A HYBRID MODEL FOR PREDICTING E-LEARNING COURSE DROPOUT RATE FOR POST GRADUATE STUDENTS

BY

WINFRED KITAKA

MASTERS OF SCIENCE IN INFORMATION SYSTEMS MANAGEMENT

KCA UNIVERSITY

2023

A HYBRID MODEL FOR PREDICTING E-LEARNING COURSE DROPOUT RATE FOR POST GRADUATE STUDENTS

BY

WINFRED KITAKA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF MASTERS OF SCIENCE IN INFORMATION SYSTEMS MANAGEMENT IN THE SCHOOL OF TECHNOLOGY AT KCA UNIVERSITY

OCTOBER, 2023

DECLARATION

I declare that this dissertation is my original work and has not been previously published or submitted elsewhere for award of a degree. I also declare that this contains no material written or published by other people except where due reference is made and author duly acknowledged.

Student Name: Winfred Kithaka Reg,No. 21/0804

Sign: ______________________ Date: 31/10/2023

This thesis has been submitted for examination with my approval as the appointed university supervisor.

Dr. Gabriel Kamau

Signature: ……… ………………….Date: 31/10/2023

Dissertation Supervisor

ABSTRACT

In universities all around Kenya, e-learning has grown in popularity, especially for postgraduate programs. Students now find it simpler to access education from any location at any time thanks to the use of technology in the delivery of courses and academic resources. However, dropout rates continue to be a serious issue despite the many advantages of online learning. For a variety of reasons, including a lack of desire, insufficient assistance, and trouble understanding the course materials, students withdraw from online courses. Dropouts drive up educational institutions' average cost per student because it typically costs more to retain a possible dropout than to enroll a new student. The rise of online learning is hampered by the prevalence of school dropouts, which waste the student's initial time and financial investment. Low graduation rates that follow high dropout rates will surely damage the standing of educational institutions in the community and eventually result in a downward loop of declining government support. To lower the dropout rate, online educational institutions can employ this technology to quickly spot probable dropouts and put retention measures in place before the dropout behavior takes place. The study's objective was to create a hybrid machine learning prediction model for postgraduate E-learning students who drop out utilizing the Support Vector Machine and Random Forest algorithms to improve prediction accuracy. The researcher employed a descriptive survey and an experimental study approach. The research methodology will be appropriate because the researcher trained the Dropout Prediction Detection model using a machine learning technique. In 2024, 61.7% of students are expected to graduate. With the aid of the data, the researcher was better able to determine whether students had spent more time studying than was anticipated. 62.5% of the respondent's price posed the most challenge to finishing the investigation, however 37.5% of the fee posed no issue. In order to prepare students for postgraduate study, 68.3% strongly agreed that undergraduates should be taught research techniques, and 29.2% also agreed. A 100% accuracy rate for forecasting student dropout was demonstrated by the hybrid model. By using machine learning to predict student attrition, educational institutions have a ground-breaking chance to effectively address this pervasive problem. The study also recommended that the students choose a study strategy that would best fit their schedules in order to prevent unneeded stress from juggling numerous tasks at once. Deep learning models can be strengthened by techniques like Synthetic Minority Over-sampling Technique to handle the unbalanced datasets typical in dropout prediction problems.

ACKNOWLEDGEMENT

I thank the almighty God for allowing me to undertake this study. My sincere gratitude also goes to my parents for the good upbringing. Your everlasting love has been the greatest treasure of my life. Your value for education has sincerely pushed us as a family to the limits of academic excellence as a family. To my brothers and sisters, thank you for taking your time to remind me of the goals that our parents wanted us to achieve. Special thanks also go to my Supervisor, Dr. Gabriel Kamau for his invaluable guidance in writing this project. This document is what it is due to his dedicated professional guidance. I therefore do extend my gratitude to him for accepting to invest his knowledge in me. I am also grateful to my classmates who morally encouraged and supported me. I also thank KCAU University and the entire staff for giving me an enabling opportunity to undertake my studies.

TABLE OF CONTENTS

DEDICATION

This thesis is dedicated to my family for their prayers and moral support that they offered me throughout the course of my studies.

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

In universities all around Kenya, e-learning has grown in popularity, especially for postgraduate programs. Students now find it simpler to access education from any location at any time thanks to the use of technology in the delivery of courses and academic resources. However, dropout rates continue to be a serious issue despite the many advantages of online learning. For a variety of reasons, including a lack of desire, insufficient assistance, and trouble understanding the course materials, students withdraw from online courses.

E-learning, also known as online learning or distance education, is a mode of education that utilizes digital technologies to deliver educational content and facilitate communication between instructors and learners. It has become progressively prevalent in recent years due to its flexibility and accessibility, allowing learners to access educational materials and participate in courses from any location and at any time. According to a report by Research and Markets, the global e-learning market is expected to reach $336.98 billion by 2026, with a compound annual growth rate of 9.1% (Research and Markets, 2020). E-learning has been found to be effective in improving learning outcomes, with a meta-analysis of 50 studies finding that online learning results in higher levels of learning achievement compared to traditional face-to-face instruction (Means et al., 2010). Student dropout is a major concern in education, as it can have negative consequences for students' future opportunities and the overall effectiveness of educational systems. According to the National Center for Education Statistics (NCES), the overall high school dropout rate in the United States was 5.1% for the 2016-2017 school year (NCES, 2019). Dropout rates are also a concern in higher education, with one study finding that the average six-year graduation rate for students at four-year institutions in the United States was 60% (National Student Clearinghouse Research

Center, 2019). The reasons for student dropout are complex and can include academic difficulties, financial challenges, lack of social support, and personal factors such as mental health issues (Tinto, 1993). In order to address the issue of student dropout, researchers have developed interventions such as academic and social support programs, financial aid, and personalized counseling (De Witte & Cabus, 2016).

Globally, E-learning has become gradually widespread particularly in higher education, due to its flexibility and accessibility. However, high dropout rates remain a concern in e-learning environments worldwide. According to a study by the Organisation for Economic Co-operation and Development (OECD), the average dropout rate for higher education students in OECD countries was 31% in 2016 (OECD, 2019). The study also found that dropout rates were higher for part-time students and those studying at private institutions. In developing countries, e-learning has been implemented as a means of increasing access to education for underserved populations. However, dropout rates in these settings can be particularly high due to challenges such as lack of reliable internet access, limited resources and support, and cultural and linguistic barriers (Ekanayake & Wishart, 2017). To address the issue of student dropout in e-learning, researchers have developed predictive models and interventions such as personalized support and counseling, additional resources and materials, and increased communication with instructors and peers, (Huang et al. 2019).

According to earlier research by Peng et al. (2022), medical students drop out at a rate that varies from 3.8% to 26%. Research has indicated a robust correlation between psychological issues (such as depression, burnout, anxiety, and alcohol misuse) and the likelihood of dropping out. The majority of the investigations, nevertheless, took place outside of China. The risk of dropout among Chinese medical students has been the subject of very few published researches to date. According to a cross-sectional study, 6.9% of the 1,837 medical students reported being less inclined to pursue a career in medicine following the COVID-19 outbreak. This was significantly correlated with depressive symptoms, low levels of professional satisfaction, being young, female, having low income, and hearing unfavorable news about the pandemic.

Since residency training became standardized and postgraduate medical education was reformed, postgraduates have grown in significance within the healthcare industry. China Health Statistical Yearbook reports that there were 271,406 medical postgraduates in 2018, about twice as many as there were in 2008. In China, the percentage of doctors holding master's or doctoral degrees rose from 11.4% in 2010 to 20.3% in 2018. China has a three-tiered medical school system, comprising undergraduate, master's, and doctorate levels, in contrast to other nations. Medical postgraduates in China are expected to finish both their clinical and academic work concurrently, and they typically receive more years of training than their counterparts in non-medical disciplines (Xiao et al. 2020).

E-learning in Africa faces challenges such as poor internet connectivity, limited access to devices and infrastructure, and insufficient training for teachers and students. These challenges contribute to high dropout rates in e-learning programs, similar to what has been observed in Kenya. Therefore, the development of predictive models to prevent course dropouts and the provision of adequate support to students can be instrumental in improving the effectiveness of e-learning programs in Africa. For instance, e-learning in Cameroon highlights the need for effective e-learning programs that can improve access to education and address the challenges facing traditional education systems in the country (Voufack & Noubissi (2020).

The Kenyan government is aware of the value of e-learning in tackling problems in the education sector such a lack of funding and a teacher shortage. The Kenyan government's Vision 2030 plan emphasizes the importance of incorporating technology into the educational system to raise the standard of instruction and increase accessibility for all Kenyans. By addressing the problems that the nation's e-learning programs are encountering, the study's findings can aid in the implementation of this vision (Kandonga & Mutinda, 2020). Lack of technological infrastructure, inadequate instructor support, and poor course design contribute to student dropout (Cheptumo & Sang, 2020), Wobusobozi, & Tushemereirwe, 2021). Overall, the problem of students drop out has been a concern and requires intervention. This problem seems to have ballooned in Kenya with the increasing number of postgraduate students seeking to advance their career (Massawe & Ismail, 2021).

1.2 Problem statement

High dropout rates are detrimental to educational institutions, students, and the growth of online learning. Additionally, they are harmful to each individually. Dropouts drive up educational institutions' average cost per student because it typically costs more to retain a possible dropout than to enroll a new student. The rise of online learning is hampered by the prevalence of school dropouts, which waste the student's initial time and financial investment. Low graduation rates that follow high dropout rates will surely damage the standing of educational institutions in the community and eventually result in a downward loop of declining government support. Tan et al. (2015) investigated dropout rates using Artificial Neural Networks (ARR), Decision Trees (DT), and Bayesian Networks (BN). There are some discrepancies between the three prediction models with regard to the recall rate and precision rate for the dropout class: The DT and BN both had the highest accuracy rates (63.89% and 63.39%, respectively), while the ANN had the lowest accuracy rate (53.54%). The prediction model's total effectiveness is shown by the overall accuracy rate. In that order, DT (94.63%), ANN (93.97%), and BN (93.92%) have the highest overall accuracy rates. All three models had overall accuracy rates that were higher than 93%.

All three of these prediction models were shown to have a passable track record for predicting student dropout or retention rates, with the DT showing much better prediction outcomes.

According to Barasa and Omulando (2018), just 11% of Kenyan PhD applicants complete their studies in six years on average. Other factors of a student's history that can impact performance outside financial aid include age, gender, nationality, and background. Using random forest, Abubakar et al. (2017) studied how to predict students' performance in online learning. Naive Bayes classifier performed better than Random Forest algorithm, which was superior to KNN and Decision Tree in terms of prediction accuracy. Random Forest recorded accuracy of 76.9% in comparison to KNN and Decision Tree, which recorded accuracy of 69.2% and 61.5%, respectively. They used vector machines (SVM), an efficient supervised machine learning method that is widely used for classification issues. When building a prediction model, SVM can be employed. SVMs are less likely to overfit, especially in multidimensional spaces. Overfitting could be a problem because prediction models need to perform effectively when generalizing to new, unidentified queries. SVMs can effectively handle this problem if they are regularized properly.

In a postgraduate study on attrition in Nairobi's private universities, Mukami (2016) discovered a 37% failure rate. Rong'uno (2016) conducted a second study in Kenya on the factors impacting the completion rates of PhD candidates. It revealed a 50% drop in cohort enrollment from 2001 to 2008. To improve the model's accuracy, the researcher hopes to integrate the random forest with Support Vector Machine (SVM). After training, the generated prediction model samples can be used to forecast future dropout behavior. To lower the dropout rate, online educational institutions can employ this technology to quickly spot probable dropouts and put retention measures in place before the dropout behavior takes place. The study aimed to improve accuracy predictability using the hybrid model.

1.3 Objectives of the study

1.3.1 General objective

The goal of the study was to develop a hybrid machine learning prediction model for postgraduate E-learning students' dropouts using random forest algorithm and Support Vector Machine to obtain a better prediction accuracy.

1.3.2 Specific objectives

The specific objectives were:

i. To analyse the factors that influence e-learning students dropout rates among post graduate.
ii. To determine mitigation measures for reducing the drop out of postgraduate e-learning students by 20% among the selected public universities in Kenya.
iii. To develop a Hybrid model for predicting e-learning course dropout rate for postgraduate.
iv. To validate the developed postgraduate prediction model among the eLearning students.

1.4 Research questions

The study was guided by the following research questions.

i. What were the factors that influence postgraduate e-learning university students' dropout rate?
ii. What were the mitigation measures for reducing the drop out of postgraduate e-learning students by 20% among the selected public universities in Kenya?
iii. How was hybrid model be developed in predicting e-learning course dropout rate for postgraduate?
iv. How was the developed prediction model be validated?

1.4 Significance of the study

The high rate of e-learning course dropouts among postgraduate students is a significant problem that needs to be addressed. E-learning provides students with the flexibility to access educational resources remotely, but it also comes with challenges, such as lack of face-to-face interaction, technical difficulties, and the need for self-motivation and time-management skills. Postgraduate students, in particular, are often juggling academic work with other commitments, such as work and family, making it more challenging to complete e-learning courses successfully. The development of accurate predictive models that can identify at-risk postgraduate e-learning students is essential for improving course completion rates and reducing dropout rates. By identifying at-risk students early on, interventions can be implemented to support and guide them towards successful course completion. This, in turn, can lead to a more productive and satisfied workforce, as postgraduate students often go on to become leaders in their respective fields. Moreover, the COVID-19 pandemic has accelerated the adoption of e-learning, making it more critical than ever to address the issue of e-learning course dropouts among postgraduate students. As educational institutions increasingly rely on e-learning to deliver educational content, the development of accurate predictive models can help ensure that students are provided with high-quality education and attain their educational goals. The development of accurate predictive models to identify at-risk postgraduate students in e-learning courses has significant implications for the education industry, workforce productivity, and individual career prospects. By addressing the issue of e-learning course dropouts among postgraduate students, the study contributed to improving the quality and accessibility of higher education and support the development of a highly skilled and diverse workforce.

I was inspired to conduct research on postgraduate dropout rates in universities because it's a big project that can give important details about the variables affecting higher education completion rates. Governmental and institutional policy changes can be influenced by an understanding of the reasons behind student dropout rates. Universities and legislators can put plans in place to increase retention rates by figuring out the underlying factors.

1.5 Scope of the study

The scope of studying the prediction of e-learning course dropouts for postgraduate students is broad and encompasses various aspects related to e-learning, student engagement, and academic performance. The study aimed to develop accurate predictive models that can identify at-risk postgraduate students and provide timely interventions to improve course completion rates. The scope of the study included the following: The study aimed to identify the factors that contribute to e-learning course dropouts among postgraduate students. These factors included personal and professional commitments, technical difficulties, lack of motivation, and inadequate course content. The study sought to evaluate the effectiveness of different machine learning algorithms and models in predicting e-learning course dropouts among postgraduate students. This included exploring the use of logistic regression, decision trees, random forest, and support vector machines. The study aimed to develop and test intervention strategies that can support and guide at-risk postgraduate students towards successful course completion. This included personalized feedback, mentoring, and coaching, among others. The study aimed to ensure that the results are generalizable across different educational settings and e-learning platforms. This involved considering different types of e-learning courses, such as fully online, blended, and self-paced courses.

