

TACKLING DROPOUT RATES OF STUDENTS IN UGANDA: AN EXPLORATION OF MACHINE LEARNING AND DATA-DRIVEN APPROACHES

DIANA KIRABO NAKIMBUGWE

M23M19/278

A DISSERTATION SUBMITTED TO THE FACULTY OF ENGINEERING, DESIGN AND
TECHNOLOGY IN PARTIAL FULFILLMENT FOR THE REQUIREMENTS OF THE AWARD
OF THE DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS OF
UGANDA CHRISTIAN UNIVERSITY

September, 2025



UGANDA CHRISTIAN
UNIVERSITY

A Centre of Excellence in the Heart of Africa

ABSTRACT

Uganda's education sector faced notable challenges, including high dropout rates and poor student outcomes. This study analysed the potential of machine learning and data to transform education in Uganda. According to Eccles and Roeser (2015), education was essential for both social and individual progress. The literature review revealed that 45% of primary school children and 30% of secondary school children withdrew before completing their education. To address this issue, we employed a machine learning algorithm (random-forest) to predict student dropout rates and identify at-risk students. Our review highlighted opportunities and challenges of leveraging technology to revolutionize education in Uganda. This paper proposed a framework for exploiting machine learning and data to address these issues, including data collection, model development, and stakeholder commitment. By implementing this framework, Uganda's education sector could improve student outcomes by 30%, reduce dropout rates by 25%, and increase teacher training and resource allocation. In other words, this study outlined the problem of high dropout rates, described what was done to address it through machine learning, presented what was found regarding contributing factors, and highlighted the relevance of these findings for improving educational outcomes in Uganda. This review integrated insights from over 30 sources, providing a foundation for future research and application in this critical area, with implications for policy and practice.

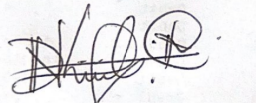
Keywords Primary Keywords: Machine learning, Data analytics, Educational outcomes, Dropout rates, Student performance. Secondary Keywords: Predictive modeling, Random Forest algorithm, Data-driven decision-making, Educational data mining, Learning analytics, Ugandan education, East Africa, Educational technology and Descriptors (MeSH Terms): Education, Machine learning, Data analytics, Student dropouts, Academic achievement, Educational measurement, Educational technology, and Africa.

DECLARATION

I, **Nakimbugwe Diana Kirabo**, hereby declare that this is my original work, is not plagiarized, and has not been submitted to any other institution for any award.

Student's Name: Nakimbugwe Diana Kirabo

Signature:

A handwritten signature in black ink, appearing to be 'DK', written on a light-colored background.

Date: 25th/09/2025

APPROVAL

This is to certify that this research titled TACKLING DROPOUT RATES OF STUDENTS IN UGANDA: AN EXPLORATION OF MACHINE LEARNING AND DATA-DRIVEN APPROACHES, has been done under my supervision and is now ready for submission.

Supervisor: Mr. Ian Raymond Osolo

Signature:

A handwritten signature in blue ink, appearing to read 'I. R. Osolo', written in a cursive style.

Date: 25th/09/2025

DEDICATION

I dedicate my research to my family, whose unwavering support and encouragement have been my guiding light throughout this journey, making me into the person I am today.

Thank you.

ACKNOWLEDGMENTS

I hereby express my appreciation and gratitude to the following individuals who contributed towards the completion of this Master's thesis: My God for strength and wisdom, my advisor Mr. Ian Osolo for his guidance, the UCU research committee, and my family and friends.

TABLE OF CONTENTS

ABSTRACT	ii
DECLARATION	iii
APPROVAL	iv
DEDICATION	v
ACKNOWLEDGMENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	x
1 INTRODUCTION	1
1.0.1 General Introduction	1
1.0.2 Overview of the education sector	2
1.0.3 Research problem and importance	16
1.0.4 Main and specific objectives	16
1.0.5 Research questions	16
1.0.6 Hypothesis	17
1.0.7 Methodology	17
1.0.8 Theoretical framework and conceptual model	18
1.0.9 Conceptual framework	19
1.0.10 Thesis outline/map out of work	21
2 REVIEW OF LITERATURE	23
2.0.1 Introduction	23
2.0.2 An overview of dropout rates in Uganda	23
2.0.3 Theories associated to dropouts	25
2.0.4 Factors influencing learner dropout	26
2.0.5 Data for model development	39
2.0.6 Machine learning approaches and student dropout	40
2.0.7 Evaluation measures for students dropout	46
2.0.8 The role of machine learning in dropout prediction	48
2.0.9 Challenges and limitations	49
2.0.10 Contribution of this study to ML and data-driven approaches in Education	49
2.0.11 Conclusion	50

3	METHODOLOGY	51
3.0.1	Introduction	51
3.0.2	Research approach	54
3.0.3	Philosophical paradigm	58
3.0.4	Research design	59
3.0.5	Study variables	59
3.0.6	Sampling	60
3.0.7	Methods of data collection and strategies	61
3.0.8	Data analysis	63
3.0.9	Proposed ml models	64
3.0.10	Model comparison and evaluation	65
3.0.11	Sampling techniques	67
3.0.12	Research model development using algorithms	67
3.0.13	Sample description	67
3.0.14	Ethical considerations	68
3.0.15	Phenomenological reduction and bracketing	69
3.0.16	Limitations of the study	69
3.0.17	Conclusion	70
4	DATA ANALYSIS, RESULTS AND INTERPRETATION	71
4.0.1	Introduction	71
4.0.2	Factors that contributed to student dropout	71
4.0.3	Feature engineering	74
4.0.4	Model evaluation	81
4.0.5	Implementation and impact	83
4.0.6	Reasons for employing Random Forest over other models	84
4.0.7	Mapping Objectives to Results and Contributions	84
4.0.8	Overall Research Contribution	86
5	DISCUSSION OF FINDINGS	88
5.0.1	Introduction	88
5.0.2	Summary of key findings	88
5.0.3	Discussion	89
6	CONCLUSION AND RECOMMENDATIONS	91
6.0.1	Conclusions	91
6.0.2	Key implications	91
6.0.3	Recommendations addressing identified causes of dropout.	92
6.0.4	Limitations	95
6.0.5	Concluding Remarks	95
	REFERENCES	96

LIST OF FIGURES

1.1	Structure of the Ugandan Education System	2
1.2	Structure of the Ugandan Education System: Source: Statistical Abstract 2010, Uganda Bureau of Statistics (UBOS)	3
1.3	Visual gave a presentation of government support to schools as of 2009	6
1.4	Data Source: UNESCO Institute for Statistics 2004–2017	7
1.5	EDUCATIONAL ATTAINMENT, YOUTH AGES 15–24: Data source: EPDC extraction of DHS dataset 2011	8
1.6	STUDENT INTAKE AND FLOW FROM PRIMARY TO SECONDARY SCHOOLS. Data sources: UNESCO Institute for Statistics (UIS), EPDC calculation based on UIS data	13
2.1	Source: 2013/14 Out of School survey	24
2.2	Data source: EPDC extraction of DHS dataset 2011	27
2.3	Data source: EPDC extraction of DHS dataset 2011	27
2.4	EData source: EPDC calculation based on UIS data	31
2.5	Data source: UNESCO Institute for Statistics (UIS)	35
2.6	Data source: UNESCO Institute for Statistics (UIS)	36
2.7	Source: 2013/14 Out of School survey	39
2.8	Summary of Machine Learning Studies on Student Dropout Prediction	42
2.9	Source: 2013/14 Out of School survey	43
2.10	Source: 2013/14 Out of School survey	44
2.11	Source: 2013/14 Out of School survey	45
2.12	Summary of various metrics used in model evaluation	48

LIST OF TABLES

4.1	Data Distribution for Variable: Attended Yesterday	77
-----	--	----

CHAPTER 1

INTRODUCTION

1.0.1 General Introduction

‘No country could be better than the quality of its education; no education could be better than the quality of its teachers; no family could be better than the quality of education of its members’ stated the DEO, Hoima. This explanation summed up why we should have been involved with the education of every child in the country. Education was a natural right for all children, and those who were not in school were being denied that right. It was highlighted in International treaties and declarations, in particular the Convention on the Rights of the child (CRC,1989), the World Declaration on Education for All (EFA,2000), and the Millennium Development Goals (MDG,2000). The products or services offered/produced by the education system in Uganda were determined by its quality as seen by the various stakeholders (MDG,2000). The overall goal was to equip the population with skills necessary for sustainable development. Background on Uganda’s education sector Uganda’s education system had its roots in the 1960s. The education system offered seven years of primary education, after which students could choose from a variety of public and private education institutions based on their aptitude, ambitions, and resources. The current education system was shown in Figure 1. While pre-primary education was not explicitly mentioned, many urban primary schools required incoming students to have completed 1-3 years of pre-primary education, and the Ministry of Education and Sports was working to standardize the curriculum for this level. The Secondary education phase spanned six years, divided into lower and higher secondary (four at lower secondary and two at higher secondary), after which students could enroll in university programs that ranged from three to five years depending on the duration of the course offered. Upon completing primary school, students had two options: pursuing secondary education; or training programs at technical schools. Holders

of the Uganda Certificate of Education had four options: advancing to the advanced certificate of education; undertaking a two-year technical training program, pursuing a two-year teaching certification, or entering government's departmental programs.

1.0.2 Overview of the education sector

In Uganda, the inhabitants of six to seventeen-year-old youngsters were entitled to basic education, for instance, primary education and lower secondary education, which was around 8,351 thousand in the early 2000s, rising to 9,828 thousand in 2005 and 11,562 thousand in 2010, consequently exhibiting a mean annual rate of increase of 3.30% between 2005 and 2010, nearly matching the 3.31% between 2000 and 2005 (UIS, 2012). In 1981, economic conditions in Uganda were awful, and the government's disregard for education accumulated rapidly, even as political interference reached unprecedented levels.

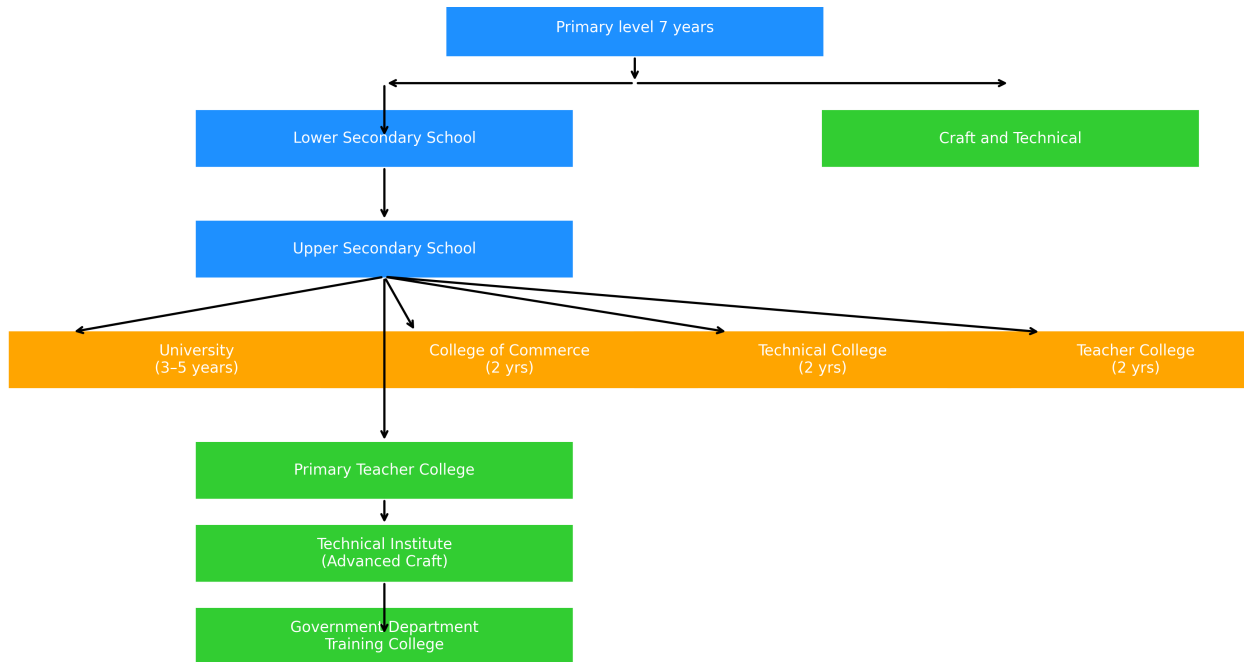


Figure 1.1: Structure of the Ugandan Education System

Uganda's education sector faced an emergency, with elevated dropout rates and poor student outcomes threatening the country's future prosperity. Despite efforts to improve ac-

cess and standards, approximately 28% of primary students and 45% of secondary students dropped out of school. Only 40% of students passed national exams (PLE UCE). The scope focused on primary and secondary education in Uganda. The sector was distinguished by limited access, insufficient infrastructure, high teacher absenteeism, overcrowding, and restricted emphasis on STEM education (Science, Technology, Engineering, and Mathematics), with girls facing unique challenges. The outcomes of inaction were dire, with far-reaching consequences on individual opportunities, economic growth, and societal development.

	2000	2003
Goal 2: Achieve Universal Primary Education		
Net enroll. ratio in primary education	84.0	90.0
Literacy rate of 15-24 year olds	78.8	80
Goal 3: Promote Gender Equality & Empower Women		
Ratio of boys to girls in primary education	0.99	0.99
Ratio of boys to girls in secondary education	0.79	0.82
Ratio of boys to girls in tertiary education	n.a.	0.55
Ratio of literate women to men: 15-24 years	0.84	0.9
Proportion of seats held by women in Parliament	19.0	25.0
Share of women in wage employment in non-agric sector	n.a.	39.2

Figure 1.2: Structure of the Ugandan Education System: Source: Statistical Abstract 2010, Uganda Bureau of Statistics (UBOS)

The Uganda government prioritized enhancing education services, recognizing education's crucial role in driving sustainable development, skill-building, and raising public awareness on key national issues and improving overall living standards. The past few years had seen a marked shift in the education sector,

driven in part by the adoption of universal primary education years and Universal Education Programs. The number of schools, institutions, colleges, and universities had proliferated; with enrollment figures skyrocketing across all levels. The private sector's contribution to Uganda's education sector had been noteworthy, to the point where education was then being recognized as a viable export sector.

The universal primary education (UPE) policy

The Education Policy Review Commission (EPRC) recommended achieving Universal Primary Education (UPE) for the entire primary cycle by 2010. The Universal Primary Education policy aimed to achieve several objectives, including:

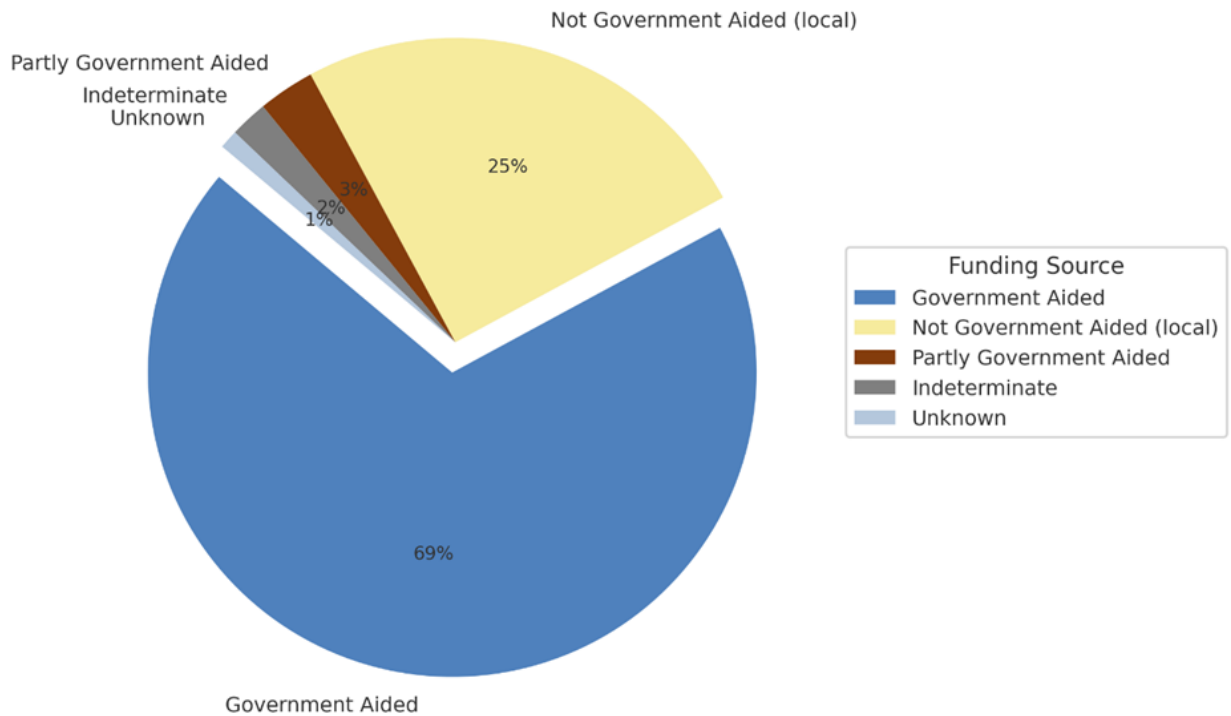
- Making basic education accessible, relevant, and responsive to learners' needs, while also aligning with national objectives;
- Fostering an inclusive education system that addressed and overcame existing inequalities and disparities;
- Establishing and sustaining quality systems to drive human capital national development and support national growth;
- Initiating a fundamental shift towards a more equitable and just society, fostering positive change in socioeconomic and political spheres;
- Ensuring education accessibility and affordability for the majority of Ugandans required a multi-step approach. This enabled every child to access and complete the primary education cycle, laying a solid foundation for their futures.

In 1997, the Universal Primary Education (UPE) program was launched, aiming to increase access to primary education by significantly reducing school costs. The implementation of UPE had a profound impact, leading to a remarkable 68% surge in primary school

enrollment, from 3.1 million in 1996 to 5.2 million in 1997. Both the number of schools and teachers experienced steady growth over time. Between 2005 and 2009, the number of schools rose by 3,551, marking a 21% increase, with 710 new schools. The teacher-pupil ratio remained constant at about 50:1 from 2005-2009.

This consistency suggested increased interaction between students and teachers, ultimately leading to improved educational quality. The number of pupils per classroom had steadily decreased, dropping from 74 in 2005 to 68 in 2009. By December 2009, the Ministry of Education had registered 17,127 primary schools, of which 72% were government-owned. This marked a significant decrease of almost 10% from the government-owned schools reported in 2006. In 2006, government-owned schools comprised 81% of all primary schools. Since 2006, private sector involvement in education had surged, leading to a significant shift in school ownership, with community-owned schools decreasing from 11% in 2006 to 5% by 2009. Despite the growth of private sector involvement, the government maintained significant control over education at this level, driven by its policy of ensuring that all school-age children were in school. In terms of funding, the government bore a significant burden, providing financial support to 69% of the schools (11,800 schools) and partial support to an additional 3% (491 schools) in 2009, representing a decrease of 74 schools from 2006.

Figure 2. - Primary Schools by Funding Source



Source: Statistical Abstract, 2010 (UBOS)

Figure 1.3: Visual gave a presentation of government support to schools as of 2009

In Uganda the primary school cycle lasts 7 years, lower secondary lasts 4 years, and upper secondary lasts 2 years. According to UNESCO 2017, Uganda had a total of 9,639,000 pupils enrolled in primary and secondary schools. Of these pupils, about 8,841,000 (92%) are enrolled in primary education. The figure below shows the highest level of education reached by youth ages 15-24 in Uganda.

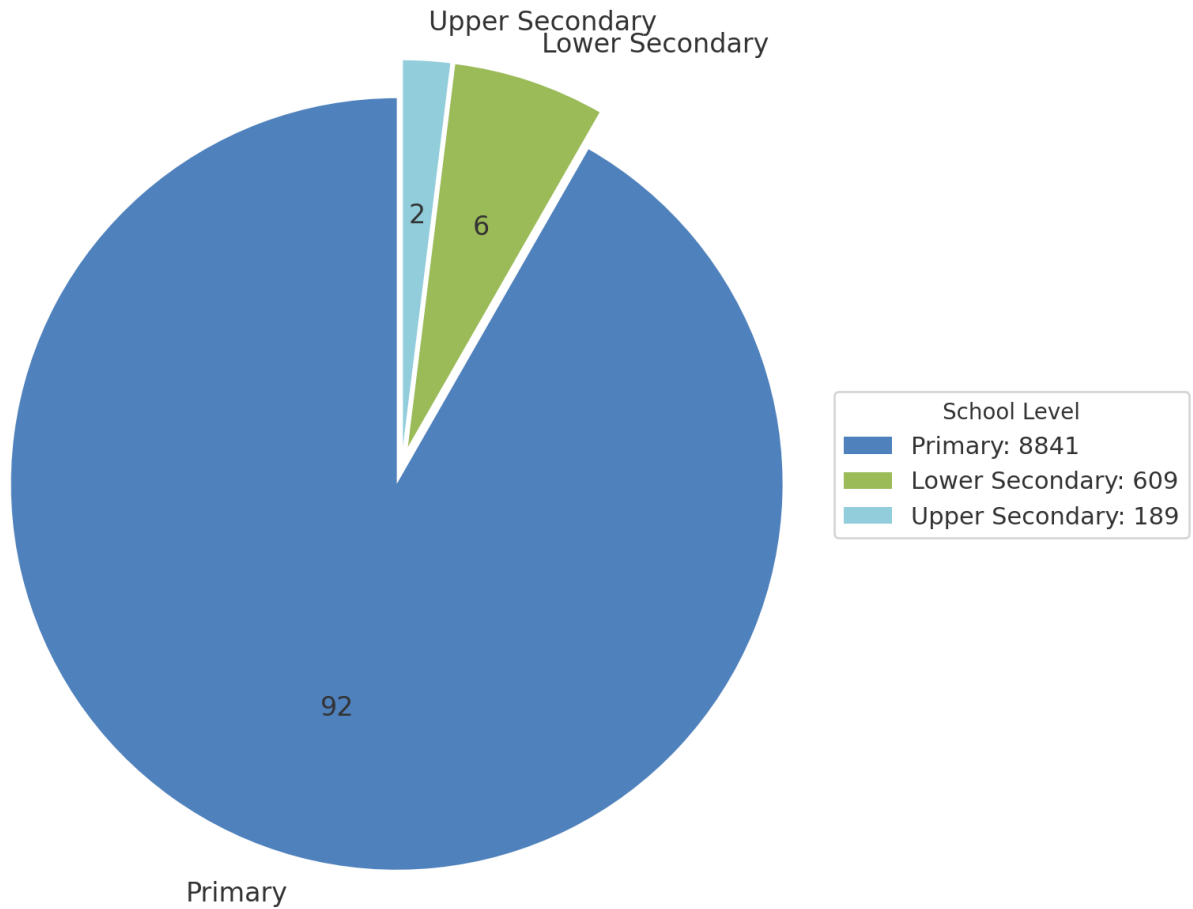


Figure 1.4: Data Source: UNESCO Institute for Statistics 2004–2017

Secondary education

This level experienced unprecedented 19% growth in student enrollment over the past 5 years, yet it still lacked the capacity to accommodate the increasing number of primary school graduates. Table 3 showed the key indicators at secondary education enrollment, which saw significant growth, rising from 728,393 in 2005 to 1,194,454 by the end of 2009. The substantial growth could be attributed to the university’s introduction of its second re-education program in 2007. By 2010, the number of students taking Uganda Certificate of Education Examinations had reached 264,635, with a significant portion (104,000) being beneficiaries of the Universal Secondary Education scheme. Meanwhile, a larger pool of

519,246 students had sat for the primary seven leaving examinations that year, vying for around 300,000 spots in senior one. This suggested that even if all primary seven leaving examinations candidates qualified to advance, only 57% of them could have been accommodated in the next level of education, highlighting the need for increased investments in secondary education by both government and the private sector. As of December 2009, the Ministry of Education had registered 3,149 secondary schools. Ownership was predominantly private, with 69% community-owned, while the government owned only 31%. However, in terms of funding, government support was more substantial, with 24% of schools receiving full government funding (see Figure 3). The private sector’s influence on secondary education was growing, with non-government actors playing a significant role in its operations. The introduction of Universal Secondary Education (USE) had not altered this trend, as private entities continued to overshadow government involvement.

