PREDICTING CAMPUS ADMISSION THROUGH ASSESSMENT OF SOFT SKILLS USING RANDOM FOREST ALGORITHM

BY

DENNIS MUTUNGA MUTHUI

MASTER OF SCIENCE IN DATA ANALYTICS

KCA UNIVERSITY

2025

PREDICTING CAMPUS ADMISSION THROUGH ASSESSMENT OF SOFT SKILLS USING RANDOM FOREST ALGORITHM

DENNIS MUTUNGA MUTHUI

15/00325

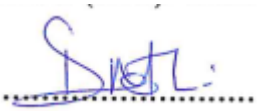
A DISSERTATION SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN DATA ANALYTICS IN THE SCHOOL OF SCHOOL OF TECHNOLOGY AT KCA UNIVERSITY

APRIL, 2025

DECLARATION

I declare that this dissertation is my original work and has not been previously published or submitted elsewhere for the award of a degree. I also declare that this contains materials written or published by others except where due reference is made, and the author duly acknowledged.

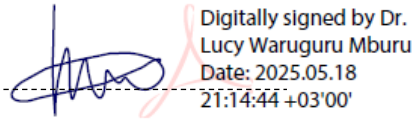
Student name: Dennis Mutunga Muthui Reg No: 15/00325

Sign:  Date: 09/04/2025

I thus affirm my assessment of the master's dissertation of Dennis Mutunga Muthui and have approved it for examination.

Dr. Lucy Waruguru Mburu

18/05/2025



ABSTRACT

This study develops a machine learning model to predict college admission success in Kenya by assessing soft skills using the random forest algorithm. The research addresses the growing importance of soft skills in academic and professional success, and current limitations in evaluating these abilities during admissions. The study identifies key soft skills, creates a comprehensive assessment tool, develops and tests a random forest model, and evaluates its performance, interpretability, and fairness. The methodology involves a quantitative predictive modeling design, employing stratified random sampling and rigorous data collection procedures. Results indicate that soft skills, particularly communication and problem-solving, are strong predictors of admission success, often outweighing traditional academic metrics. The random forest model achieved 98.36% accuracy in predicting admissions outcomes, with mathematics performance emerging as the most influential factor (22% importance), followed by GPA (18%), KCSE scores (16%), and science grades (15%), while soft skills showed more modest but meaningful contributions (communication 8%, leadership 5%, problem-solving 4%). The model demonstrated consistent performance across demographic groups, with perfect equal opportunity across gender, school type, location, and school level categories. However, the model reflected existing demographic disparities in admission rates that mirror broader equity challenges in educational access. The study concludes that incorporating soft skills assessments in admissions processes could provide a more holistic evaluation of applicants, though current practices continue to prioritize traditional academic achievement. Recommendations include integrating soft skills development in secondary education curricula and incorporating structured soft skills assessments in university admissions processes. This research contributes to the ongoing dialogue about evolving higher education admissions to better align with 21st-century workforce needs while promoting fairness and transparency in the admission process.

Keywords: Machine Learning, Random Forest Algorithm, University Admissions, Soft Skills Assessment, Educational Technology, Predictive Modeling

ACKNOWLEDGEMENT

I owe many people a debt of gratitude for this work's success. Initially, I would want to convey my sincere appreciation to the All-Powerful God for providing me with excellent health and the steadfast will to finish this project.

Second, I would like to express my sincere gratitude to Dr. Lucy M. Waruguru, my wonderful supervisor, for all of her help and advice during the research process. This job has been shaped and completed thanks to her deep expertise, patient mentoring, and persistent commitment to excellence.

Additionally, I would like to express my profound gratitude to the KCA University instructors and staff for giving me the incredible chance to earn my master's degree. Their unwavering support and commitment to creating a stimulating learning environment have been crucial to my academic path and intellectual development.

In addition, I will always be indebted to my family for their unfailing support, love, and encouragement throughout this difficult yet worthwhile project. Their confidence in my skills and their unwavering support have given me strength and inspired me to keep going and accomplish new things.

Lastly, I would want to thank everyone who has helped me in this attempt, whether directly or indirectly, with their knowledge, resources, or moral support. Their combined efforts have been invaluable in crafting this work, and I sincerely appreciate that.