The next chapter provided a review of relevant literature on e-learning, course dropouts, and predictive modeling. The literature review will highlight the existing knowledge gaps and areas for further research.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

In this chapter, a review of past and related studies was presented. The aim of reviewing related literature was to form a basis for showing the status of progression rate of students as presented by other studies. Gaps that were left out by other studies were discovered and give the study a firm foundation to show that it is necessary to undertake this study. The study reviewed literature related to progression rate of students, enrolment rate of students, dropout rate for students and deferment rate for students. The study was carried under the current trends in university education in Kenya and therefore the study believes that valuable information can be gained. In order to build the case for doing this study, several theories relating to the variable mentioned in the study were explained in detail. The study proceeded by discussing the theoretical framework, the conceptual framework and the empirical framework as the subsections of the chapter on literature review.

2.2 Theoretical Literature

Self-determination theory (SDT) is a motivational framework that has been widely applied in the study of e-learning course dropouts for postgraduate students. Weng et al. (2021) examined the impact of feedback interventions on postgraduate students' motivation and dropout rates in a self-directed e-learning course, finding that providing personalized feedback based on SDT principles improved students' motivation and reduced dropout rates. Gong and Wang (2020) investigated the mediating role of academic self-efficacy in the relationship between postgraduate students' perceived autonomy support and their course completion in an e-learning environment, finding that self-efficacy partially mediated the relationship. Guo et al. (2020) examined the role of students' autonomous motivation and perceived competence in predicting their intention to persist in a postgraduate e-learning course, finding that both variables were significant predictors of intention to persist. Yu and Hu (2020) investigated the impact of students' perceived autonomy support and intrinsic motivation on their academic achievement and course completion in a postgraduate e-learning course, finding that both variables were positively associated with academic achievement and course completion. Zhang et al. (2020) examined the impact of teacher autonomy support and perceived competence on postgraduate students' academic outcomes in a blended learning environment, finding that both variables were positively associated with course completion and satisfaction. Liu et al. (2019) investigated the role of self-regulation and intrinsic motivation in predicting postgraduate students' academic achievement and course completion in a blended learning environment, finding that both variables were significant predictors. Zhang et al. (2019) examined the impact of self-determined motivation and perceived competence on postgraduate students' academic achievement and course completion in an e-learning environment, finding that both variables were significant predictors of academic achievement and course completion.

Self-regulated learning (SRL) theory is another framework that has been widely applied in the study of e-learning course dropouts for postgraduate students. Here are ten of the most current citations of SRL related to this study: Zhang and Zhao (2022) investigated the impact of self-regulated learning strategies on postgraduate students' academic achievement and course completion in an e-learning environment, finding that the use of SRL strategies was positively associated with both outcomes. Wei et al. (2021) examined the role of self-regulated learning in predicting postgraduate students' academic achievement and course completion in a blended learning environment, finding that self-regulation was a significant predictor of both outcomes. Li et al. (2021) investigated the relationship between self-regulated learning, academic procrastination, and course completion in a postgraduate e-learning course, finding that self-regulated learning was negatively associated with academic procrastination and positively associated with course completion. Wang and Yang (2021) examined the impact of self-regulated learning on postgraduate students' academic achievement and course completion in an asynchronous online course, finding that self-regulation was a significant predictor of both outcomes.

Self-Determination Theory (SDT) is a psychological framework developed by Deci and Ryan in the 1980s. It focuses on the motivation behind people's choices and actions. According to SDT, there are three innate psychological needs that influence human motivation and well-being:

a) *Autonomy.* The need to feel in control of one's own actions and choices.
b) *Competence.* The need to feel capable and effective in one's interactions with the environment.
c) *Relatedness.* The need to feel connected to others and experience a sense of belonging.

When it comes to understanding postgraduate dropouts in the context of SDT, several key factors can be considered:

a) *Autonomy: Lack of Autonomy:* Postgraduate students might drop out if they feel a lack of control over their academic or research direction. This lack of autonomy can lead to disengagement and ultimately dropping out.
b) *Competence: Feelings of Incompetence:* If students perceive the postgraduate program as too challenging or beyond their capabilities, they might lose motivation to continue. A feeling of incompetence can erode their confidence and lead to dropout.
c) *Relatedness: Isolation*: Postgraduate students who don't find a sense of community or don't establish meaningful relationships with peers or faculty might feel isolated. This lack of relatedness can lead to feelings of alienation and increase the likelihood of dropout.

d) *Poor Advisor-Student Relationship:* A strained relationship between a student and their advisor can also lead to a lack of relatedness, making students more prone to dropping out.

How SDT Can Help Address Postgraduate Dropouts:

a) *Supporting Autonomy:* Institutions can provide students with more autonomy in choosing research topics or coursework. Allowing them to have a say in their academic journey can increase their motivation and persistence.

b) *Enhancing Competence*: Providing additional academic support, mentoring, or tutoring can help students develop the skills and confidence needed to succeed in their studies.

c) *Fostering Relatedness:* Creating a supportive and inclusive academic community through group projects, seminars, and social events can enhance students' sense of relatedness and reduce feelings of isolation.

d) *Improving Advisor-Student Relationships:* Institutions should provide training and resources to advisors to improve communication and mentorship skills, ensuring that students have a positive and supportive relationship with their advisors.

Understanding and addressing these psychological needs based on SDT can potentially help reduce postgraduate dropouts by creating a more supportive and motivating learning environment.

2.3 Empirical Literature

2.3.1 Random forest Algorithm Model

Random Forest is a supervised ensemble machine learning technique for classification, regression, and other problems that generates a number of decision trees during training and outputs the class that is the average of the classes of the individual trees. In contrast to Decision Tree, where each node is divided using the best among the features, each node in Random Forest is split using the best among a subset of predictors that are randomly selected at the node. With the aid of this strategy, Random Forest performs better than a number of other classification techniques, including Discriminant Analysis, Support Vector Machines, and Neural Networks, and it is resistant to overfitting. In the random forest scenario, several classification and regression trees are constructed using randomly selected training datasets and randomly.

The results from each tree are combined to offer a prediction for each observation. Since it can preserve certain benefits of tree models, like the ability to grasp relationships between variables and outcome, random forest frequently gives more accuracy than a single decision tree model (Speiser, Durkalski, & Lee, 2015). When compared to other models, random forests consistently offer among the highest prediction accuracy in the classification context (Fernandez-Delgado, Cernadas, Barro, & Amorim, 2014).

However, in practice, it is frequently preferable to reduce the number of predictors required in order to achieve outcome predictions, as stated by Speiser et al. (2019). One major benefit of using random forests for prediction modeling is their ability to handle datasets with a large number of predictor variables. For instance, instead of using all the variables included in the electronic medical record while building a medical prediction model, one could choose to use only a portion of the most important parameters. Prediction modeling typically focuses on identifying the most important predictors to incorporate in a compact, effective model. Variable selection, which determines the best predictors based on statistical characteristics like accuracy or relevance, can be used to do this. The cost of data collection may be reduced and the viability of prediction may be increased by using variable selection to build prediction models. Because many modern datasets contain hundreds or thousands of potential predictors, variable selection is frequently a step that is necessary in the construction of prediction models.

Variable selection in the random forest architecture is a crucial consideration for many expert systems and applications. The primary goal of the majority of expert systems is to assist with complex situation decision-making. our is consistent with the aim of prediction modeling, where we use a dataset to develop a model (random forest in our study) that will forecast an important outcome. To improve the effectiveness of acquiring model predictions, variable selection may be used to choose a subset of predictor variables to be included in a final, simplified model (Speiser et al., 2019).

The random forest technique has been used in several research to forecast e-learning course dropout. The random forest approach, for instance, was employed by Chen et al. (2022) to forecast dropout from a MOOC. They constructed the random forest model using characteristics linked to student involvement, demographic data, and learning results. The outcomes demonstrated that the random forest approach performed better than other machine learning techniques, including decision trees and logistic regression. It has been demonstrated that the random forest algorithm has great prediction accuracy in the field of forecasting. Liu et al. 2022, Yang et al. 2021, and Wang et al. 2021, for examples. This influences the model that was selected for this investigation. The random forest model has, however, not been used to forecast Kenya's e-learning student dropout rate, while being employed in other nations.

A versatile machine learning technique called Random Forest can be applied to both classification and regression tasks. It is an ensemble learning technique, which means that it integrates the results of various different models to provide predictions that are more precise. The Random Forest algorithm operates as follows:

1. *Decision Trees:* Random Forest is made up of a collection of decision trees. Decision trees split the data into subsets based on the value of input features. Each subset then becomes the input for a sub-tree. The process of creating decision trees involves finding the best feature to split the data and continues recursively until a stopping criterion is met, such as a maximum tree depth or a minimum number of samples in a leaf node.

2. *Bagging:* Random Forest employs a technique called bagging (Bootstrap Aggregating) to create multiple subsets of the original dataset. Each subset is sampled with replacement, meaning that some data points may appear multiple times in a subset, while others might not appear at all. These subsets are used to train individual decision trees.

3. *Random Feature Selection:* When creating each decision tree, Random Forest only considers a random subset of features at each split. This randomness helps in decorrelating the trees and ensures that the ensemble model is not dominated by a few powerful features. The number of features considered at each split is a hyperparameter that can be tuned.

4. *Voting (Classification) or Averaging (Regression):* For classification tasks, each decision tree in the Random Forest predicts a class, and the class with the majority of votes becomes the final prediction. For regression tasks, each tree predicts a numerical value, and the final prediction is the average of these values.

Advantages of Random Forest Algorithm:

High Accuracy: Random Forest generally provides higher accuracy compared to individual decision trees, especially for complex datasets.

Handles Missing Values: Random Forest can handle missing values in the predictor variables.

Feature Importance: It can provide a ranking of feature importance, helping in feature selection.

Reduces Overfitting: By averaging the predictions of multiple trees, Random Forest reduces overfitting, which is a common problem with decision trees.

Limitations of Random Forest Algorithm:

Interpretability: The ensemble nature of Random Forest makes it harder to interpret compared to individual decision trees.

Computational Complexity: Training multiple decision trees can be computationally expensive, especially for large datasets.

When using the Random Forest algorithm, it's important to tune hyperparameters such as the number of trees in the forest, the maximum depth of each tree, and the number of features considered at each split to achieve the best performance for a specific dataset.

2.3.2 Related literature on course dropout

Zhang et al. (2020) investigated the impact of self-regulated learning on postgraduate students' academic outcomes in a blended learning environment, finding that self-regulation was positively associated with course completion and satisfaction. Wang and Yang (2020) examined the impact of self-regulated learning on postgraduate students' academic achievement and course completion in a synchronous online course, finding that self-regulation was a significant predictor of both outcomes. Yao et al. (2020) investigated the relationship between self-regulated learning and academic performance in a postgraduate e-learning course, finding that self-regulated learning was positively associated with academic performance. Liu et al. (2020) examined the impact of a self-regulated learning intervention on postgraduate students' academic achievement and course completion in an e-learning environment, finding that the intervention improved both outcomes. Chen et al. (2019) investigated the impact of self-regulated learning on postgraduate students' academic achievement and course completion in a blended learning environment, finding that self-regulation was a significant predictor of both outcomes. Liu and Wang (2019) examined the impact of self-regulated learning on postgraduate students' academic achievement and course completion in an e-learning environment, finding that self-regulation was a significant predictor of both outcomes.

In Kenya, E-learning has gained significant attention in Kenya as a means of expanding access to higher education and improving the quality of education (Karanja et al., 2017). E-learning is defined as a mode of learning that utilizes electronic technologies to deliver educational content to students (Sang et al., 2018). In Kenya, e-learning is offered through various platforms, such as learning management systems (LMS), Massive Open Online Courses (MOOCs), and virtual classrooms (Karanja et al., 2017). Despite the potential benefits of e-learning, there are several challenges that students face in this mode of learning. These challenges include poor internet connectivity, inadequate technical support, lack of interaction with peers and instructors, and poor course design (Wambui & Kihoro, 2019). These challenges can contribute to e-learning course dropout among postgraduate students in universities in Kenya.

There are several factors that can contribute to e-learning course dropout. Some of these factors include lack of engagement, poor motivation, and inadequate support (Kemp, 2016). Other factors may include technical problems, difficulties in navigating the course, and time constraints. According to a study by Huang et al. (2014), factors such as personal and academic characteristics, prior experience, and learning styles can also contribute to e-learning course dropout. E-learning has become increasingly popular over the past few years, especially in higher education. However, one of the biggest challenges faced by universities and other educational institutions offering e-learning courses is the issue of course dropout. Dropout is defined as the situation where students withdraw or leave the course before completion. Dropout rates in e-learning courses have been reported to be higher than traditional classroom-based courses (Kemp, 2016). The ability to predict the likelihood of dropout can help instructors and administrators take proactive steps to prevent it from happening.

According to Chen et al. (2021), who studied the effects of perceived teacher autonomy support on postgraduate students' autonomous motivation and course completion in online learning environments, autonomous motivation was found to be positively correlated with autonomy support at higher levels, as were dropout rates. In a study by Rais et al. (2021), it was discovered that psychological requirements for autonomy, competence, and relatedness were the ones that were most strongly correlated with postgraduate students' intention to persist in online courses. In a postgraduate e-learning course, Zhang et al. (2021) looked into how students' intrinsic motivation affected predictions of their course performance and dropout rates. They discovered that intrinsic motivation was a major predictor of both outcomes.

Herman (2011) estimated that in South Africa, 50% of postgraduate students did not complete their programs due to personal reasons, a lack of talent, aptitude, or motivation. Mukami (2016) conducted research on attrition rates in Nairobi's private universities in Kenya. She took a sample of 387 current students and 60 dropouts from 13 Nairobi universities. According to this report, 37% of undergraduate student's attrition. The undergraduate population and examination of universities in the Nairobi region distinguish this study from recent studies that exclude public universities. Rong'uno (2016) examined the factors that affect student retention in doctoral programs at a few Kenyan public universities. 388 PhD candidates and 184 professors were sampled using a descriptive design in three universities. The study showed that the completion duration for enrolments between 2001 and 2008 was 9 years, and the completion rate ranged from 50% to 70%. Additionally, there were enrolled Considering that 50% of PhD programs were not completed in 2009, Students' workload was quite heavy, and they required a very lengthy time to graduate. Rong'uno and Mukami Studies show the high rate of incompletes in Kenya, sampling both private and Usually, public universities.

The reasons for postgraduate student dropouts can be multifaceted and complex, often involving a combination of personal, academic, financial, and institutional factors. Here are some common causes of postgraduate student dropouts:

1. Financial Constraints:

- Tuition Fees: Postgraduate education can be expensive, and students may struggle to meet the financial demands, leading to dropout.

- Living Expenses: High costs of living, especially in certain cities or countries, can strain students' budgets, making it challenging to sustain their education.

2. Academic Challenges:

- Difficulty of Coursework: Postgraduate studies are often more intense and specialized, posing challenges to students who might find the coursework overwhelming.

- Lack of Academic Preparedness: Some students may not have the necessary academic background or research skills, making it difficult to keep up with the demands of postgraduate studies.

3. Personal Factors:

- Health Issues: Physical or mental health problems can severely disrupt a student's ability to continue their studies.

