, course failure, or program delay prior to the actual event.

Dropout Risk The predicted probability of a student discontinuing enrolment before completing their degree requirements.

Course Failure Risk The predicted likelihood of a student failing a course or becoming exam-ineligible due to attendance falling below the mandatory 67% threshold at Strathmore University.

Program Delay Risk The probability that a student will exceed the standard programme duration due to academic probation, course repetition, or accumulated failed semesters.

Academic Management System (AMS) The primary institutional database for official student records, including demographic information, enrolment status, and academic history.

Learning Management System (LMS) A digital platform used to deliver course content and track student engagement through behavioural data such as login frequency, resource access, and assignment submissions.

Early Warning System An integrated tool that uses data and predictive modelling to trigger timely alerts for students identified as requiring academic intervention.

Engagement Metrics Quantitative measures of student participation derived from institutional systems, such as LMS login frequency, assignment submission timeliness, and attendance rate trajectories.

Ensemble Learning A machine learning technique that combines multiple algorithms—such as Random Forest, Gradient Boosting, and XGBoost—to produce predictions that are more accurate and stable than any single model alone.

Feature Engineering The process of transforming raw institutional data into optimised input variables for machine learning model training, including the creation of derived indicators such as attendance flags and GPA thresholds.

Intervention A proactive support action—such as academic mentoring, tutoring referral, or advisor outreach—triggered by a student’s identified risk status and designed to prevent academic failure or withdrawal.

Predictive Analytics A branch of data analytics that uses statistical modelling and machine learning techniques applied to historical data to forecast future outcomes. In this study, predictive analytics refers specifically to the use of student historical records to generate probabilistic risk classifications for dropout, course failure, and programme delay.

Explainability (Model Explainability) The capacity of a predictive model to provide human-interpretable reasons for its predictions. In this study, explainability is implemented through SHAP (SHapley Additive exPlanations) values, which quantify each feature's contribution to an individual student's risk score, enabling advisors to understand why a student has been flagged and what specific factors require intervention (Lundberg & Lee, 2017).

Historical Cohort A defined group of students whose complete academic trajectories—from enrolment to graduation, withdrawal, or dropout—are known and documented. In this study, the historical cohort comprises undergraduate students who enrolled at Strathmore University in 2021 and whose outcomes are fully recorded by 2024, providing the labelled training data required for supervised machine learning.

Model Accuracy The proportion of correct predictions (both at-risk and not-at-risk) made by a model relative to total observations. Accuracy is one of five evaluation metrics used in this study; it is most informative when class distributions are balanced.

Precision The proportion of students predicted as at-risk who are genuinely at-risk. High precision minimises false alarms and ensures that limited intervention resources are directed towards students who actually need support.

Recall (Sensitivity) The proportion of genuinely at-risk students who are correctly identified by the model. High recall is the primary optimisation target in this study, as failing to identify a student who subsequently drops out (a false negative) carries greater institutional cost than a false alarm.

F1-Score The harmonic mean of Precision and Recall, providing a single balanced metric that accounts for both false positives and false negatives. Particularly useful in this study given the class imbalance between at-risk and non-at-risk students.

ROC-AUC (Receiver Operating Characteristic – Area Under Curve) A threshold-independent metric measuring a model’s ability to discriminate between at-risk and non-at-risk students across all possible classification thresholds. A value of 1.0 indicates perfect discrimination; 0.5 indicates performance no better than random chance.

1.8 Significance of the Study

1.8.1 Theoretical Contribution

This study advances scholarly understanding of educational data mining in resource-constrained environments along two theoretical dimensions. First, it operationalises Tinto’s (Tinto, 1975, 1993) Student Integration Model within an African institutional context by mapping the theoretical constructs of academic and social integration—attendance behaviour, academic performance, and LMS engagement—to quantifiable data elements extracted from existing institutional systems. While Tinto’s framework has been extensively validated in Western residential universities, this study provides empirical evidence of its applicability within a Kenyan private university, contributing to the cross-cultural transferability of established retention theory.

Second, the study makes a methodological contribution to the Educational Data Mining (EDM) framework by demonstrating that model explainability is not merely a technical enhancement but a prerequisite for sustainable deployment in resource-constrained settings. Existing literature predominantly evaluates predictive systems on accuracy metrics alone (Turkmenbayev et al., 2025; Wang et al., 2025). This study extends that paradigm by integrating SHAP-based explainability as a core design requirement, empirically testing the proposition that interpretable models generate higher stakeholder adoption and more targeted interventions than opaque high-accuracy alternatives—a theoretical claim with significant implications for future implementations across African higher education.

1.8.2 Practical Contribution

For Institutions The platform provides Strathmore University with the analytical infrastructure to transition from reactive to proactive student support, directing limited counselling and mentoring resources toward students most likely to benefit before crises occur. Beyond Strathmore, the study delivers a documented, open-source implementation blueprint directly applicable to comparable institutions across East Africa and the broader continent. Critically, the platform is designed to address the sustainability failures documented in prior African implementations—specifically the individual-dependency collapse observed at Kenyatta University (Wachira & Kimani, 2024) and the absence of real-time intervention capacity at the University of Pretoria (Moloi et al., 2024)—by embedding institutional processes, distributed ownership, and automated data pipelines rather than relying on individual champions.