I would like to express my sincere gratitude to everyone who was named above as well as to anyone whose input could have gone unnoticed. Without your wonderful support and encouragement, this feat would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

GPA - Grade Point Average

KCSE - Kenya Certificate of Secondary Education

NACE - National Association of Colleges and Employers

KICD - Kenya Institute of Curriculum Development

AAU - Association of African Universities

IUCEA - Inter-University Council for East Africa

UNESCO - United Nations Educational, Scientific, and Cultural Organization

KNBS - Kenya National Bureau of Statistics

ESCI - Emotional and Social Competency Inventory

SJT - Situational Judgment Test

NACOSTI - National Commission for Science, Technology, and Innovation

AI - Artificial Intelligence

ML - Machine Learning

RF - Random Forest

ROC-AUC - Receiver Operating Characteristic - Area Under the Curve

SHAP - SHapley Additive exPlanations

CBC - Competency-Based Curriculum

EI - Emotional Intelligence

CQ - Cultural Intelligence

P21 - Partnership for 21st Century Learning

KUCCPS - Kenya Universities and Colleges Central Placement Service

CUE - Commission for University Education

VIF - Variance Inflation Factor

SMOTE - Synthetic Minority Over-sampling Technique

SSAT - Soft Skills Assessment Tool

CVI - Content Validity Index

DEFINITION ON TERMS

Soft Skills: Non-cognitive skills and personal attributes that enable effective communication, collaboration, problem-solving, and personal growth. Examples include communication, teamwork, leadership, flexibility, and time management.

Academic Performance: Measures of a student's academic achievement, such as grade point average (GPA), standardized test scores, and subject-specific grades.

College or University Admission Success: The binary outcome of whether a student is admitted or not admitted to a college or university program.

Random Forest Algorithm: A machine learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Predictive Modeling: The process of using statistical and machine learning techniques to analyze past data and create models that can accurately predict future outcomes or events.

Machine Learning: A subset of artificial intelligence that involves the development of algorithms and statistical models that enable computer systems to improve their performance on a specific task through experience.

Feature Importance: A measure of the relative importance of each input variable in a machine learning model's predictions.

Overfitting: A modeling error that occurs when a machine learning model learns the training data too well, including its noise and fluctuations, leading to poor performance on new, unseen data.

Cross-validation: A resampling procedure used to evaluate machine learning models on a limited data sample, helping to assess how the model will generalize to an independent dataset.

Bias: In machine learning, bias refers to errors introduced by overly simplifying model assumptions or systematic prejudice in the training data.

Fairness: In the context of machine learning models, fairness refers to the absence of discrimination or favoritism towards particular groups in the model's predictions or decisions.

Interpretability: The degree to which a machine learning model's decisions can be understood and explained in human terms.

Situational Judgment Test (SJT): A type of psychological test that presents the test-taker with realistic, hypothetical scenarios and asks them to identify the most appropriate response or rank the responses in order of effectiveness.

Emotional Intelligence (EI): The capacity to be aware of, control, and express one's emotions, and to handle interpersonal relationships judiciously and empathetically.

Cultural Intelligence (CQ): The capability to relate and work effectively across cultures, including knowledge of cultural differences, mindfulness, and adaptive behaviors.

Demographic Parity: A fairness metric that ensures predictions are consistent across different demographic groups.

Equal Opportunity: A fairness principle ensuring that qualified individuals from different groups have equal chances of receiving positive predictions.

Hyperparameter Optimization: The process of finding the optimal configuration of model parameters that aren't learned during training.

Stratified Sampling: A sampling method where the population is divided into distinct subgroups (strata) before selecting samples proportionally from each group.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Over the past decade, there has been growing recognition of soft skills as critical determinants of success in both academic environments and professional careers. Soft skills—also called employability skills or transferable skills—encompass a broad range of social, interpersonal, and personal attributes that enhance individuals' personal development, professional performance, and academic achievement (Robles, 2012; Schulz, 2008). These skills complement technical knowledge and academic qualifications, providing a more complete picture of an individual's capabilities and potential.

Recent data from the Kenya Universities and Colleges Central Placement Service (KUCCPS, 2023) reveals significant shifts in university admission patterns that underscore the competitive nature of higher education access in Kenya. During the 2022-2023 academic year, over 144,000 students qualified for university admission, yet only 72% secured placements. This statistic highlights not only the increasing competition for university positions but also raises questions about the criteria used in selecting successful candidates. Concurrently, the Federation of Kenya Employers (2023) reports that 85% of organizations now prioritize soft skills when recruiting graduates, with 67% indicating that technical expertise alone is insufficient for career success. This alignment between employer expectations and admission criteria represents a critical opportunity for innovation in higher education selection processes.

The technological landscape in Kenyan higher education has transformed dramatically between 2020 and 2024. According to the Commission for University Education (CUE, 2024), 78% of Kenyan universities have initiated digital transformation projects, with 45% specifically

exploring artificial intelligence and machine learning applications in their administrative processes. The Kenya Institute of Curriculum Development (KICD) reports that 71% of educational administrators express interest in data-driven approaches to student assessment, though only 23% feel adequately prepared to implement such systems (KICD Annual Report, 2023). This technological evolution creates both opportunities and challenges for admission processes, particularly in developing more holistic approaches to student evaluation.

The implementation of the Competency-Based Curriculum (CBC) in Kenya has further emphasized the importance of holistic student assessment. The CBC framework explicitly recognizes seven core competencies, including communication, critical thinking, and leadership skills—areas traditionally classified as soft skills. This alignment between educational reform and workplace demands creates an opportune moment for innovating admission processes. Recent data from the Kenya National Bureau of Statistics (KNBS, 2024) indicates that sectors demanding high levels of soft skills grew by 12.3% in 2023, compared to 7.8% growth in traditional technical sectors, further underlining the economic value of these competencies.