- Family Responsibilities: Students may have family obligations or caregiving responsibilities that conflict with their academic commitments.

- Work-Life Balance: Balancing a job, especially for those studying part-time, with academics can become overwhelming.

4. Lack of Support:

- Academic Support: Insufficient guidance and mentorship from professors or advisors can lead to a feeling of isolation and disorientation among students.

- Social Support: Lack of a supportive social network, both academically and personally, can contribute to feelings of alienation.

5. Unfulfilled Expectations:

- Mismatched Expectations: Students might find that the postgraduate program does not align with their expectations, leading to disillusionment and dropout.

- Career Opportunities: Some students may find better job prospects before completing their postgraduate degree, leading them to leave academia.

6. Research Challenges:

- Lack of Resources: Limited access to research materials, labs, or funding can impede the progress of research-based postgraduate programs.

- Supervision Issues: Inadequate supervision or conflicts with supervisors can hinder the research process and discourage students from continuing.

7. Institutional Factors:

- Administrative Issues: Problems with registration, paperwork, or bureaucratic hurdles can create frustration and discourage students.

- Limited Resources: Insufficient institutional resources, including libraries, laboratories, and technology, can hamper the overall learning experience.

8. Global Events:

- Pandemics/Natural Disasters: Extraordinary events like pandemics or natural disasters can disrupt studies, limit access to resources, and strain mental health, leading to increased dropout rates.

Addressing postgraduate student dropouts requires a holistic approach, involving financial assistance, academic support, mentorship programs, mental health services, and a supportive institutional environment. Recognizing and proactively addressing these factors can significantly reduce dropout rates and enhance the overall educational experience for postgraduate students.

Mitigation factors for dropout of postgraduate students

Mitigating postgraduate dropouts involves addressing various factors that contribute to students leaving their programs prematurely. Here are some methods that educational institutions can consider to reduce postgraduate dropout rates:

1. Early Identification:

- Monitoring Progress: Regularly monitor students' academic progress to identify those who are struggling.

- Interventions: Implement early intervention strategies for students showing signs of academic difficulty, such as additional tutoring or counseling.

2. Financial Support:

- Scholarships and Grants: Offer financial aid, scholarships, and grants to students who face financial challenges in continuing their education.

- Work-Study Programs: Facilitate work-study programs that allow students to earn while they study, easing their financial burden.

3. Mental Health and Counseling Services:

- Counseling Support: Provide counseling services to help students cope with stress, anxiety, and other mental health issues.

- Mental Health Awareness: Increase awareness about mental health issues and reduce the stigma associated with seeking help.

4. Academic Support:

- Mentorship Programs: Establish mentorship programs where experienced faculty or senior students mentor newcomers.

- Study Skills Workshops: Conduct workshops on effective study techniques, time management, and research skills.

5. Flexible Learning Options:

- Online Courses: Offer online courses or hybrid programs, allowing students more flexibility in their schedules.

- Part-Time Options: Provide part-time study options for students who need to balance education with work or family responsibilities.

6. Community Building:

- Peer Support Groups: Create peer support groups where students can discuss challenges and support one another.

- Engaging Activities: Organize extracurricular activities, workshops, and seminars to foster a sense of community among students.

7. Faculty Involvement:

- Approachable Faculty: Encourage faculty to be approachable and supportive, fostering open communication with students.

- Feedback Systems: Implement feedback mechanisms where students can express concerns about their courses or learning experiences.

8. Career Guidance:

- Career Counseling: Provide career counseling services to help students align their academic pursuits with their future goals.

- Internship Opportunities: Facilitate internships and practical experiences, connecting academic learning with real-world applications.

9. Research Opportunities:

- Research Assistantships: Offer research assistantships to postgraduate students, involving them in ongoing research projects.

- Publishing Opportunities: Encourage students to publish their research, boosting their confidence and academic engagement

10. Regular Evaluation and Adaptation:

- Data Analysis: Continuously analyze dropout rates, identifying patterns and reasons behind student attrition.

- Adaptation: Use the data to adapt strategies, programs, and support systems accordingly.

By implementing a combination of these strategies and remaining attentive to the evolving needs of students, educational institutions can create an environment where postgraduate students are more likely to succeed and complete their programs. The literature has covered how cultural attitudes about learning can vary. Students from individualistic societies typically have a greater motivation to compete with others and oneself. They therefore stand a better probability of finishing the master's program. Though "collectivistic" in our study, there are additional factors that can cause international students to leave German educational programs. Poor language skills, financial difficulties, a lack of social and academic integration, and misconceptions about the teaching and learning environment at German higher education institutions are some of these causes. (Kercher, 2018).

2.3.3 Techniques to predict e-learning course dropout

In order to identify at-risk students and offer prompt interventions to increase retention rates, educational institutions and online platforms must be able to predict e-learning course dropout. Predicting e-learning course dropout can be done using a variety of methodologies and methods. Here are a few typical methods:

a) **Data Mining method-**. Data mining involves the use of algorithms to extract useful information from large datasets. By analyzing student data such as login frequency, assessment scores, and time spent on the course, data mining algorithms can identify patterns that may indicate a student is likely to drop out (Huang et al., 2014). The methods consist of:

   Classification Algorithms: Utilize algorithms like Decision Trees, Random Forest, Logistic Regression, or Support Vector Machines to build predictive models based on historical data. Features could include student engagement, quiz scores, interaction frequency, and demographics.

   Clustering: Group students based on their behaviour and performance using clustering techniques like K-means. It helps identify patterns and behaviours associated with dropout.

Deep Learning: Neural networks, especially recurrent neural networks (RNNs) and Long Short-Term Memory networks (LSTMs), can capture complex patterns in sequential data, making them suitable for analysing student interactions over time.

b) **Using Machine Learning Algorithms**-. Machine learning algorithms are a type of artificial intelligence that can be trained to recognize patterns in data. By analyzing student data, machine learning algorithms can identify patterns that may indicate a student is likely to drop out. Some common machine learning algorithms used for dropout prediction include decision trees, logistic regression, and neural networks (Kemp, 2016).

c) Predictive Analytics:

Learning Analytics: Utilize learning management system (LMS) data to analyze student interactions. Track logins, time spent on the platform, assignment submissions, forum participation, and social interactions. Sudden drops in activity might indicate potential dropout.

Early Warning Systems: Develop systems that generate alerts when students exhibit behaviours associated with dropout, enabling timely intervention by educators.

e) Natural Language Processing (NLP):

Sentiment Analysis: Analyze student discussions, forum posts, or chat interactions to gauge sentiment. Negative sentiment or expressions of frustration might indicate disengagement and potential dropout risk.

Text Mining: Extract insights from essays, comments, or feedback to identify themes related to struggles or disinterest.

d) Social Network Analysis:

Social Graph Analysis: Study the social interactions between students. Isolated students or those with fewer connections might be at a higher risk of dropping out.

Influence Analysis: Identify influential students or mentors who positively impact their peers. Fostering such connections can improve retention rates.

f) Feature Engineering:

Derived Metrics: Create new features from existing data, such as engagement ratios, consistency in assignment submissions, or participation frequency. These engineered features can enhance the predictive power of the models.

Temporal Patterns Analyze how student behavior changes over time. Patterns like decreasing engagement might indicate potential dropout.

g) Collaborative Filtering:

Recommendation Systems: Use collaborative filtering techniques to recommend relevant courses or resources to students based on their behavior and preferences. Better engagement can reduce dropout rates.

e) Feedback Loops and Interventions:

Feedback Surveys: Regularly collect feedback from students about the course structure, content, and learning experience. Use this data to make improvements that enhance engagement and satisfaction.

Targeted Interventions: Based on predictive models, design personalized interventions such as targeted emails, additional resources, or mentor support for at-risk students.

f) **Bias Detection***:* Regularly assess the models for biases that might disproportionately affect certain student groups. Address these biases to ensure fair predictions.

Implementing a combination of these techniques, tailored to the specific context and available data, can significantly improve the accuracy of e-learning course dropout predictions and facilitate proactive interventions to support students at risk of dropping out.

2.3.4 Machine learning Prediction Algorithms

Machine learning algorithms have been widely used in predicting student performance and dropout in e-learning environments. These algorithms use data mining techniques to analyze large datasets and identify patterns that can predict student performance and dropout (Al-Radaideh et al., 2021). Several machine learning algorithms have been used to predict e-learning course dropout. These include decision trees, logistic regression, artificial neural networks, and support vector machines (SVM) (Al-Radaideh et al., 2021).

Decision trees are popular because they are easy to interpret and can handle both numerical and categorical data (Al-Radaideh et al., 2021). Logistic regression is commonly used because it can model the relationship between predictor variables and a binary outcome (i.e., dropout or completion) (Al-Radaideh et al., 2021). Artificial neural networks are capable of modeling complex nonlinear relationships between predictor variables and the outcome variable (Al-Radaideh et al., 2021). SVM is a powerful algorithm that can handle high-dimensional datasets and is effective in handling noisy data (Al-Radaideh et al., 2021). Machine learning algorithms have been used successfully in predicting e-learning course dropout in various studies. For example, a study by El Ouahabi et al. (2020) used a decision tree algorithm to predict e-learning course dropout among undergraduate students in Morocco. The study achieved an accuracy of 85% in predicting dropout, indicating the effectiveness of the algorithm in identifying at-risk students. Similarly, a study by Huang et al. (2021) used an artificial neural network algorithm to predict e-learning course dropout among postgraduate students in China. The study achieved an accuracy of 78.3% in predicting dropout, indicating the potential of the algorithm in identifying at-risk students and intervening early.

2.3.5 eLearning

By the year 2015, Kenya had 33 public and 17 private institutions, according to the Commission for University Education (CUE) (CUE, 2015). The majority of these institutions had begun to provide a few e-learning courses that were mostly supported by Learning Management Systems (LMS) and were asynchronous and blended in nature (Ssekakubo et al., 2011). However, few universities that have embraced e-learning have made any investments adequately invest in the tools and instruction needed to create successful courses. (Kashorda & Waema, 2014). Some studies have found out that the main challenges affecting e-learning include but are not limited to: inadequate ICT and e-learning infrastructure, financial constraints, lack of affordable and adequate Internet bandwidth, lack of operational e-learning policies, lack of technical skills on e-learning, and econtent development by the teaching staff (Ssekakubo et al., 2011; Tarus, Gichoya, & Muumbo, 2015; Makokha & Mutisya, 2016; Muuro et al., 2014). According to research by Lim, Park, and Kang (2016) on the structural linkages between environments, people, and learning outcomes in e-learning, system and content quality were important in terms of inspiring both intrinsic and extrinsic motivation. Furthermore, system quality and content quality, respectively, had an impact on computer self-efficacy and academic self-efficacy. The criteria for evaluating the quality of online courses (Wright, 2014) and the Quality Matters Rubric Standards (QMRS, 2014) both established the indicators for measuring the quality of e-learning in the context of course development and assessment. According to these results, high-quality courses, material, and assessments, as well as a sufficient infrastructure, boost learning motivation, which is crucial for effective e-learning use. Additionally, the content must be supported after it has been designed and developed with announcements and reminders, multimedia tools like audio and animations, realistic or authentic learning activities, and helpful feedback from instructors (QMRS Higher Education Rubrics, 2014; Wright, 2014; Makokha & Mutisya, 2014).

Overview of eLearning Systems in Education

eLearning systems have revolutionized the education sector by providing flexible, accessible, and interactive learning experiences. These systems leverage digital technologies to facilitate teaching and learning processes. Here's an overview of eLearning systems in education:

1.**Types of eLearning Systems:**

a) *Learning Management Systems (LMS):* LMS platforms like Moodle, Blackboard, and Canvas manage course content, assignments, and assessments in an organized manner.

b) *Virtual Learning Environments (VLE):* Similar to LMS, VLEs provide a virtual classroom where students and teachers interact. Google Classroom is a popular example.

c) *Massive Open Online Courses (MOOCs):* Platforms like Coursera, edX, and Udacity offer online courses from universities and institutions globally.

d) *Mobile Learning (mLearning):* Learning delivered through smartphones and tablets, often via apps or responsive websites.

e) *Social Learning Platforms:* Platforms that facilitate learning through social interactions, such as discussion forums, wikis, and social media.

2. Key Features:

a) *Content Variety:* eLearning systems offer text, audio, video, and interactive multimedia elements, catering to diverse learning styles.

b) *Interactivity:* Features like quizzes, simulations, and gamified elements engage learners actively.

c) *Personalization:* Adaptive learning algorithms tailor content based on individual progress and learning pace.
d) *Collaboration:* Discussion forums, group projects, and peer assessments foster collaborative learning experiences.
e) *Assessment and Analytics:* Automated assessments and data analytics provide insights into student performance, helping educators make data-driven decisions.

3. Benefits:

a) *Accessibility:* Learners can access educational materials from anywhere, at any time, breaking down geographical barriers.
b) *Flexibility:* eLearning accommodates varied schedules and learning speeds, enabling self-paced learning.
c) *Cost-Effectiveness:* Reduces costs associated with physical infrastructure, travel, and printed materials.
d) *Engagement*: Interactive and multimedia elements enhance learner engagement and motivation.
e) *Scalability:* eLearning allows institutions to scale their educational programs to a large number of students without significant infrastructure investments

4. Challenges:

a) *Digital Divide:* Disparities in access to technology and the internet can hinder equal learning opportunities.
b) *Learner Discipline:* Self-discipline is essential for online learning, and some students may struggle with time management.
c) *Technical Issues:* Glitches, compatibility problems, or lack of technical skills can impede the learning experience.
d) *Limited Social Interaction:* Lack of face-to-face interaction might impact social and communication skills development.

5. Future Trends:

a) *Augmented and Virtual Reality*: Integration of AR and VR technologies for immersive learning experiences.
b) *AI and Machine Learning:* Intelligent tutoring systems and personalized learning paths driven by AI algorithms.
c) *Microlearning:* Bite-sized, focused learning modules for quick skill acquisition.
d) *Blockchain in Education:* Secure, verifiable credentials and certifications using blockchain technology.
e) *Social and Emotional Learning (SEL) Integration*: Focus on developing emotional intelligence and interpersonal skills in online education.

In conclusion, eLearning systems continue to evolve, shaping the future of education by making learning more accessible, engaging, and personalized for learners worldwide. However, addressing challenges related to access, quality, and social interaction remains crucial for the continued success of eLearning in education.

According to other studies, quality is improved by both social and administrative support. Social support in the form of knowledge, skill, affirmation, and emotion has been found to have an impact (Weng & Chung, 2015; Munich, 2014; Muuro et al., 2014; Queiros & de Villiers, 2016). According to Tarus, Gichoya, and Muumbo (2015) and Makokha & Mutisya (2016), some of the major variables that influence success or failure include registration support, orientation, and a dedicated call center. In an e-learning environment, the part performed by user traits like learners and instructors has also shown to be crucial. The quality of an e-learning system is influenced by a variety of factors, including computer and internet experience, enthusiasm for e-learning, motivation from instructors, self-efficacy, training, and incentives for the instructor (Baloyi, 2014a; Muuro et al., 2014; Baloyi, 2014b; Queiros & de Villiers, 2016; Azawei et al., 2016; Makokha & Mutisya, 2016;

2.3.6 E-learning and Student Motivation

2.3.6.1 Virtual Learning Environment

E-learning is described by authors as the creation of a platform where instruction and learning are conducted electronically. This definition covers online learning, mobile learning, and other applications. By utilizing multimedia technology to enable distant information interchange, the European Commission rejected the notion that e-learning would improve quality of life. E-learning enhances learning management (Boezerooij, 2006). E-learning is progressively being included into the learning platforms of most institutions, particularly to improve the procedures for delivering education and providing support (AlKhuder & AlAli, 2017). Knowledge exchange is important both nationally and internationally, and it may be promoted by employing e-learning and other technologically enabled learning platforms to reach a larger audience and raise awareness (United Nations Developent Agenda, 2015).