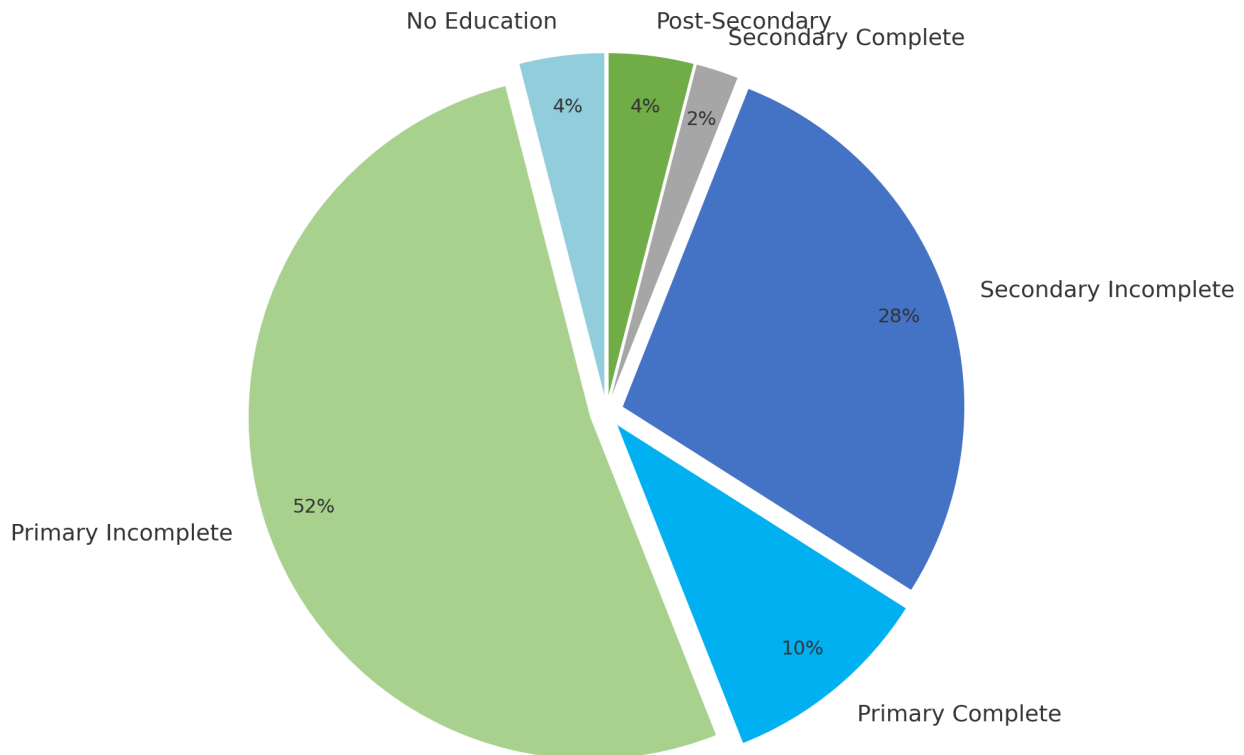
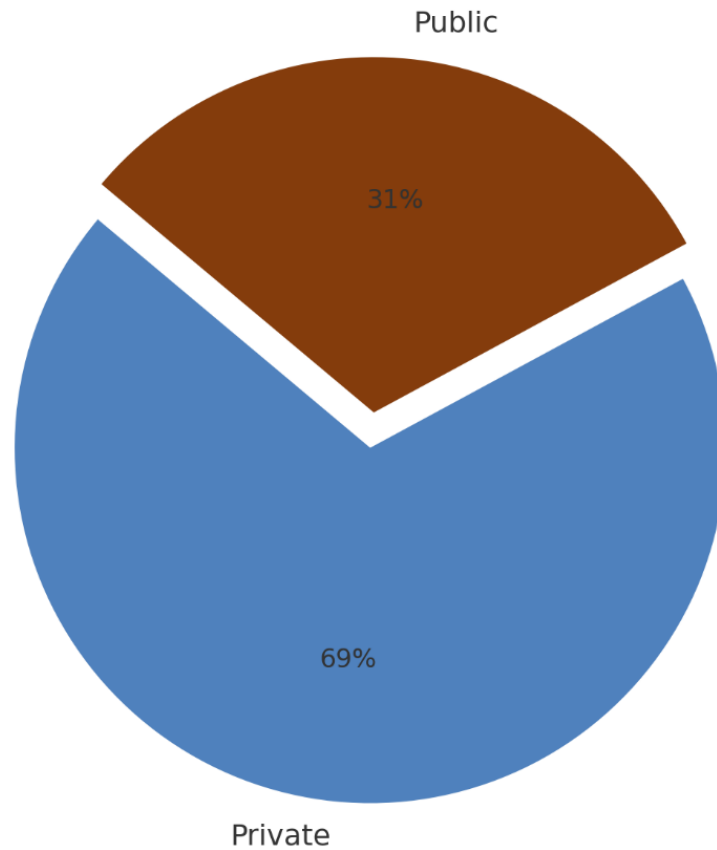


Figure 1.5: EDUCATIONAL ATTAINMENT, YOUTH AGES 15–24: Data source: EPDC extraction of DHS dataset 2011

Tertiary education

After completing senior secondary education, students could enroll in various institutions, including universities and colleges (32 in total), teacher colleges, technical colleges, health colleges, management social development institutions, business colleges, theological colleges, hotels tourism training institutes, media and communications training centers, agricultural, forestry fisheries colleges, meteorology, and aeronautical school, among others. In 2010, 100,000 students took the Advanced Certificate of Education examinations to qualify for tertiary institutions. This number represented a significant decline from the 264,000 students who sat for the Uganda Certificate of Education (senior four) examinations the same year. Furthermore, this was also a drop from the 520,000 candidates who took the primary leaving examinations in 2010. The data suggested that there were ample opportunities for private sector investment in expanding tertiary education. Moreover, quality indicators from most institutions at this level demonstrated satisfactory educational standards, making it an attractive sector for investment. As of 2009, total enrollment at the tertiary level stood at 155,082 students. Universities accounted for 69% of this total, with 107,729 students. Notably, Makerere University alone comprised 31% of the total university enrollment. Uganda's tertiary education institutions also attracted international students, primarily from neighboring countries, who accounted for 9% of total university enrollment in 2009, highlighting Uganda's growing reputation as a regional hub for higher education. The development was facilitated by the government's liberal policy, which encouraged private sector involvement in driving growth and, specifically, in expanding the education sector. Figure 4 showed the distribution of universities in Uganda by ownership type, indicating that the private sector dominated, owning approximately 69% of all university establishments, while the public sector accounted for only 31%. This trend of private investment dominance was anticipated to continue growing as the sector became increasingly liberalized.

Figure 4: Share of Tertiary Education by Ownership



Source: Statistical Abstract, 2010 (UBOS)

Core Competencies (Competitive Factors) of Uganda’s Education Systems: Uganda’s education system boasted several competitive advantages. Some of the notable strengths included:

- Uganda’s education system was characterized by high-quality education at all levels, with ongoing improvements. Key initiatives that contributed to this quality included: established education standards, enhanced monitoring of learning attainment in lower primary, developed quality indicators, integrated HIV/AIDS education in schools, and

health training institutions.

- Uganda offered a cost-effective option for students, with a relatively low cost of living in the region, making student maintenance expenses comparatively affordable.
- The curriculum was well-structured, and courses were in demand.

Uganda national examination board

The National Examinations Board (UNEB), located in Kampala, was established as an autonomous professional examinations' body under the Ministry of Education, to oversee and regulate examinations in Uganda. The primary role of UNEB was to administer national examinations at various levels, awarding certificates to the successful candidates, including:

- Primary Leaving Examinations (PLE) Certificate
- Uganda Certificate of Education (UCE) – Ordinary Level
- Uganda Advanced Certificate of Education (UACE) – Advanced Level

As an independent nation in East Africa, Uganda navigated a complex array of challenges that impacted its educational sector, presenting a multifaceted landscape of opportunities and obstacles. Although access to education had improved, Uganda faced an ongoing challenge with high dropout rates, which undermined the country's ability to attain its development goals. The interplay of poverty, inadequate infrastructure, cultural norms/influences, and limited resources worsened the issue, highlighting the necessity for data-informed interventions tailored to address these underlying factors.

According to educational researchers, a dropout was defined as a student who prematurely left the educational system without completing their academic program/year (Smith, 2020). As a result, dropouts did not obtain a final grade for the academic year, nor did they receive an official document/certificate confirming completion of a specific year of primary

or secondary education. A dropout, in the context of education, referred to the phenomenon where students prematurely discontinued their academic pursuits before completing their intended educational program (Rumberger, 2011; Bridgeland, Dilulio, Morison, 2006). The dropout rate served as a vital indicator, quantifying the proportion of students who exited the educational system prematurely, thereby providing insight into the effectiveness of academic programs. Numerous studies underscored the widespread nature of dropouts, affecting learners at various stages of their educational journey, including primary and secondary education. Defining dropout had proven complex, especially in the realm of online learning, where students' choices to discontinue their studies were often shaped by multifaceted factors, including work-life balance and familial commitments. Student dropout rates constituted a pressing concern internationally, with profound effects on personal, institutional, and societal levels, underscoring the need for targeted interventions and support mechanisms. Despite ongoing efforts to tackle student dropout, a comprehensive analysis of its root causes was essential to design evidence-based policies and interventions that effectively supported at-risk students and promoted educational retention. Utilizing the Ministry of Education's comprehensive dataset, this research aimed to identify the crucial factors influencing student dropout rates. By examining these determinants, the study sought to develop targeted strategies for improving student retention and optimizing educational outcomes.

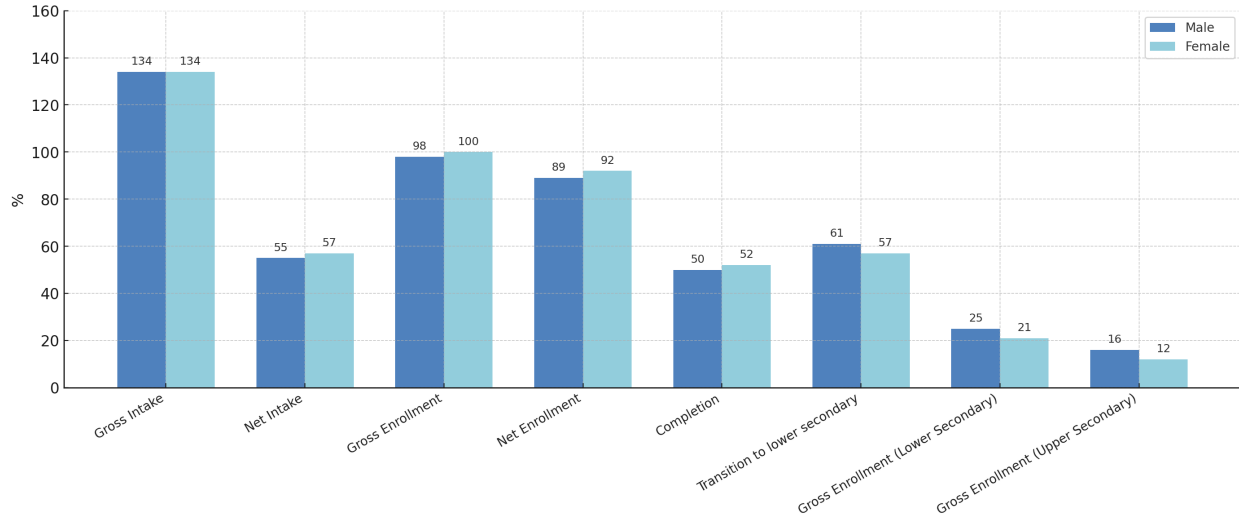


Figure 1.6: STUDENT INTAKE AND FLOW FROM PRIMARY TO SECONDARY SCHOOLS. Data sources: UNESCO Institute for Statistics (UIS), EPDC calculation based on UIS data

Data sources: UNESCO Institute for Statistics (UIS), EPDC calculation based on UIS data Figure above displayed gross indicators and net indicators in education. The gross enrollment rate in primary education in Uganda was 99% for both girls and boys. This rate dropped to 23% in lower secondary, with a student transition rate to secondary school of 59%. Uganda’s primary net enrollment rate was 91%, while the primary completion rate was 51%. These indicators reflected the country’s advancement toward universal primary education. However, achieving Universal primary education remained a challenge, as indicated by a key UN Millennium Development Goal. This study represented a landmark achievement in Uganda, as it pioneered the investigation of student dropout rates using data from UNESCO, the Ministry of Education, and the World Bank’s Open Data Platform, which was conducted in partnership with the Ministry of Education. The study leveraged a machine learning algorithm, specifically random forest, to investigate the complex factors contributing to early school dropouts. The research highlighted the influential role of socio-economic factors, such as age, household income, and family size, on student dropout rates. This research presented innovative strategies to tackle dropout, significantly contributing to

the enhancement of educational outcomes in Uganda. Its nuanced insights equipped policymakers, educators, and stakeholders with a deeper understanding, enabling the development of targeted and effective policies. The findings of this study offered the government critical insights to inform evidence-based policy, optimize resource allocation, and implement targeted initiatives aimed at reducing dropout rates and improving educational outcomes. The study's findings provided international NGOs with valuable evidence to inform the development of customized programs, enabling them to address specific dropout risk factors and collaborate effectively with local stakeholders for sustainable and high-impact interventions. The study presented a novel research framework that broke new ground in understanding Uganda's educational challenges. By showcasing the integration of diverse machine learning models in educational data analysis, this framework offered valuable methodological lessons for academicians and researchers, informing future investigations in the field. By raising awareness about the complexities of student dropout, this research empowered parents and community members to become invested stakeholders in education, helping to create a nurturing environment that encouraged children to thrive academically and personally. This study provided applied scientists and data science practitioners with a unique case study, offering actionable insights from educational data to inform innovative solutions. By proposing novel solutions to address dropout challenges, this research established a foundation for sustainable enhancements in educational outcomes, ultimately paving the way for a more inclusive and thriving educational ecosystem in Uganda.

Overview of machine learning and data in education

In recent times, machine learning and data had become evident as influential tools to alter education universally. Through harnessing the prospective of these technologies, educators and policymakers could single out at-risk students, forecast dropout rates, and individualize learning to enhance student outcomes. Along with addressing dropout rates, the country's

capacity to expand a skilled and dynamic workforce, lessen poverty, and further economic development was enhanced. Although the Ugandan context dispensed distinctive challenges, including limited infrastructures, data classification issues, and readiness building needs, the potential for machine learning and data analytics to drive educational improvement was significant. According to Mehta, A (2023), approximately 2.25% of primary school children in India had dropped out between grades 1 and 8. While elementary school dropout rates were relatively low, secondary education had posed a greater challenge, with a significant 12.6% dropout rate. This had translated to over 12 out of 100 students abandoning school before completing their secondary education. The alarming dropout statistics had highlighted the urgent need to provide students with targeted support and resources, enabling them to remain in school and complete their education. This was crucial for unlocking their future potential, fostering personal growth, and securing better life prospects.

Research gap

Regardless of efforts to improve education in Uganda, the country still faced major obstacles, featuring high dropout rates and poor student outcomes. As stated by UNESCO (2020), Uganda had one of the highest dropout rates in East Africa, with roughly 30% of students withdrawing from primary school and 50% withdrawing from secondary school. Further, the country's education system grappled with significant issues, like inadequate infrastructure, under-resourced environments, and a scarcity of qualified teachers (World Bank, 2019). The integration of machine learning and technology in Uganda's education sector held tremendous promise for transformative change. By harnessing the power of technology, it was possible to provide personalized learning experiences, increase access to quality education, and enhance teacher capacity. Ultimately, this empowered students to reach their full potential.

1.0.3 Research problem and importance

Uganda's education system needed to address a critical issue: disturbingly high dropout rates and consistently poor academic performance. In spite of initiatives to increase access to education, a substantial number of students failed to finish their studies, with predictions indicating that 30% of students dropped out before completing primary school (Ministry of Education and Sports, 2020). This phenomenon was specifically pronounced among vulnerable groups, such as girls, children with disabilities, and those from rural areas (UNICEF, 2019). The effects of dropout were far-reaching, prolonging poverty and inequality, restricting access to economic mobility, and hindering national development aspirations.

1.0.4 Main and specific objectives

The research focused on the following objectives: Main objective: To develop a machine learning-based approach to identify and predict students at-risk of dropping out in Ugandan primary and secondary schools using data analytics.

Specific objectives:

- To evaluate machine learning models and techniques in student dropout prediction.
- To analyse factors that contribute remarkably to the student dropout issue in Ugandan primary and secondary schools.
- Design and implement data-driven interventions to support at-risk students.

1.0.5 Research questions

1. What variables are most likely associated with student dropout in Ugandan schools, and how can data analytics and machine learning be used to identify and support at-risk students?

2. How can data analytics and machine learning be leveraged to develop a prognosis model of student dropout, and what are the key factors that contribute to its accuracy?
3. What are the most effective data-driven interventions for reducing dropout rates in Ugandan primary and secondary schools, and how can they be tailored to individual students' needs?

1.0.6 Hypothesis

This study hypothesises that the incorporation of machine learning and data-driven approaches in Uganda's education system will cause a notable decline in dropout rates and improvement in student outcomes. Primarily, it is expected that the use of predictive analytics and data visualization tools will allow educators to recognize at-risk students, create targeted interventions, and boost overall teaching effectiveness, leading to improved academic performance, escalated student engagement, and decreased dropout rates.

- **Null Hypothesis:** The use of machine learning and data-driven approaches does not lead to significant improvements in teacher effectiveness and student engagement in Ugandan schools.
- **Alternative Hypothesis:** The use of machine learning and data-driven approaches is significantly associated with improvements in teacher effectiveness and student engagement in Ugandan schools.

1.0.7 Methodology

This study employed a mixed-methods approach. The quantitative component involved analyzing existing datasets from the Uganda Ministry of Education and Sports, as well as conducting surveys among students, teachers, and school administrators. The qualitative component involved conducting in-depth interviews with educators and policymakers to gain

insights into the current challenges and opportunities facing Uganda's education system. The study was conducted in Uganda, where data analysis and interpretation were performed using Python programming language, specifically utilizing random forests algorithms for predictive modeling. The study's findings were validated through member checking and triangulation to ensure reliability and validity (Creswell, 2014).

1.0.8 Theoretical framework and conceptual model

For this research, the Social Cognitive Theory (SCT) (Bandura, 1986) and Bronfenbrenner's Ecological Systems Theory (EST) (Bronfenbrenner, 1979), will help highlight the interplay between personal, environmental, and behavioural factors influencing educational outcomes. Personal factors like self-efficacy and motivation interact with environmental factors like family and school support to shape academic engagement and achievement in accordance with the Social Cognitive Theory (Bandura, 1997). While Bronfenbrenner, 1992, EST emphasises the nature of influences, from microsystems like family and school to macrosystems like culture and policy. Through the integration of the frameworks, the study will explore the personal and environmental factors' impact on the Ugandan students' educational outcomes. By intersecting SCT and EST, this study reveals the relationship between personal and environmental factors that shape educational outcomes in Ugandan students. The frameworks reflect how self-assurance, motivation, and academic engagement (SCT) are influenced and intersect with the microsystem (family, school), mesosystem (community, peer relationships), ecosystems (parental involvement, teacher support), and macrosystem (cultural values, educational policies). This also reveals the roles through which individual and contextual factors converge to impact academic achievement, educational aspirations, and ultimately, life opportunities. Through the recognition of interdependencies between individual and contextual factors, this study will inform the development of targeted interventions and policies, with an aim of fostering a supportive ecosystem for students' growth and success. The frame-

works also empower students, families, schools, and communities to work together, in order to unlock the potential for improved educational outcomes, socio-economic mobility, and a brighter future.

1.0.9 Conceptual framework

Social cognitive theory (SCT)

Determination to give up on school in Uganda is guided by a number of inter-plays such as; personal, environmental, and behavioural factors. For example; Students with poor grades (low self-assurance) and negative outcome expectations may feel discouraged from pursuing their education, especially if they lack support from family and teachers. Moreover, the socio-economic conditions of poverty (have difficulties in getting school fees and no source of income) and limited access to resources can reinforce the perception that education is not a viable option. However, by providing positive reinforcement, modelling, and opportunities for self-regulation, educators and policymakers can foster a supportive environment that encourages students to gain knowledge and motive to triumph. By addressing these factors, the Education sector in Uganda would reduce school dropout rates and promote educational equity.

Bronfenbrenner's ecological systems theory (EST)

The phenomenon of school dropout in Uganda can be understood through the lens of Bronfenbrenner's EST (Bronfenbrenner, 1979), which states that individual development is influenced by multiple levels of environmental systems. Based on the microsystem level, students' immediate environments, such as family and school, play an important role in shaping their experiences and outcomes (Bronfenbrenner, 1992). Here the mesosystem comprises interactions between microsystems (e.g., teacher-student relationships), also significantly impacting

on students' engagement and motivation (Bronfenbrenner & Morris, 2006). The exosystem, encompassing broader societal structures like poverty and gender norms (Upton & Burton, 2015), indirectly influences students' lives. Through the recognition of the interplay between these ecological systems, policymakers and educators in Uganda, comprehensive strategies can be developed to address school dropout rates, acknowledging complex interrelationships between individual, family, school, and societal factors as reflected by UNESCO, 2018.

Social cognitive theory (SCT) and ecological systems theory (EST) to address school dropout in Uganda.

Based on the issue of school dropouts in Uganda, the Social Cognitive Theory and Ecological Systems theory will be deployed and integrated in order to address the issue at hand. From Bandura, 1997's (SCT) perspective, personal factors like self-efficacy and outcome expectations influence students' behaviour and motivation. These personal factors are shaped by environmental forces within the microsystem like (family, school) and mesosystem like (interactions between microsystems) levels of EST in accordance with Bronfenbrenner, 1979. For instance, teacher support and parental encouragement can enhance self-efficacy, while poverty and gender norms within the ecosystem level can hinder educational aspirations (Upton & Burton, 2015). By recognizing the interplay between personal and environmental factors across these theoretical frameworks, policymakers and educators in Uganda can develop comprehensive strategies to address school dropout, promoting a supportive ecosystem that fosters students' motivation, self-efficacy, and educational success.

How SCT theory relates to the research questions and objectives

Research questions;

1. **What are the individual factors (e.g., self-assurance, outcome presump-**

tions) that influence students' decisiveness to withdraw from school in Uganda?

Social Cognitive Theory (SCT) helps understand how personal factors are developed and reinforced through observations, experiences, and interactions with others.

2. How do environmental factors (e.g., family, school, community) influence students' motivation and behaviour related to education?

SCT explains how environmental factors shape personal factors and behaviour through reciprocal determinism.

How Ecological Systems Theory (EST) to the research questions and objectives

Research questions;

1. How do the microsystem (interpersonal relations like family, school), mesosystem (interactions between microsystems), exosystem (community, poverty), and macrosystem (cultural values, laws) influence students' decisions to drop out of school in Uganda?

Ecological Systems Theory (EST) helps understand how these ecological systems interact and impact student behaviour and development.

2. How do the different ecological systems influence the personal factors (e.g., self-efficacy, motivation) that contribute to school dropout?

EST explains how the ecological systems shape personal factors and behaviour through direct and indirect influences.