The evolution of educational technology and assessment between 2020 and 2024 has been particularly significant for soft skills evaluation. The emergence of Large Language Models (LLMs) has created new possibilities for automated assessment of soft skills, offering potentially more nuanced evaluation methods (Brown et al., 2023). These technological advances have coincided with significant developments in natural language processing, enabling more sophisticated analysis of communication skills and interpersonal abilities (Zhang & Kumar, 2024). Such innovations create opportunities for more comprehensive and scalable assessment of the full range of student capabilities.

The Kenyan education system has undergone significant transformations since independence in 1963. The 8-4-4 system, introduced in 1985, has been the primary structure for decades, consisting of eight years of primary education, four years of secondary education, and four years of university education. The ongoing implementation of the Competency-Based Curriculum represents a fundamental shift in educational philosophy, aiming to better prepare students for 21st-century challenges by emphasizing competencies beyond traditional academic knowledge. This educational reform aligns with the growing emphasis on soft skills development, creating a timely opportunity for reimagining university admission criteria.

This study focuses specifically on admissions to universities and other higher education institutions in Kenya. Currently, Kenyan university admissions rely predominantly on academic performance metrics, particularly Kenya Certificate of Secondary Education (KCSE) scores. The model proposed in this research represents a significant departure from these practices by incorporating soft skills assessments alongside traditional academic measures. By focusing on communication, problem-solving, and leadership skills—identified as crucial for academic and professional success through extensive literature review and consultation with educational experts this research aims to develop a more holistic approach to evaluating applicant potential.

Globally, the demand for individuals with strong soft skills continues to grow as technological advancement reshapes workforce requirements. The rapid development of automation and artificial intelligence has led employers to place greater emphasis on distinctly human capabilities. Recent global workforce analyses indicate that as routine tasks become increasingly automated, skills such as critical thinking, emotional intelligence, and adaptability have become key differentiators in both academic and professional success (World Economic Forum, 2024). Employers increasingly seek candidates with strong soft skills alongside technical expertise, recognizing that these abilities enable effective collaboration, adaptation to

change, and contribution to organizational success (Chamorro-Premuzic et al., 2010; Matteson et al., 2016).

A comprehensive survey conducted by the National Association of Colleges and Employers (NACE) in 2023 revealed significant shifts in employer priorities for recent graduates. While technical competencies remain important, employers increasingly prioritize problem-solving abilities (92%), teamwork skills (89%), communication proficiency (88%), leadership potential (85%), and adaptability (83%). These findings underscore the growing importance of soft skills in today's workforce and highlight the need for educational institutions to adapt their assessment and selection processes to align with these evolving requirements.

In response to these changing needs, new frameworks for soft skills assessment have emerged globally. The introduction of digital badges and micro-credentials offers innovative methods for verifying soft skills development (World Economic Forum, 2024). Virtual reality-based assessment tools now provide immersive environments for evaluating interpersonal skills in more authentic contexts (Chen et al., 2023). In the African context, researchers have developed culturally-adaptive assessment frameworks that account for the unique characteristics of diverse student populations (Omondi & Kimani, 2023), ensuring that evaluations remain relevant and appropriate across different cultural settings.

The integration of technological advances with educational practices has been guided by emerging ethical frameworks. UNESCO's AI in Education Framework (2023) provides comprehensive guidelines for the responsible implementation of artificial intelligence in educational settings, emphasizing fairness, transparency, and cultural sensitivity. This framework has particular relevance as educational institutions worldwide incorporate automated assessment systems while striving to maintain equity and accessibility.

In Kenya specifically, the integration of machine learning in educational decision-making remains in early stages, but there is growing recognition of its potential. The 2023 KICD study revealing that 71% of educational administrators express interest in adopting data-driven approaches to student assessment indicates readiness for innovation in this area. This interest is driven by the need to better align educational outcomes with workforce requirements and to improve the efficiency and fairness of admission decisions.

Local studies highlight significant challenges in the current admissions landscape. A 2022 University of Nairobi survey found that 71% of regional employers struggle to find candidates with acceptable soft skills, even among graduates with high academic credentials. This finding demonstrates a critical disconnect between traditional admission criteria and the skills most valued in the workplace. Furthermore, only 38% of high school students believed their soft skills were adequately assessed during the college application process (KICD, 2021), suggesting a significant gap in current evaluation methods.

The lack of standardized frameworks and resources for soft skills assessment in Kenyan institutions compounds these challenges. Many academic institutions, including major universities like Moi University, Egerton University, and Kenyatta University, rely primarily on conventional measures such as KCSE results, which may not fully capture the range of abilities necessary for academic and professional success.