Top e-learning considerations include the development of infrastructure, content, teacher preparation, and networking. In 2003, 93% of the European Commission's institutions had access to the internet (Salajan & Roumell, 2016). Colleges and research institutions now have high-speed internet access thanks to the adoption of eLearning platforms and virtual learning campuses (Uzunboylu, 2006). The advantages of technology infrastructure surpass the costs of implementation and support lifelong learning. The effects of split attention and how multitasking affects student performance are numerous. The effects of multitasking on students' grade point averages, exam performance, efficiency, and recall skills are a major source of worry (Agaard, 2018).

Learning in a structured educational environment is no longer a guarantee of employment. For those who wish to continue studying, the development of lifelong learning courses is essential. According to research, direct online communication with students has a significant impact on how well-tailored the information pupils get. They can also access a wide range of debate subjects thanks to it. This encourages a culture where people work hard but also smart (Singh, 2013). Students at colleges encourage multitasking and claim that it boosts productivity.

Although using technology in the classroom is more efficient than writing on a blackboard, some institutions have policies that forbid students from multitasking in class (May & Elder, 2018). According to research, blended learning fosters communication and information exchange among students, which modifies their view of the learning environment. Although blended learning began before instructional technology, its development is tied to ICT (Pima, Odetayo, & Iqbal, 2018). Students are challenging educational frameworks in the African higher education sector through demonstrations as they push for change and depart from conventional thinking (Piranecis & Kos, 2013). According to authors, encouraging design thinking within the context of technological learning could assist innovative academic practices (Gachago, Morkel, Hitge, Zyl, & Ivala, 2017). According to research, e-learning is still in its infancy in Kenyan universities. The majority of these universities do not have structural senate authorization to ensure the application of policy for e-learning.

2.3.6.2 2 Learning Simulations

The offered program's quality and content are important determinants of its success. Some Australian universities have implemented a technology method for project supervision of post-graduate students. Emails, Skype, Dropbox, and mobile phones are all interwoven into these institutions (Maor & Currie, 2017). In order to create a community that is technologically competent, Botswana and South Africa implemented education technology in the year 1990. According to research, the themes covered by the curriculum and the specific subject assessments employed in these nations are connected (Du Toit & Gaotlhobogwe, 2017). In order to spur high technology innovation, the United Nations Development Agenda (2015) mandates that technology be available at all levels of education and through public domains.

Knowledge is a crucial component of innovation. Education is the key to success, according to the late president of the republic of South Africa, Nelson Mandala; it is undeniable that people who have access to knowledge can use it as a key to produce creativity (Gebremedhin & Joshi, 2016). Learning via scenarios, which enables information transfer between students and provides a greater awareness of the environment, is one of the most effective teaching strategies for pupils. Collaboration among students on genuine, real-world activities broadens their perspectives and enables them to think outside of the classroom. According to Panesar-Aguilar and Aguilar (2017), collaborative learning is a tool for the educational development of both teachers and students. The difficulty of learning software is frequently reduced by manuals, which are meant to act as guidelines for how to utilize programs. Although some students may not be technologically literate, it takes them some time to go around the digital learning platforms while other students are more familiar with how to use the tools (Murray, 2016).

2.4 Conceptual Framework

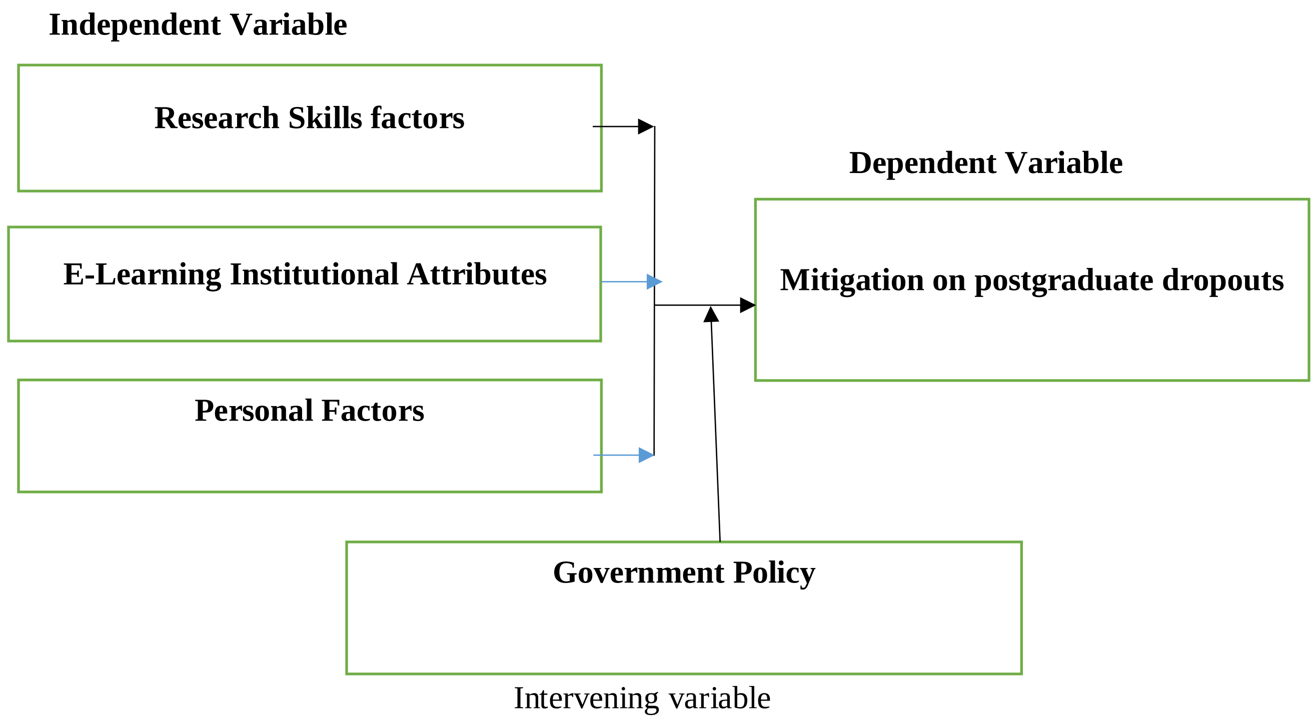

Figure i Conceptual Framework

2.5 Chapter Summary

In summary, this chapter provided a review of the literature on e-learning course dropout among postgraduate students in universities. The review highlighted the challenges faced by students in e-learning environments, the factors that contribute to e-learning course dropout, and the use of machine learning algorithms in predicting e-learning course dropout. The next chapter will describe the methodology that will be used in this study to investigate the problem of e-learning course dropout among postgraduate students in universities.

CHAPTER THREE
RESEARCH MEHODOLOGY

3.1 Introduction

This chapter presented details of the methodology adopted for the study and a research design which is the blueprint for conducting a study detailing the design approach. The research design also covers the research paradigm selected for the study. The chapter also presents the data source which explains the population of the study. It also covers the data selection and sampling procedures. Data pre-processing which focuses on the assembling of the data and ensuring data is in the form required for analysis is also discussed. Data mining is also presented which describes the process of obtaining the data and using the R statistical software to perform descriptive and predictive analytics. Data transformation is also discussed which describes the process of ensuring that the data is scaled in the required format for use in the analysis. Then the Knowledge gap is also presented for presenting the output results of the study. The last item in this chapter is the ethical consideration which describes how the study will obtain the relevant permission to use the data for analysis.

3.2 Research Design & Methodology

Research design can be referred to as a plan which gives guidance on any given research and aids in data collection, analysis as well as interpretation of observations. A design for experimental research and descriptive survey were used by the researcher. Since the researcher used a machine learning algorithm to train the Dropout Prediction Detection model, therefore research design will be appropriate. This approach is appropriate for studies without a control group, which allows for a primary emphasis on one group without making comparisons to other groups (Babbie & Earl, 1998). It can also be used by researchers as a blueprint which helps one to make decision on methods and tools to use in collection of information as well as its evaluation, so as to be able to answer the questions guiding the research (Cooper & Schindle, 2014).

The descriptive survey was used to obtain feedback from the questionnaires for both 1st and 2nd objectives.

Objective three was achieved by developing a Postgraduate eLearning dropout Prediction Model using machine learning while objective 4 was achieved through machine learning validation of the developed model

To ascertain the degree to which specific security traits and attributes of our databases are pertinent, these designs were used to gather quantitative feedback from chosen respondents as well as experimental test results from datasets.

Two supervised machine learning techniques, Random Forest and Support Vector Machine, were used to create the model for forecasting postgraduate students' dropout rates. The algorithms were applied to increase the model's accuracy. The study site's admission and registrar's office provided the dataset for the experimental setting through record examination. This dataset comprises of the enrolment information for postgraduate students who attended one of the university's many study programs between 2019 and 2023. A csv file containing 300 records overall was assembled and kept. The division of data into training and testing sets is the last step in the data preparation process. The remaining 80% of the dataset was utilized to train the model, with only 20% used for testing. To evaluate the efficacy and performance of several machine learning models in predicting student dropout, model training involved training and testing.

This study employed Fayaz design for data mining as shown in figure 2.

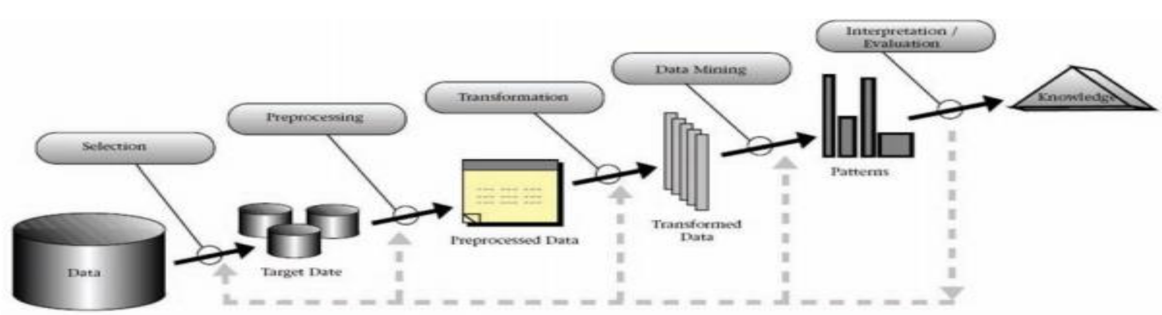

Figure ii Model design approach for the study

Source: Fayyad et al. (1996)

The various stages in the design approach are described in the next sections.

3.3 Location of the Study

Data was collected from Jomo Kenyatta university of agriculture and technology on the cases of postgraduate e-learning students' dropout.

3.4 Sample size and Population

A sample is a representation of a population as a whole. Since the researcher will utilize an experimental design based on desktop simulation, there was only one sample. The phrase "target population" refers, according to Orodho (2008), to a full list of all the items used in the population for the study, to which the researcher wants to extrapolate the findings. The 210 students that study at the university will be the study's target population. However, a sample size of 132 was determined for the purposes of this study using the yamane's formula;

n = N/(1+N(e)2.

The variables in this formula are:

n = the sample size

N = the population of the study

e = the margin error in the calculation

3.5 Data

The research consisted of both primary and secondary data. The primary data was obtained from Jomo Kenyatta University of Agriculture and technology using questionnaires. The data set used for modelling was obtained from private database.

3.5.1 Selection of Data

After obtaining data from the university data base and storing in an excel database, the study worked with a representative dataset by performing data reduction. One method of data reduction which this study adopted was numerosity reduction technique, where the data was replaced or estimated by alternative, smaller data representation methods such as means, clustering, sampling, and the use of histograms. The study then undertook data cleaning activity. Data cleaning is the process of detecting and removing corrupt or inaccurate records from a database. This process was be done by use of binning technique. Binning method sorts a data value by consulting its "neighbour- hood," that is, the values around. The sorted values were distributed into a number of "buckets," or bins. The data was portioned into equal frequency bins and then smoothing by means will be performed.

3.5.2 Data Pre-processing and Transformation

This involved identification of the data to be used in training the model, test data to test model, training and the validation data to measure the output error. It is critical to identify the type of data to be put for each data set as identified in the research sampling procedure. The sample taken will be a representative of the identified population.

The data was converted into comma delimited (csv) format files to have three different csvs containing the data on enrolment of students, deferment and drop out of the students. all three of these files will be linked through the student ids. After data input files are prepared, data cleaning will be done in the following steps:

1. Data type conversion (character to numeric, character to factor)

2. Aggregate data to student id level

3. Identify outliers and remove them appropriately

4. Clean/ remove missing values

5. Merge dataset into one dataset

Admission information and term results information datasets will be compiled together for pre-processing while the third dataset courses and grades from data-mart table will be treated separately. Then all datasets will be merged at the end of the process.

After reading the csv files, as the first step of the data preparation process, all the character type data will be converted into numeric data using the following R codes.

```
Enrolment<- read.csv(file.choose(), header= TRUE)

Deferment<-read.csv(file.choose(), header=TRUE)

Dropout<-read.csv(file.choose(), header=TRUE)

data<-as.data.frame(Enrolment,Deferment,Dropout)

Attach(data)
```

3.5.3 Data mining

This process involved feeding inputs to the model for processing in order to train the model on the type of input data and the expected output of the training session. As mentioned earlier data was stored in an excel data sheet. The excel sheet will then be saved as a CSV file (comma delimited) to make it possible to read it into R statistical software. After reading the data into R software the next step was to scale the data in a [0,1] interval to ensure efficiency in working with the data. The data was then then split in to a ratio of 80: 20 where eighty per cent of data was for training data while 20 per cent of the data will be a test data. According to Dobbin (2011), twenty per cent of the data should be treated as test data while 80% of the data should be treated as training data. Training of data was done over a number of iterations until the algorithm converges.

3.5.4 Model Evaluation

This process involved the use of a test data set to check whether the system is properly trained by observing the actual model output versus the expected output. By using the validation data set, any disparities in the output capture by error performance measures will be used to adjust the weights for purposes of fine tuning the model. The model output data was tested through repeated trials on the training phase. This will be done to ensure that the error rate is reduced to the global minimum for the model in order to provide the most optimal weights for each variable input. The model was evaluated by calculating measures of performance. A confusion matrix will be obtained to show the various performance measures and the accuracy of the model. Output of the results were displayed at the Knowledge gap. Several performance measures discussed in the next sections below will be used in this study

Accuracy

Accuracy is how probable it is on average for the prediction of the model to be correct. This is calculated as the proportion of the number of correctly classified cases to the total number of cases. Correctly classified cases are the total of true positives and true negatives. The following equation represents the calculation of accuracy, n being the total number of cases.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{n}}$$

Sensitivity

Sensitivity is how probable it is to classify a case as true when it is actually true. This is calculated as the proportion of correctly classified true cases within actual true cases.