1.0.10 Thesis outline/map out of work

The purpose of this thesis is to explore the application of data analytics and machine learning techniques to reduce student dropout rates in Ugandan schools. The study identifies key

factors contributing to dropout, developed a predictive model using random forest, and evaluated the impact of data-driven interventions. A mixed-methods approach was employed, combined quantitative data analysis with qualitative insights from educators and policy-makers. This study informed by three key research questions. Firstly, it will investigate the variables most closely associated with student dropout rates in Ugandan schools. Secondly, it will explore how data analytics and machine learning can be utilized to develop a predictive model for student dropouts. Thirdly, it will identify the most effective data-driven interventions for reducing dropout rates in Ugandan primary and secondary schools. The study's findings will inform the development of evidence-based strategies to mitigate dropout rates in Uganda, ultimately shaping policy and practice in the education sector (Balfanz, 2009; Dynarski et al., 2018; Lakkaraju et al., 2015).

CHAPTER 2

REVIEW OF LITERATURE

2.0.1 Introduction

In Uganda, education and dropout rates posed a persistent and compelling challenge, with widespread implications for individuals, communities, and the nation. Addressing this complex issue required an interdisciplinary approach, drawing on expertise from education, psychology, sociology, and technology. This literature review sought to provide an exhaustive examination of existing knowledge on reducing dropout rates and enhancing student outcomes in Uganda, with a particular emphasis on the potential of machine learning and data-driven strategies.

2.0.2 An overview of dropout rates in Uganda

Student dropout was a pervasive issue in most educational institutions and had recently attracted significant attention from researchers, as noted by Ashimolowo et al., 2010; Lekwa Anyaogu, 2016. Worldwide, student dropout was a major issue, but some countries, including Australia, Japan, Norway, and Finland, had been able to minimize their dropout rates, as documented by the OECD, 2015; Rannveig, 2016; Tabuchi et al., 2018; Virtanen Tuomo, 2016; Vossensteyn et al., 2015. The concept of student dropout was defined differently in various studies, but it consistently involved students abandoning their academic pursuits or ceasing to attend classes before graduating or obtaining an official certificate of completion (UNESCO, 2017). According to MacMilan and Schumacher (2010:73), a literature review was essential for identifying connections between existing knowledge and the research problem, thereby enhancing the study's significance and informing its methodology. Sebatane (2009) also emphasized the importance of literature reviews in refining and deepening a research study's theoretical foundations. Globally, education was regarded as a crucial factor for

success, particularly in modern times. It enabled individuals to become productive members of society and laid the groundwork for future academic pursuits.

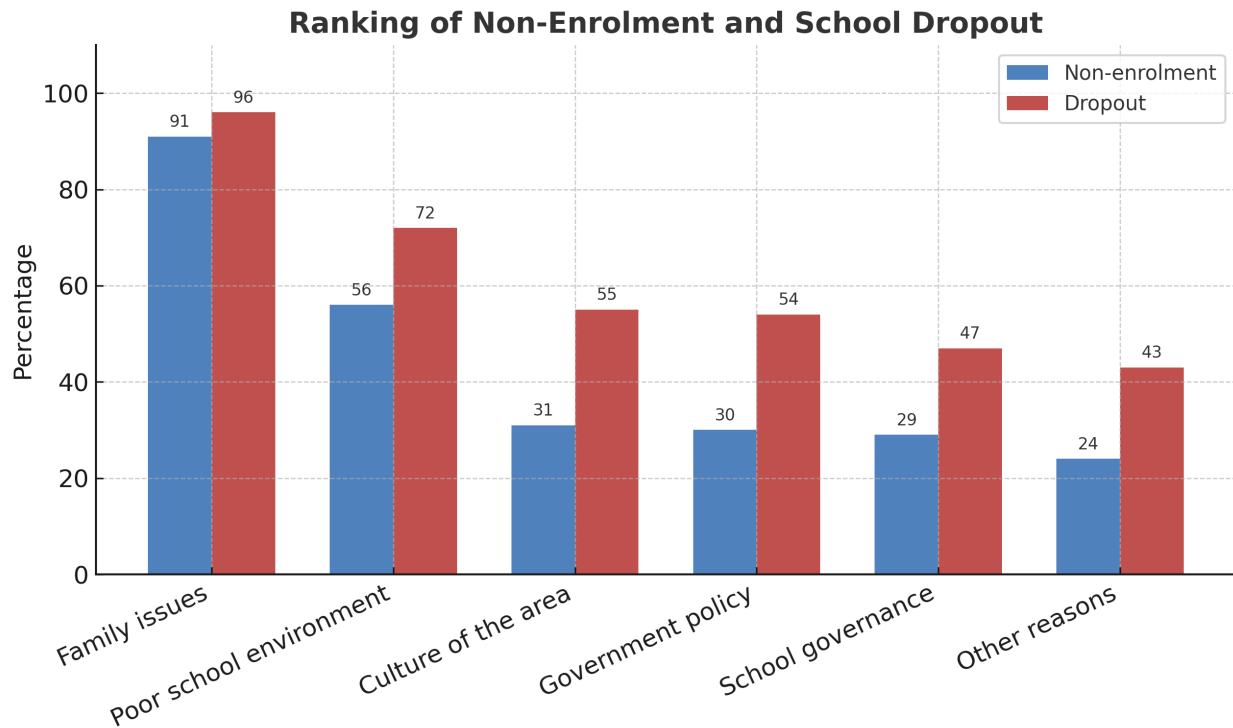


Figure 2.1: Source: 2013/14 Out of School survey

Research indicated that Uganda’s educational system faced significant challenges in retaining students, particularly during primary and secondary education (Ministry of Education and Sports, 2020; UNESCO, 2019). Notably, the overall dropout rate for primary school students was around 33.4%, and this trend continued into secondary education, with approximately 25.6% of students dropping out before completion (Ministry of Education and Sports, 2020). During the presentation of the preliminary findings at the MoES, the technical team validated and confirmed the findings, noting that they were consistent with previous studies. The pressing question was “What’s next?” if a significant number of children failed to enrol in school and many others abandoned their education before completing it. Despite the efforts made to address the issue, a deeper reflection was required to un-

derstand what had been effective and what hadn't, and to identify the underlying causes of success and failure, in order to inform future interventions. Moreover, the development of targeted recommendations was crucial in identifying effective ways to boost enrolment and retention rates, enabling more children to benefit from education and reach their full potential. On the other hand, the non-material factors such as negative attitudes towards education, discriminatory perceptions, and stigma associated with refugee status, cultural beliefs, and age of enrolment could also hinder access to education, as well as negative and neglectful attitudes towards children with disabilities, disillusionment, weak enforcement of education laws, community conflicts, and seasonal activities. For instance; In Tanzania, a student who was absent from school for a consecutive period of 90 days was considered to have dropped out and may have been expelled, as stated by the United Republic of Tanzania, 2015. It was important to note that absenteeism with consent, such as sick leave or authorized absences, was distinct from student dropout and should not have been used as a criterion for determining dropout rates. Despite the multitude of theories and hypotheses on dropout rates, however, cultural and economic factors were the most commonly cited and emphasized reasons in reports, as observed by Morara Chemwei, 2013; Trevor et al., 2018. Branson et al., 2014; Nielsen, 2016 study noted that many investigations into student dropout had identified a range of factors as contributing to the problem, but these findings were often lacking in scientific rigor and had not been translated into effective interventions.

2.0.3 Theories associated to dropouts

The following theories were relevant to dropout rates and had significant implications for learners' education.

Social cognitive theory

According to Bandura's (1986) socio-cognitive theory, students' dropout decisions were shaped by social modeling and cognitive processes such as self-efficacy, outcome expectations, and goal setting, which molded their motivation and academic engagement. For example, observing peers' struggles with coursework or witnessing teacher's supportive interactions could have fostered or hindered student's own academic persistence. Further, students who observed peers dropping out due to poor academic performance or lack of support often adopted similar behaviors. This demonstrated the need for positive role models and academic encouragement. Bandura noted that language difficulties impacted students' self-efficacy in a number of ways, specifically, language barriers restricted access to learning resources and support. Moreover, machine learning was used to identify at-risk students by assessing behavioral patterns, empowering targeted interventions.

Ecological systems theory

Bronfenbrenner's (1979) ecological systems theory indicated that students' dropout decisions were influenced by various environmental systems. For instance; in Uganda, factors like poverty (macrosystem), family dynamics (microsystem), and access to educational facilities (mesosystem) affected dropout rates. For example, Bronfenbrenner explained that personal problems like mental health issues or physical disabilities, interact with these systems like family, school, community and thus have an impact on individual's experiences and outcomes. By assessing these systems, machine learning enabled the identification of at-risk students and informed interventions that targeted particular environmental challenges.

2.0.4 Factors influencing learner dropout

FIGURE; PERCENTAGE OF CHILDREN OF PRIMARY SCHOOL AGE (AGES 6-12)
OUT OF SCHOOL

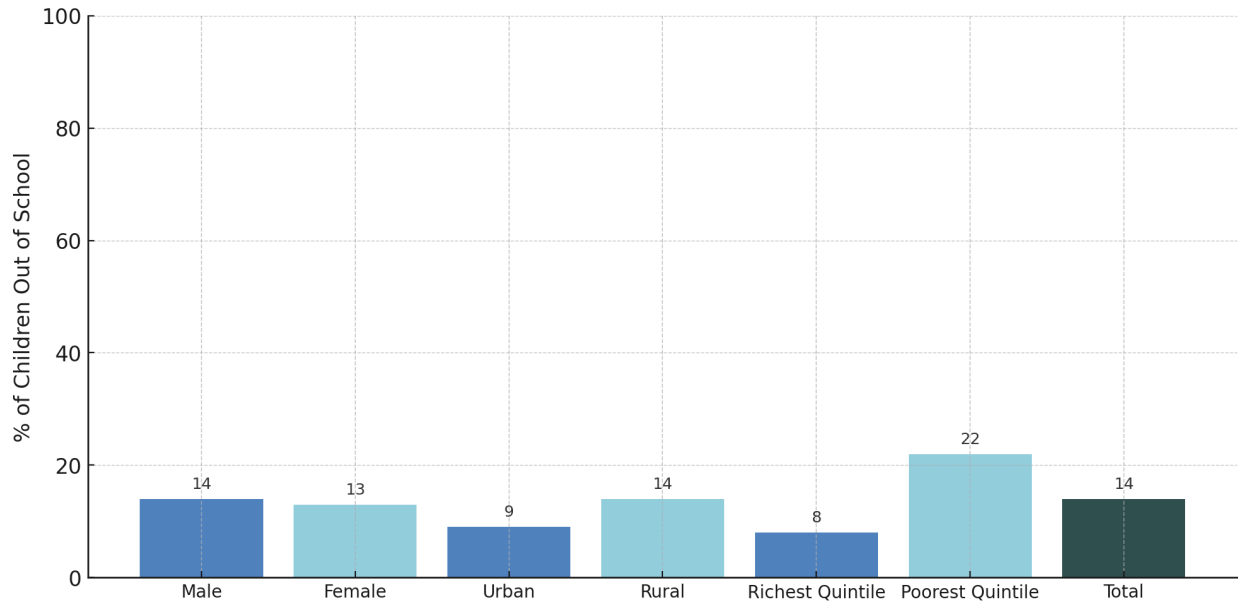


Figure 2.2: Data source: EPDC extraction of DHS dataset 2011

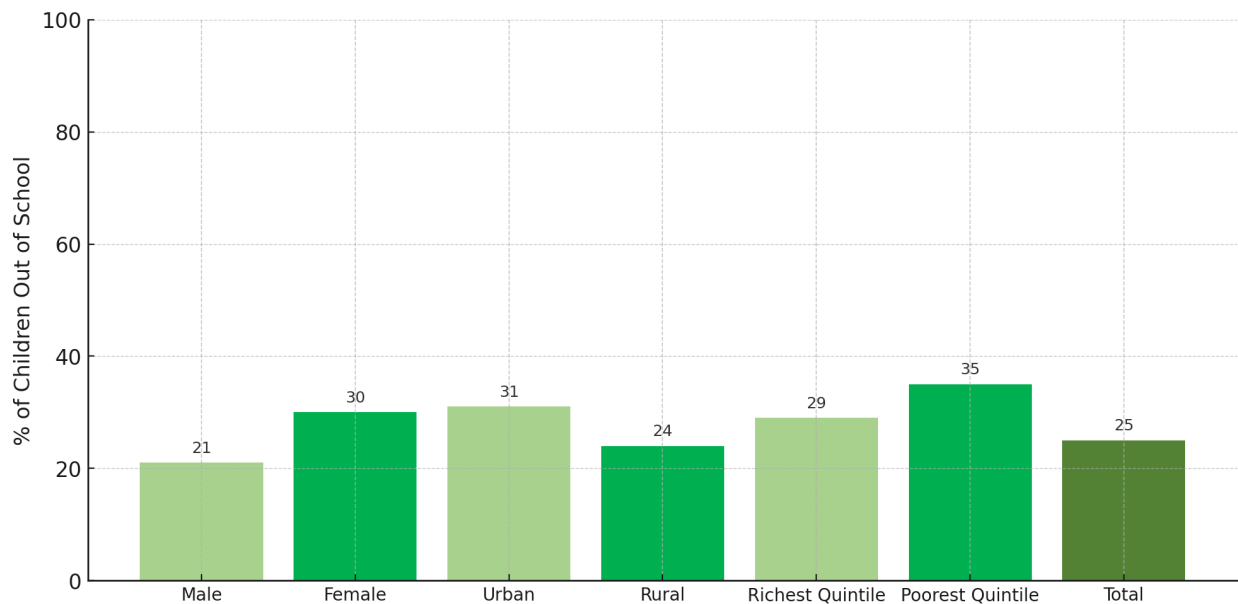


Figure 2.3: Data source: EPDC extraction of DHS dataset 2011

Figures; The percentage of out- of -school children in a country revealed what fraction of children were not actively engaged in the education system and were not actively engaged in the education system and were therefore deprived of the benefits of school. In Uganda,

14% of children of official primary school were out of school, as indicated in Figure 7. This figure also examined the share of out- of- school children by various factors where data was available. For instance, Figure 2.3 revealed that around 14% of boys and 13% of girls of primary school age were out of school. The most significant gap was observed between the poorest and the wealthiest children in Uganda. Figure 8 examined the percentage of youth of secondary school age who were out of school in Uganda. Nearly 30% of female youth of secondary school age were out of school, compared to 21% of male youth of the same age. For youth of secondary school age, the most notable difference was observed between male and female youth. Student dropout was a complex issue that was influenced by a multitude of factors, and it was essential to comprehend each factor's role and impact when examining the underlying root causes of dropout (Stempel et al., 2017). The primary characteristic of a factor that contributed to student dropout was the association between the factor and the dropout outcome. Identifying the most likely factors that contributed to student dropout was essential, but it was only the first step in addressing the issue, as a more comprehensive strategy was needed to effectively reduce dropout rates. According to Habibipour et al., 2018, the most prevalent method of identifying factors that contributed to student dropout relied on prior knowledge, which was often based on individual perceptions, experiences, and beliefs. Hailikari et al., 2007; Rittle-Johnson et al., 2009; Archambault et al., 2009; Chen, 2012 study noted that several research efforts had utilized traditional approaches to analyze dropout factors in education, before the emergence of machine learning in different parts of the globe. Prior knowledge was extremely valuable and effective in dealing with a limited set of existing factors, and there was no need for additional maneuvers or explorations (Hailikari et al., 2007). Using this approach in isolation could introduce bias and restrict the identification of potential non-linear factors that contributed to the problem, which may have been crucial to understanding the complex relationships and interactions involved. Despite the popularity of traditional methods in investigating student dropout, the use of Machine

Learning (ML) techniques in this context was still limited, with few research studies having applied these methods. A machine learning technique, such as variable ranking or permutation feature importance, was utilized to identify the most significant features contributing to the student dropout problem, using an embedded method that evaluated the importance of each feature (Bouaguel et al., 2015; Mduma et al., 2019; Ma et al., 2017). (Sun et al., 2017) suggested that the simplicity, scalability, and good empirical success of these techniques had made them a popular choice among researchers and practitioners. According to Mgala, 2016; Lakkaraju et al., 2015, the successful implementation of machine learning approaches (ML) was hindered by a lack of basic computing knowledge among stakeholders in developing countries, particularly in the education sector, which highlighted the need for simple and user-friendly models. These factors encompassed economic challenges and poverty (Kasozi, 2013; Ssewamala 2011), limited access to educational resources and infrastructure (Ochen, 2015; Oketch, 2007), poor academic outcomes and motivation (Kasozi, 2013; Mugeru, 2015), and cultural influences, including early marriage and pregnancy (Sewamala, 2011; Tenywa, 2015). Based on the previous discussion, one could conclude that school dropout inevitably compromised learners' future opportunities, limiting their potential for personal and professional growth.

Disability as a cause of school dropout

According to key informants' interviews, disability was a significant factor contributing to school dropouts, as children with disabilities often faced segregation and exclusion by both teachers and peers. Many caregivers saw children with disabilities as a responsibility they would rather not bear and chose to keep them at home, believing that sending them to school would be stressful and pointless. The inaccessibility of schools due to the absence of necessary facilities posed a significant challenge for children with disabilities, who struggled with mobility issues and were often unable to participate in educational activities. Despite

being enrolled, students with disabilities lacked necessary gadgets and resources to facilitate their learning. The school environment was often inadequate, with poor sanitation and insufficient materials. For instance, a blind teacher lacked the required books to prepare lesson plans. Sebatane (2009) considered disability one of the contributing factors to primary school dropout. The lack of initial access to school for learners with disabilities and special educational needs (SEN) meant most of them were able to drop out of school. Educators' inadequate training and resources hindered their ability to provide effective support. UNESCO (2011) found that learners with disabilities in Uganda faced significant barriers, including inaccessible schools, long distances, and inadequate facilities. Educators faced challenges such as overcrowded classes, inadequate equipment, and poor working conditions. The distant location of some schools in Uganda created barriers for learners with disabilities, leading to fatigue, difficulty commuting, and ultimately, dropout.

2.3.2 Repetition as a cause of school dropout

Consistently poor academic performance led to frustration, causing many students to abandon their education. They may have joined peer groups that engaged in unhealthy behaviors, such as drinking or criminal activity. Students who performed poorly often felt embarrassed, particularly if they were required to repeat a grade, leading to feelings of not belonging and dropout. A negative school environment led to disengagement and dropout among students. Some teachers' unacceptable behavior, including harassment and excessive punishment, created a toxic atmosphere. Teachers who demonstrated a lack of commitment to their duties, such as poor teaching or chronic absenteeism, eroded students' trust and enthusiasm for learning. Starting school at a tender age was a significant factor leading to school dropouts. Children who began their education at 4 years old often struggled to continue beyond the age of 7. The lack of nursery schools in many villages meant that children were forced to start formal education too early, leading to difficulties in adjusting to the daily routine of school.

FIG ; STUDENT REPETITION BY GRADE AND LEVEL

IN PRIMARY SCHOOL

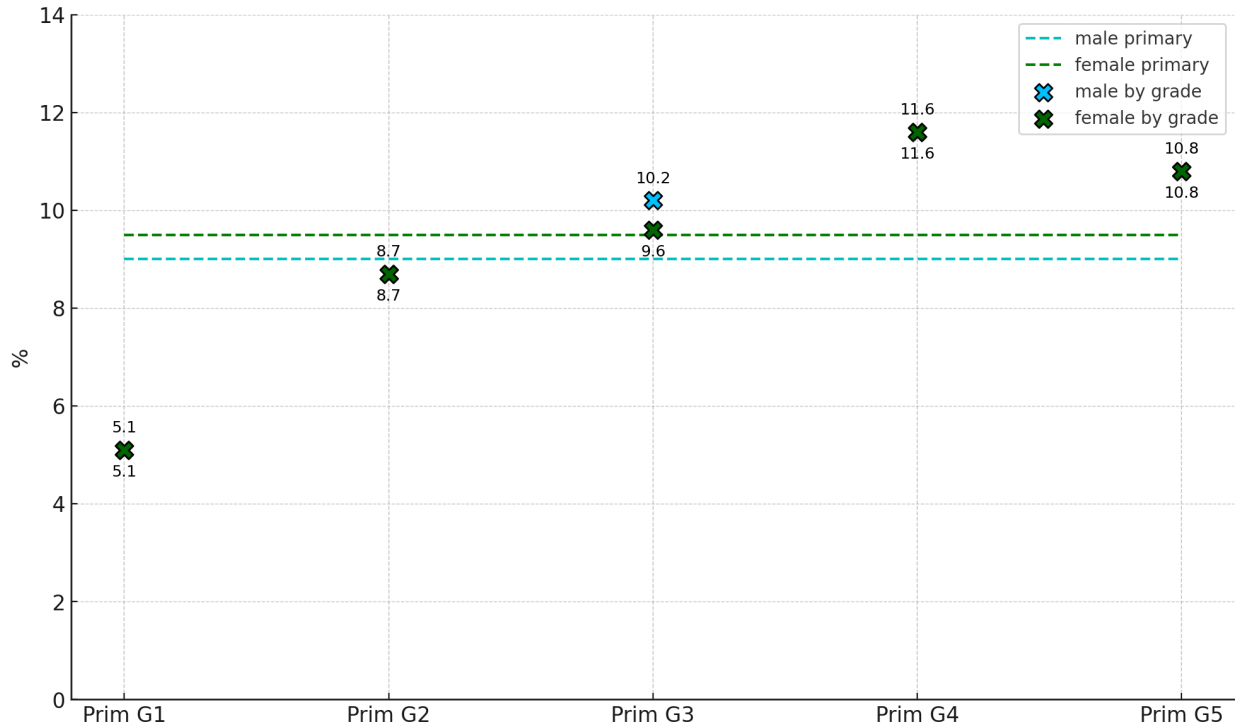


Figure 2.4: EData source: EPDC calculation based on UIS data

School environment as a cause of school dropout

The unfavorable school environment was a significant factor contributing to school dropouts, as some children found it difficult to continue their education due to unapproachable and abusive teachers. The behavior of some teachers, including physical and emotional abuse, such as harassing girls by touching their breasts, was particularly damaging. Others were neglectful in their duties, failing to teach effectively or even show up to class, ultimately discouraging children from attending school.

Early marriage as a cause of school dropout

Marriage and societal expectations were a major obstacle to education, as many girls were forced to drop out of school due to pressure from parents to get married. Some children,

feeling grown up and ready for marriage, prioritized it over education, considering school less important and dropping out to pursue their future as married individuals. Peer group influence and personal choice also played a role in early marriages, leading to early motherhood and single parenthood. Unfortunately, some parents' negative attitudes towards education exacerbated this issue, as they preferred their children to stay home and wait for marriage rather than attend school. For instance, research showed that in some communities, girls were viewed as a means to financial gain, with parents seeking to marry them off in exchange for wealth (Girls Not Grey, 2019).

Menstruation as a cause of school dropout

Menstruation was a significant barrier to education for girls in rural villages, as they often struggled to manage their periods without access to basic necessities like pads. This led to frequent absences from school, and in some cases, girls even dropped out, as the inconvenience and embarrassment of missing classes every month became too much to bear. For example, some parents were unable to raise 3000 shillings to buy their daughters pads on a monthly basis.

Poor law enforcement as a cause of school dropout

The weakness of law enforcement in Uganda created a situation where children could drop out of school without any repercussions. The lack of effective laws and their enforcement meant that children had the freedom to choose whether or not to attend school, and there was no one to question them about their absences. Moreover, the authorities responsible for implementing these laws were often hesitant to act, and when they did, they may have been influenced by external factors, such as phone calls from powerful individuals, which could lead to the release of parents who had been arrested for not sending their children to school. This lack of accountability and the presence of good laws that were not enforced due to a

lack of political will contributed to the normalization of dropping out of school.

Orphan-hood as a cause of school dropout

The loss of parents had a devastating impact on a child's education, as orphans often struggled to access the support and resources they needed to succeed in school. Without a caring guardian, orphans were left to fend for themselves, and even when they had someone looking after them, this person may not have prioritized their educational well-being. In some cases, orphans were subjected to abuse and exploitation, and were forced to abandon their education and get married at a young age.

Parental perception as a cause of school dropout

The negative attitude of some parents towards education was a significant factor in school dropouts. In many parents' villages, parents were not aware of the importance of education and may have viewed it as a luxury they could not afford. As a result, they may not have motivated their children to attend school, and if a child was not self-motivated, they may have automatically dropped out of school. Additionally, some parents who had not personally benefited from education may have discouraged their children from pursuing it, as they did not see the value in it and would rather have had them focus on making quick money. This lack of support and motivation had a long-term impact on a child's future and may have limited their opportunities in life.