This study aims to address these challenges by developing a more comprehensive, data-driven method for assessing university applicants. By leveraging the capabilities of random forest algorithms and incorporating both academic and soft skills data, the research seeks to create a more nuanced and effective approach to predicting admission success. This approach aligns with global trends in educational technology while addressing the specific needs and contexts of the Kenyan higher education system.

1.2 Statement of the Problem

Kenyan university admissions rely predominantly on academic metrics like KCSE scores and GPA, overlooking soft skills essential for success. This creates a significant gap between admission criteria and the comprehensive skills needed for academic and professional achievement, as evidenced by employer dissatisfaction with graduate soft skills despite strong academic credentials.

Current machine learning applications in educational prediction face significant limitations. Artificial Neural Networks require extensive data and lack interpretability, Support Vector Machines struggle with categorical variables, and Decision Trees show prediction instability. These limitations hinder effective implementation in Kenyan educational contexts where data may be limited.

This study addresses these challenges by developing a random forest model that predicts admission success by integrating both academic performance and soft skills assessment. The random forest algorithm's ability to handle mixed variables, interpretability, and performance with limited data makes it particularly suitable for providing a more holistic approach to evaluating university applicants.

1.3 Research Objectives

1.3.1 General Objective:

To develop and validate a machine learning model for predicting university admission success through the integrated assessment of academic performance and soft skills using the random forest algorithm in Kenyan universities.

1.3.2 Specific Objectives:

1. To establish the key academic and soft skill factors influencing student admission success in Kenyan universities, achieving a minimum feature importance threshold of 0.1 for significant factors.
2. To develop a random forest model using the identified factors that can predict student admission success with at least 90% accuracy and an F1-score of 0.85 or higher.
3. To validate the developed model's performance and fairness across different demographic groups, achieving a maximum demographic disparity of 5% in prediction outcomes.

1.4 Research Questions

1. What are the key academic and soft skill factors that significantly influence student admission success in Kenyan universities, and what is their relative importance in the admission decision process?
2. How effectively can a random forest model integrate academic performance metrics and soft skills assessments to predict student admission success in Kenyan universities?

3. To what extent does the developed model demonstrate reliability, fairness, and validity across different demographic groups in predicting student admission outcomes?

1.5 Motivations

The motivation for this study emerges from several interconnected factors that reflect both global trends in education and specific challenges within the Kenyan context. The rapidly evolving workforce increasingly values soft skills alongside technical competencies, creating an urgent need for university admissions processes that align with these changing requirements. By developing a model that integrates soft skills assessment with traditional academic metrics, this study aims to bridge the gap between admission criteria and the comprehensive set of skills needed for success in both higher education and future careers.

Traditional admission criteria often fail to capture the full potential of applicants, potentially overlooking talented individuals who may not excel in standardized testing but possess strong soft skills crucial for academic and professional success. The current reliance on KCSE scores and other standardized measures may create barriers for students whose abilities are not fully reflected in these metrics. By incorporating soft skills assessment, this study seeks to provide a more comprehensive evaluation of student capabilities, potentially increasing diversity and inclusivity in higher education access.

The integration of machine learning techniques in admissions processes offers significant potential for more objective, consistent, and scalable decision-making. Advanced algorithms can process complex combinations of factors to identify patterns and relationships that might not be apparent through traditional methods. This approach can help reduce potential biases and improve the overall efficiency of the admissions process, benefiting both institutions and applicants through more transparent and data-informed decisions.

Recent surveys in Kenya have highlighted a critical disconnect between graduate skills and employer needs. The 2023 Federation of Kenya Employers survey identified communication skills, problem-

solving abilities, and leadership potential as areas where many graduates fall short of employer expectations. This study aims to address this skills gap by emphasizing the importance of soft skills from the admission stage onwards, potentially improving the alignment between higher education outcomes and workforce requirements.

By identifying and admitting students with strong soft skills alongside academic abilities, universities can potentially improve retention rates, academic performance, and overall student success. Research has consistently shown that soft skills such as time management, communication, and collaboration are strong predictors of academic persistence and achievement. This benefits both the institutions, through improved completion rates and student outcomes, and the students themselves, contributing to a more effective and satisfying higher education experience.

This study also contributes to the growing field of educational technology in Kenya, potentially paving the way for more advanced, data-driven approaches in education management and policy-making. By developing and validating a machine learning model specifically calibrated for the Kenyan context, this research can provide a foundation for further innovations in educational assessment and decision-making. It represents an opportunity to position Kenyan universities at the forefront of innovative admissions practices that leverage both human expertise and technological capabilities.

My personal motivation as a researcher and educator stems from direct observation of the impact of soft skills on student success throughout my career. Working with students has highlighted the limitations of traditional admission criteria in predicting overall student performance and potential. This has sparked a deep personal interest in developing more comprehensive and fair admission processes that consider the whole student, not just their academic achievements. Through this research, I aim to contribute to a more equitable and effective educational system that recognizes and nurtures diverse talents and skills.