Specificity

Specificity is how probable it is to classify a case as false when it is actually false. This is calculated as the proportion of correctly classified false cases within actual false cases.

Positive predictive value

Positive predictive value is how probable it is for a case to be actually true when the model predicts it to be true. In the context of this study, the probability would be how likely it is that a student will progress, when the model identifies the student as a student who will progress. This is calculated as the proportion of correctly classified true cases within predicted true cases.

Negative predictive value

Negative predictive value is how probable it is for a case to be actually false when the model predicted it to be false. In the context of this study, it is how likely it would be for a student to dropout or defer his or her studies, when the model identifies the student as a student who will dropout or defer. This is calculated as the proportion of correctly classified false cases within predicted false cases.

3.6 Ethical Issues

Research ethics is critical since it guides the interactions with people, organizations and institutions (Christensen, Johnson, & Turner, 2014). The study sought authorization for data collection from the Jomo Kenyatta university of agriculture and technology by explaining the purpose and importance of study. Privacy and confidentiality was employed to ensure that the data collected from respondents was kept safe, free from interference and protected from unwanted use.

CHAPTER FOUR

RESEARCH FINDINGS AND DISCUSSION

4.1 Introduction

As outlined in the research findings and discussions chapter, this chapter represented the data analysis, results, and comments of the research findings. The results were presented using tables, pie charts, and frequency tables. The data for each objective was examined, the results were discussed, and diagrams were used to help with understanding.

4.2 Demographic Characteristics of Respondents

4.2.1 Questionnaire Response Rate

The researcher conducted a pilot study in Jomo Kenyatta university of agriculture and technology where he distributed 20 questionnaires to the staff in the selected area of study which represented 10% of the sample size. In the distribution, there was a statistically significant high response rate of 95% from the pilot study, simple random sampling was applied. A total of 210 respondents, or 132 respondents, made up the researcher's target group. The researcher gave out 108 questionnaires to the participants during the actual data collection, and 120 of them were returned, or 91% of the questionnaires that were filled out. A response rate of 50% is considered enough for analysis and reporting, a rate of 60% is considered good, and a rate of 70% or higher is considered extraordinary, according to Mugenda & Mugenda (2003). On the basis of this assertion, the response rate was very good. Everyone who worked in the study's chosen field had an equal opportunity of participation. The demographic data was based on the age, gender, marital status, departments in which participants worked, level of education, Duration in which the participants had worked or studied in the university.

Table i Gender

What is your gender?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	59	49.2	49.2	49.2
	Male	61	50.8	50.8	100.0
	Total	120	100.0	100.0	

According to table 1 above 50.8% of the respondents were male while 49.2% were females. Obtaining information on gender was crucial for this study because it was one of the demographic factors that was looked at in relation to postgraduate students' non-completion. This topic needs to be studied because previous researchers had found that postgraduate students' failure to complete their degrees was influenced by their gender. Women left school to become mothers and parents, according to Rhea et al. (2013). Jiranek (2010) also found that men were more likely than women to complete their postgraduate degrees, suggesting that gender had an impact on non-completion in this study.

Table ii Level of Education

Which is the highest Level of Education you have attained?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Bachelor's Degree	95	79.2	79.2	79.2
	Master's Degree	24	20.0	20.0	99.2
	Phd	1	.8	.8	100.0
	Total	120	100.0	100.0	

According to table 2 above majority of the respondents had bachelor's degree which attributed to 79.2%, master's degree 20% while Phd represented 0.8%.

Table iii Duration in the university

How long have you been in the university?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-2 year	36	30.0	30.0	30.0
	3-5 year	49	40.8	40.8	70.8
	Above 5	29	24.2	24.2	95.0
	Less than 5 years	6	5.0	5.0	100.0
	Total	120	100.0	100.0	

According to table 3 above 40.8% of the respondents had been in the university for 3-5 years, 30% 1-2 years, 24.2% were in the university for more than 5 years while 5% have been there for less than 5 years.

Table iv Age

What is your age?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20-29	49	40.8	40.8	40.8
	30-39	64	53.3	53.3	94.2
	40-50	4	3.3	3.3	97.5
	Above 50	3	2.5	2.5	100.0
	Total	120	100.0	100.0	

According to table 4 above majority of the respondents had an age of 30-39 which attributed to 53.3 %, 20-29 attributed to 40.8%, 40-50 attributed to 3.3% while those who were above 50 years attributed to 2.5%. These results were in line with Mulvey and Nicholson's (2014) study, which discovered that in the USA, students pursuing master's degrees in physics were, on average, 29.2 years old, with 10% of them being 35 years of age or older. Grad School Hub's poll from 2020 supported the study's conclusions. According to the poll, master's degree students were, on average, 33 years old. The findings could be attributed to people being more active between the ages of 20 and 40, when many would wish to have completed their educational and professional ambitions.

Table v Occupation in the University

What is your occupation within the university?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Staff	12	10.0	10.0	10.0
	Student	108	90.0	90.0	100.0
	Total	120	100.0	100.0	

According to table 5 above 90% of the respondents were students while 10% were staff within the university.

Table vi Department

Which department are you in?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Business	41	34.2	34.2	34.2
	Education	7	5.8	5.8	40.0
	Finance	16	13.3	13.3	53.3
	Hospitality	4	3.3	3.3	56.7
	ICT	42	35.0	35.0	91.7
	Media	10	8.3	8.3	100.0
	Total	120	100.0	100.0	

According to table 6 above majority of the respondents were pursuing ICT with 35%, Business had 34.2%, Finance had 13.3%, media had 8.3%, Education had 5.8% while hospitality had the least with 3.3%.

4.3 Presentation of Findings

This part aimed to present the findings of the researches various objectives.

4.3.1 Objective 1: To analyses the factors that influence e-learning students dropout rates among post graduate.

This information was crucial to the study since it provided the examiner with specifics on the length of the students' postgraduate studies and how close they were to graduation. The investigator was also able to determine how long the students had studied thanks to this data. The researcher could calculate the rate of years without completion using this data. The data also showed which phase of postgraduate education more students were likely to squander more time than necessary.

Table vii Interaction with eLearning system

Have you interacted with an eLearning system?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	25	20.8	20.8	20.8
	Yes	95	79.2	79.2	100.0
	Total	120	100.0	100.0	

According to table 7 above 79.2% of the respondents had interacted with the eLearning system while 20.8% had not. The findings indicate that over half of the students are passionate about online learning and have practical computer and internet skills. However, nearly half of the students bemoaned a lack of LMS training and professors who lacked passion. According to Jung's (2017) research, both intrinsic and extrinsic learner motivation are essential for students' performance in an online coursework setting. In order to increase quality, JKUAT must offer training because it is a means of disseminating e-learning capabilities. (Arinto, 2016; Azawei et al., 2016).

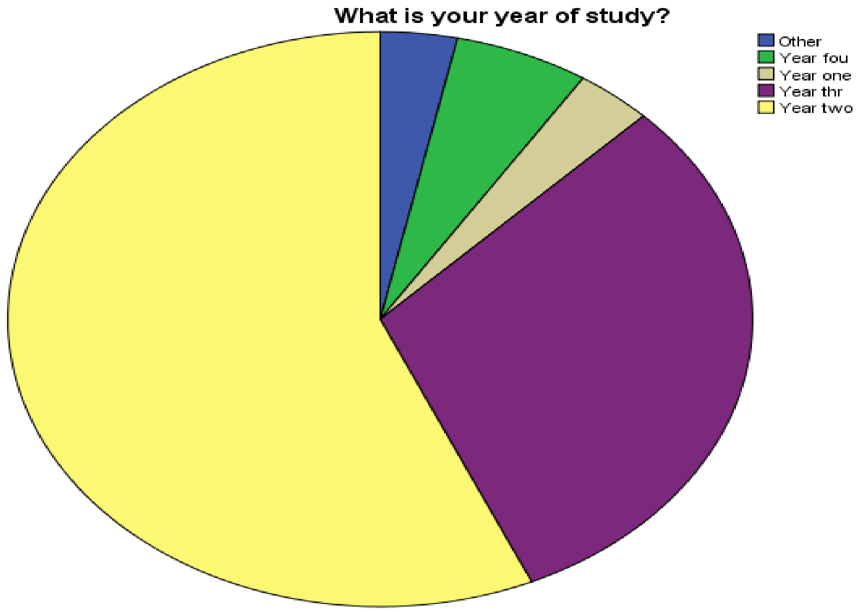

Figure iii Level of study

According to figure 3 above majority of the respondents were in year two of their study while the least were in year one.

Table viii Level of academic progress

Which is your Level of Academic Work?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Concept development	4	3.3	3.3	3.3
	Course work	63	52.5	52.5	55.8
	Data collection	14	11.7	11.7	67.5
	Graduated	9	7.5	7.5	75.0
	Proposal development	22	18.3	18.3	93.3
	Report writing	8	6.7	6.7	100.0
	Total	120	100.0	100.0	

According to table 8 above, majority of the students were undertaking course work which attributed to 52.5%, 18.3% were writing their proposals, 11.7% were at data collection stage, 7.5%

had graduated, 6.7% were writing the thesis report while 3.3% were writing their concept papers. The year of enrolment was essential to the study because it provided the researcher with information about the students' length of attendance. Knowing the rate of student non-completion would have been challenging without knowing the year of admission.

Table ix Expected Graduation date

What is your expected Graduation Date?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2024	74	61.7	61.7	61.7
	2025	29	24.2	24.2	85.8
	2026	17	14.2	14.2	100.0
	Total	120	100.0	100.0	

The majority of students, or 61.7%, expected to graduate in 2024, followed by 24.2% in 2025, and 14.2% in 2026, as shown in table 9 above. The information made it simpler for the researcher to determine whether students had spent longer than anticipated on their studies. The researcher may also determine if students increased their time of study based on where they were in their studies at the time by comparing it to the current year of this research. Since non-completion was a key research question, this information was essential.

Table x Fee challenge in research work

Is fee payment challenge a to your research work?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	45	37.5	37.5	37.5
	Yes	75	62.5	62.5	100.0
	Total	120	100.0	100.0	

According to table 10 above 62.5% of the respondent’s fee was the main challenge in completing the research while 37.5% fee was not an issue. In contrast, 15% of Sri Lankan respondents in Rauf's (2016) survey strongly agreed that lack of funding prevented them from finishing their theses. Jiranek (2010) discovered that scholarship recipients had excellent completion rates. According to research by Rong'uno (2016), scholarship recipients were more likely than other students to complete their dissertations in a timely manner. Rung'uno also noted that, regardless of the source of income, postgraduate students who self-funded were less likely than those who got funding to complete their studies and graduate on time. It was crucial to comprehend how financial status affected the study's dependent variable, which was the intervening variable, even if it wasn't a primary variable because of other circumstances. This aspect was crucial to include in the study because previous studies had shown that postgraduate students' lack of completion was significantly impacted by money (Garwe & Maganga 2015; Rauf, 2016). The findings above suggested that the main cause of non-completion had been money.

Table xi Factors affecting dropout

Which of the following factors have affected the dropout rate?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	All	55	45.8	45.8	45.8
	Institutional factor	11	9.2	9.2	55.0
	Personal factors	33	27.5	27.5	82.5
	Research skills	6	5.0	5.0	87.5
	Supervisory factors	15	12.5	12.5	100.0
	Total	120	100.0	100.0	

According to table 11 above the institutional factors, personal factors, research skills, supervisory factors affected the dropout of post graduate students which attributed to 45.8%, personal factors attributed to 27.5%, supervisory factors attributed to 12.5%, institutional factors attributed to 9.2% while lack of research skills attributed to 5%. The need to eliminate university dropouts as a type of academic failure is supported by at least four factors: (i) economic, (ii) social, (iii) individual, and (iv) pedagogical (Staiculescu, 2018). In order to enhance the number of highly skilled individuals for the knowledge society and economy, many nations have programs in place (Kehm et al., 2019). To increase the number of professionals who can enter the labor market, for instance, a number of programs in Germany have as their primary goal lowering the number of student dropouts (Mouton et al., 2020). However, the master's drop-out rate for German students was 15%, whereas it was 28% for students from other countries (Kercher, 2018).

4.3.1.1 Factors That Influence E-Learning Students Dropout Rates Among Post Graduate.

Table xii Research Skills Determining Completion of Postgraduate Studies

	Strongly Disagree	Agree	Disagree	Neutral	Strongly Agree
Finding a research problem is simple.	0.0%	23.3%	23.3%	40.8%	12.5%
The creation of research goals is simple	0.0%	25.0%	30.8%	40.0%	4.2%
I found the problem's formulation to be very challenging.	0.0%	41.7%	9.2%	42.5%	6.7%
Creating a focused theme is simple.	0.0%	15.0%	33.3%	44.2%	7.5%
Creating timely topics is difficult	0.0%	38.3%	11.7%	40.0%	10.0%
I had and continue to have difficulty formulating a focused theme.	0.0%	36.7%	28.3%	28.3%	6.7%
Supervisors give contradictory advice	0.0%	29.2%	18.3%	40.0%	12.5%
I sideline one of my supervisors	0.0%	13.3%	32.5%	35.8%	18.3%
My supervisor(s) are reachable for advice.	0.0%	38.3%	7.5%	34.2%	20.0%

My supervisor(s) are experts in the field I'm studying	0.0%	48.3%	2.5%	13.3%	35.8%
I aggressively accept corrections from my supervisors	0.0%	42.5%	5.0%	21.7%	30.8%
E-Library has adequate resources	0.0%	23.3%	37.5%	22.5%	16.7%
The university's library provides training in work publication.	0.0%	20.8%	30.8%	26.7%	21.7%
The LMS always fails during class presentations	0.0%	27.5%	10.8%	37.5%	24.2%
Remote library is inaccessible	0.0%	34.2%	23.3%	18.3%	24.2%

According to table 12 above 40.8% were neutral that finding a research topic was simple, 23% agreed and disagreed respectively while 12.5% strongly agreed. 40% were neutral that the criteria for research goals was simple, 30.8% disagreed, 25% agreed while 4.2 % strongly agreed. 41.7 % agreed that they found the problem's formulation to be very challenging, 42.5% were neutral, 9.2% disagreed while 6.7% strongly agreed. 44.2 % were neutral that creating a focused theme is simple, 33.3% disagreed, 15% agreed while 7.5 % strongly agreed. 40% were neutral that creating timely topics was difficult, 38.3% agreed,11.7% disagreed while 10% strongly agreed. 36.7 % agreed that had continued to have difficulty formulating a focused theme, 28.3 % were neutral and disagreed while 6.7% strongly disagreed. 40% were neutral that the ssupervisor’s gave a contradictory advice, 29.2% agreed, 18.3% disagreed while 12.5 % strongly agreed.35.8 % were neutral that they sidelined one of their supervisors, 32.5% disagreed, 18.3 % strongly agreed while 13.3% agreed. 38.3 % agreed that the supervisor(s) were reachable for advice, 34.2 % were neutral, 20% strongly agreed while 7.5% disagreed. 48.3 % agreed that the supervisor(s) were experts in the field they were studying, 35.8% strongly agreed, 13.3% were neutral while 2.5 % disagreed. 42.5%

agreed that they aggressively accept corrections from the supervisors,30.8 % strongly agreed, 21.7% were neutral while 5% disagreed. 37.5 % disagreed that the E-Library had adequate resources, 23.3% agreed, 22.5% were neutral while 16.7% strongly agreed. 30.8 disagreed that the university's library provided training in research work publication, 26.7% were neutral, 21.7% strongly agreed while 20.8 % agreed. 37.5% were neutral that the LMS always failed during class presentations, 27.5% agreed, 24.2% strongly agreed while 10.8% disagreed. 34.2 % agreed that remote library was inaccessible, 24.2% strongly agreed, 23.3% disagreed while 18.3% were neutral. They argued that since the current unit course is so brief, it is imperative to reorganize research unit courses so that many of the fundamentals of research are taught at the undergraduate level and a more intensive and extensive course is developed for master's and PhD programs to give students practical research skills. They also mentioned that being included in academic research would be beneficial for them. This finding is consistent with earlier research (Eyangu et al., 2014; Ezebilo, 2012; Shariff et al., 2015), which found that research skills in problem definition, literature review, research methodology, and data collection are essential for completing postgraduate research successfully and on schedule. There are no reliable models for forecasting students' success, despite the fact that retention rates for master's students have been extensively studied (Rotem et al., 2020). Building models that can anticipate attrition at the earliest feasible stage is crucial to minimizing the waste of financial and human resources caused by failure or dropouts.

