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

Chapter 1

Introduction

1.1 Background of the Study

Student retention and academic success represent critical challenges confronting higher education institutions worldwide. Recent data indicates that while persistence rates have rebounded to pre-pandemic levels, reaching 76.5% nationally in 2024, nearly one in four undergraduate students still fail to return after their first year (**Hanover2025**; Modern Campus, 2025). This persistent attrition represents not only significant personal and economic costs for individual students but also substantial challenges for institutional sustainability (**Element4512024**).

The advent of artificial intelligence and machine learning (ML) has revolutionized approaches to student success analytics. Contemporary platforms have evolved into comprehensive ecosystem monitors that analyze hundreds of data points to forecast outcomes (Mapademics, 2025). Georgia State University's implementation tracks over 800 risk factors daily, generating 90,000 targeted interventions annually (Mapademics, 2025). ML algorithms now identify patterns that precede student disengagement with accuracy levels reaching 88–92% (**Ahmed2024**; Scientific Reports, 2025).

In the African context, Kenyan institutions face distinctive challenges. While postgraduate retention improved to 87% in 2024, significant attrition remains (**AJER2025**). Universities confront multifaceted hurdles including inadequate funding (covering only 57% of students), infrastructure deficits, and staffing shortages (**Visualdo2024**; **Musasia2025**; KIPPRA, 2024).

Strathmore University, despite collecting extensive data through SIS and LMS platforms, lacks integrated, automated tools to proactively identify students at risk. The current reactive approach—addressing difficulties only after grades decline—represents a missed opportunity. Recent research demonstrates that ensemble methods like Random Forest and XGBoost can achieve prediction accuracies exceeding 88% (**Turkmenbayev2025**; Scientific Reports, 2025). Successful implementation requires transparent, interpretable models to provide clear insights

for interventions (**Frontiers2025**).

1.2 Statement of the Problem

Despite the availability of student data, many institutions lack automated systems to detect early warning signs of academic struggle. This leads to high attrition rates that could be mitigated through timely intervention.

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

KIPPRA. (2024). *The state of university funding in kenya*. Kenya Institute for Public Policy Research and Analysis.

Mapademics. (2025). Evolution of predictive analytics in higher education. *Student Success Ecosystems*.

Modern Campus. (2025). State of student retention 2025. *Higher Ed Reports*.

Scientific Reports. (2025). Machine learning patterns in student disengagement. *15*, 123–145.