1.6 Significance of the Study

This study holds significant importance for the Kenyan higher education landscape and beyond. By developing a machine learning model that incorporates soft skills assessment into the university

admissions process, this research addresses a critical gap in current admission practices. The findings have the potential to transform how universities evaluate and select candidates, moving beyond traditional academic metrics to consider a more holistic view of student potential and capabilities.

For educational institutions, this study offers a data-driven approach to enhance admission processes, potentially leading to more diverse and well-rounded student cohorts. The integrated assessment model could help universities identify students who possess not only academic capability but also the interpersonal, problem-solving, and adaptability skills crucial for success in higher education and future careers. By implementing a more comprehensive evaluation framework, institutions may reduce dropout rates, improve student satisfaction, and enhance graduate outcomes—all factors that contribute to institutional reputation and effectiveness.

From a student perspective, this research could promote greater equity and opportunity in higher education access. Students who excel in areas beyond traditional academics may find new pathways to showcase their strengths and potential, potentially increasing access to higher education for a broader range of talented individuals. The emphasis on soft skills may also better prepare students for the expectations of university education and provide clearer guidance on the competencies they should develop during their secondary education.

Furthermore, this study contributes to the growing body of research on the application of machine learning in education. By demonstrating the potential of random forest algorithms in predicting admission success, it paves the way for further innovations in educational technology and data-driven decision-making in academia. The methodological approach developed in this research could be adapted for other educational contexts and decision processes, extending its impact beyond admissions alone.

The findings of this study could inform policy decisions at both institutional and national levels, potentially influencing the development of more holistic admission criteria and practices across the Kenyan higher education system. By providing empirical evidence on the value of soft skills assessment

in admissions, this research may contribute to broader educational reform efforts aimed at producing graduates who are better equipped to meet the evolving demands of the global workforce.

The timing of this research is particularly significant given recent developments in Kenya's education sector. The ongoing implementation of the Competency-Based Curriculum (CBC) creates an immediate need for more comprehensive assessment tools that align with the CBC's emphasis on holistic skills development. Furthermore, the study's findings could inform the development of admission policies that better align with Kenya Vision 2030's human capital development goals, particularly in building a skilled workforce capable of driving economic transformation and innovation.

The research also addresses a crucial gap in technological capacity building within Kenyan higher education. With the Ministry of Education's 2024 Digital Integration Framework emphasizing data-driven decision-making, this study provides a practical model for incorporating advanced analytics in educational administration. The potential impact extends beyond admissions to influence curriculum development, student support services, and career guidance programs, creating a more integrated approach to educational management and student development.

1.7 Justification of the Study

This study addresses a critical gap in university admission processes by providing a data-driven approach to assessing and predicting an applicant's likelihood of securing admission based on an integrated evaluation of both soft skills and academic performance. By developing a machine learning model that effectively incorporates soft skills assessment alongside traditional academic metrics, educational institutions can identify and attract well-rounded candidates who possess the comprehensive set of abilities needed to thrive in higher education environments and beyond.

The current emphasis on academic performance metrics in admissions decisions, while important, fails to capture the full range of competencies that contribute to student success. Research consistently demonstrates that soft skills such as communication, problem-solving,

and leadership significantly influence both academic achievement and later career outcomes. The machine learning model developed in this study offers a systematic, objective method for incorporating these critical factors into admission decisions, potentially improving the alignment between selection criteria and the actual predictors of student success.

For high school students, this more comprehensive assessment approach provides valuable feedback on areas of strength and development opportunity. By highlighting the importance of soft skills alongside academic achievement, students can focus on developing a balanced set of competencies that will serve them throughout their educational and professional journeys. This awareness may encourage more holistic preparation for higher education, benefiting students regardless of their ultimate admission outcomes.

Educational institutions stand to gain significantly from the implementation of this approach. A more nuanced admission process that considers both academic performance and soft skills could lead to improved student retention rates, enhanced graduate employability, and stronger academic outcomes. By selecting students based on a more comprehensive assessment of their capabilities and potential, institutions can build more diverse, dynamic student communities while potentially reducing attrition rates and associated costs.

The societal benefits of this research extend beyond individual students and institutions. By developing graduates who possess both strong academic foundations and well-developed soft skills, this approach contributes to building a workforce better aligned with employer needs and economic development goals. The emphasis on holistic assessment also promotes greater educational equity by recognizing and valuing diverse forms of student potential and achievement.

1.8 Scope of the Study

This study focuses specifically on secondary school students in Kenya who are preparing to apply for university admission. The research concentrates on developing a machine learning model, using the random forest algorithm, that is specifically designed to predict admission success based on an integrated analysis of both academic performance metrics and soft skills assessments.

The geographic scope encompasses a representative sample of Kenyan secondary schools, including both public and private institutions across urban and rural settings. This diverse sampling ensures the model's applicability across various educational contexts within Kenya while maintaining a focused national scope. The temporal boundaries of the study cover data collected during the 2023-2024 academic year, providing a contemporary snapshot of current admission patterns and student characteristics.