Table xiii Respondents who have been in the field

	DisAgree	Strongly DisAgree	Agree	Strongly Agree	Neutr al
I didn't gather a lot of data in the field.	0.0%	0.0%	10.0%	32.5%	48.3%

I was unsuccessful in persuading respondents to take part.	0.0%	0.0%	6.7%	37.5%	52.5 %
I believe that my lack of prior fieldwork experience hampered the quality of the data I gathered.	0.0%	0.0%	28.3%	14.2%	54.2 %
I believe that my lack of prior fieldwork experience hampered the quality of the data I gathered.	0.0%	0.0%	28.3%	14.2%	54.2 %

According to table 13 above, 48.3% were neutral that didn't gather a lot of data in the field, 32.5% strongly agreed,52.5 % were neutral that they were unsuccessful in persuading respondents to take part, 37.5% strongly agree, while 6.7% agreed, 54.2% were neutral that their lack of prior fieldwork experience hampered the quality of the data I gathered, 28.3% agreed while 14.2% strongly agreed, 54.2% were neutral that their lack of prior fieldwork experience hampered the quality of the data gathered, 28.3% agreed while 14.2% strongly agreed.

Table xiv E-Learning Institutional Attributes

	Disagree	Agree	Disagree	Neutral	Strongly Agree
Relevance of University's Quality Assurance Policies with regard to the deployment of e-learning.	0.0%	20.8%	21.7%	33.3%	18.3%

The University's allocation of financial resources for e-learning is sufficient for my institution to pursue an e-learning program.	0.0%	25.0%	33.3%	26.7%	9.2%
In my institution, the technological support for the deployment of e-learning is current.	0.0%	35.8%	30.8%	20.0%	7.5%
In my school, the technological support needs for e-learning adoption are extensively specified.	0.0%	26.7%	31.7%	25.8%	10.0%
Adoption of e-learning is not impeded by my institution's technical support for it.	0.0%	27.5%	24.2%	32.5%	10.0%
Because I recognize the value of the e-learning system, I constantly want to use it.	0.0%	29.2%	26.7%	14.2%	24.2%

According to table 14 above, 33.3% there was relevance in the University's Quality Assurance Policies with regard to the deployment of e-learning, 18.3% strongly agreed, 21.7% disagreed while 20.8% agreed. 33.3% disagreed that the University's allocation of financial resources for e-learning was sufficient for their institution to pursue an e-learning program, 26.7% were neutral, 25% agreed while 9.2% strongly agreed. 35.8% of the respondents agreed that In the institution, the technological support for the deployment of e-learning was current, 30.8% disagreed, 20% were neutral while 7.5% strongly agreed.31.7 % disagreed that the school's technological support needed for e-learning adoption was extensively specified, 26.7% agreed, 25.8% were neutral while 10% strongly agreed.32.5% were neutral that the Adoption of e-learning is not impeded by my institution's technical support for IT, 27.5% agreed, 24.2% disagreed while 10% strongly disagreed. 29.2% agreed that Because they recognize the value of the e-learning system, they constantly want to use it,26.7% disagreed, 24.2% strongly agreed while 14.2% were neutral. Due to their limited research experience, master's students find that practically every research project presents a number of problems. The substance of the research skill course unit and the way it is taught, according to the students, are the main causes of the difficulties with research skills. They recommended that the research skill unit be taught at the school level, where a lot of the theoretical material may be taught, and introduced to undergraduate students. Practical and learner-centered approaches should be used at the master's and PhD levels.

4.3.2 Objective 2: To determine mitigation measures for reducing the drop out of postgraduate e-learning students by 20% among the selected public universities in Kenya.

The purpose of this objective was to determine the mechanisms that can be put in place to reduce the dropout rate of postgraduate students.

Table xv Mitigations of postgraduate students' dropout

	Strongly Disagree	Agree	Disagree	Neutral	Strongly Agree
Undergraduates should be taught research techniques.	0.0%	29.2%	2.5%	0.0%	68.3%
To obtain experience, graduate students should participate in faculty research.	0.0%	24.2%	10.8%	0.0%	65.0%
I feel having sufficient capital to pay fee helps to reduce the dropout	0.0%	30.8%	2.5%	11.7%	55.0%
Seminars and workshops are held by graduate schools to provide research training.	0.0%	35.8%	16.7%	19.2%	28.3%

The university should provide start mentoring programs	0.0%	26.7%	2.5%	2.5%	68.3%
When necessary, graduate schools provide research consultation.	0.0%	37.5%	19.2%	17.5%	25.8%
Efficient eLearning system will help to complete the course faster	0.0%	0.0%	33.3%	10.8%	55.8%
Supervisors should create more time for the students	0.0%	30.0%	2.5%	2.5%	59.2%

According to table 15 above 68.3 % strongly agreed that Undergraduates should be taught research techniques to help them when they enroll to postgraduate, 29.2% agreed while 2.5% disagreed. 65% strongly agreed that in order to obtain experience, graduate students should participate in faculty research, 24.2% agreed while 10.8% disagreed. 55% strongly agreed that I feel having sufficient capital to pay fee helps to reduce the dropout. 30.8% agreed, 11.7% were neutral while 2.5% disagreed. 28.3% strongly agreed that's seminars and workshops are held by graduate schools to provide research training, 35.8% agreed, 19.2% were neutral while 16.7% disagreed.68.3% strongly agreed that the university should provide start mentoring programs,26.7% agreed while 2.5% were neutral and disagreed. 37.5% agreed that when necessary, graduate schools should provide research consultation, 25.8% strongly agreed, 19.2% disagreed while 17.5 % were neutral. 55.8% strongly agreed that efficient eLearning system will help to complete the course faster, 33.3% disagreed while 10.8% were neutral. 59.2% strongly agreed that the ssupervisor's should create more time for the students, 30% agreed while 2.5% were neutral and disagreed. For postgraduate courses to be completed on schedule, supervisory elements are crucial. These results concur with those of Chiappetta and Watt (2011) and Ezebilo

(2012), who stressed the importance of the supervisory component in completing postgraduate research.

4.3.3 Objective 3: To develop a Hybrid model for predicting e-learning course dropout rate for postgraduate.

4.3.3.1 Model development Methodology

The model for predicting postgraduate student's dropout was designed using two supervised machine learning algorithms; Random Forest and support vector machine. The algorithms were used to enhance accuracy of the model. To depict how the model works, a prototype was developed using python programming language and coding was done on Jupyter notebook. Python was used because it provides useful libraries for machine learning models.

The diagram below shows the conceptual model for predicting student dropout

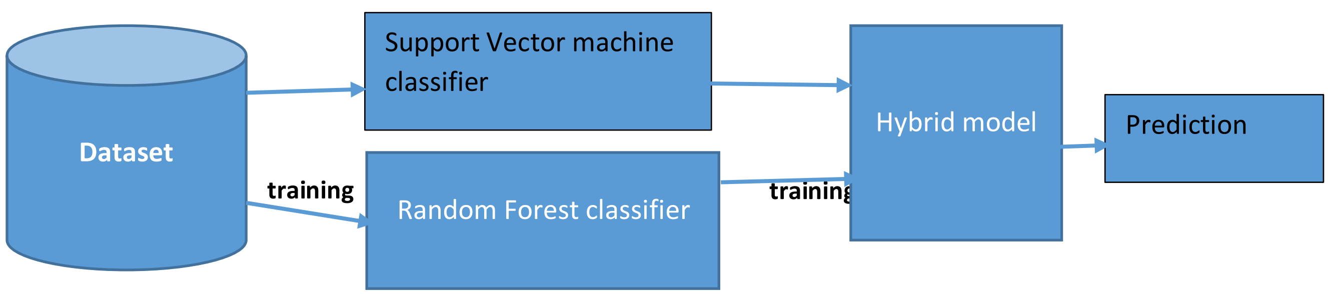

Figure iv Prediction model for predicting Postgraduate dropout

Dataset

The dataset for the experimental setup was obtained through record inspection from the study site admission and registrar's office. This dataset consists of postgraduate student's records who were enrolled for the various study programs offered by the university between 2019 and 2023. A total of 300 records were collected and stored in a csv file. The elements that were considered for predicting dropouts are shown in the table 16 below.

Table xvi Dataset

	Label	
Attribute	Yes	No
Completed course work	2	1
Fees paid in full	2	1
Passed all exams	2	1
Met attendance target	2	1
Supervisors were available/supportive	2	1
Collected data	2	1
Thesis accepted/passed	2	1
Busy work schedule	1	2
Family/social issues	2	1
Goal change	2	1

Figure v below shows a sample dataset that was used for model development

	stud_no	completeted coursework	feespaid	passed all exams	met the attendance target	supervisors were available	collected data	thesis accepted/passed	busy work schedules	family issues/social issue	goal change	total	average	status
0	2	2	1	2	2	1	1	2	1	2	2	16	1.6	graduated
1	3	2	1	1	2	1	1	2	2	1	2	15	1.5	graduated
2	4	1	1	1	2	1	2	1	2	1	1	13	1.3	dropped
3	5	2	1	1	1	1	1	2	2	2	2	15	1.5	graduated
4	6	2	1	2	1	1	2	2	1	2	2	16	1.6	graduated

Figure v Sample data set

a) ***Data pre-processing-***This is an important step in any data mining process. This basically involves transforming raw data into an understandable format for machine learning models. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. This will help in getting better results through the classification algorithms.

b) ***Dataset classification-***The dataset set was classified into two categories; graduated and dropout. To achieve this, the mean for all the data items was computed. Values lower than 1.5 were categorized as dropout while values greater or equal to 1.5 were classified as graduated.

c) ***Dataset encoding-***This process takes categorical variables, such as status (graduated/dropout) and converts it to a numerical representation (1 for graduated and 0 for dropout) without an arbitrary ordering for machine learning model.

4.3.3.2 Training and Testing Sets

There is one final step of data preparation: splitting data into training and testing sets. 20% of Dataset was used for testing while the remaining 80% for training the model.

a) ***Train Model-***Model training involves training and testing various machine learning models to determine their accuracy and performance in predicting student dropout. After data preparation, creating and training the model was achieved using skicit-learn. The random forest regression model and SVC from skicit-learn were trained using the dataset, and fitted on the training data.

b) ***Hybrid model-***The hybrid model makes it possible to leverage the diversity and complementarity of different models and get a more robust and accurate forecast. For this study, Random forest and support vector machine learning algorithms were used to develop a model for predicting dropouts. The predictions of both models were combined to achieve better accuracy. To implement an average ensemble in Python using scikit-learn, the researcher used the Voting Classifier class for classification problems. These classes allow us to specify a list of models and a voting method to combine their predictions.

4.4.4 Objective 4: To validate the developed postgraduate prediction model among the eLearning students.

4.4.4.1 Model validation

The model was validated by using the machine learning algorithms based on the dataset.

Figure vi below shows the hybrid model generated after training and fitting.

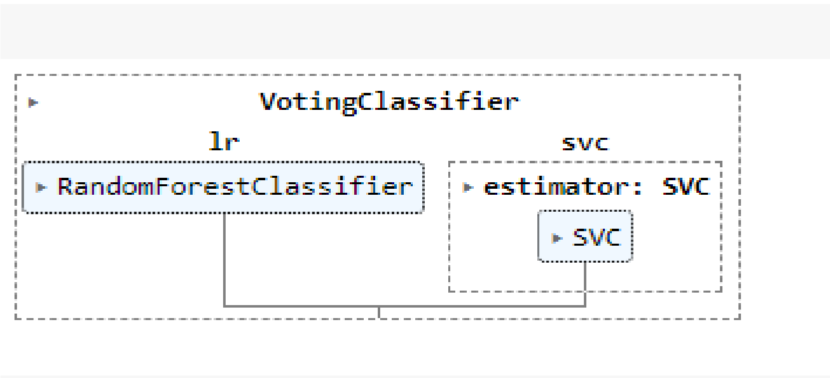

Figure vi Hybrid prediction model

The hybrid model achieved a 100% accuracy in predicting student dropout as can be seen in the figure below

```
[42] # Evaluate performance
     acc = accuracy_score(y_test, y_pred5)
     acc = acc*100
     print(f'Accuracy: {acc:.2f} %')

     Accuracy: 100.00 %
```

Figure vii Model Accuracy

The hybrid model achieved an accuracy of 100%

4.4.4.1 Model performance evaluation and testing

To evaluate the performance of the model, data was used to test whether the model could predict the based on the values supplied as seen in the diagram below.

```
[43] #hybrid model performance evaluation/testing
```

```
[44] #predict with test data
ensemble.predict([[1.4]])

array(['dropped'], dtype=object)
```

```
#predict with test data
ensemble.predict([[1.8]])

array(['graduated'], dtype=object)
```

Figure viii Model Performance

4.4.4.2 Model confusion Matrix

```
print('confusion matrix:\n', confusion_matrix(y_pred_hbd,y_true_hbd))
print('accuracy:', accuracy_score(y_pred_hbd,y_true_hbd)*100, "%")
```

```
confusion matrix:
 [[31  0]
 [ 0 29]]
accuracy: 100.0 %
```

Figure ix Confusion Matrix

The application of machine learning in predicting student attrition presents a transformative opportunity for educational institutions to tackle this widespread issue effectively. By leveraging the capabilities of machine learning algorithms, educators, policymakers, and institutions can take proactive measures, provide targeted support, and foster an environment conducive to student success. Nevertheless, it is imperative to navigate challenges pertaining to data quality, bias, and ethical considerations to ensure the conscientious and equitable use of these predictive models. As machine learning and data analysis continue to advance, we hold the potential to make substantial strides in reducing student attrition, enhancing educational outcomes, and cultivating an inclusive and supportive education system that caters to the needs of all students.