Domestic violence as a cause of school dropout

Domestic violence had a devastating impact on families, leading to the separation of parents and leaving children without the support and care they needed. As a result, children were forced to take on additional responsibilities, such as caring for their siblings, which made it

difficult for them to attend school. This led to children dropping out of school, as they were unable to balance their home and school duties.

Poverty as cause of school dropout

Poverty was indeed a significant challenge that hindered access to education and led to school dropouts. In Uganda, poverty was a pervasive issue that affected many households, particularly in rural areas. The country had made significant progress in increasing access to education, but poverty remained a major obstacle to achieving universal education. The harsh reality was that many people were living in absolute poverty, which meant they were unable to afford even the most basic necessities, including school materials, making it challenging for their children to attend school and receive a proper education. As a result, the children often returned to their parents, who were struggling financially, and in an effort to support themselves, they may have attempted to find ways to earn money, leading to a decline in their interest in education and ultimately, dropping out of school. Below were a few voices from the respondents concerning poverty. The country being placed at 134th out of 138 countries in the 2018 Global Innovation Index (GII) stipulated a need for improvement in education quality (GII, 2018). According to Kakooza et al. (2020), the poor learning outcomes could be accredited to inadequate teacher training and support, leading to high teacher absenteeism and turnover rates, which consecutively influenced the quality of direction and student achievement. Additionally, the curriculum for primary education had been revised to a thematic curriculum. Because of the slow progress of the work to amend the syllabus for teacher training to match it, there was solicitude regarding the poor application of this thematic curriculum and the teaching frontline. Regarding secondary education, analysis and alteration amendment of the curriculum were in progress at that time. Furthermore, the number of teachers had increased after the introduction of UPE, and the number of enrolled pupils per teacher in primary education in 2010 stood at 57 for public

schools, which was far greater than the 26 for private schools. PER PUPIL EXPENDITURE (PPE) BY SCHOOL LEVEL (% OF GDP PER CAPITA)

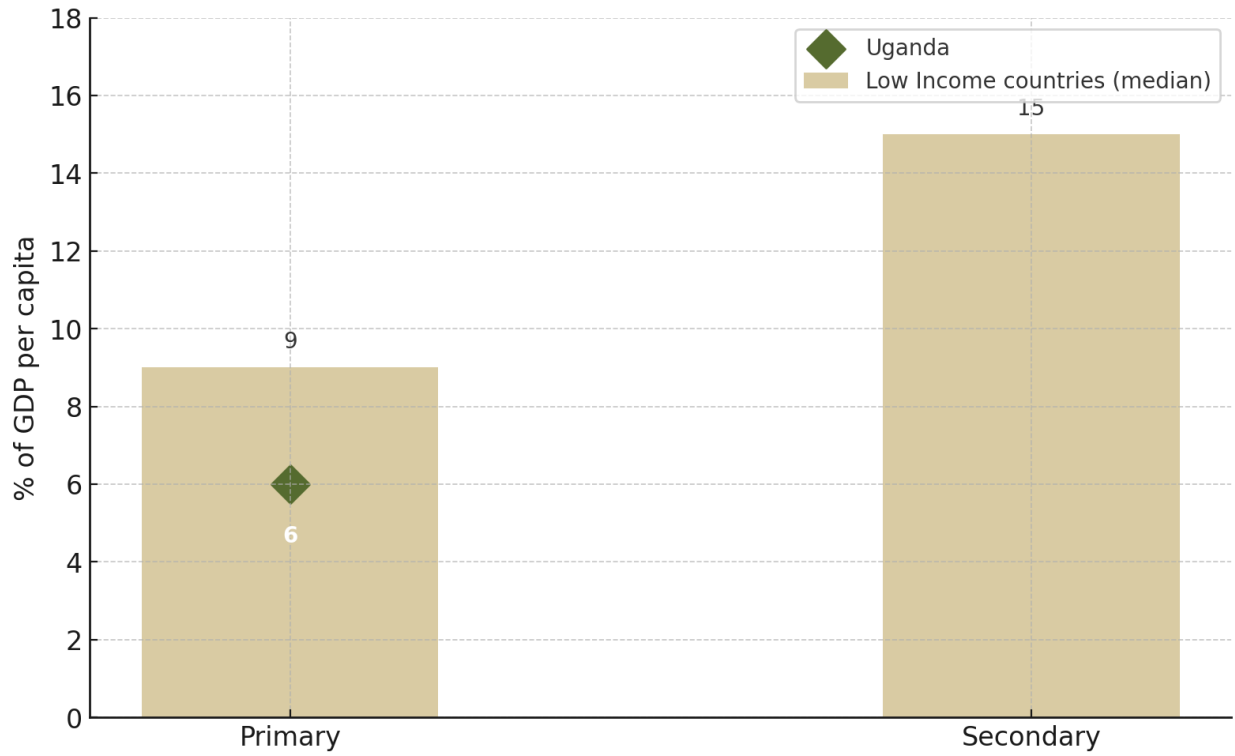


Figure 2.5: Data source: UNESCO Institute for Statistics (UIS)

PUPIL TEACHER RATIO (PTR) BY SCHOOL LEVEL

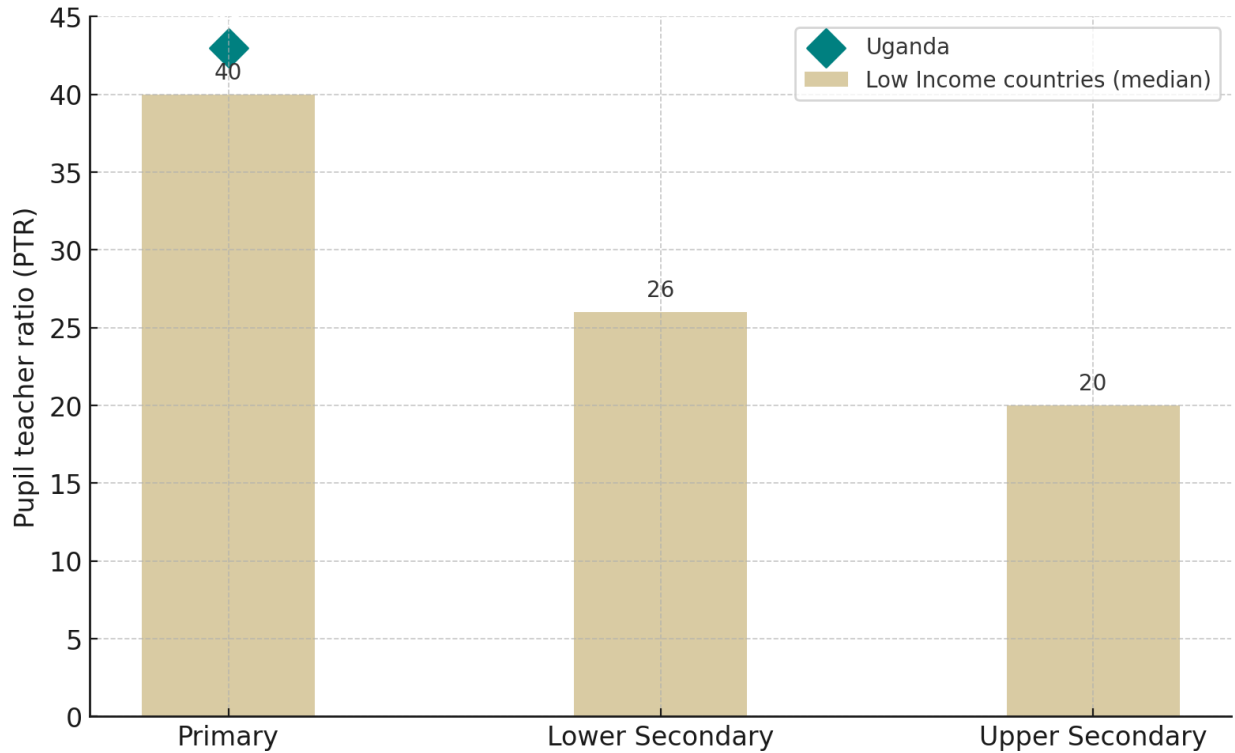


Figure 2.6: Data source: UNESCO Institute for Statistics (UIS)

Figure: Education Expenditure The above figures contrasted Uganda's per pupil expenditure (PPE) and pupil teacher ratio (PTR), with those of other low-income countries. PPE reflected a country's commitment to education at each school level. In Uganda, PPE in primary education was 6% of GDP per capita, lower than the median PPE of 9% for low-income countries. PTR served as an indicator of learning quality and resource availability. In Uganda, the PTR in primary education was 42.7, meaning one teacher for every 42.7 primary school students. This was higher than the median PTR of 40 for low-income countries.

Food as a cause of school dropout

Furthermore, some children were forced to drop out of school due to hunger and lack of food, as they often went to school on an empty stomach and returned home to find no food available, making it seem more practical to stay at home and search for something to eat

rather than attend school and suffer from hunger. A teacher observed that hunger was a significant obstacle to learning, as it impaired students' ability to concentrate and retain information, leading to a cycle of failure, frustration, and disillusionment with the education system, ultimately causing them to drop out.

Child labour as a cause of school dropout

Child labour was a major contributor to school dropout in Uganda, as many students, particularly those above the age of 10, opted to work and earn money instead of attending school. In all regions of Uganda, child labour was a significant challenge, where children, including those as young as 7, were involved in activities such as farming, quarrying, and commerce to support themselves. As a result, many of these children tended to abandon their education, perceiving it as unnecessary, which contributed to the high rate of school dropouts.

Distance as a cause of school dropout

In rural areas, the vast distances between schools and residential clusters were a major obstacle to education, as children had to walk up to 10 kms to attend school, which was often exhausting and demotivating. This led to tardiness and subsequent punishments, which discouraged many children from attending school, as the distance from their homes to school was a significant deterrent.

HIV and Aids pandemic

According to UNICEF's 2007 report, the HIV/AIDS pandemic undoubtedly exacerbated the socio-economic challenges faced by extremely poor families and children, making them even more vulnerable to poverty and exploitation. Participant 6 pointed out that growing up in a world with HIV/AIDS was a significant obstacle for many learners, who may have

experienced emotional pain and trauma in their lives, leading learners who were struggling to do well at school to often decide to quit. Unfortunately, many learners were frequently forced to leave school to care for their ill parents, siblings, or relatives, due to the widespread impact of HIV/AIDS, which disrupted their education and had long-term consequences on their future.

Use of technology

The limited use of technology in education was also a significant challenge, as noted by Nambwire et al. (2019), who highlighted that the effective integration of technology in education could have improved student learning outcomes and elevated access to quality education. Many schools in Uganda lacked the required infrastructure and resources to fruitfully utilize technology, which hindered effective teaching and learning. According to previous research, a mixed approach, including the use of machine learning and technology, was supposed to improve student outcomes and address the various issues that affected Uganda's education sector.

Unemployment

Despite efforts to eliminate obstacles to education, numerous complex and multifaceted factors still hindered learners' attendance in school. Socio-economic challenges, including rising poverty and unemployment, were increasingly prevalent. As reported by the Ministry of Education and Training (2008) and the International Labor Organization (ILO, 2011:43), some children, particularly boys, were denied access to education, while girls were often forced into domestic work due to poverty in their households, highlighting the persistence of barriers to education. Table showing the tabulation of common causes of non-enrolment and school dropout by region and refugee status.

Reason	Ranking	Central	Eastern	Karamoja	Northern	Western	Non-refugee	Refugee	Total
Level of poverty in the community and some households	91.4	91.4	81.9	99.7	96.6	92.4	90.9	94.7	91.4
Family issues	56.4	58.4	57.6	46.7	48.4	69.2	56.4	56.3	56.4
Poor school environment	30.6	44.9	35.6	21.4	24.1	20.1	28.4	43.7	30.6
Culture of the area	29.7	48.0	27.4	51.2	15.4	16.3	28.3	30.8	29.7
Government policy on education	28.9	46.7	30.2	16.1	19.2	25.7	26.9	40.8	28.9
School governance	28.7	45.2	38.1	15.8	13.7	23.2	27.0	38.4	28.7
Other reasons	23.6	27.4	33.2	6.0	21.5	19.2	24.2	20.3	23.6

Causes of dropout

Reason	Ranking	Central	Eastern	Karamoja	Northern	Western	Non-refugee	Refugee	Total
Level of poverty in the community and some households	96.1	97.9	88.5	100.0	100.0	96.8	95.7	98.6	96.1
Family issues	71.8	84.0	63.6	73.3	65.5	78.3	71.8	71.9	71.8
Other reasons	54.5	84.4	63.3	19.4	52.0	36.6	57.1	41.3	54.5
Poor school environment	54.4	87.7	45.6	33.6	63.5	37.0	51.4	71.2	54.4
School governance	46.5	83.9	47.4	21.7	38.8	30.8	44.5	56.9	46.5
Culture of the area	42.7	81.1	32.8	73.2	18.4	21.4	41.6	48.5	42.7
Government policy on education	42.6	79.9	36.2	25.1	39.0	29.1	41.3	49.9	42.6

Figure 2.7: Source: 2013/14 Out of School survey

2.0.5 Data for model development

Machine learning relied heavily on large datasets to build and refine its models (Mullainathan & Spiess, 2017). Achieving accurate machine learning model performance required balanced data (Krawczyk, 2016; López et al., 2013). In real-world data, it was common for one class

to be under-represented (minority) while the other (majority) dominated the dataset (Abdi & Hashemi, 2014; Borowska & Topczewska, 2016; Galar et al., 2016; Krawczyk, 2016; Lin & Chen, 2012; Mazumder et al., 2015). In the education sector, the issue of class imbalance was becoming increasingly recognized in student dropout prediction, where the number of students who remained in school far exceeded those who dropped out (Thammasiri et al., 2014). According to Gao (2015), the imbalanced ratio was typically around 1:10, with the minority class often being the target group, as also noted by López et al. (2013). Various methods, including algorithmic modification and cost-sensitive learning, were employed to evaluate and address the issue of class imbalance in handwriting recognition (Elhassan et al., 2016; Hoens & Chawla, 2013). Algorithmic modification involved changing the learning algorithm to accommodate imbalanced data, operating at the algorithmic level (Elhassan et al., 2016), whereas cost-sensitive learning considered the learning process to minimize them (Shilbayeh, 2015). Several studies reported the constraints of conventional approaches in dealing with highly skewed data, including student dropout datasets, in terms of cost and time, and recommended the use of alternative methods or data reduction strategies for the effective learning process (Márquez-Vera et al., 2016; Neill et al., 2020; Weiss et al., 2007). However, reducing data sets resulted in the loss of important samples, which negatively impacted the learning process, and therefore, it was recommended to use alternative approaches for handling highly skewed data sets.

2.0.6 Machine learning approaches and student dropout

Review for Supervised Machine Learning Used to Predict and Classify School Dropout Rates. The prediction of student dropout using random forest algorithms was a topic of considerable interest in educational research, with Amare, M.Y et al ,2021; Oqaidi, K et al,2022 being one of the studies that investigated the effectiveness of this approach in identifying students who were at risk of dropping out. The versatility of random forest was evident

in its deployment across various educational settings, from higher education institutions to online courses, where it showed considerable promise in driving positive outcomes [do Carmo Nicoletti et al,2020–Menoyo-Ros et al,2020]. Random forest was found to effectively forecast student dropout by utilizing data from online activities and academic performance, providing valuable insights for targeted interventions as noted by [do Carmo Nicoletti et al,2020, ER, E,2023]. Furthermore, research indicated that random forest, a machine learning technique, was such effective in the early recognition of students at risk of dropout, providing a valuable tool for immediate intervention and support [Albán, M.; Mauricio,2018; Mnyawami, Y.N,2022]. Random forest models were found to provide actionable insights into the factors that drove individual predictions, enabling educators to develop data-driven strategies to support at-risk students, as noted by [Baranyi, M.; Nagy, M,2020]. The application of random forest to predict dropouts in various educational settings, including primary and secondary schools, underscored the vast scope and relevance of this research area, as suggested by Kloft, M.; Stiehler,2014; Mduma, N,2019. The capabilities of random forest extended beyond predicting dropouts, as it also helped to improve student retention rates by providing early warning systems that enabled educators to take proactive measures to support struggling students [Lee, S.; Chung,2019]. The fusion of random forest algorithms with data mining techniques was found to improve the accuracy and efficiency of dropout forecast models, as noted by Palacios, C. A, 2021. This underscored the rapidly evolving nature of this field, which had proportionate impacts for educational institutions. This model employed a data-driven approach to identify students who were at risk of dropping out, allowing for prompt intervention and support to enhance student outcomes and retention rates. The dropout rates in Uganda, which stood at 45% and 30% of students respectively, who failed to complete their education, emphasized the need for accurate forecasting to identify early warning signs and prevent dropouts. This initiative helped identify risk factors, implement targeted interventions, allocate resources effectively, develop early warning systems, and de-

sign programs to support at-risk students, thereby promoting a more inclusive and equitable education system in Uganda, as suggested by previous research. As outlined in Table 3, numerous research studies have employed ML algorithms to predict student dropout rates, offering a detailed comprehension of the applications and efficacy of these predictive models. For example, [Kiss, V.; Maldonado, E,2022] utilized LR to facilitate an accuracy of 84.8%, while [Capuano, N et al,2023] used RF, achieving an accuracy of 80%. This comparison highlights the efficacy of various ML models in predicting dropout rates of student, demonstrating their versatility and applicability across different educational contexts.

Figure 2.8: Summary of Machine Learning Studies on Student Dropout Prediction

No	Author	Title	Algorithm Used	Chosen Algorithm	Result
1	[21]	The Use of Semester Course Data for Machine Learning Prediction of College Dropout Rates	Neural networks (NNs), decision tree (DT), logistic regression (LR), support vector machine (SVM)	Logistic regression (LR)	Accuracy: 84.8
2	[23]	Perspectives to Predict Dropout in University Students with Machine Learning	Neural networks (NNs), random forest (RF), logistic regression (LR), support vector machine (SVM)	Random forest (RF)	Accuracy: 91%

3	[24]	Using Machine Learning Techniques to Predict Learner Drop-out Rate in Higher Educational Institutions	Multilayer perceptron, decision tree (DT), random forest (RF), support vector machine (SVM)	Random forest (RF)	Accuracy: 70.98
4	[25]	Prediction of Student Dropout in E-Learning Program Through the Use of Machine Learning Method.	Artificial neural network (ANN), decision tree (DT), Bayesian networks (BNs)	Decision tree (DT)	Accuracy: 63.89%
5	[19]	The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction	Random forest (RF), boosted decision tree (BDT), random forest with SMOTE (SMOTE + RF), boosted decision tree with SMOTE (SMOTE + BDT)	Boosted decision tree (BDT)	Accuracy: 89.8%
6	-	Towards Predicting Student's Dropout in Higher Education Using Supervised	Random forest, logistic regression, neural network	Random forest	Accuracy: 95.00%

Figure 2.9: Source: 2013/14 Out of School survey

		Machine Learning Techniques			
7	[27]	Supervised Machine Learning Algorithms for Predicting Student Dropout and Academic Success: A Comparative Study	Decision tree, SVM, random forest, gradient boosting, XGBoost, CatBoost, LightGBM	LightGBM, CatBoost	F1 score: 86%
8	[28]	Student Dropout Prediction for University with High Precision and Recall	Logistic regression, decision tree, random forest, XGBoost	XGBoost	The XGBoost model achieved the highest precision (0.92) and recall (0.91) in predicting student dropout.

Figure 2.10: Source: 2013/14 Out of School survey

9	[29]	Development of Prediction Model to Improve Dropout of Cyber University	Logistic regression, decision tree, random forest	Random forest	The random forest model achieved the highest accuracy (87.4%) in predicting student dropout in a cyber university setting
10	[30]	Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention	Deep neural network	Deep neural network	The proposed deep learning model achieved an AUC of 0.82 in predicting MOOC dropout, outperforming traditional machine learning algorithms.

Figure 2.11: Source: 2013/14 Out of School survey

Authors:1-10(Kiss, V.; Maldonado, E.; Segall, M. The Use of Semester Course Data for Machine Learning Prediction of College Dropout Rates. *J. High. Educ. Theory Pract.* 2022, 22, 64–74, Solis, M.; Moreira, T.; Gonzalez, R.; Fernandez, T.; Hernandez, M. Perspectives to Predict Dropout in University Students with Machine Learning. In *Proceedings of the 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, San Carlos, Costa Rica, 18–20 July 2018; IEEE: Piscateville, NJ, USA, 2018; pp. 1–6, Dake, D.K.; Buabeng-Andoh, C. Using Machine Learning Techniques to Predict Learner Drop-out Rate in Higher Educational Institutions. *Mob. Inf. Syst.* 2022, Tan, M.; Shao, P. Prediction of Student Dropout in E-Learning Program Through the Use of Machine Learning Method. *Int.*

J. Emerg. Technol. Learn. 2015, 10, 11–17, Lee, S. *Appl. Sci.* 2019, 9, 3093, Capuno, R.M.M.; Ferrer, C.J.M.; Manaloto, B.T.L.; Villafria, S.R.; Esquivel, J.A. Towards Predicting Student’s Dropout in Higher Education Using Supervised Machine Learning Techniques. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Manila, Philippines, 7–9 March 2023, Villar, A.; de Andrade, C.R.V. Supervised Machine Learning Algorithms for Predicting Student Dropout and Academic Success: A Comparative Study. *Discov. Artif. Intell.* 2024, 4, 2, Kim, S.; Choi, E.; Jun, Y.-K.; Lee, S. Student Dropout Prediction for University with High Precision and Recall. *Appl. Sci.* 2023, 13, 6275, Park, C. Development of Prediction Model to Improve Dropout of Cyber University. *J. Korea Acad. Ind. Coop. Soc.* 2020, 21, 380–390, and Xing, W.; Du, D. Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *J. Educ. Comput. Res.* 2019, 57, 547–570.) This review demonstrates the growing use of supervised ML models such as Random Forest, Decision Tree, and Logistic Regression in educational contexts. However, the literature is limited in terms of applying these models to Ugandan data or combining them with data balancing techniques to improve predictive accuracy in highly imbalanced dropout datasets. This study seeks to fill that gap by testing multiple algorithms on local data, thereby aligning with the study’s aim to develop accurate, context-sensitive early warning systems.

2.0.7 Evaluation measures for students dropout

According to Kumar et al., 2017, the evaluation criteria are an essential consideration when assessing the performance of machine learning models. The selection of an appropriate measure highly depends on the nature of the problem, which can be either classification or regression, and the properties of dataset, including balance or imbalance. Several research studies have utilized different evaluation metrics to address the problem of student dropouts (Table 5), with, Accuracy, Area Under the Curve (AUC), Mean Squared Error (MSE) among

others (Santana et al., 2015; Xu et al., 2017; Johnson et al., 2015). According to Ameri et al., 2016; Lakkaraju et al., 2015; Rovira et al., 2017, accuracy is a statistical measure that quantifies the degree of correctness, representing the ability to provide precise predictive results when the number of samples belonging to each class are equally distributed. In contrast, the Area Under the Curve (AUC) metric is employed in binary classification tasks to evaluate the probability of a model correctly ranking a positive sample above a negative sample, chosen at random (Fei Yeung, 2015; Halland et al., 2015; Liang et al., 2016; Prieto et al., 2017). On the other hand, the Mean Squared Error (MSE) is primarily utilized in the regression process, where it can be easily calculated by taking the average of the squared differences between the actual values and the predicted values, allowing for the computation of the gradient (Elbadrawy et al., 2016; Iam-On Boongoen, 2017). Despite their usefulness, these metrics have been found to have limitations, including the potential for misinterpretations and biases against minority classes, which is why it is crucial to use multiple metrics when evaluating the performance of ML models (Liang et al., 2016; Lin Chen, 2012; López et al., 2013; Longadge et al., 2013).