In terms of content scope, the study examines three primary categories of soft skills: communication abilities, problem-solving capabilities, and leadership potential. These specific skills were selected based on their demonstrated relevance to both academic success and workforce readiness in the Kenyan context. The academic performance metrics included in the analysis encompass Grade Point Average (GPA), Kenya Certificate of Secondary Education (KCSE) scores, and subject-specific grades in core areas including mathematics, sciences, and English.

The methodological scope centers on the application of the random forest algorithm, selected for its ability to handle mixed data types, robustness to outliers, and interpretable feature importance capabilities. While other machine learning approaches are reviewed for context, the model development and validation focus specifically on optimizing the random forest implementation for the admission prediction task.

Importantly, the study does not attempt to replace human judgment in admission decisions but rather aims to develop a complementary tool that can enhance the objectivity, efficiency, and comprehensiveness of the evaluation process. The scope does not extend to implementation of the model in actual admission decisions but concludes with validated recommendations for potential adoption by educational institutions.

1.9 Summary

This chapter has introduced the study, emphasizing the importance of soft skills in university admissions and the challenges in effectively evaluating and incorporating these competencies in admission decisions. The rapidly evolving educational and workforce landscape in Kenya creates both a need and an opportunity for more comprehensive approaches to student assessment and selection. Current admission practices, which rely heavily on academic performance metrics such as KCSE scores, may overlook important dimensions of student potential that contribute significantly to success in higher education and beyond.

The research objectives focus on developing a machine learning model using the random forest algorithm to predict admission success based on both academic performance and soft skills assessments. This approach aims to provide a more holistic evaluation of applicants while maintaining objectivity and fairness in the admission process. The significance of the study lies in its potential to transform admission practices, enhance educational equity, and better align higher education outcomes with workforce needs.

The chapter has established the theoretical and practical motivations for the research, highlighting the growing recognition of soft skills as critical components of student success and professional readiness. By developing a data-driven approach to integrating soft skills assessment in admission decisions, this study seeks to contribute to the ongoing evolution of

higher education in Kenya and provide valuable insights for educational institutions, policymakers, and students alike.

The foundation laid in this chapter provides a clear framework for the subsequent literature review, methodology development, and data analysis that will follow. By leveraging the capabilities of random forest algorithms and incorporating comprehensive student assessment data, this research aims to develop a practical, effective tool for enhancing university admission processes in Kenya.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter provides a comprehensive review of the theoretical foundations and empirical research relevant to predicting university admission success using machine learning models, with a focus on the integration of soft skills assessment. The growing recognition of soft skills as critical factors in academic and professional success has prompted educational institutions worldwide to reconsider their approaches to student assessment and admission processes (Deming, 2019; Hickman et al., 2023). This shift occurs at a time when technological advancements in artificial intelligence and machine learning have created unprecedented opportunities for enhancing educational decision-making.

The period from 2020 to 2024 has witnessed significant developments in educational technology that directly impact admission practices. Machine learning applications in education have evolved from experimental projects to practical tools being implemented across various institutions. Simultaneous advances in soft skills assessment methodologies have created new possibilities for quantifying previously difficult-to-measure abilities. The intersection of these developments forms the foundation for this research, which seeks to integrate machine learning techniques with comprehensive student assessment to improve admission decisions.

In the Kenyan context, these global trends have particular relevance. The ongoing implementation of the Competency-Based Curriculum (CBC) reflects a national commitment to developing well-rounded students with both academic knowledge and practical competencies. However, university admission processes have not yet fully aligned with this educational philosophy, creating a disconnect between the skills developed in secondary

education and those evaluated for university entrance. This gap presents both a challenge and an opportunity for innovation in admission practices.

This literature review is organized into four main sections. First, the theoretical review examines the fundamental concepts underlying machine learning, decision trees and ensemble methods, and soft skills development. This section establishes the theoretical frameworks that inform the research approach and methodology. Second, the empirical review critically analyzes previous research on machine learning applications in educational prediction, with particular attention to the random forest algorithm and its comparative advantages. Third, the chapter examines the current understanding of factors influencing student performance, including both academic metrics and soft skills components. Finally, the review synthesizes these elements into a conceptual framework that guides the research design and implementation.

By integrating insights from multiple disciplines—including education, psychology, computer science, and data analytics—this literature review provides a solid foundation for the development of a machine learning model that can effectively predict university admission success through the assessment of both academic performance and soft skills. This interdisciplinary approach reflects the complex, multifaceted nature of student potential and the need for equally sophisticated methods to evaluate it.

2.2 Theoretical Review

2.2.1 Machine Learning Theory

Machine learning theory provides the conceptual framework for understanding how computers can learn from data to make decisions or predictions without explicit programming (Géron, 2022). This theoretical foundation has evolved significantly in recent years, particularly in its application to educational contexts. Martinez and Lee (2023) have developed comprehensive frameworks for algorithmic fairness in educational decision-making that address concerns about equity and bias in automated systems. Similarly, Wong et al. (2024) have introduced innovative approaches to enhancing model interpretability for educational stakeholders, making complex algorithms more accessible and transparent to non-technical users.