For Educators The platform equips academic advisors, school administrators, and mentors with evidence-based insights that move decision-making beyond intuition. Role-based dashboards surface specific risk drivers—declining attendance trajectories, GPA below threshold, disengagement from LMS—enabling targeted, timely outreach rather than generic support. By providing SHAP-based explanations alongside risk scores, the platform enables advisors to conduct substantive, reason-led conversations with students rather than simply delivering an algorithmic verdict, increasing both the quality and credibility of interventions.

For Students Although undergraduate students are not direct platform users, they are its primary beneficiaries. Early identification of attendance non-compliance enables proactive outreach before the 67% examination eligibility threshold is irreversibly breached. Early flagging of GPA decline creates space for academic recovery before the 2.0 good-standing threshold triggers academic probation. More broadly, the system transforms Strathmore’s attendance and GPA policies from purely punitive enforcement mechanisms into supportive compliance frameworks—ensuring students receive guid-

ance precisely when data indicates they need it most, rather than only after they have already failed.

1.8.3 Policy Contribution

This research carries direct implications for higher education policy at both institutional and national levels. At the institutional level, the study demonstrates that a cost-effective predictive analytics platform—operable within the budget and staffing constraints of a mid-sized African private university—can generate measurable improvements in student support without proportional increases in expenditure. This provides Strathmore University’s leadership with an evidence base for formalising data-driven student monitoring as a standing institutional policy, rather than a one-off research project.

At the national level, the study responds directly to the Republic of Kenya Ministry of Education’s (Republic of Kenya Ministry of Education, 2024) Sessional Paper No. 1 of 2024, which identifies the inadequate utilisation of technology and unreliable institutional data as core bottlenecks in higher education. By demonstrating how existing institutional data can be systematically transformed into actionable early warnings, this research offers a practical model for addressing that bottleneck in a manner consistent with Kenya’s transition to Competency Based Education and Training (CBET), which emphasises evidence-informed teaching and learning environments. The platform’s open-source architecture and documented implementation methodology further position this study as a replicable policy instrument: other Kenyan universities and, more broadly, comparable institutions across East Africa facing similar funding constraints and enrolment pressures can adapt the framework without the prohibitive costs associated with commercial analytics platforms.

1.9 Summary

This chapter established the context, rationale, and direction of the study. Beginning at the global level, it demonstrated that student dropout remains a critical challenge in higher education and that machine learning-based predictive analytics has proven capable of identifying at-risk students weeks before conventional indicators surface—yet these benefits remain largely inaccessible to African universities. Across the continent, the few implementations attempted have been technically feasible but institutionally fragile, often collapsing when individual champions depart, as Kenyatta University’s experience most clearly illustrated.

At the institutional level, Strathmore University typifies this unresolved challenge. Rich data flows through its Academic Management System (AMS), Learning Management System (LMS), and attendance systems daily, yet these operate in isolation. This siloing forces the institution into a reactive support paradigm where students are identified only after breaching the 67% attendance threshold or 2.0 GPA floor—thresholds whose breach often signals a crisis already beyond easy recovery. This gap, situated within Kenya’s national call for greater technology utilization in education, defines the problem this study addresses.

The study responds by developing an explainable, machine learning-based predictive analytics platform tailored to Strathmore’s policies and resource context. The platform predicts dropout, course failure, and programme delay using ensemble algorithms trained on the 2021–2024 historical cohort. Its contributions span theory, practice, and policy: advancing retention frameworks in African contexts, providing a sustainable and replicable institutional tool, and demonstrating a cost-effective model for comparable universities across the region. Chapter Two now reviews the literature underpinning these contributions, examining theoretical frameworks, documented system implementations with their successes and limitations, and the precise knowledge gaps this study is designed to fill.

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>
- Bouihi, B., Bousselham, A., Aoula, E., Ennibras, F., & Deraoui, A. (2024). Prediction of higher education student dropout based on regularized regression models. *Engineering, Technology & Applied Science Research*, 14(6), 17811–17815. <https://doi.org/10.48084/etasr.8644>
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- 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>
- Gathani, S., Faiz, H., Hazilan, W. F., Khandhadia, R., & Fenn, C. J. (2026). Exploring the effectiveness of Machine Learning-based early warning systems in education: A scoping review [Accepted: February 4, 2026]. *Global Journal of Educational Thoughts*, 3(1), 32.
- 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/>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NIPS)*, 4765–4774.
- Moloi, K., Matabane, M. E., & Lumadi, M. (2024). Predictive analytics for student success in south african universities: Opportunities and challenges. *African Journal of Science, Technology, Innovation and Development*, 16(1), 45–57. <https://doi.org/10.1080/20421338.2024.2015483>
- Olawoyin, A., et al. (2024). A systematic review of learning analytics in sub-saharan africa: Gaps and opportunities. *African Journal of Higher Education Studies*, 8(1), 112–130.
- Republic of Kenya Ministry of Education. (2024). *Sessional paper no. 1 of 2024 on reforming education and training for sustainable development* (tech. rep.). Government Printer, Nairobi.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89–125. <https://doi.org/10.3102/00346543045001089>
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd). University of Chicago Press.
- 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>
- Wachira, P. M., & Kimani, E. N. (2024). Predictive analytics for student retention in kenyan universities: A case study approach. *Journal of Education and Practice*, 15(8), 42–51.
- Wang, L., et al. (2025). Deep learning for student success: Performance metrics vs. practicality. *Journal of Artificial Intelligence in Education*.