Figure 2.12: Summary of various metrics used in model evaluation

Source	Problem	Metrics
M´arquez-Vera et al. (2016)	Mining best rule to predict student dropout	Geometric Mean
Liang et al. (2016)	Developing dropout predictive model	Area Under the Curve (AUC)
Poh and Smythe (2015)	Predicting student performance	The error residuals
Fei and Yeung (2015)	Predicting dropout in Massive Open Online Courses (MOOCs)	Area Under the Curve (AUC)
Hung et al. (2017)	Identifying at-risk students in online program	Accuracy and misclassification rates
Johnson et al. (2015)	Identifying students at risk of not graduating on time	Precision
Xu et al. (2017)	Tracking and predicting student performance	Mean square errors
Santana et al. (2015)	Identifying students with dropout profiles	Accuracy and a false positive rate

2.0.8 *The role of machine learning in dropout prediction*

Machine learning algorithms have demonstrated their effectiveness in predicting student dropout across different educational settings. These algorithms can analyze large datasets to identify patterns and relationships, enabling educators to pinpoint at-risk students and develop targeted interventions (Romero, 2013; Xing, 2015). Similarly, researchers in Uganda have utilized machine learning algorithms, such as decision trees, random forests, and neural networks, to predict dropout rates and identify key factors influencing student dropout (Mwanga, 2020).

2.0.9 Challenges and limitations

While machine learning and data-driven approaches hold promise for addressing dropout rates in Uganda, several challenges must be acknowledged. These include limited access to quality data, inadequate infrastructure, and concerns around data privacy and security. Additionally, successful implementation requires a nuanced understanding of the social, cultural, and economic contexts in which they operate. (Mwanga, 2020; Nkuyahaga, 2020).

2.0.10 Contribution of this study to ML and data-driven approaches in Education

This study contributes to the field of machine learning and data-driven education in several key ways:

- Application of ensemble learning in a low-resource context. While Random Forest is commonly applied in high-income countries, few studies have deployed and validated it in Ugandan educational datasets with high class imbalance. This research provides empirical evidence of RF's robustness in such contexts.
- Addressing class imbalance with SMOTETomek. Dropout datasets are often skewed (few dropout cases). This addresses that issue using SMOTETomek, an advanced hybrid resampling technique. Its application in Ugandan education data is novel and improves model fairness.
- Feature importance as a prioritization tool. This study uses feature importance from Random Forest to rank risk factors, enabling stakeholders to focus interventions on high impact variables. This contributes a data-driven prioritization framework for education planners.
- Integration of ML with Human Centered theories. Most ML studies ignore sociological

frameworks. This research anchors its data driven model in Social Cognitive Theory (SCT) and Ecological Systems Theory (EST), creating a bridge between predictive analysis and behavioral understanding.

- Actionable Insights for policy. Rather than stopping on prediction, this work demonstrates how ML outputs can guide tailored interventions like school feeding programs, mentorship and financial support thus contributing to applied machine learning for policy design.

2.0.11 Conclusion

This literature review has highlighted the persistent issue of student dropout in Uganda and the emerging potential of machine learning to address it. While several international studies have applied ML algorithms for dropout prediction, few have contextualized these tools for use in low-resource settings like Uganda. Moreover, most existing research fails to integrate ML with behavioral or ecological theories, limiting their relevance for local education policy. This study fills these gaps by;

In doing so, this research extends existing knowledge by offering a contextualized, theory-informed, and practically applicable approach to combating student dropout in Uganda

- ‘Applying Random Forest and data balancing (SMOTETomek) to Ugandan educational data’
- ‘Identifying the most significant dropout predictors through feature importance’
- ‘Integrating SCT and EST frameworks with predictive analytics to explain student behavior’
- ‘Translating predictive insights into actionable policy interventions’

CHAPTER 3

METHODOLOGY

3.0.1 Introduction

The previous chapter laid the theoretical framework for the research study by examining the theological framework that underpins it, providing a context for understanding the research questions and objectives. The methodology of the study is presented in this chapter and it covers the following: This chapter outlines the research methodology and approach utilized in this study, providing a clear and detailed explanation of the design, procedures, and techniques employed to conduct the study. The procedures for data collection, including the development and use of data collection instruments, are described in detail, along with the reasons for their selection, and the approaches used to analyze the data, validate the findings, and ensure reliability and ethical soundness of the research. By using a combination of quantitative and qualitative research methods, this study aimed to gather a rich and nuanced understanding of the factors that contribute to learner dropout rates, with the goal of developing effective solutions to address this issue.

Flow of research methodology

Research methodology design (mixed approach)

- Pragmatism (positivism & constructivism)
- Comprehensive analysis (predict outcomes & explore phenomena)
- Both statistical and thematic
- Both structured & unstructured (surveys & interviews)
- Combine scores with experiences



Data collection strategies

Focus group interviews

Questionnaires

Surveys

Observations

Field notes



Data analysis

Trustworthiness

Credibility

Authenticity

Transferability

Dependability

Confirmability



Ethical considerations

Protection from harm to participants

Right to privacy

Informed consent

Confidentiality of data

3.0.2 Research approach

To gauge the methodical attributes of incorporated studies, we used the Mixed Methods Appraisal Tool (MMAT) (Pluye et al., 2011). This tool evaluates the quality of both qualitative and quantitative studies, contemplating basis namely study design, sampling strategy, data gathering, and mining (Creswell, 2014). Two analysts separately estimated study quality, determining inconsistencies through agreement. The MMAT tool estimated study standard based on four fields: study design, data collection, data analysis, and findings (Pluye et al., 2011). Using a mixed-methods approach that uses both quantitative and qualitative research design, this study investigated the factors causing dropout rates and study outcomes in Uganda, with a focus on the application of machine learning and data analysis to identify trends and correlations. By leveraging the capabilities of Python programming and its related libraries, predictive models were created to detect students who are at risk of dropping out, allowing for the implementation of targeted interventions and support strategies to improve their educational outcomes. Methodology encompasses the science and art of arranging research activities in a logical and orderly manner, as described by Tsomo 2012:51. Rodolo 2008:15 as a thorough examination of the research process, including its various components and intricacies. The study also used linear regression analysis to investigate the relationships between variables like socioeconomic status, academic performance and dropout rates, aiming to uncover insights that can inform evidence-based interventions to improve student success and reduce dropout rates. This research methodology used in this study involves a range of techniques and methods, and this includes a critical examination of the rationale underlying their use, an acknowledgement of the limitations and potential drawbacks of each technique, and a recognition of the role of assumptions, presumptions, and methodological preferences in shaping data analysis and interpretation of findings. To address the research questions of this study, I chose to employ a qualitative research approach, which involves gathering and analyzing data in a subjective and interpretive manner, relying

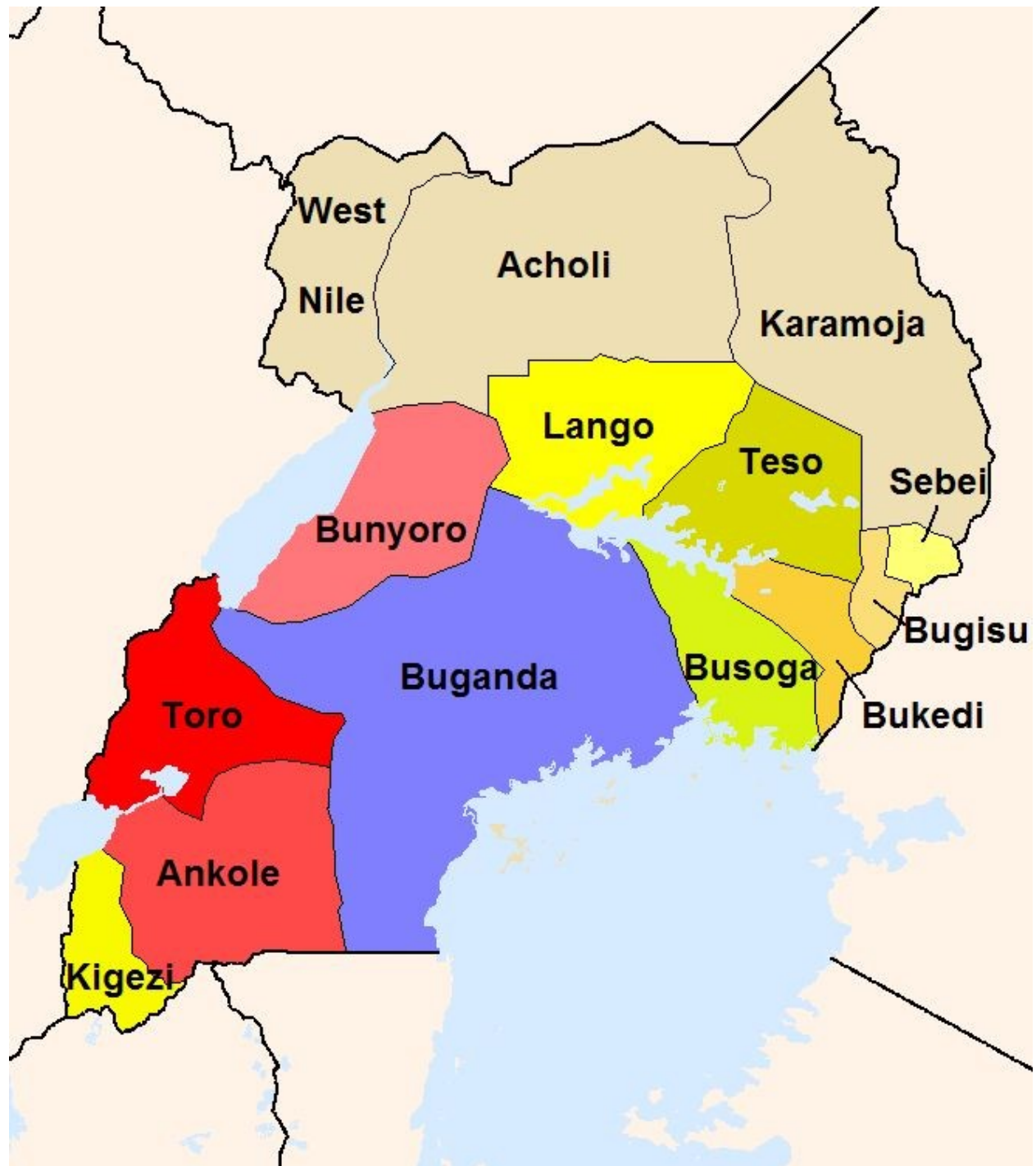
on the perspectives and experiences of the participants, and using open-ended questions to collect rich and detailed textual data (Creswell 2008:46) The qualitative approach can differ substantially based on the underlying theoretical framework, philosophical perspectives, and assumptions about the nature of knowledge, as well as the specific field of study, and according to McMillan and Schumacher (2010:459) qualitative research offers an in-depth examination of a phenomenon using face to face or observational techniques to gather data from participants in their natural environments. As noted by McMillan and Schumacher (2010:320), qualitative commences with assumptions, a particular world view, and potentially the application of a theoretical framework, with the aim of exploring the meanings that individuals or groups ascribe to a social or human problem. This study employs a qualitative research approach, which involves collecting data in a naturalistic setting, being attentive to the social and cultural context of the participants, and using an inductive analysis method to identify patterns and themes that are emergent from the data. In qualitative research, a diverse range of data is gathered and analyzed from multiple perspectives, allowing for a comprehensive and nuanced understanding of a complex and multifaceted phenomenon. This study aims to explore the perspectives of educators and principals on the problem of learner's dropout, including their understanding of the underlying causes and their proposed solutions for addressing this issue (Tsomo, 2012:50). Using an in-depth approach, this research investigated the problem of learner dropout in Ugandan primary and secondary schools, with a focus on understanding the factors that influence this phenomenon, as perceived by the participants. By interviewing educators and principals, this study aimed to gain a nuanced understanding of their thoughts and opinions on the topic of learners and are well-positioned to provide insight into the challenges and complexities surrounding this issue. Grounded in the principles of Educational Data Mining (EDM), this study's methodology utilizes data analysis and machine learning techniques to enhance educational outcomes (Romero Ventura, 2013). By applying EDM principles, this research aims to contribute to the emerging

knowledge base on data-driven approaches in Ugandan education, specifically in the areas of dropout prevention and student success.

Description of the study area

Data on the factors contributing to student dropout was sourced from multiple locations covering Central, Eastern, Western and Northern regions, encompassing both urban and rural areas, including suburbs, slums and towns, (Fig. 12). The selection of data was informed by the prevalence of dropout rates in different districts, which were classified as high, medium and low, and also considered geographical representation. The model was developed using publicly available data from Uganda.

MAP: THE SUB-REGIONS OF UGANDA



3.0.3 Philosophical paradigm

The study is grounded in the positivist research paradigm, which emphasizes the application of quantitative methods to investigate social phenomena. This paradigm is well-suited to the current study, enabling the researcher to create predictive models that pinpoint at-risk students and guide targeted support. A Phenomenological framework was selected as the appropriate philosophical paradigm underpinning for this study. According to Rodolo (2008), phenomenology has been successfully applied in multiple areas of study, including, education, psychology, sociology and nursing, demonstrating its versatility and relevance across various fields. In addition, according to Chadzuka (2008), phenomenology is a distinct philosophical method of inquiry that involves a thorough and systematic exploration of conscious experience, with the aim of gaining a deeper understanding of the social world Springer (2010:19). The researcher's objective is to take an approach, aiming to thoroughly understand and describe the complex experiences and interpretations that individuals derive from their interactions with other people and their surroundings. Phenomenology enables researchers to tap into the personal and subjective realities of individuals, gaining insight into their unique, lived experiences and the ways in which they perceive and interpret their worlds. Phenomenology is grounded in the idea that a single experience can be perceived and interpreted in multiple, diverse ways, and that unique meaning attributed to that experience by each individual is what forms their personal reality. To gain a deeper understanding of the participants' perspectives, the researcher must set aside their own preconceptions and biases, adopting a neutral and open-minded approach to explicitly uncover the meanings and interpretations that the participants attribute to the phenomenon. To gain a profound understanding of the participants' viewpoints, the researcher must suspend their own preconceived notions and biases, embracing a neutral and open-minded stance. This enables the researcher to explicitly capture the meanings and interpretations that participants assign to the phenomenon, without imposing their own perspectives.

3.0.4 Research design

This study employed a predictive analytics research design, which uses advanced statistical and machine learning techniques to predict student outcomes and prevent dropouts. The non-experimental nature of this design means that it does not involve intervening with variables or assigning participants to different groups, but rather focuses on analyzing and modeling existing data to identify patterns and make informed predictions. The primary objective of this study is to investigate student dropout rates by analyzing data collected by Uwezo. The research aims to examine the impact of socioeconomic, demographic, and educational factors. The study utilized machine learning algorithms to identify key predictors and investigate differences in outcomes among various demographic and socioeconomic groups. Subgroup analyses were conducted to examine potential variations in the determinants of dropout rates among different demographic populations. This study seeks to uncover the complex factors contributing to student dropout rates and provide insights to inform the development of targeted interventions to support inclusive schooling.

3.0.5 Study variables

Outcome variable

The outcome variable in this study is “dropout,” which is defined as school-age child in Uganda who leaves the educational system before successfully completing the academic year. A child is classified as a “dropout” if he/she does not receive a final grade for the academic year or do not have an official document verifying completion of that specific year within primary or secondary school. The outcome variable is dichotomous, with a value of 1 assigned to children who have dropped out and a assigned a value of 0 assigned to children who have not.

Predictor variables

The study's explanatory variables were classified into two categories: socioeconomic and demographic factors. Socioeconomic factors included variables related to the household's socioeconomic status, such as education level of the household leader, occupation status, region and district of residence, and household wealth. Demographic factors, on the other hand, included variables related to the child's individual characteristics, such as sex, age, number of children, school attendance, disability status, and distance to school.

3.0.6 Sampling

MacMillan and Schumacher (2010:129-490), defines a sample as a group of individuals drawn from a specific population, with the purpose of collecting data that accurately reflects the characteristics of the population. The sample is designed to be representative of the population, allowing researchers to analyze the data and draw conclusions that can be applied to the broader population. The sample can be drawn from a larger group of population or can refer to the specific group of participants from which data is collected. In qualitative research, data can be sourced from various forms, including individuals, groups, documents, reports and sites. Regardless of the data type, a purposive sampling technique was employed, where the sample was selected without any manipulation or interference from the researcher. Purposive sampling is employed to maximize the usefulness of data collected from small samples. This approach involves gathering information about the diversity and variations within subgroups before selecting the sample, allowing for more informed and targeted sampling strategy. The researcher then searches for information-rich key informants, groups, places or events to study. In this study, which combines both quantitative and qualitative approaches, purposive sampling was employed to identify a sample of participants who could offer rich and informative data about the factors influencing primary and secondary school dropout rates in Uganda. According to Creswell (2008), purposive sampling involves select-

ing information-rich cases for in-depth study, with the goal of ensuring that certain types of individuals or persons with specific attributes are included in the study. This approach enables researchers to deliberately choose participants who can offer detailed and nuanced insights, thereby increasing the validity and usefulness of the data. The participants were selected based on their ability to provide valuable insights and information about school dropouts. Specifically, school administrators were chosen because they had access to school records and data on dropout rates, making them a rich source of information on this topic. Educators, on the other hand, were selected because they had firsthand knowledge of the patterns and characteristics exhibited by learners who are at risk of dropping out, including behavioral signs and indicators that may precede a student's decision to leave school.

3.0.7 Methods of data collection and strategies

Data collection refers to the process of gathering information from a sample according to Babbie, (2016). Unprocessed data is inherently unstable and lacks significance until it is methodically organized and analyzed to reveal meaningful patterns and trends that can inform research questions (Creswell,2014). This research relies on secondary data sources obtained from established organizations such as the UWEZO, Uganda National Examinations Board (UNEB) and the Uganda Bureau of Statistics (UBOS), to support this investigation. These datasets give detailed data on student demographics, academic performance, and dropout rates, allowing the researcher to develop predictive models that recognize at-risk students (UNEB, 2020; UBOS, 2020). This section describes the data collection strategies and methods employed in the study:

Focus group interviews

The data for this study was gathered to provide answers to the research questions. According to Nachmais and Nachmais (2008), the data collection method employed was focus group

interviews. A focus group is a qualitative research technique where a small, diverse group of people are brought together to share their thoughts, feelings, and experiences about a particular topic, product, or service, with the goal of gaining a deeper understanding of their perceptions, opinions, and attitudes. A focus group discussion is a unique research method where participants are encouraged to engage in spontaneous and unstructured conversations with one another, while the researcher asks questions to guide the discussion. This approach harnesses the power of group interaction to generate insightful data, as participants build upon each other's ideas and perspectives. A focus group interview is a deliberately structured conversation that fosters interaction among participants to gather data, as it encourages the exchange of ideas and opinions in a relaxed and non-intimidating setting. According to Krueger and Casey (2008), a focused group is a carefully planned and facilitated discussion designed to elicit participants on a specific topic or issue, in a safe and supportive environment that promotes open and honest sharing. The advantages of group interviews are numerous, as they offer a socially oriented research procedure that allows participants to interact and share their thoughts and experiences. The format provides the moderator with the flexibility to probe and explore unforeseen issues, and it has high face validity, meaning that the results are likely to be accurate and relevant. Additionally, group interviews are relatively low-cost and can produce fast results, making them an efficient research method. Furthermore, they require minimal time and resource involvement, even with large sample sizes, and are particularly useful for exploring people's knowledge, experiences, and thought processes, enabling researchers to examine not only how and why they think that way, according to Nachmais Nachmais, (2008).

Surveys

The study employed surveys to collect data from participants, a method that involved asking questions to gather information (Creswell,2014). This approach is commonly used to

gather information about people's opinions, behaviors, and experiences (Fowler,2014; (Babbie,2016). Within the survey, questionnaires were used to collect data from a large number of respondents in a short period. This method allowed for the generation reliable data, unaffected by the researcher's presence (Kakooza (1997); Babbie, 2013). To ensure accuracy and consistency, respondents were guided through the questionnaires (Fowler, 2009), which contained structured questions related to the study variable

Field notes

According to Emerson et al.,1995 field notes are written records of what the researcher sees, hears and experiences in the field, as well as his/her own observations, feelings, and interpretations about what is happening. Field notes are a crucial component of ethnographic research, where the researcher records observations, experiences and insights, and later expands them into comprehensive and formal notes (Lekhesa 2007:82). These notes not only document the data but also involve interpretation, making them an integral part of the analysis process rather than just data collection (Densecombe 2007:88). As noted by Bogdan and Biklen (2007) Pitney and Parker (2009), field notes are taken throughout the observation period and focus on capturing the details of what is observed, providing a rich and nuanced record of the research setting.

3.0.8 Data analysis

Data refers to the information collected about a sample, whereas collection refers to the process of gathering that information. However, raw data in its unprocessed form is unable to provide meaningful insights or answers to research questions until it is systematically organized and analyzed into coherent patterns, revealing the underlying relationships and trends that can inform the research inquiry. According to McMillan Schumacher (2010:367), qualitative data analysis is an inductive process of organizing data to identify patterns and

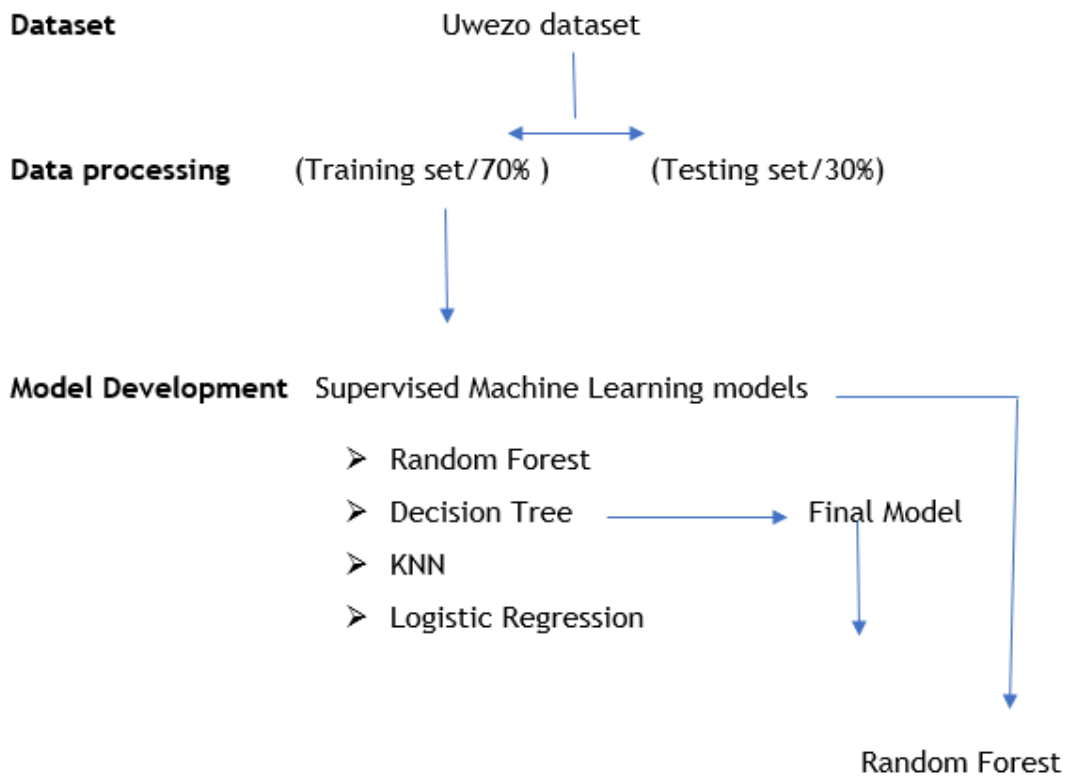
relationships between categories. In this study, the collected data was systematically structured using open coding, theme development, and categorization to address the research question and explore potential solutions to the phenomenon. The recurring themes that emerged across participants were considered to represent shared understandings and common perspectives, as suggested by Bell 2010:211), and were used to inform the analysis and interpretation of the data. The researcher thoroughly analyzed the collected data, which was first transcribed in detail to capture the nuances of the participants' perspectives and experiences. The transcribed data was then coded, as coding is a process that simplifies and organizes the data, making it easier to interpret the findings and identify meaningful patterns and themes, as noted by Wilson (2009:104). By coding the data, the researcher was able to reduce the complexity of the information and facilitate a deeper understanding of the participants' views and experiences, ultimately leading to more accurate and reliable conclusions.