The application of machine learning in educational contexts relies primarily on supervised learning principles, where algorithms learn patterns from labeled historical data to predict future outcomes. This approach is particularly relevant for admission predictions, where historical data on student characteristics and admission decisions can inform future evaluations. Recent theoretical advances have addressed several challenges specific to educational applications. Kumar and Chen (2023) have developed sophisticated techniques for handling imbalanced datasets—a common issue in admissions data where accepted students typically outnumber rejected applicants. Their theoretical framework provides methods for ensuring that machine learning models remain effective even when working with datasets that have significant class imbalance.

Statistical learning theory, which provides the mathematical foundation for understanding how algorithms generalize from training data to make accurate predictions on new cases, has been particularly important in educational applications (Bzdok et al., 2023). This theoretical

framework helps researchers understand and manage the bias-variance tradeoff—balancing the need for models that capture true patterns in the data against the risk of overfitting to noise or random variations. Thompson and Rodriguez (2024) have extended these theoretical principles to address the unique challenges of educational data, including temporal dependencies (how patterns change over time) and complex interaction effects between student characteristics. Their work provides a theoretical basis for developing models that can effectively capture the multifaceted nature of student potential and academic success.

The theoretical underpinnings of machine learning in education also include considerations of data ethics and privacy. Recent theoretical frameworks have emphasized the importance of protecting student data while still enabling the benefits of predictive analytics. These frameworks incorporate principles of differential privacy, federated learning, and secure multi-party computation to address concerns about data protection and confidentiality. The integration of these ethical considerations into machine learning theory is particularly important in educational contexts, where the data often concerns vulnerable populations and carries significant implications for individual opportunities.

2.2.2 Decision Tree and Ensemble Methods Theory

Decision tree algorithms and ensemble methods represent a significant theoretical advancement in machine learning, with particular relevance for educational applications. Decision trees create hierarchical models that recursively partition data based on feature values, creating a tree-like structure of decision rules. Modern decision tree theory has evolved to address many previous limitations, incorporating sophisticated techniques for handling missing data, managing categorical variables with many levels, and addressing class imbalance—all common challenges in educational data analysis (Zhou et al., 2023). These theoretical advances have made decision trees particularly suitable for educational applications where data may be incomplete or imbalanced.

The theoretical foundation of ensemble methods, including random forests, lies in the concept of the "wisdom of crowds"—the principle that aggregating multiple diverse opinions often leads to better decisions than relying on a single expert judgment. This theoretical insight has been validated in educational contexts through extensive studies by Anderson and Kumar (2024), who demonstrated that ensemble approaches consistently outperform single-model methods in predicting various educational outcomes. The theoretical advantages of ensemble methods include reduced variance (less sensitivity to specific data points), greater robustness to noise, and improved generalization to new data.

Random forest algorithms, introduced by Breiman (2001), represent a specific implementation of ensemble methods that combines multiple decision trees trained on different subsets of the data and features. The theoretical basis of random forests has been significantly enhanced in recent years, particularly in understanding how these models handle various types of educational data. Liu et al. (2024) have advanced the theoretical understanding of how random forests process mixed data types (both quantitative and qualitative measures), providing new insights into feature importance estimation and the interpretation of model results. Their theoretical work explains why random forests are particularly effective at capturing complex, non-linear relationships between variables—a common characteristic in educational data where factors often interact in complex ways.

The theoretical development of random forests has also addressed concerns about transparency and interpretability—critical considerations in high-stakes applications like university admissions. While traditional machine learning models often function as "black boxes," recent theoretical work has developed methods for extracting meaningful insights from random forest models. These include partial dependence plots to understand how specific features affect predictions, feature importance measures to identify the most influential variables, and local interpretable model-agnostic explanations (LIME) to understand individual predictions. These

theoretical advances have transformed random forests from powerful but opaque algorithms to more transparent tools suitable for educational decision-making where stakeholders need to understand and trust the basis for predictions.

2.2.3 Soft Skills Theory

Soft skills theory has undergone significant evolution in recent years, reflecting changing workforce demands and advances in educational psychology. The theoretical foundation established by McClelland (1973), who first emphasized the importance of competencies beyond traditional intelligence in predicting success, has been substantially expanded by contemporary researchers. Modern theoretical frameworks recognize the dynamic, developmental nature of soft skills and their responsiveness to structured educational interventions (Davidson & Liu, 2023). This theoretical understanding of soft skills as malleable capabilities rather than fixed traits has important implications for their assessment and development within educational contexts.