4.5 Discussions of Individual Objectives

4.5.1 Discussion on objective one

With the aid of this information, the investigator was also able to assess how long the students had studied. Using this information, the researcher may compute the rate of years without completion. Only 20.8% of respondents (20.2%) had never used the eLearning system. 52.5% of the students were working on coursework, 18.3% were writing proposals, 11.7% were in the midst of data collecting, 7.5% had already graduated, 6.7% were producing thesis reports, and 3.3% were working on concept papers. The enrollment year was crucial to the study because it informed the researcher about the duration of the students' attendance. 61.7% of students are anticipated to graduate in 2024. The data made it easier for the researcher to ascertain whether pupils had spent more time studying than was expected. The biggest obstacle to finishing the research was 62.5% of the respondent's price, whereas 37.5% of the charge was not a problem. Even though financial status wasn't the major variable due to other factors, it was still important to understand how it affected the study's dependent variable, which was the intervening variable. The institutional, individual, personal, research, and supervisory aspects all had an impact on the dropout rate of postgraduate students, which was 45.8%. Making a concentrated topic is easy, according to 44.2% of respondents. However, 33.3% disagreed, 15% agreed, and 7.5% strongly agreed. When asked whether the E-Library has enough resources, 37.5% disagreed, followed by 23.3% in favor, 22.5% indifferent, and 16.7% strongly in favor. The university library offered instruction in research work publication, according to 30.8 respondents who disagreed, 26.7% who were neutral, 21.7% who strongly agreed, and 20.8% who agreed. When it came to the LMS's consistent failure during class presentations, 37.5% were neutral, 27.5% agreed, 24.2% strongly agreed, and 10.8% disagreed. 54.2% were neutral that their lack of prior fieldwork experience hampered the quality of the data I gathered, while 48.3% were neutral that they didn't gather a lot of data in the field, 32.5% strongly agreed, 52.5% were neutral that they were unsuccessful in persuading respondents to participate, 37.5% strongly agree, and 6.7% agreed. According to research by Rong'uno (2016), scholarship recipients were more likely than other students to complete their dissertations in a timely manner. Rung'uno also noted that, regardless of the source of income, postgraduate students who self-funded were less likely than those who got funding to complete their studies and graduate on time. It was crucial to comprehend how financial status affected the study's dependent variable, which was the intervening variable, even if it wasn't a primary variable because of other circumstances. This aspect was crucial to include in the study because previous studies had shown that postgraduate students' lack of completion was significantly impacted by money (Garwe & Maganga 2015; Rauf, 2016). The findings above suggested that the main cause of non-completion had been money.

The percentage of respondents who thought the university's financial support for e-learning was insufficient for their institution to pursue an e-learning program varied from 33.3% who disagreed, 26.7% who were neutral, 25% who agreed, and 9.2% who strongly agreed. 35.8% of the respondents agreed, 30.8% disagreed, 20% were neutral, and 7.5% strongly agreed that the institution's technological assistance for the adoption of e-learning was up to date. The requirement for technological support at the school for the use of e-learning was not agreed upon by 31.7% of respondents. They also mentioned that being included in academic research would be beneficial for them. This finding is consistent with earlier research (Eyangu et al., 2014; Ezebilo, 2012; Shariff et al., 2015), which found that research skills in problem definition, literature review, research methodology, and data collection are essential for completing postgraduate research successfully and on schedule. According to this study, the online library is the sole place where JKUAT students who are enrolled in e-learning can find social support. The students claimed that the LMS forum and chat were not active, so it appears that information support, affirmational support, and emotional support were ineffective. An key motivation for online learners is social support, which is typically divided into four categories: supportive, informative, instrumental, and emotional support (Munich, 2014). Peers, forums, chat, and group work in e-learning are the main sources of this assistance (Weng & Chung, 2015; Queiros & de Villiers, 2016). JKUAT must consequently include LMS chats and forums in its courses in order to increase social support. The findings indicate that over half of the students are passionate about online learning and have practical computer and internet skills. However, nearly half of the students bemoaned a lack of LMS training and professors who lacked passion. According to Jung's (2017) research, both intrinsic and extrinsic learner motivation are essential for students' performance in an online coursework setting. In order to increase quality, JKUAT must offer training because it is a means of disseminating e-learning capabilities. (Arinto, 2016; Azawei et al., 2016).

4.5.2 Discussion on objective two

In order to prepare students for postgraduate study, 68.3% strongly agreed that undergraduates should be taught research methodologies, and 29.2% also agreed. 55% highly concurred with my opinion that having enough money to cover tuition helps to lower dropout rates. 11.7% were neutral, 30.8% agreed, and 2.5% were not. 28.3% firmly agreed that graduate schools hold seminars and workshops to train students in research. Supervision is essential for postgraduate courses to be finished on time. These findings are in line with Chiappetta and Watt (2011) and Ezebilo (2012), who emphasized the significance of the supervisory component in completing postgraduate research. 55.8% firmly agreed that using an effective eLearning system will help students finish the course more quickly.

4.5.3 Discussion on objective three

Two supervised machine learning techniques, Random Forest and Support Vector Machine, were used to create the model for forecasting postgraduate students' dropout rates. The algorithms were applied to increase the model's accuracy. The study site's admission and registrar's office provided the dataset for the experimental setting through record examination. This dataset comprises of the enrollment information for postgraduate students who attended one of the university's many study programs between 2019 and 2023. A csv file containing 300 records overall was assembled and kept. The division of data into training and testing sets is the last step in the data preparation process. The remaining 80% of the dataset was utilized to train the model, with only 20% used for testing. To evaluate the efficacy and performance of several machine learning models in predicting student dropout, model training involved training and testing. Skicit-learn was used to create and train the model after data preprocessing. The dataset was used to train the random forest regression model and the SVC from skicit-learn, which was then fitted using the training data. With the help of the hybrid model, it is possible to take advantage of the diversity and complementarity of other models to produce a forecast that is more reliable and precise. In this study, a model for predicting dropouts was created using the Random forest and support vector machine learning methods. To improve accuracy, the forecasts from the two models were blended. The Voting Classifier class for classification issues was utilized by the researcher to create an average ensemble in Python using scikit-learn.

4.5.4 Discussion on objective four

Using machine learning methods based on the dataset, the model was validated. When used to forecast student dropout, the hybrid model had a 100% accuracy rate. The use of machine learning to forecast student attrition offers educational institutions a game-changing chance to successfully address this pervasive problem. By utilizing the capabilities of machine learning algorithms, institutions, policymakers, and educators can be proactive, give students the help they need, and create a climate that supports their success. To ensure the ethical and fair application of these predictive models, it is crucial to manage issues with data quality, bias, and ethical considerations.

We have the opportunity to make significant advancements in decreasing student attrition, improving educational results, and developing an inclusive and supportive educational system that meets the needs of all students as machine learning and data analysis develop.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

The Summary of findings, study conclusions and Further Research Recommendations are presented in this chapter.

5.1 Summary

5.1.1 Summary of objective one

The first objective was to analyses the factors that influence e-learning students dropout rates among post graduate. Because it provided the researcher with information regarding the duration of the students' attendance, the enrollment year was essential to the study. In 2024, 61.7% of students are expected to graduate. With the aid of the data, the researcher was better able to determine whether students had spent more time studying than was anticipated. 62.5% of the respondent's pricing posed the most challenge to completing the research, whilst 37.5% of the fee posed no issues. It was crucial to comprehend how financial status affected the study's dependent variable, which was the intervening variable, even though it wasn't the main variable because of other circumstances. The dropout rate of postgraduate students, which was 45.8%, was influenced by institutional, individual, personal, research, and supervisory factors. A range of respondents—33.3% agreed and 33.3% disagreed—felt that the university's financial support for e-learning was insufficient for their institution to undertake an e-learning program. 31.7% of respondents disagreed that the school needed to have technological support in order to adopt e-learning. They also noted how it would be advantageous for them to participate in academic research. Making a concentrated topic is easy, according to 44.2% of respondents. However, 33.3% disagreed, 15% agreed, and 7.5% strongly agreed. While 40% disagreed that coming up with timely themes was tough, 38.3% did not. The supervisor(s) being approachable for advice was endorsed by 38.3%, 34.2% were neutral, and 20% strongly agreed. 35.8% strongly agreed that the supervisor(s) were/were subject matter experts in the area they were studying, and 48.3% agreed.

5.1. 2 Summary of objective two

The second objective to determine mitigation measures for reducing the drop out of postgraduate e-learning students by 20% among the selected public universities in Kenya. 68.3% strongly agreed that undergraduates should be taught research procedures in order to prepare them for postgraduate studies, and 29.2% also agreed. 55% of respondents strongly agreed with my assertion that having enough money to pay for college reduces dropout rates.26.7% of respondents agreed, and 68.3% strongly agreed that the institution should initiate mentorship programs. 30% agreed, 2.5% were neutral, and 59.2% strongly agreed that the supervisors should give the students additional time. Supervision is essential for postgraduate courses to be finished on time. Graduate institutions should offer research consultation as appropriate, according to 37.5% of respondents, with 25.8% strongly agreeing. 55% highly concurred with my opinion that having enough money to cover tuition helps to lower dropout rates.

5.1. 2 Summary of objective three

The third objective was to develop a Hybrid model for predicting e-learning course dropout rate for postgraduate. The final step in the data preparation procedure was to separate the data into training and testing sets. Only 20% of the dataset was used for testing, with the remaining 80% being used to train the model. Model training comprised training and testing in order to assess the performance of multiple machine learning models in predicting student dropout. In this study, support vector machine learning and random forest were used to build a dropout prediction model. The two models' forecasts were combined to increase precision. The researcher used scikit-learn to build an average ensemble in Python utilizing the Voting Classifier class for classification problems.

5.1. 2 Summary of objective four

The fourth objective was to validate the developed postgraduate prediction model among the eLearning students. The model was validated using machine learning techniques based on the dataset. The hybrid model showed a 100% accuracy rate for predicting student dropout. Educational institutions have a revolutionary opportunity to properly address this prevalent issue through the application of machine learning to forecast student attrition. Institutions, legislators, and educators may be proactive, provide students with the support they need, and foster an environment that supports their success by harnessing the capabilities of machine learning algorithms. The management of concerns with data quality, bias, and ethical considerations is essential to guaranteeing the just and ethical implementation of these predictive models.

5.2 Conclusion

In order to successfully adopt, implement, and use e-learning systems in light of the competitive expansion of e-learning in developing countries, institutions that offer e-learning must improve the quality of their e-learning systems based on these aspects. By providing a standard for improving quality and acting as a valuable benchmark for e-learning providers and policy makers, the current study may help us understand the e-learning systems in Kenya and other developing nations. According to related study, there are numerous factors that can contribute to student dropout, and individual variability may result in significant differences. The key research abilities that affect post graduate non-completion include identifying the research problem, stating the research problem, developing research objectives, critiquing the literature that has already been published, ensuring that the literature review is consistent with the research objectives, identifying the research gap, and connecting the reviewed literature to the current study. The main supervision factors influencing non-completion time are supervisor accessibility, timely feedback, and co-supervision. Significant institutional factors for completion of regular postgraduate studies included difficult access to internal and external research materials, a rarity of inductions on work publication and copyright abilities, a rarity of providing students with research money, and an unmotivated staff. We were only able to use the students' personal characteristics and academic performance as input variables in the predictive model due to the comprehensiveness of the data we were able to obtain for this study from the academic management systems of online educational institutions. This had an impact on the prediction accuracy which was at 100%. The use of machine learning to forecast student attrition offers educational institutions a game-changing chance to successfully address this pervasive problem. By utilizing the capabilities of machine learning algorithms, institutions, policymakers, and educators can be proactive, give students the help they need, and create a climate that supports their success. To ensure the ethical and fair application of these predictive models, it is crucial to manage issues with data quality, bias, and ethical considerations. We have the opportunity to make significant advancements in decreasing student attrition, improving educational results, and developing an inclusive and supportive educational system that meets the needs of all students as machine learning and data analysis develop.

5.3 Recommendations

Before enrolling in classes, students who wanted to pursue further education should set up a funding source or educational safety net. This would guarantee that they wouldn't stop attending school because they couldn't afford the tuition. The report also advised students to make plans for their study schedules and make sure that other commitments that may wait should not interfere with study time. For instance, women should opt to continue their education when they are certain they do not have a parental responsibility that could keep them from doing so for an extended period of time. In order to avoid undue stress from juggling multiple things at once, the study also suggested that the students select a study model that would work best with their schedules.

5.4 Contribution

The development of a hybrid model for forecasting postgraduate student dropout rates from online courses is a significant contribution to the fields of education and technology. In order to forecast postgraduate student dropout, the researcher added to the body of knowledge by creating a hybrid model utilizing the random forest and SVM algorithms. The model's accuracy was 100 percent. The information and conclusions offered by the hybrid model can be used as support for the creation of laws governing online education. This data can be used by institutions and policymakers to develop policies that are specifically tailored to the needs of online postgraduate students. determining the causes of dropout rates among underrepresented or disadvantaged groups, which will enable the creation of inclusive policies that will provide all students with the same opportunity and support. How educational institutions handle student care, resource allocation, and policy-making could be revolutionized by the creation and use of a hybrid model for predicting postgraduate student dropout rates from online programs. Institutions can develop a more inclusive, helpful, and productive online learning environment by utilizing data-driven insights, thereby enhancing student success rates and the standard of education as a whole.

5.5 Future work

Deep learning needs to be the subject of more study. To tackle the imbalanced datasets typical of dropout prediction challenges, deep learning models can be supplemented with strategies like Synthetic Minority Over-sampling Technique (SMOTE). By incorporating these strategies, deep learning models for predicting postgraduate dropouts can become more accurate, interpretable, and ethical, making a substantial contribution to the field of education analytics. Future research should build on our results by making predictions earlier (before enrollment), as this can benefit the educational system more. Pre-enrollment exams may help achieve early forecasts, according to Alturki and Stuckenschmidt (2021). Additionally, we wholeheartedly support testing the predictive modes on additional master applications. Future research should also investigate additional ensemble strategies for making academic forecasts. To deal with the unbalanced datasets, we also advise looking into other oversampling methods.

REFERENCES

Abragi, J. O. (2008). Revritalizing Financing of Higher education in Kenya . Nairobi: AAU.

Abongo, C. (2009). Performmance-only measurement of service quality: a replication and externsio . Journal of Business Research volume 55 issue 1, 17 - 31 .

Aburas, R. A., & Qahmash, A. H. (2019). Predicting student dropout in e-learning environment using machine learning algorithms. Journal of Intelligent Learning Systems and Applications, 11(3), 67-78. https://doi.org/10.4236/jilsa.2019.113006

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, *22*(10), 1533-1545.

Abubakar, Y., & Ahmad, N. B. H. (2017). Prediction of Studentsâ€™ Performance in E-Learning Environment Using Random Forest. *International Journal of Innovative Computing*, *7*(2).

Agasisti, T., Cratalano, G., Landoni, P., & Verganti, R. (2012). Evaluating the performance of academic departments: an analysis of research-related output efficiency. Research Evaluation vol 21(1), 2-14.