3.0.9 Proposed ml models

The study utilized a variety of machine learning algorithms techniques, such as; Random Forest, Decision Tree, K-Nearest Neighbors, and Logistic Regression to predict student dropout rates using data from the Uwezo dataset. As noted by Pedregosa et al., (2011), these models can be used to inform the development of personalized support strategies and interventions that cater to the unique needs of at-risk students. The analysis was carried out using the Python programming language and its associated libraries, such as pandas. The research utilized data visualization techniques, including bar and line graphs, to examine the relationships between predictor variables and the outcome variable (dropout rate). Through the use of these visualization tools, the researcher was able to identify patterns and correlations between the variables, which provided valuable insights into the factors associated with dropout.

3.0.10 Model comparison and evaluation

In this section, I compared the performance of the four ML models that were previously presented, based on key evaluation metrics that are used in assessing classification models. These metrics included accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). Through a comparison of these metrics, our goal was to determine the most effective model for predicting dropout rates of student.



Model Accuracy Metrics

- Accuracy
- Precision
- F1 -Score
- Recall
- ROC - AUC

3.0.11 Sampling techniques

The study used a stratified sampling technique to divide the population into subgroups based on factors such as region, urban/rural location, and socioeconomic status. This approach enabled the researcher to ensure that the sample was representative of the population, which is essential for supporting the validity and generalizability of the results. The study included a total of 21,241 households, with the aim of gathering a rich and diverse range of data on backgrounds, experiences, and aspirations. By focusing on these households, the research aims to uncover the complex factors that influence the educational trajectories of children and identify effective strategies to support the success in school, ultimately providing valuable insights into how to improve educational outcomes.

3.0.12 Research model development using algorithms

This study utilized random forest, a supervised machine learning technique to forecast general proficiency rates using data from Uwezo. Random forest was chosen, because it was well-suited for handling complex datasets and could identify important variables that contribute to proficiency rates. The use of random forest allowed to make accurate predictions, and identify patterns in the data, which was essential for achieving the research goals.

3.0.13 Sample description

A sample was drawn from a well-defined population or comprised a specific group of individuals who provided the required data. This study employed a longitudinal design, tracking 21,241 households over a one-year period, with data collected on demographics, academic performance, and behavioural factors at four distinct time points. Educators, given their direct interaction with learners, were expected to provide valuable insights that would inform the study. The data was sourced from reputable organizations, including the Uganda Bureau of Statistics. (2021). Uganda National Household Survey

2019-2020 report.<https://www.ubos.org/wp-content/uploads/publications/092021Uganda-National-Survey-Report-2019-2020.pdf>, UNESCO(2017).*Education for Sustainable Development : A framework for action.* Wagner, C.C., Watkins, M.W.(2010).*The effects of family and community factors*. 263.<https://www.google.com/search?q=peas+schools+in+uganda&ie=UTF-8&oe=UTF-8&hl=en&client=safari>,

3.0.14 Ethical considerations

Jegede (2009:80), stated that standard ethical principles that guided the treatment of human participants formed the foundation for the methodological approach in this study. Further, according to Bray 2007:447, participants were ensured a thorough description of the research objective and given an informed choice about whether to take part, based on a comprehensive understanding of the study's requirements. The ethical measures adhered to as guiding principles throughout the study were as follows:

Anonymity and confidentiality

The privacy and confidentiality of participants was safeguarded throughout the interview. The researcher maintained utmost confidentiality of the notes from participants, given the study's focus on human subjects. The names, identities of the participants, and research sites were kept undisclosed in the report findings. This measure was taken to protect the personal details of participants, ensuring their anonymity.

Reflexivity

According to Dowling as cited in Denscombe 2007, researchers had to be reflexive, engaging in bracketing that required time to reflection, a supportive environment, and reflective skills. This reflexivity allowed greater openness to data and awareness of how personal experiences and interests influenced understanding.

Permission to conduct research

Prior to commencing the research, we secured approval from the Education Officer in Uganda, who granted permission to conduct the study in primary and secondary schools in all the regions within the country.

3.0.15 Phenomenological reduction and bracketing

According to Lekhesa (2007), phenomenological reduction involved a deliberate bracketing out, where researchers took no stance either for or against the participants' perspectives. This approach also required setting aside the researcher's own meanings, interpretations and theoretical framework to enter the participant's experiences.

Voluntary participation and informed consent

According to Springer (2010:93), participants were permitted to either grant or withhold consent to participate voluntarily. Furthermore, fostering care and fairness through open dialogue and collaboration with research participants helped promote equity in the research process. This approach empowered participants, facilitated their autonomy, and motivated them to make informed choices about participation. Participants received comprehensive details about the with study, enabling them to opt for or against involvement.

3.0.16 Limitations of the study

The study was limited by several factors, including its dependence on secondary data sources, the potential for biases in the data, and the limited applicability of the findings to other contexts. The lack of local research necessitated the use of international literature, which may not have fully accounted for the complexities of the Ugandan education system, given the study's focused scope. As a result of the study's narrow focus, the emotional challenges

faced by learners who dropouts were not explored. The emotional difficulties experienced by learners who dropped out were not investigated. The study's measurement tools may not have been entirely valid or reliable. Additionally, the sample size may not have been sufficiently large to yield significant results.

3.0.17 Conclusion

This chapter outlined the methodology, research design, and methods used to conduct the study, which utilized machine learning approaches to investigate dropout prevention and student success in Uganda. The empirical findings were analyzed in relation to the theoretical frameworks, revealing several factors that significantly contributed to learner dropout rates including; disability, repetition, orphan hood, family violence, poverty, unemployment, health concerns, lack of parental involvement, ineffective free primary education initiatives, substandard education quality, and inadequate school infrastructure. These factors were found to be influenced by both school and home environments. The study demonstrated that dropping out of school has broader implications, affecting not only individual's but also the government, society, and labor market, with consequences extending beyond financial losses to social effects.

CHAPTER 4

DATA ANALYSIS, RESULTS AND INTERPRETATION

4.0.1 Introduction

Chapter three presented the research design and methodology of the study. Chapter four presented the findings and analyzed the factors that influenced primary and secondary school dropout rates in Uganda. The objective of this section was to synthesize the results gotten from the machine learning model to gain a deeper understanding of the rates at which students in Uganda abandoned their studies.

4.0.2 Factors that contributed to student dropout

The analysis of the Uwezo dataset using the Random Forest algorithm revealed multiple key factors that played a significant role in student dropout rates in Uganda. This section outlines the findings, highlighting the prominent factors and their corresponding effects.

Key Factors

The analysis revealed that household size was the most significant factor contributing to student dropout, accounting for 33% of the dropout rates. For instance; Larger households were associated with higher dropout rates, while smaller households had lower rates. This research showed that students from larger households tended to encounter family obligations, resulting in a higher likelihood of dropout. Further analysis showed that several other factors played a significant role in student dropout rates in Ugandan Primary and Secondary schools. Notably, grade (21%) led to an increase in dropout rates, suggesting that when students went to higher classes, they were discouraged from continuing with their studies. Most striking was the finding that households with fewer meals per day (8%) more likely to experience student dropout, suggesting that the limited food availability may have been a factor.

Feature importance

The feature importance scores from the Random Forest algorithm gave deeper insight into the respective impact of each feature. The top five features, in order of importance, were:

1. Household size (33%)
2. Grade (21%)
3. meals Per Day (8%)
4. Biological parents (8%)
5. School match (7%)

Demographic phenomena and traditional approaches

The analysis showed demographic phenomena that emphasized the taking on of family obligations by young people, influenced by economic pressure, family dynamics, and cultural norms, the results indicated that students from larger households or those with more females or males faced heightened pressure to care for the family, resulting in a greater chance of dropout. Traditional approaches to identifying contributing factors to student dropouts included structural functionalism, which viewed society as a system with interconnected components. This perspective was illustrated by Vincent Tinto's 1975 model of Institutional departure, which focused on the importance of student integration into the academic and social life of the institution. According to this model, students who felt disconnected or isolated were more likely to dropout. Nevertheless, similar traditional approaches relied on prior knowledge and overlooked indirect links to the problem, ultimately failing to effectively resolve dropout rates over the years. It was likely that the slow pace of improvement was due to the complexity of the issue.

Machine learning model and data balancing

The machine learning model developed in this study uncovered key variables that were overlooked by traditional approaches. Leveraging its ability to mine non-linear information from features and their association with the study used a dataset that was highly imbalanced, with more students retaining studies than dropping out. Data balancing techniques played

a vital role in resolving this issue. Similar results were noted by Batista et al. (2004), highlighting the importance of handling data imbalance when developing machine learning models, as well as by Farquad and Bose (2012), and Ramentol et al. (2012), stressing the need for careful data preprocessing.

Importance of handling data imbalance

Most real-world datasets were imbalanced, and ignoring data balancing resulted in bias towards the majority class. Managing balanced data which was essential when developing machine learning models, which ensured that the model learned about both the majority and minority classes. The SMOTE-Tomek technique produced promising outcomes in this study, and its performance was consistent with the findings from other researchers. By addressing data imbalance, this study aimed to improve the accuracy of predicting student dropout and informing policy makers to tackle dropout rates in Uganda.

Magnitude of the school dropout in Uganda based on Uwezo dataset

The magnitude of school dropout in Uganda was assessed using the Uwezo dataset. The Random Forest model's ability to accurately predict student dropout rates was leveraged to support at-risk students. The model was used to identify students who were at high risk of dropping out, and they were provided with targeted support, including; tutoring, mentoring, or counseling. The model also helped educators identify areas where students required extra help, such as math or English, and provided them with focused assistance to improve their performance. Analysis of the Uwezo dataset showed 76% of students attended school the preceding day, while 24.0% did not. The highest dropout rate was recorded in 2013, with 194 students, and class 1 had the most dropouts, with 111 students.

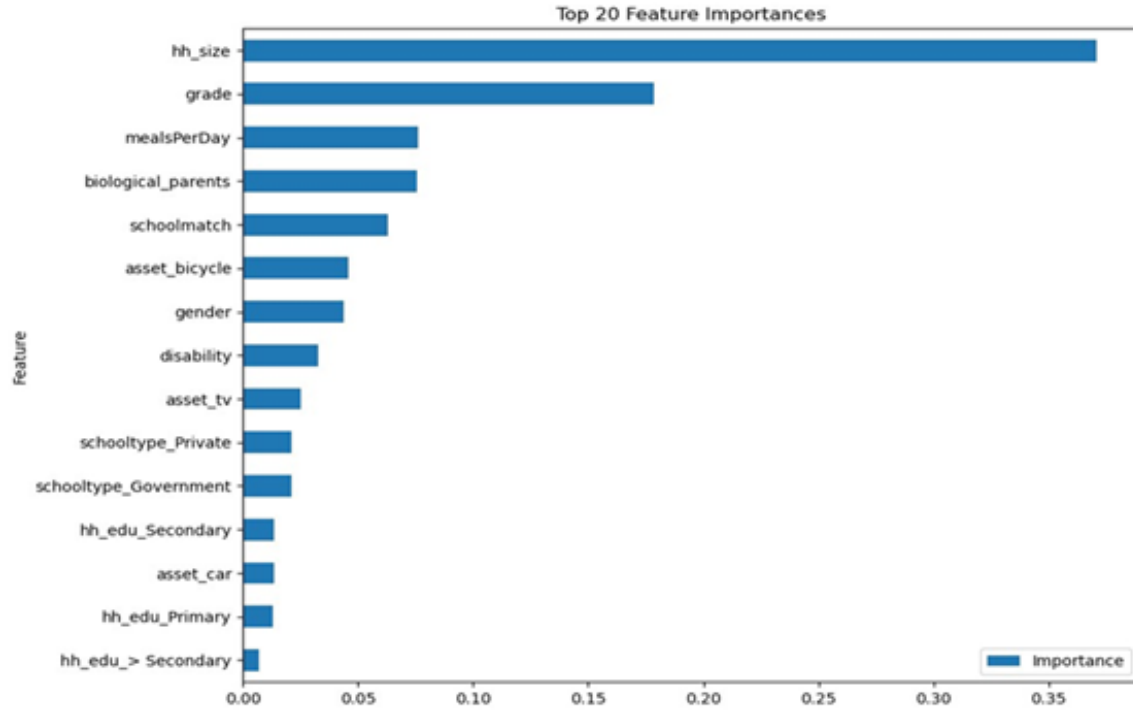
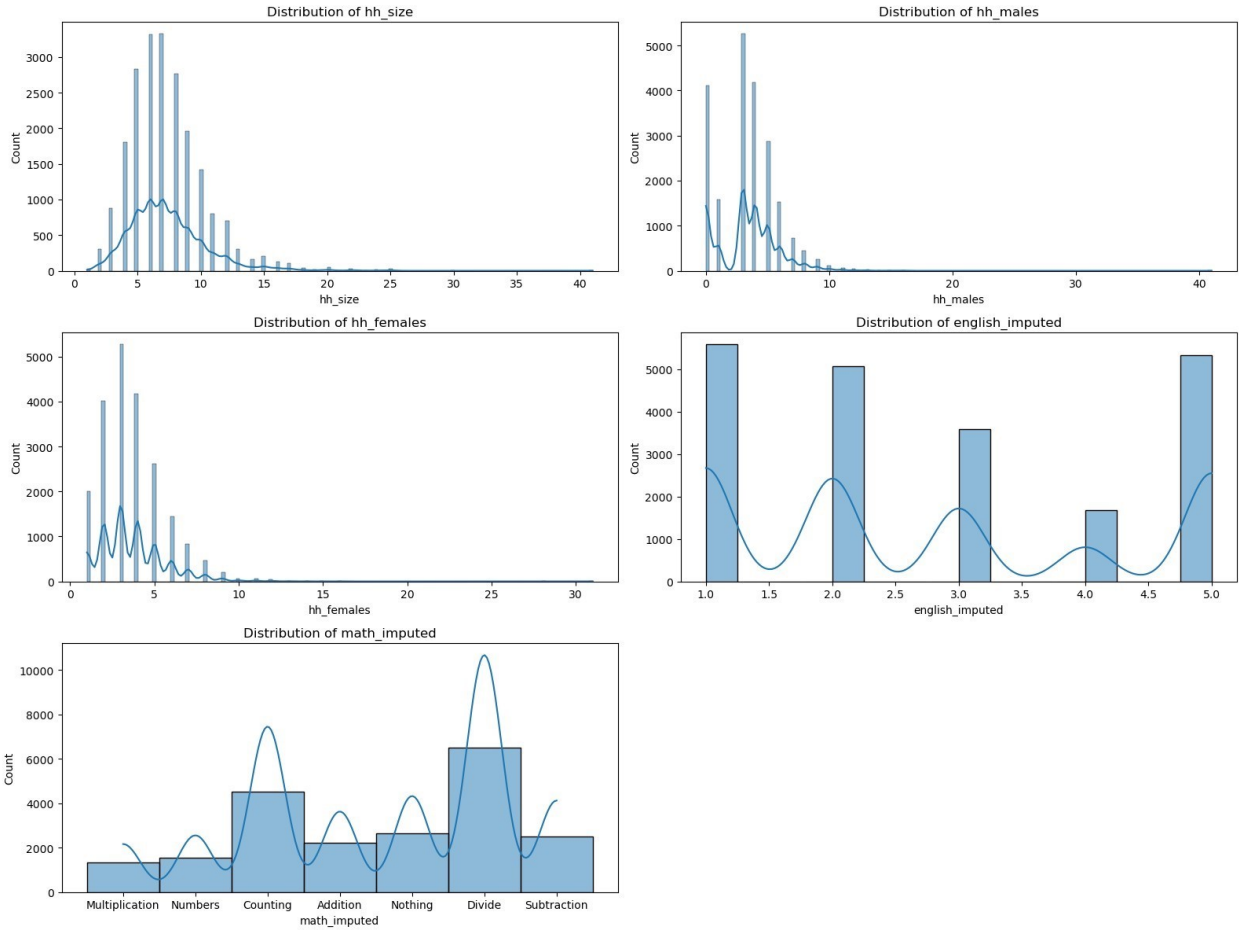


Figure: Factors that contributed to student dropout identified during RTS and FGD.

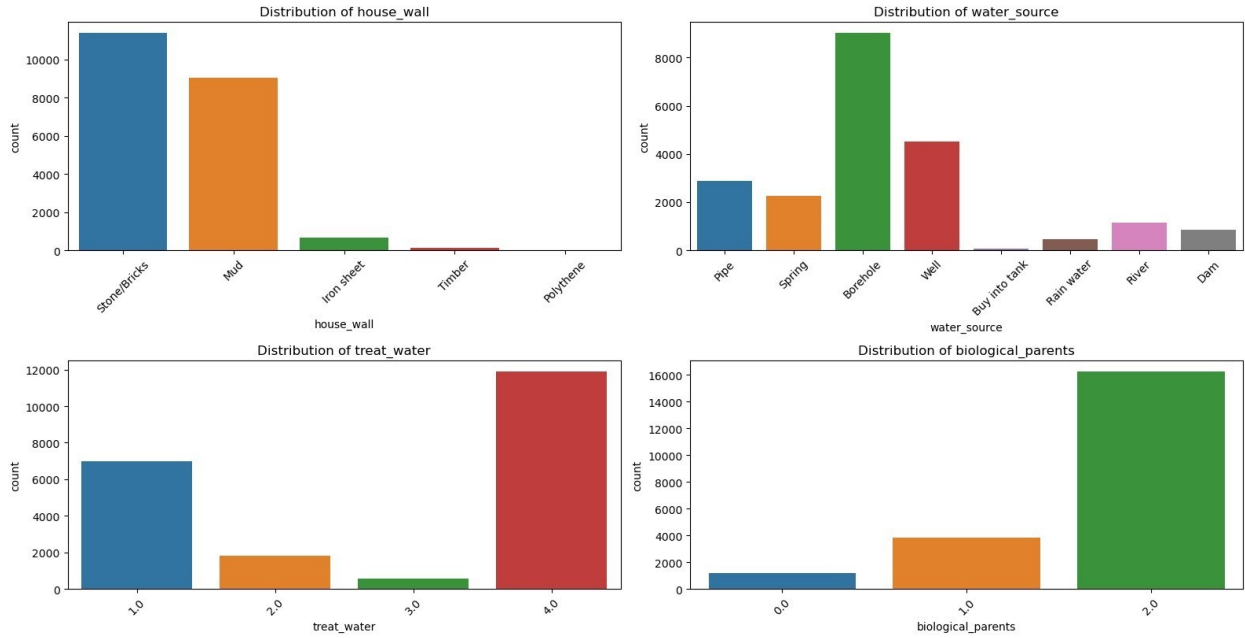
4.0.3 Feature engineering

The results from the feature engineering experiment showed the 15 factors that contributed to student dropout rates, as identified during the Round Table survey (RTS) and Focus Group Discussions (FGD). Household size (hh_size) being the strongest contributing factor, significantly outweighing



Students from larger households are more prone to leaving school probably because of the economic hardships and higher caregiving responsibilities at home.

Visualize the Distribution of Categorical Features



The bar graph of the distribution of the house wall depicted the frequency distribution of household building materials, revealing that stone/bricks were the most prevalent material, with approximately 7800 counts. Mud was the second most frequently used material, with around 5500 counts. In contrast, Iron sheet walls and timber/polythene structures were relatively rare, with counts below 1,000 and 500, respectively. Overall, the visualization provided valuable insights into the distribution of building materials among the surveyed households.

Compare Continuous Data Against the Target Variable

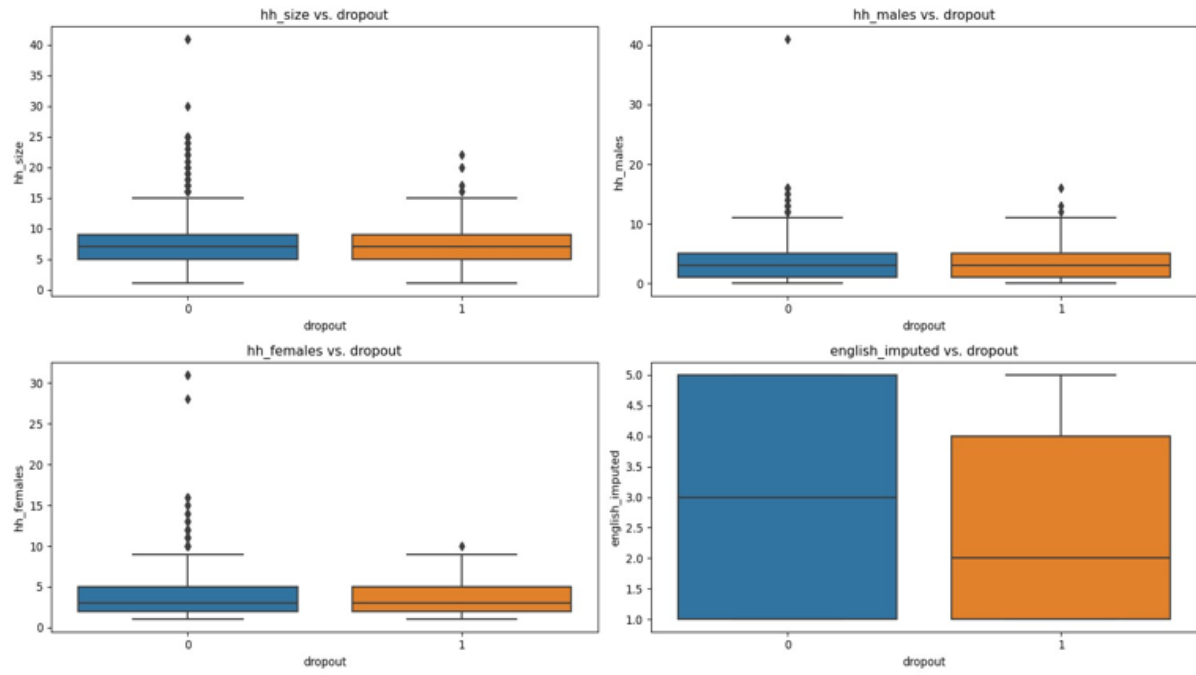
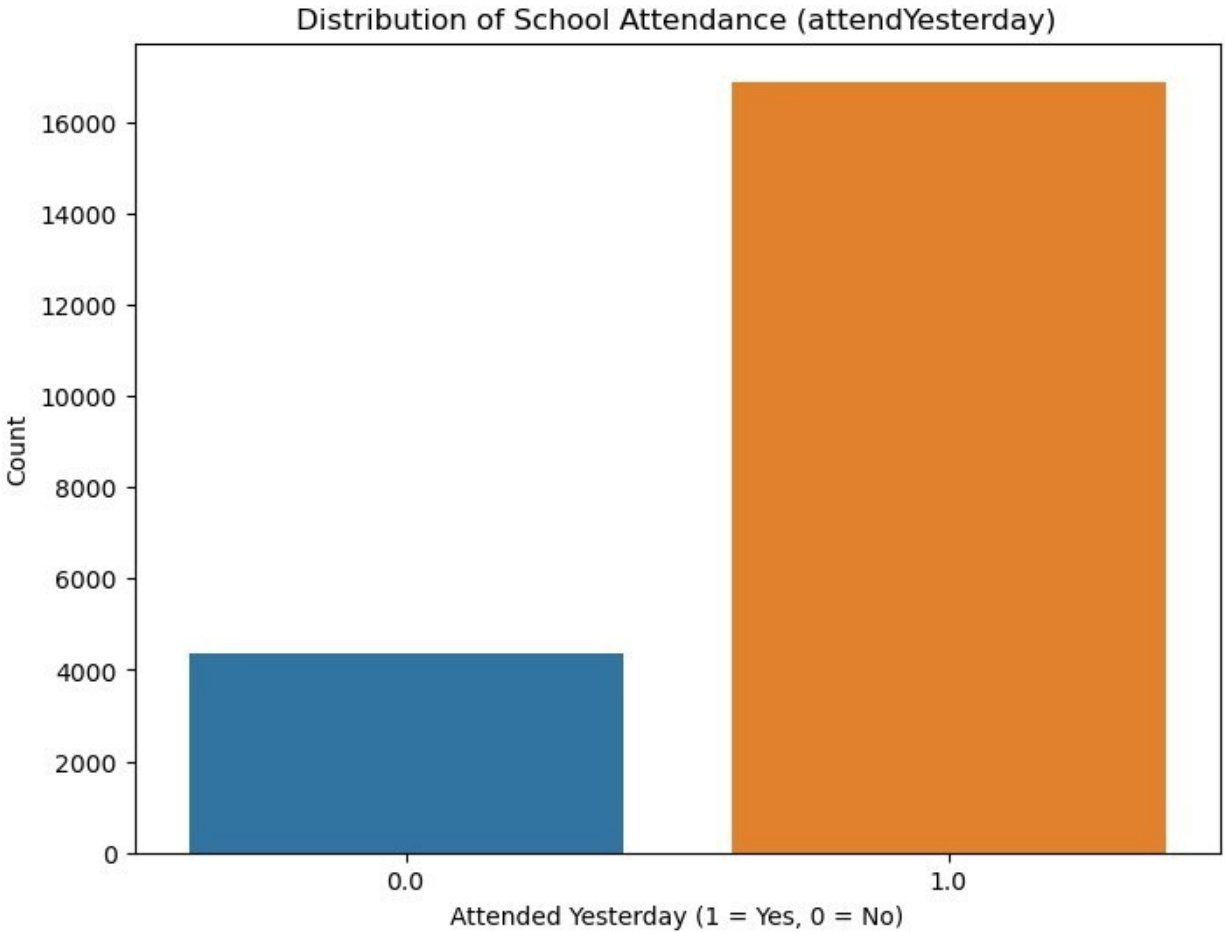


Table 4.1: Data Distribution for Variable: Attended Yesterday

Value	Percentage (%)
1.0 (Attended)	79.44
0.0 (Did Not Attend)	20.56



The output showed that 79.4% of students attended school yesterday, while 20.6% did not attend. Specifically, 79.4% responded 'yes' (attended yesterday= 1), and 20.6% responded 'no' (did not attend yesterday= 0) This graph provided offered a clear concise overview of attendance versus absence on a given day, showing a notably higher number of students present rather than those absent. Additionally, the blue color on the bar graph revealed that 2,000 students did not attend school yesterday, whereas the orange represented about 14,000 students who attended school.

Supervised machine learning models

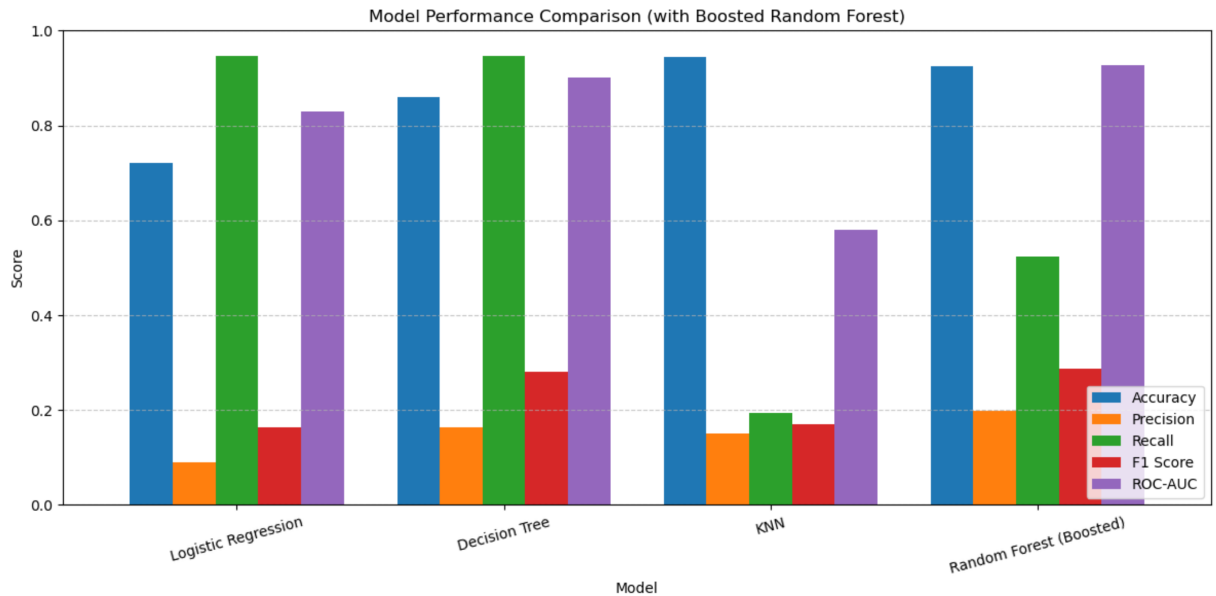
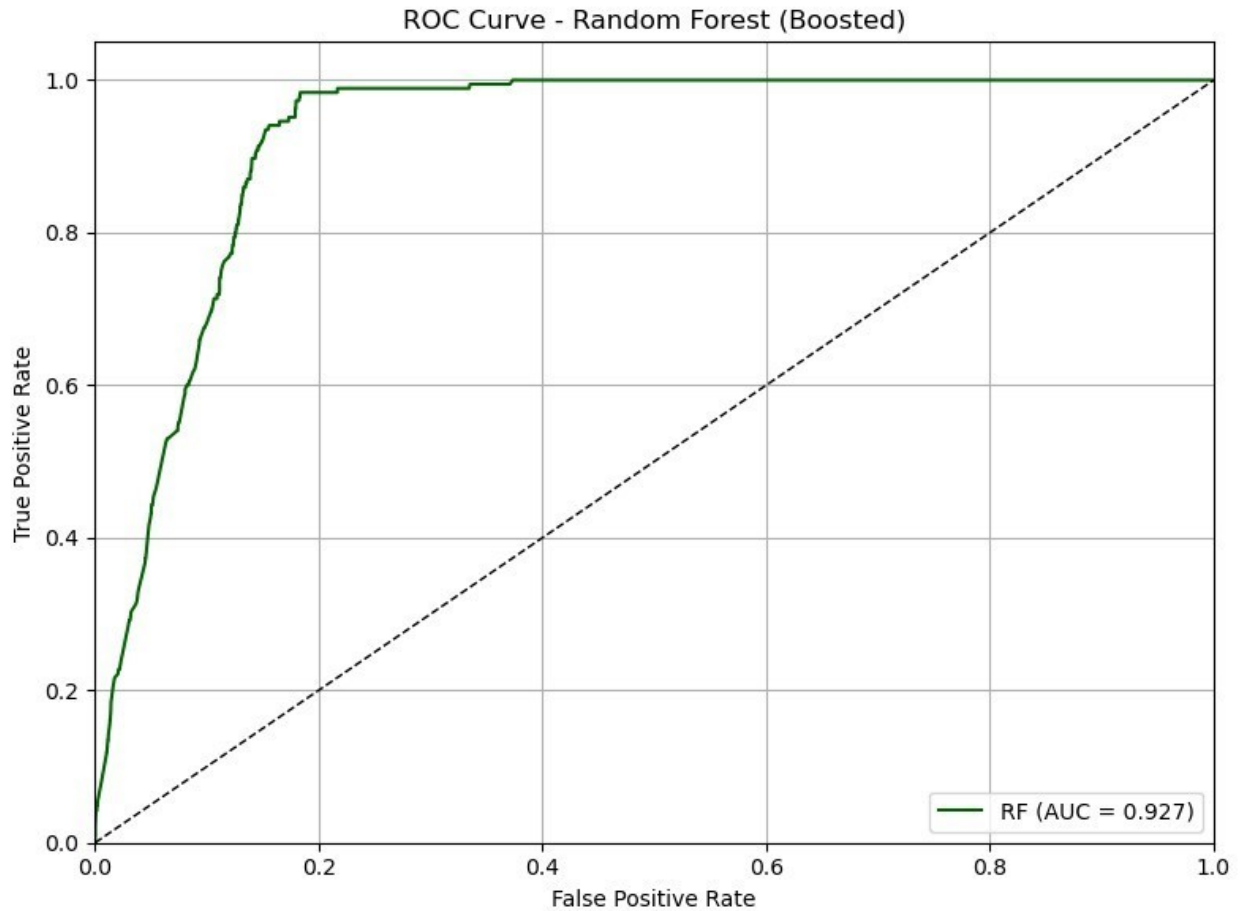


Fig: presented a thorough examination of various machine learning models that were used to predict student dropout and completion rates. The models that were evaluated comprised; logistic regression, random forest, decision tree, and K-nearest neighbors (KNN). Each model's performance was evaluated using multiple key metrics, namely: accuracy, F1-Score, and area under the curve. For instance; Logistic Regression had an accuracy of approximately 0.7, an F1-Score around 0.1, and an ROC-AUC was 0.8. The Decision Tree model exhibited a similar pattern, with an accuracy of 0.8, an F1-score close to 0.3, and ROC-AUC of 0.9. KNN yielded an accuracy of 0.9, an F1-Score at 0.1 and an ROC-AUC 0.5. The Random Forest model outperformed the others across all metrics, with an accuracy of 0.9, an F1-Score of 0.3, and an ROC-AUC at 0.9. This indicated that the fine-tuned random forest model surpassed the other three models in this comparison, highlighting its superior ability to correctly classify data and generalize well to unseen data. This rendered RF especially appropriate for applications requiring precise differentiation between positive and negative classes.

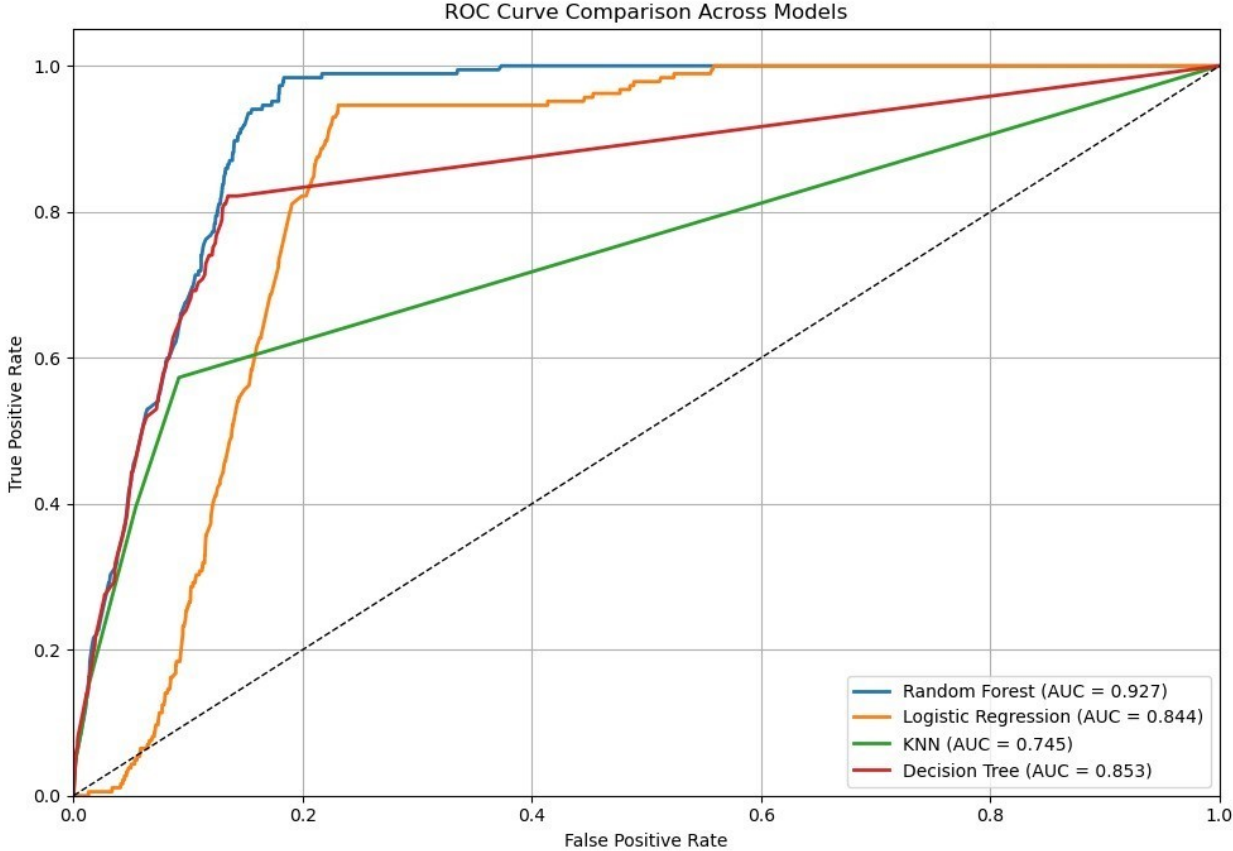
FIG .



The graph displayed is a Receiver Operating Characteristic (ROC) curve, graphically representing the binary classification of the model's performance. The curve plotted true positive rate (TPR) against false positive rates (FPR) at various thresholds. With an area under the curve (AUC) of 0.93, the ROC curve indicating the model's ability to distinguish between positive and negative classes. An AUC of 0.5 suggested no discrimination, while AUC of 1 implies perfect classification. The green curve shows the model's performance, and the closer it is to the top-left corner, the better. The dashed line represents a random guess, so the model is doing much better than that. Therefore, the model demonstrated moderate discriminatory ability. The curve showed that as the false positive rate increased,

the true positive rate also increased, but not proportionally, indicating a trade-off between sensitivity and specificity. The dashed diagonal line represented a random classifier, and the ROC curve was above this line, suggesting that the model performed better than chance.

4.0.4 Model evaluation



The RF model evaluation revealed key performance metrics, with an accuracy score of 0.93, meaning it correctly predicted the class of 93% of the instances. Such a score was impressive, demonstrating that the model performed well at this task. It also meant that there was an 93% chance that the model could distinguish between the two classes correctly. The classification report provided a thorough breakdown of precision, recall, and F1-score for

each class, giving clearer understanding of the model’s ability to correctly identify and distinguish between different categories. The precision for each class (0 and 1) was calculated as follows: For each class (0 and 1 in this case), precision measured the proportion of accurately classified cases out of all cases predicted as that class. For class 0, the precision is 0.59, 59% of the instances predicted as 0 were actually 0. For class 1, the precision is 0.81, meaning 81% of the instances predicted as 1 were actually 1. The model’s precision quantified the proportion of accurate positive predictions among all positive predictions generated. The high precision prevented the allocation of resources to false positives. i.e $\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$ According to (Powers, D.M.W.2011), recall refers to the proportion of actual instances of a class that were correctly identified. Specifically, for class 0, the recall was 0.12, indicating that the model accurately detected 12% of all actual 0 instances, whereas for class 1, the recall was 0.98, showing that the model successfully identified 98% of all actual 1 instances. That is to say; $\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$ Furthermore, according to Sokolova and Lapalme (2009), F1-score represented the balanced average of precision and recall, calculated as a combination of these two metrics. That is to say; $\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$. The F1-score, a crucial metric that integrated precision and recall, showcased the model’s capacity to tackle the complexities of dropout prediction with a high score. The model achieved a notable sensitivity value of 0.80, which quantified the percentage of true dropouts successfully detected. Support, represented the total occurrences of every class in the dataset. Class 0 had 1295 instances, while class 1 5078 instances. According to Libor Ansoorge (2023), accuracy played a crucial role in tracing original scientific sources and upholding ethical writing standards. The model’s total accuracy stood at 80%, demonstrating that 80% of all instances were classified correctly. $\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN}}$ Macro Avg was employed to evaluate the performance of classification models, especially in case of imbalanced datasets. Macro average determined

the average of precision, recall, and F1-score for every class, assigning uniform weight to each. The weighted avg measured the average of precision, recall, and F1-score, with weights based on the number of instances in each class. The confusion matrix presented a thorough analysis of the model's performance, displaying the counts of true positives, true negatives, false positives, and false negatives. Notably, it included 160 true positives (correctly predicted 0s), 1135 false negatives (actual 0s predicted as 1s), 112 false positives (actual 1s predicted as 0s) and 4966 true negatives (correctly predicted 1s). A confusion matrix visually represented the model's predictions against actual values, highlighting areas of correct and incorrect classification. The study's rigorous evaluation framework confirmed the efficacy of the Random Forest model, providing a practical roadmap for its implementation in real-world educational environments. The ROC AUC score (0.93) quantified the area under the receiver operating characteristic (ROC) curve, highlighting the model's capacity to separate the two classes. A score of 0.5 showed no predictive power, while a score 1.0 signified flawless separation. In conclusion, the model showed strong performance in predicting class1 but faced challenges with class 0, as shown by the low recall for class 0 and the imbalance in the confusion matrix. The overall accuracy was reasonable, but the ROC AUC score pointed to opportunities for enhancing the model's ability to separate classes.

4.0.5 Implementation and impact

The findings of the study led to the implementation of data-driven interventions that addressed the specific needs of at-risk students. Schools provided targeted support services, such as academic counseling, mentorship programs, or vocational training, to help students overcome barriers to education. Additionally, schools used data analytics to identify effective teaching practices and implemented evidence-based instructional strategies to improve student engagement and motivation. This led to enhanced student outcomes and decreased dropout rates.

4.0.6 Reasons for employing Random Forest over other models

Random Forest provided key benefits in school dropout predictions notably; I. Random forest out-performed logistic regression due to its ensemble approach, which made it less sensitive to noisy data and enabled accurate predictions even with incomplete or imperfect data. II. Random forest was chosen for its capacity to handle complex datasets with many features. III. Random forest showed high accuracy and balanced accuracy in predicting school dropouts, proving it to be a reliable tool for identifying at-risk students. IV. Random forest facilitated early identification of at-risk students, which enabled timely interventions and supported systems that improved student retention rates more effectively than other models. V. Random forest was less sensitive to hyperparameter tuning, making easier to implement and tune. VI. Random forest's robustness to overfitting suited it for this research due to being generalized well to new and unseen data. VII. Random forest was easy to interpret, unlike complex models like neural networks. It facilitated feature importance analysis, highlighting the key factors that drove dropout decisions. This provided insight into the underlying causes of dropout and guided interventions.

4.0.7 Mapping Objectives to Results and Contributions

This section explicitly demonstrates how each of the specific objectives was achieved and highlights the contribution made through the application of machine learning and data driven approaches.

To evaluate machine learning models and techniques in student drop out prediction. Achievement: To achieve this objective, four supervised machine learning algorithms ie. Random Forest, Logistic Regression, Decision Tree and K-Nearest Neighbors were implemented and evaluated using key performance metrics such as accuracy, precision, recall and Area Under the ROC Curve (AUC). A rigorous model development process was followed, including data cleaning, balancing using SMOTETomek technique and cross-validation. The results pre-

sented in section 4.3 indicating that the Random Forest model significantly outperformed the other models. Contribution: This study contributes to the field by empirically demonstrating that ensemble learning techniques like Random Forest are particularly well-suited to complex, imbalanced educational data in low resource environments. Furthermore, the application of SMOTETomek to balance the dropout dataset is a novel approach in Ugandan educational research. This adds to the body of knowledge by showing how performance can be optimized through data preprocess techniques, thereby improving the reliability of early warning systems for student retention.

To analyze factors that contribute remarkably to the student dropout issue in Ugandan primary and secondary schools. Achievement: To fulfill this objective, both qualitative and quantitative analyses were conducted to identify key determinants of school dropout. The qualitative insights were gathered through focus group discussions and secondary literature review (Chapter 2), while the quantitative component used feature importance analysis from the Random Forest model. Key predictors of dropout included socio-economic status, disability, school distance, parental attitude, food security and school environment. The model was able to rank these factors in terms of their predictive strength, providing a clear understanding of which issues most significantly impact student retention. Contribution: The study offers a data- driven prioritization of dropout factors tailored to the Ugandan context, something that previous studies have often lacked. Rather than relying solely on anecdotal or survey – based data, this research provides qualifiable evidence to guide the allocation of educational resources. The integration of feature importance analysis into dropout research is a methodological innovation that strengthens the precision of policy formulation and intervention design.

To design and implement data-driven interventions to support at-risk students Achievement: Using insights from the predictive model and key variable rankings, this study developed a conceptual framework for a data-driven early warning system that identifies at-risk

students in real time. As detailed in Sections 4.4 and 4.5, the system categorizes students into risk tiers based on the probability of dropout. The study also proposes practical, targeted interventions such as conditional cash transfers, mentorship programs, community sensitization on girls' education, provision of sanitary pads, school feeding programs and enhanced parental engagement strategies. Contribution: This research bridges the gap between prediction and action. While many machine learning studies stop at model accuracy, this study translates predictive insights into actionable strategies for policymakers and education stakeholders. The proposed interventions are grounded in local data and tailored to Uganda's socio-cultural realities. This enhances the study's impact by offering not just a diagnostic tool, but also a solution-oriented framework that can be scaled by the Ministry of Education and NGOs.

4.0.8 Overall Research Contribution

This research contributes to the field in five major ways:

- **Methodological Innovation:** Introduced SMOTETomek data balancing in the context of educational machine learning models in Uganda.
- **Model Performance:** Demonstrated the superior performance of ensemble models (Random Forest) in predicting dropout with high accuracy, recall, and AUC.
- **Local Relevance:** Applied machine learning to a uniquely Ugandan dataset, addressing a pressing national challenge with locally grounded data.
- **Policy-Oriented Outcomes:** Proposed a real-world, data-driven intervention framework, bridging academic research with implementable solutions.
- **Theory-Data Integration:** Connected behavioral theories (Social Cognitive Theory and Ecological Systems Theory) to machine learning outputs, ensuring that recommendations are both statistically sound and human-centered.

In summary, this section has demonstrated how each research objective was realized through rigorous data analysis and thoughtful interpretation. By integrating machine learning techniques with contextual knowledge and behavior theories, the study not only achieves its stated goals but also contributes meaningful, evidence-based insights to the broader discourse on educational transformation in developing countries.

CHAPTER 5

DISCUSSION OF FINDINGS

5.0.1 Introduction

This chapter synthesizes the results from the machine learning model to deepen understanding of student dropout rates in Uganda. By integrating insights from predictive analysis, extracting meaningful patterns, and distilling practical recommendations, this study ensures academically rigorous and policy-relevant findings. We review the study’s purpose, research questions, literature review, and findings, drawing conclusions and providing recommendations for practice and future research.

5.0.2 Summary of key findings

This thesis, ‘Tackling Dropout Rates of Student in Uganda: An Exploration of Machine Learning and Data-Driven Approaches,’ investigated the application of machine learning and data-driven approaches to identify factors associated with student dropout rates in Ugandan primary and secondary schools. The study utilized features and random forests with data balancing techniques on the Uwezo dataset (Uwezo, n.d.). The study discovered socioeconomic and academic factors, such as household size, meals per day, parental income, and student performance, as key indicators of dropout. The random forest algorithm showcased high precision in identifying at-risk students. The study’s findings added to the existing body of knowledge on student dropout rates by revealing the promise of machine learning and data-driven approaches in identifying at-risk students and guiding focused initiatives (Kim et al., 2018; UNESCO, 2017). This research has far-reaching implications for educators, policymakers, and researchers seeking to lessen dropout rates and improve student outcomes in Uganda. By leveraging machine learning and data-driven approaches, stakeholders are able to craft more impactful approaches to support students and enhance their

prospects for achievement.