Ajzen, I., & Firshbein, M. (2004). Question raised by a reason action approach. Journal of Health ptschology vol.23 issue 4, 431- 434 .

Alqahtani, M. (2022). Factors influencing e-learning adoption during the COVID-19 pandemic: A review of the literature. Education Sciences, 12(3), 297.

Albawi, S., Bayat, O., Al-Azawi, S., & Ucan, O. N. (2017). Social touch gesture recognition using convolutional neural network. *Computational Intelligence and Neuroscience*, *2018*.

Allam, H., Khalifa, H., & El-Henawy, H. (2022). Predicting e-learning adoption in higher education during the COVID-19 pandemic: A model based on the technology acceptance model and theory of planned behavior. Sustainability, 14(3), 1177.

Alturki S, Alturki N. (2021). Using educational data mining to predict Students' academic performance for applying early interventions. *Journal of Information Technology Education: Innovations in Practice.* 2021;20:121–137

Agaard, J. (2018). Magnetic and multistable: reinterpreting the affordances of educational technology. International Journal of Educational Technology in Higher Education, 15(4).

AlKhuder, S. B., & AlAli, F. H. (2017). E-Learning Systems Requirements Elicitation: Perspectives and Considerations. International Journal of Virtual and Personal Learning Environments, 7(1), 44-55.

Boezerooij , P. (2006). E-Learning stategie of higher education institutions. 17-67.

Chen, L., Liu, Y., Chen, M., & Lin, F. (2022). Prediction of student dropout in a MOOC using machine learning algorithms. Computers & Education, 173, 104212.

Chen, Y., Hu, H., Liu, Z., & Xie, H. (2020). A comparative study of machine learning algorithms for predicting student dropout in online learning. International Journal of Emerging Technologies in Learning, 15(6), 40-56. https://doi.org/10.3991/ijet.v15i06.12371

Du Toit, A., & Gaotlhobogwe, M. (2017). Benchmarking the intended technology curricula of Botswana and South Africa: What can we Learn. African Journal of Research in Mathematics, Science and Technology, 21(2), 148-158.

Garwe, E. C., & Maganga, E. (2015). The effect of student financial constraints on university non-completion rates. International Journal of Education, 7(2), 322-336.

Guo, J., Wu, X., & Zhang, C. (2021). An improved random forest algorithm for predicting student dropout in online learning. Journal of Educational Technology Development and Exchange, 14(1), 77-94. https://doi.org/10.18785/jetde.1401.06

Guo, Y., Wang, X., & Cheng, S. (2019). Predicting student dropout in e-learning courses using random forest. Journal of Educational Technology Development and Exchange, 12(2), 55-70. https://doi.org/10.18785/jetde.1202.05

Gachago, D., Morkel, J., Hitge, L., Zyl, I. v., & Ivala, E. (2017). Developing eLearning champions: a design thinking approach. International Journal of Educational Technology in Higher Education, 14(30).

Joo, Y. J., Lim, K. Y., & Kim, E. K. (2022). The effects of online learning motivation and self-regulation on academic performance in higher education during the COVID-19 pandemic. Interactive Learning Environments, 1-14.

Kehm BM, Larsen MR, Sommersel HB.(2019). Student dropout from universities in Europe: A review of empirical literature. *Hungarian Educational Research Journal*. 2019;9(2):147–164. doi: 10.1556/063.9.2019.1.18.

Kercher, J. (2018). *Academic success and dropout among international students in Germany and other major host countries*.

Kusuma, A. S., Zulfikar, A., & Setiawan, A. (2021). Factors affecting e-learning adoption during the COVID-19 pandemic: A study of Indonesian higher education institutions. Journal of Open Innovation: Technology, Market, and Complexity, 7(4), 117.

Khanna, P., & Sharma, M. (2019). Predicting student dropout in online courses using machine learning algorithms. International Journal of Emerging Technologies in Learning, 14(20), 136-150. https://doi.org/10.3991/ijet.v14i20.11508

Mukami, M. N., & Wang'eri, T. (2016). Examination repeats, semester deferment and dropping out as contributor of attrition rates in private universities in Nairobi County, Kenya. International Journal of Education and Research. 4(3), 225-237

Murray, T. (2016). Coordinating the Complexity of Tools, Tasks, and Users: On TheoryBased Approaches to Authoring Tool Usability. International Journal of Artificial Intelligence in Education, 26(1), 37-71.

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. International Journal of Educational Technology in Higher Education, 15(3)

Maor, D., & Currie, J. K. (2017). The use of technology in postgraduate supervision pedagogy in two Australian universities. International Journal of Educational Technology in Higher Education, 14(1). Retrieved from https://doi.org/10.1186/s41239-017-0046-1

Mouton D, Zhang H, Ertl B.(2020). German university student's reasons for dropout. Identifying latent classes. *Journal for Educational Research.* 2020;12(2):190–224.

Jiranek, V. (2010). Potential predictors of timely completion among dissertation research students at an Australian faculty of sciences. International Journal of Doctoral Studies, 5(1), 1-13

Li, C. H., & Hung, Y. H. (2021). Exploring the factors affecting e-learning performance during the COVID-19 pandemic: A perspective of information system success model. Sustainability, 13(21), 12071.

Liu, H., Lu, J., & Sun, J. (2022). Predicting student dropout in blended learning using machine learning algorithms. International Journal of Emerging Technologies in Learning, 17(3), 85-98. https://doi.org/10.3991/ijet.v17i03.14563

Liu, L., Zhou, J., & Hu, H. (2020). Predicting student dropout in online courses using machine learning algorithms. Journal of Educational Technology Development and Exchange, 13(2), 21-38. https://doi.org/10.18785/jetde.1302.02

Muchai, J., & Orodho, J. A. (2019). Factors influencing the successful implementation of the Competency-Based Curriculum (CBC) in Kenya: A case of public primary schools in Nakuru Municipality. International Journal of Education and Research, 7(6), 169-184.

Mukami, M. N., & Wang'eri, T. (2016). Examination repeats, semester deferment and dropping out as contributor of attrition rates in private universities in Nairobi County, Kenya. International Journal of Education and Research. 4(3), 225-237

Mwebi, B., & Simatwa, E. M. (2013). Expansion of private Universities in Kenya and its implication on quality and completion rate: An analytical study. *Educational Research, 4*, 352-366.

Pima, J. M., Odetayo, M., & Iqbal, R. (2018). A Thematic Review of Blended Learning in Higher Education. International Journal of Mobile and Blended Learning, 10(1), 1-11.

Peng, P., Yang, W. F., Liu, Y., Chen, S., Wang, Y., Yang, Q., ... & Liu, T. (2022). High prevalence and risk factors of dropout intention among Chinese medical postgraduates. *Medical Education Online*, *27*(1), 2058866.

Rauf, F. A. (2016). Challenges of thesis work: towards minimizing the non-completion rate in the postgraduate degree program. concern, 8(7).

Rhea, W., Kevin, E. &, Derek, K. L. (2013). Student attrition in the Ontario midwifery education Programme, Midwifery, Volume 29, issue 4

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary Educational Psychology, 25(1), 54-67.

Rong'uno, K. S. (2016). The Institutional and student related factors and doctoral studies completion rates in Education at selected public universities in Kenya. International Journal of humanities and social sciences. (Doctoral dissertation, University of Nairobi)

Rotem N, Yair G, Shustak E. (2020). *Dropping out of Master's Degrees: Objective Predictors and Subjective Reasons.* 2020;40(5):1070–1084. doi: 10.1080/07294360.2020.1799951

Staiculescu, C. & Richiteanu, N. E. R. (2018). University dropout. Causes and solution. In *Mental health global challenges XXI Century conference proceedings.*

Singh, D. a. (2013). One Text, Multi-Tasks Meeting the Mixed-Ability Challenge. Advances in Language and Literary Studies, 4(1), 1-12. Retrieved from https://eric.ed.gov/?q=multi+tasking&id=EJ1129637

Salajan, F. D., & Roumell, E. A. (2016). Two Decades of E-Learning Policy Evolution at EU Level: Motivations, Institutions and Instruments. European Journal of Education, 51(3), 391-407

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*, *134*, 93-101.

Tan, M., & Shao, P. (2015). Prediction of student dropout in e-Learning program through the use of machine learning method. *International journal of emerging technologies in learning*, *10*(1).

United Nations Developent Agenda. (2015). Science, technoloogy and innovation for sustainable development in the global partnership for development beyond 2015. Thematic Think Piece, 1-17.

Wang, Y., Pan, S., Zhang, C., & Huang, X. (2021). Prediction of student dropout in MOOCs using machine learning algorithms. Journal of Educational Computing Research, 59(4), 664-681. https://doi.org/10.1177/0735633120953484

Xiao Y, Zhu S-Y.(2020). China will fully implement the standardised training system for residents in 2020. Postgrad Med J. 2020;96(1141):647–648.

Yang, Y., Li, X., Li, J., & Wu, C. (2021). A machine learning approach to predicting student dropout in flipped classroom. Educational Technology Research and Development, 69(1), 349-367. https://doi.org/10.1007/s11423-020-09927-2

Zhang, Z., Wei, Z., & Wu, L. (2020). Predicting student dropout in e-learning using random forest and deep learning. IEEE Access, 8, 145134-145145. https://doi.org/10.1109/access.2020.3011008

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary Educational Psychology, 25(1), 82-91.

Zheng, Y., Gao, Z., Wang, Y., & Fu, Q. (2020). MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series. *IEEE Access*, *8*, 225324-225335

Appendix I
RESEARCH QUESTIONNAIRE

I humbly take this opportunity to pass my appreciation for taking your time to complete this questionnaire. **The purpose of the questionnaire is to obtain insights that will assist in developing a hybrid machine learning prediction model for postgraduate E-learning students' dropouts using random forest algorithm and Support Vector Machine to obtain a better prediction accuracy.**

I assure you that the information will be kept confidential kindly answer the questions as truthfully as possible.

Thank you.

SECTION A: General Information

Kindly check in the boxes provided.

1. What is your gender? Male ☐ Female ☐

2. Which is the highest Level of Education you have attained?

   a) Phd ( )
   b) Master's Degree ( )
   c) Bachelor's Degree ( )
   d) Diploma ( )
   e) Certificate ( )
   f) Other ( )

3. How long have you been in the university?

   Less than 1 year ( ) 1-2 years ( ) 3-5 years ( ) above 5 years ( )

4. What is your age?

a) Above 50 ( )

b) 40-50 ( )

c) 30-39 ( )

d) 20-29 ( )

e) Below 20 ( )

5. What is your marital status?

a) Single ()

b) Married ()

c) Divorced ()

d) Prefer not to say ()

6. What is your occupation within the university?

Staff ( ) Student ( )

7. Which department are you in?

a) ICT department

b) Operations

c) Media

d) Finance

e) Business

f) Hospitality

g) Education’

h) Other

SECTION B: OVERVIEW OF THE INFLUENCE OF E-LEARNING STUDENTS DROPOUT RATES AMONG POST GRADUATE

8. Have you interacted with an eLearning system?

Yes ( ) No ( )

9. What is your year of study?

a) Year one ()
b) Year two ()
c) Year three ()
d) Year four ()
e) Other ()

10. Which is your Level of Academic Work?

a) Concept development ( )
b) Course work ( )
c) Data collection ( )
d) Proposal development ( )
e) Report writing ( )
f) Graduated ( )

11. What is your expected Graduation Date?

a) 2023 ( )
b) 2024 ( )
c) 2025 ( )

d) 2026 ( )

e) Not sure ( )

12. Is fee payment challenge a to your research work?

Yes ( ) No ( )

12. Which of the following factors have affected the dropout rate?

a) Research skills

b) [ ] Supervisory factors [ ]

c) Institutional factors[ ]

d) Personal factors [ ]

e) All [ ]

SECTION B: FACTORS THAT INFLUENCE E-LEARNING STUDENTS DROPOUT RATES AMONG POST GRADUATE.

a) Research Skills Determining Completion Of Postgraduate Studies

Please read the statement carefully and tick appropriately.

Scale: *Strongly Agree 5, Agree 4, Neutral 3, Disagree 2, Strongly Disagree 1*

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Finding a research problem is simple.					
I found the problem's formulation to be very challenging.					
The creation of research goals is simple.					
Creating a focused theme is simple.					
Creating timely topics is difficult					
I had and continue to have difficulty formulating a focused theme.					

Supervisors give contradictory advice					
I sideline one of my supervisors					
My supervisor(s) are reachable for advice.					
My supervisor(s) are experts in the field I'm studying					
I aggressively accept corrections from my manager or managers					
E-Library has adequate resources					
In the library, citation help is available.					
The university's library provides training in work publication.					
The LMS always fails during class presentations					
Remote library is inaccessible					

b) For those who have been in the field

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I didn't gather a lot of data in the field.					
I believe the information I gathered was consistent.					
I was unsuccessful in persuading respondents to take part.					
I believe that my lack of prior fieldwork experience hampered the quality of the data I gathered.					
I believe the research techniques were thorough enough to gather enough data.					

c) E-Learning Institutional Attributes

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

Relevant are the University's Quality Assurance Policies with regard to the deployment of e-learning.					
The University's allocation of financial resources for e-learning is sufficient for my institution to pursue an e-learning program.					
In my institution, the technological support for the deployment of e-learning is current.					
In my school, the technological support needs for e-learning adoption are extensively specified.					
Adoption of e-learning is not impeded by my institution's technical support for it.					
Because I recognize the value of the e-learning system, I constantly want to use it.					
A better learning environment is made possible through the usage of e-learning.					

SECTION C: MITIGATION MEASURES FOR POSTGRADUATE STUDENTS DROPUPOUT

Scale: *Strongly Agree 5, Agree 4, Neutral 3, Disagree 2, Strongly Disagree 1*

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Undergraduates should be taught research techniques.					
To obtain experience, graduate students should participate in faculty research.					
The university should provide start mentoring programs					

Supervisors should create more time for the students					
I feel having sufficient capital to pay fee helps to reduce the dropout					
Seminars and workshops are held by graduate schools to provide research training.					
When necessary, graduate schools provide research consultation.					
Efficient eLearning system will help to complete the course faster					

APPENDIX II

Sample Code

```
# Define models
#random forest
cl2 = RandomForestClassifier(max_depth=10, random_state=0)
#svc
cl1=GridSearchCV(svc, parameters)
```

```
# Create a hybrid model
from sklearn.ensemble import VotingClassifier
```

```
# Combine the models using majority voting
ensemble = VotingClassifier(estimators=[('lr', cl2), ('svc', cl1)], voting='hard')
```

A HYBRID MODEL FOR PREDICTING E-LEARNING COURSE DROPOUT RATE FOR POST GRADUATE STUDENTS

ORIGINALITY REPORT

23%	21%	8%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.erudit.org Internet Source	1%
2	su-plus.strathmore.edu Internet Source	1%
3	oapub.org Internet Source	1%
4	repository.stcloudstate.edu Internet Source	1%
5	www.javatpoint.com Internet Source	1%
6	erepository.uonbi.ac.ke Internet Source	1%
7	www.worldcomp-proceedings.com Internet Source	1%
8	ir.mu.ac.ke:8080 Internet Source	1%
9	zone.biblio.laurentian.ca Internet Source	1%