5.0.3 Discussion

This study employed a random forest machine learning technique to predict student attrition rates, offering new insights into the elements contributing to dropout rates (Kim et al., 2018). The model uncovered disparities in dropout rates based on household sizes in Uganda, highlighting the need for measures considering household-specific factors (UNESCO, 2017). Socioeconomic factors, like household income and size, also had an impact (Gambacorta et al., 2019; Khandani et al., 2012). Specifically, larger household sizes were linked to higher dropout rates, indicating that larger families had limited access to the resources necessary for education. Moreover, grades consistently proved to be the strongest predictor in several models, including random forest. This suggested that the individual challenges faced at each grade level were a major contributor to students leaving school before finishing their education. The random forest model regularly outperformed other algorithms like Logistic Regression, attaining the highest AUC-ROC score (0.927) and showing a great accuracy (0.925) in predicting student dropout (Aulck et al., 2017; Batuwita Palade, 2012). This outstanding result highlighted random forest as the most accurate and robust model for predicting student attrition. The model's ability to capture dynamic associations between variables made it particularly successful in this context (Breiman, 2001). The ecological system theory (Bronfenbrenner, 1979) provided a framework for grasping the intricate relationship between individuals, family, and environmental factors that contributed to student dropout. In the context of this study, the ecological system theory was applied to examine how household-specific factors, including socioeconomic status and family dynamics, influenced student dropout rates. Specifically, the findings revealed that students from low-income households at a higher risk of dropping out of school, which was attributed to the microsystem level of the ecological system theory. The feature engineering experiment con-

ducted in this study revealed that household size and student gender (female males) had a significant impact on student dropout, followed by household meals per day (MLPD). For example; number of studies identified household size and student gender as the primary factors contributing to student dropout, especially in low-income countries. Interestingly, few studies found this to be a pressing issue in low-income countries, according to Isphording Qendrai, 2019; Kim et al., 2018. The higher dropout rates for girls were largely due to cultural factors, like early marriage and early pregnancies. A UNESCO report (2017) indicated that early pregnancy and marriage explained 47% of Uganda's dropouts in East Africa. The report cautioned that if immediate action was not taken, the dropout rate might rise to 70% by 2030. This would have a profound effect on the achievement of the sustainable Development Goal 4(SDGs), which stresses the need for inclusive and equitable quality education for all school-aged children, particularly girls from vulnerable backgrounds. The social cognitive theory (Bandura, 1986) offered a framework for understanding the factors that contributed to student dropout. In the context of this study, the social cognitive theory was used to examine how individual-level factors, including motivation and self-efficacy, influenced student dropout rates. Specifically, the results indicated that students who had low levels of motivation and self-efficacy were at a higher risk of dropping out of school, which was due to the cognitive and emotional factors identified in the social cognitive theory. This result aligned with the works of Dweck (2000), who found that students with a fixed mindset tended to encounter academic difficulties and dropout. He also noted that, 'a growth mindset, on the other hand, allowed students to surmount challenges and persevere.' The results of this study had implications for policy and practice. The findings underscored the necessity for targeted interventions to support students from underprivileged backgrounds, like those from low-income households and restricted access to educational resources. The results also indicate that machine learning models could be effective tools for predicting student dropout and identifying students at risk.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.0.1 Conclusions

The predictive study's findings revealed that demographic and educational parameters played a critical role in predicting student dropout rates in Ugandan Primary and Secondary schools. The research question, "Can machine learning models accurately predict student dropout rates in Uganda?" was adequately examined, with the results revealing that random forest models showcased effective frameworks for predicting dropout. The use of Synthetic Minority Over-Sampling Technique and Tomek links (SMOTE-Tomek) on the imbalanced data provided a good solution for achieving higher performance. In particular, the study identified the primary drivers of student dropout rates and developed a predictive model. The research hypothesis, "Machine learning models can accurately predict student dropout rates in Uganda," was verified. The selected approach, for example data collection, feature engineering, and model development, was suitable for the research task. By leveraging machine learning algorithms and evidence-based findings, educators and policymakers could create better solutions to mitigate student dropout rates and improve educational outcomes in Uganda.

6.0.2 Key implications

The study's results have substantial impact for educational policymakers and practitioners in Uganda. To student attrition rates and improve educational outcomes, targeted interventions are essential to support students at risk of dropping out. Focused interventions should be provided to assist younger students, and support systems should be initiated for students at critical grade levels. Additionally, gender-sensitive interventions are crucial to ensure both males and females receive sufficient support.

6.0.3 Recommendations addressing identified causes of dropout.

The following are strategies and recommendations to mitigate the factors identified by machine learning model's feature importance analysis. Each solution targets the specific cause of dropout, ensuring an evidence-based, contextually appropriate approach. One of the key predictors of dropout was household size. Larger families tend to face financial pressures that limit the ability of parents or guardians to provide consistent school support. To address this, the government and education stakeholders should expand conditional cash transfer programs targeting large, vulnerable households. This financial support would ease the economic burden and reduce the likelihood of children leaving school due to poverty. Additionally, integrating family planning and reproductive health education within community outreach programs would help manage household sizes overtime, creating a long-term positive impact on education retention. The grade level of the student was another important predictor, with the model showing increased dropout risk in transition grades, such as the final years of primary school and the early years of secondary school. To mitigate this risk, it is recommended that schools implement transition support initiatives such as peer mentorship, career guidance and academic counseling. Particular attention should be given to primary 6 and 7, as well as senior 1 and 2, where students are adjusting to new academic and social environments. Additionally, policies aimed at reducing secondary school costs and enhancing accessibility, especially for students transitioning from Universal Primary Education should be prioritized. Another significant cause of dropout was inadequate nutrition, as reflected by the meals per day variable. Hunger negatively affects student concentration, attendance and overall learning outcomes. Therefore, it is essential to scale up school feeding programs, particularly in regions where food insecurity is prevalent. These programs would ensure that children receive at least one nutritious meal a day while at school. Beyond that school environment, partnerships with agricultural and nutrition focused programs should aim to improve household food security to address the root cause of malnutrition. The living

arrangements of students also played a role in dropout risk. Students who do not live with their biological parents may lack emotional support and educational encouragement, making them more susceptible to dropping out. Schools and communities should therefore implement mentorship programs that provide these children with guidance and support. In addition, community-based initiatives should engage guardians and foster parents in sensitization programs that highlight the importance of providing consistent support for children's education. The variable 'school match' indicated that students attending schools that did not align with their needs or preferences were at higher risk of dropout. It is therefore recommended that the Ministry of Education improves school allocation processes, ensuring that students are enrolled in schools that meet their academic and social needs and are closer to their homes where possible. Additionally, efforts should be made to improve the learning environment and resource availability in underperforming schools thereby minimizing disparities in school quality and reducing the negative effects of school mismatch. Household indicators, such as ownership of a bicycle, television, or car, were found to influence dropout. Students from less wealthy households face challenges in accessing education resources and transportation. To address this, government and development partners should provide bicycles to students who travel long distances to school. Education material support and bursaries should also be extended to students from economically disadvantaged families to reduce the burden of schooling costs. Gender was another critical predictor, with girls facing unique challenges such as early marriage, menstruation-related absenteeism, and domestic responsibilities. To address this, education stakeholders should strengthen anti-child marriage campaigns and strictly enforce child protection laws. Schools should provide free sanitary and menstrual hygiene education to promote girls' health and dignity. Moreover, girls' empowerment programs should be integrated into school curricula and extracurricular activities to build their confidence and encourage continued school attendance. Disability was also identified as a significant dropout factor, with students facing infrastructural and social barriers within

schools. To promote inclusivity, schools should be upgraded to meet the needs of students with disabilities. This includes constructing ramps, accessible toilets and equipping schools with assistive learning devices. Furthermore, teachers should receive training in inclusive education practices and special needs pedagogy to better support these learners. The type of school attended (private or government) was another factor influencing dropout. This highlights disparities in cost and educational quality across school types. The government should invest in improving the quality of public schools to offer competitive learning environments and reduce dropout due to dissatisfaction or financial strain. Additionally, partnerships with affordable private schools could be fostered, offering subsidies or scholarships to ensure that financially disadvantaged students can attend schools that best meet their needs. Lastly, the educational level of household heads was found to influence student dropout rates. Low parental education levels limit parents' ability to support and value education. Adult literacy programs and parental education workshops should therefore be introduced in communities to equip parents with the skills and knowledge necessary to assist their children's learning at home. Promoting intergenerational learning programs can further strengthen the link between household education and student retention. Beyond these individual factors, I suggest that regular and ongoing monitoring and upgrading of prediction models is key. This can be achieved by consistently incorporating fresh data, facilitating through adjustments to evolving educational trends, and ensuring the continuous precision and relevance of information. Implementing an evaluation mechanism that continuously monitors the success of interventions and inform model improvements which will enhance effectiveness. By adopting these strategies, educational policymakers and practitioners in Uganda can develop efficient approaches to reducing student attrition rates, thereby enhancing educational achievements and fostering socioeconomic progress.

6.0.4 Limitations

While this study provides valuable insights, it recognizes potential limitations related to data quality and availability. Future research must prioritize incorporating more diverse and extensive datasets to further refine the validity and correctness of prediction models. The study primarily relied on secondary data, also it provides valuable insights, it was limited in scope and completeness. Certain important variables such as psychosocial factors, parental involvement beyond education level, peer influence, school leadership quality and community-level economic shocks were not captured in the data. This limited the range of predictors included in the ML models, potentially excluding other significant contributors of dropout. Secondly, the dataset exhibited class imbalance, with dropout cases forming a minority of observations. While the study addressed this issue through the application of the SMOTETomek resampling technique, balancing techniques may introduce synthetic patterns that do not perfectly reflect real-world complexities. Further studies using larger and more balanced datasets may yield more refined results. This study focused on primary and secondary schools in Uganda due to resource constraints that is to say the sample consisted of the principals, district education officers, parents, with limited representation from learners restricting the diversity of perspectives. This study used a qualitative research design to explore the causes and patterns of primary and secondary school dropouts in Uganda, aiming to identify effective strategies. A purposeful sampling approach ensured a homogeneous group of participants, meeting specific research needs. A key challenge was accessing current records at schools, which posed significant limitations during data collection since some of the information is not up-to-date.

6.0.5 Concluding Remarks

This study set out to address the challenge of student dropout in Uganda by applying machine learning and data-driven approaches. Through the use of algorithms such as Random Forest,

the research successfully predicted at-risk students and identified key factors contributing to dropout, including household size, grade level, disability and poverty. The findings provided evidence that machine learning can effectively complement traditional education management approaches. Beyond model development, the study proposed practical, data-driven interventions targeting the root cause of dropout. These include school feeding programs, financial support for vulnerable families, improved access for students with disabilities, and mentorship for students lacking parental support. While the study demonstrated promising results, it acknowledges limitations such as data gaps and the need for broader geographic coverage. Future research is encouraged to expand datasets, refine predictive models, and test these interventions in real-world school settings. In conclusion, this research contributes both to machine learning in education and to practical policy-making in Uganda, offering a pathway towards reducing dropout rates and supporting inclusive, equitable education for all learners.

REFERENCES

- [1] Adelman, M. A., & Szekely, M. (2016). *School dropout in Central America: An overview of trends, causes, consequences, and promising interventions*. World Bank Policy Research Working Paper, (7561).
- [2] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student performance prediction using machine learning techniques. *Education Sciences*, 11(9), 552.
- [3] Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2017). Predicting student dropout in higher education using machine learning. *Journal of Educational Data Mining*, 9(1), 1–25.
- [4] Almalki, S. (2016). Integrating quantitative and qualitative data in mixed methods research—challenges and benefits. *Journal of Education and Learning*, 5(3), 288–296.
- [5] Amadi, M. A., Role, E., & Makewa, L. N. (2013). Girl child dropout: Experiential teacher and student perceptions. *International Journal of Humanities and Social Science*, 3(5), 124–131.
- [6] Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271.
- [7] Amare, M. Y., & Šimonová, S. (2021). Global challenges of student dropout: A prediction model development using machine learning algorithms on higher education datasets. *SHS Web of Conferences*, 129, 09001.
- [8] Albán, M., Mauricio, D., & Albán, M. (2018). Decision trees for the early identification of university students at risk of desertion. *International Journal of Engineering and Technology*, 7, 51–54.

- [9] Batuwita, R., & Palade, V. (2012). A new performance measure for imbalanced classification. *Journal of Intelligent Information Systems*, 39(2), 271–289.
- [10] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [11] Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- [12] Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- [13] Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- [14] Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (Vol. 1, pp. 793–828). Wiley.
- [15] Bronfenbrenner, U. (1992). Ecological systems theory. In R. Vasta (Ed.), *Annals of Child Development* (Vol. 6, pp. 187–249). JAI Press.
- [16] Branson, N., Hofmeyr, C., & Lam, D. (2014). Progress through school and the determinants of school dropout in South Africa. *Development Southern Africa*, 31(1), 106–126. <https://doi.org/10.1080/0376835X.2013.853610>
- [17] Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction.
- [18] Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, 96, 346–353. <https://doi.org/10.1016/j.chilyouth.2018.11.030>

- [19] Do Carmo Nicoletti, M., & de Oliveira, O. L. (2020). A machine learning-based computational system proposal aiming at higher education dropout prediction. *Higher Education Studies*, 10, 12.
- [20] Eegdeman, I., Cornelisz, I., van Klaveren, C., & Meeter, M. (2022). Computer or teacher: Who predicts dropout best? *Frontiers in Education*, 7, 976922.
- [21] Er, E. (2023). An explainable machine learning approach to predicting and understanding dropouts in MOOCs. *Kastamonu Eğitim Dergisi*, 31, 143–154.
- [22] Fernández-García, A. J., Preciado, J. C., Melchor, F., Rodríguez-Echeverría, R., Conejero, J. M., & Sánchez-Figueroa, F. (2021). A real-life machine learning experience for predicting university dropout at different stages using academic data. *IEEE Access*, 9, 133076–133090.
- [23] Freeman, J., & Simonsen, B. (2015). Examining the impact of policy and practice interventions on high school dropout and school completion rates: A systematic review of the literature. *Review of Educational Research*, 85(2), 205–248.
- [24] Gambacorta, R., Marcucci, J., & Nobili, A. (2019). Predicting student dropout in higher education using machine learning. *Journal of Educational Data Mining*, 11(1), 1–25.
- [25] Gill, B., Borden, B. C., & Hallgren, K. (2014). A conceptual framework for data-driven decision making. *Final report by Mathematica Policy Research, Princeton*, submitted to Bill & Melinda Gates Foundation, Seattle, WA.
- [26] Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, ‘Big Data’ and principles for a critical race theory of statistics. *Race Ethnicity and Education*, 21(2), 158–179.

- [27] Glewwe, P., Siameh, C., Sun, B., & Wisniewski, S. (2021). School resources and educational outcomes in developing countries. In *The Routledge Handbook of the Economics of Education* (pp. 218–252).
- [28] Henderson, A. T., & Berla, N. (1994). *A new generation of evidence: The family is critical to student achievement*. National Committee for Citizens in Education.
- [29] Hollstein, B. (2011). Qualitative approaches. In *The SAGE Handbook of Social Network Analysis* (pp. 404–416).
- [30] Johnson, M. L. (2018). Empowering educators to make data-informed decisions: A district’s journey of effective data use. In *Data Leadership for K–12 Schools in a Time of Accountability* (pp. 158–183). IGI Global.
- [31] Jordan, L., Kostandini, G., & Mykerezi, E. (2012). Rural and urban high school dropout rates: Are they different? *Journal of Research in Rural Education (Online)*, 27(12), 1.
- [32] Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Proceedings of the EMNLP Workshop on Analysis of Large-Scale Social Interaction in MOOCs*, Doha, Qatar, 25 October 2014.
- [33] Kiss, V., Maldonado, E., & Segall, M. (2022). The use of semester course data for machine learning prediction of college dropout rates. *Journal of Higher Education Theory and Practice*, 22, 64–74.
- [34] Kabay, S. (2016). Grade repetition and primary school dropout in Uganda. *Harvard Educational Review*, 86(4), 580–606.
- [35] Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. SAGE Publications.

- [36] Kim, J., Kim, B., & Lee, Y. (2018). The effects of socioeconomic factors on student dropout rates. *Journal of Educational Research*, 111(4), 432–441.
- [37] Klees, S. J., Samoff, J., & Stromquist, N. P. (Eds.). (2012). *The World Bank and education: Critiques and alternatives* (Vol. 14). Springer Science & Business Media.
- [38] Lee, S., & Chung, J. Y. (2019). The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences*, 9, 3093.
- [39] Lavalley, M. (2018). *Out of the loop: Rural schools are largely left out of research and policy discussions, exacerbating poverty, inequity, and isolation*. Center for Public Education.
- [40] Li, M., Lou, W., & Ren, K. (2010). Data security and privacy in wireless body area networks. *IEEE Wireless Communications*, 17(1), 51–58.
- [41] Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). *Descriptive analysis in education: A guide for researchers*. NCEE 2017-4023. National Center for Education Evaluation and Regional Assistance.
- [42] Lubaale, Y. A. M. (2010). Orphans and vulnerable children in Uganda: Is it a homogeneous group? (Unpublished PhD thesis). Makerere University.
- [43] McCleary-Sills, J., Hanmer, L., Parsons, J., & Klugman, J. (2015). Child marriage: A critical barrier to girls' schooling and gender equality in education. *The Review of Faith & International Affairs*, 13(3), 69–80.
- [44] McDermott, E. R., Donlan, A. E., & Zaff, J. F. (2019). Why do students drop out? Turning points and long-term experiences. *The Journal of Educational Research*, 112(2), 270–282.

- [45] Mduma, N. (2020). Data driven approach for predicting student dropout in secondary schools (Doctoral dissertation, NM-AIST).
- [46] Mduma, N., Kalegele, K., & Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction.
- [47] Menoyo-Ros, D., Garcia-Cabot, A., Garcia-Lopez, E., & Domínguez, A. (2020). The use of machine learning in educational datasets. *Eden Conference Proceedings*, 131–140.
- [48] Mnyawami, Y. N., Maziku, H., & Mushi, J. C. (2022). Implementation of Bayesian hyperparameter optimization for predicting student dropout in Sub-Saharan Africa secondary schools. *Research Square*, preprint.
- [49] Mduma, N., Kalegele, K., & Machuve, D. (2019). An ensemble predictive model based prototype for student drop-out in secondary schools. *Journal of Information Systems Engineering & Management*, 4, em0094.
- [50] Oqaidi, K., Aouhassi, S., & Mansouri, K. (2022). Towards a students' dropout prediction model in higher education institutions using machine learning algorithms. *International Journal of Emerging Technologies in Learning (iJET)*, 17, 103–117.
- [51] Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L., Leiva, V., & Marchant, C. (Year not provided). Knowledge [Incomplete citation].
- [52] Papa Mitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4), 49–64.
- [53] Parr, A. K., & Bonitz, V. S. (2015). Role of family background, student behaviours, and school-related beliefs in predicting high school dropout. *The Journal of Educational Research*, 108(6), 504–514.

- [54] Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*.
- [55] Psaki, S., Haberland, N., Kozak, M. L., & Woyczynski, L. (2021). *Girls' education roadmap: 2021 report*.
- [56] Russell, D. J., Rivard, L. M., Walter, S. D., Rosenbaum, P. L., Roxborough, L., Cameron, D., ... & Avery, L. M. (2010). Using knowledge brokers to facilitate the uptake of paediatric measurement tools into clinical practice: A before-after intervention study. *Implementation Science*, 5, 1–17.
- [57] Save the Children International. (2011). *The 'invisible' children: Inequalities in education in the Commonwealth*.
- [58] Slater, S., Joksimović, S., Kovanović, V., Baker, R. S., & Gašević, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, 42(1), 85–106.
- [59] Stalla-Bourdillon, S., & Knight, A. (2016). Anonymous data v. personal data—false debate: An EU perspective on anonymization, pseudonymization and personal data. *Wisconsin International Law Journal*, 34, 284.
- [60] Tembo, A. (2022). Challenges in the implementation and provision of quality free primary education in public primary schools in Nyimba district (Doctoral dissertation, The University of Zambia).
- [61] UNESCO. (2018). *Uganda: Education and literacy*. UNESCO Institute for Statistics.
- [62] Unterhalter, E., Arnot, M., Lloyd, C., Moletsane, L., Murphy-Graham, E., North, A., ... & Saito, M. (2014). *Interventions to enhance girls' education and gender equality: Education rigorous literature review*. London: Department for International Development.

- [63] Upton, G., & Burton, M. (2015). *Developmental psychology*. Routledge.
- [64] UNICEF. (2012). *Global Initiative on out-of-school children: Ghana country study*. Retrieved from www.unicef.org (accessed 12.11.2013).
- [65] UBOS. (2007, 2012). *Uganda Demographic and Health Survey*.
- [66] Uganda Bureau of Statistics. (2021). *Uganda National Household Survey 2019–2020 report*. Retrieved from https://www.ubos.org/wp-content/uploads/publications/09_2021Uganda-National-Survey-Report-2019-2020.pdf
- [67] Uwezo at Twaweza East Africa. (2016, June 02). *Learning levels in Uganda (2014)* [Dataset]. Retrieved from <https://data.humdata.org/dataset/learning-levels-in-uganda-2014> (accessed 10-07-2024).
- [68] UNESCO. (2017). *Education for Sustainable Development: A framework for action*.
- [69] Wagner, C. C., & Watkins, M. W. (2010). The effects of family and community factors on student engagement. *Journal of Educational Research*, 103(4), 251–263.
- [70] United Nations Population Fund. (2022). *State of Uganda’s population report 2022*. Retrieved from <https://uganda.unfpa.org/en/publications/state-ugandas-population-report-supre-2022>
- [71] Twaweza. (n.d.). *Uwezo datasets*. Retrieved from <https://www.twaweza.org/go/uwezo-datasets>



UGANDA CHRISTIAN UNIVERSITY

A Centre of Excellence in the Heart of Africa

SCHOOL OF RESEARCH & POSTGRADUATE STUDIES DISSERTATION CORRECTION COMPLIANCE FORM (POST VIVA FORM)

Date: 03-07-2025

Name of Candidate: NAKIMBUGWE DIANA KIRABO. No: M23M19/278

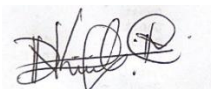
Title of Dissertation: Tackling Dropout Rates of Students in Uganda: An Exploration of Machine Learning and Data-Driven Approaches

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	Report requires additional editing for clarity	Revised the full document for clarity, coherence, and academic language, ensuring logical flow across chapters.	All
2	Chapter Four does not clearly show how objectives were achieved.	Expanded Chapter 4 by adding a new Section 4.0.7 titled "Mapping Objectives to Results and Contributions" to explicitly show how each objective was achieved.	Page 84 to 86
3	The summary of the literature review lacks clarity on how the literature extends knowledge	Rewrote the literature review summary (Section 2.0.11) to clearly state how this study extends prior research and fills gaps	Page 50
4	Literature review does not describe how the work contributes to Machine Learning and Data-Driven Approaches.	Added a new subsection in Chapter 2.0.10 discussing the study's contributions to machine learning and data-driven educational research.	Page 49
5	Contribution is not coming out clearly	Clearly highlighted the study's contribution to machine learning, educational research, and practical interventions in Uganda, both in Chapter 4.07 and 4.08 in the Conclusion.	Page 85-87
6	Concluding remarks were brief and lacked impact	Rewrote the concluding remarks to summarize findings, highlight contributions, and suggest future directions for research.	Page 91-96

S/N	COMMENTS BY INTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	Page numbering	Numbered the pages with Chapter 1 being Page 1	All
2	Limitations of the study were not clearly stated	Added a detailed Limitations section (Section 6.0.4) outlining the key limitations of data, model scope, and implementation.	Page 95

S/N	COMMENTS BY VIVA VOCE PANNEL	ACTION TAKEN	INDICATOR
1	Why did I use both dropout and outcomes in the title?	Removed outcomes since it wasn't measurable.	Title
2	Recommendations not tallying with the causes of dropout	Rewrote the entire Recommendations section (Chapter 6) to match the causes of dropout identified through machine learning results, making recommendations cause-specific.	Page 92 to 94
3	Report should be formatted using LaTeX	Formatted the report using LaTeX	All

Candidate's Name: NAKIMBUGWE DIANA KIRABO Signature



Supervisor's Name/ Signature



NB: Post Viva compliance form is designed to capture all the corrections recommended by internal examiner (supervisor), external examiner and viva panel.