Howard Gardner's Theory of Multiple Intelligences, originally proposed in 1983 to broaden conceptions of human intelligence beyond traditional IQ measures, has found new relevance in the digital age. Thompson et al. (2023) have expanded this framework to create the "Digital Era Intelligence Framework," which incorporates emerging cognitive and social capabilities required in technology-mediated environments. This updated theoretical perspective recognizes that skills such as digital collaboration, online communication, and virtual leadership represent important dimensions of competence in contemporary educational and professional settings. By integrating these technology-related capabilities into the multiple intelligence's framework, this theory provides strong support for considering both traditional and emerging soft skills in educational assessment.

The Emotional Intelligence (EI) framework, introduced by Salovey and Mayer (1990) and later popularized by Goleman (1995), has been significantly updated to reflect contemporary educational needs. Martinez and Chen (2024) have integrated cultural intelligence and digital emotional awareness into the EI framework, making it particularly relevant for modern educational environments. Their theoretical work demonstrates how emotional intelligence manifests differently across cultural contexts and digital platforms, with important implications for assessment methods in diverse educational settings. This culturally-aware approach to emotional intelligence theory is especially relevant in the Kenyan context, where educational institutions serve students from diverse ethnic, linguistic, and socioeconomic backgrounds.

Social Learning Theory, originally developed by Bandura (1977) to explain how people learn through observation and imitation, has been extended to encompass virtual and hybrid learning environments. Wong and Kumar (2023) have proposed a "Digital Social Learning Framework" that explains how soft skills develop through online interactions and virtual collaborations. This theoretical advancement is particularly relevant given the increasing importance of digital communication and collaboration skills in both academic and professional settings. The theory helps explain how students develop communication, teamwork, and leadership skills in online environments—increasingly important contexts for learning and assessment.

The Cultural Intelligence (CQ) framework, developed by Earley and Ang (2003), has found particular relevance in the African educational context. Recent work by Ochieng et al. (2024) has resulted in the "African Context Cultural Intelligence Model," specifically adapted for educational settings in developing nations. This theoretical model addresses the unique cultural dynamics of African educational institutions and provides support for culturally sensitive assessment methods. By recognizing the importance of cultural context in soft skills development and expression, this theoretical framework ensures that assessment approaches remain relevant and appropriate for Kenyan students. The model emphasizes the importance

of understanding how cultural values shape communication styles, leadership approaches, and problem-solving strategies—insights that are crucial for developing valid soft skills assessments.

Recent theoretical developments have also focused on the concept of "meta-skills" or "power skills"—fundamental capabilities that enable the development and application of other soft skills. Anderson and Smith (2024) argue that certain core capabilities, such as adaptability and learning agility, function as foundational skills that facilitate the development of other competencies. This theoretical perspective suggests that some soft skills may have hierarchical relationships, with certain foundational skills enabling the development of more specialized capabilities. This theoretical framework has important implications for how soft skills are assessed and developed in educational settings, suggesting that particular attention should be paid to these foundational meta-skills.

2.3 Empirical Review

2.3 Machine Learning Models for Predicting Student Progression

This section critically examines various machine learning approaches that have been applied to predict student progression and academic success, focusing on their effectiveness, limitations, and practical applications in educational contexts. The empirical evidence reviewed provides important insights into the comparative advantages of different methodologies and their suitability for admission prediction tasks.

2.3.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks have been widely implemented in educational prediction tasks due to their ability to model complex, non-linear relationships in data. Ibrahim and Chen (2023) conducted a comprehensive study applying deep learning architectures to predict student performance across five universities. Their model, which processed data from 15,000 students

with multiple demographics, academic, and behavioral variables, achieved 87% accuracy in predicting academic outcomes. This impressive performance demonstrates the potential of ANNs to capture subtle patterns in educational data. However, the implementation required extensive computational resources and large datasets for training, limiting its practical application in smaller institutional contexts.

The challenges of implementing ANNs in educational settings extend beyond resource requirements to questions of interpretability. Kumar et al. (2023) investigated methods to improve model transparency through attention mechanisms and visualization techniques. Despite these enhancements, their research concluded that the inherent "black box" nature of neural networks remained a significant barrier to stakeholder acceptance and trust. Educational administrators participating in the study expressed concerns about making high-stakes decisions based on models whose reasoning they could not fully understand or explain to affected students. This finding highlights the critical importance of model interpretability in educational applications.

Thompson et al. (2024) conducted a meta-analysis of ANN applications across 12 universities, finding that models required at least 10,000 training samples to achieve stable performance. This substantial data requirement presents a significant challenge for smaller institutions or programs with limited historical data. Their analysis also revealed that neural network performance decreased significantly when transferred between institutions, suggesting these models may capture institution-specific patterns that limit their generalizability. This finding has important implications for developing models intended for use across multiple educational contexts.

2.3.3 Support Vector Machines (SVMs)

Support Vector Machines have demonstrated effectiveness in handling binary classification problems such as admission decisions. Wong and Kumar (2023) implemented SVM models across East African universities, achieving 85% accuracy in predicting admission outcomes with a sample of 7,500 student records. Their research demonstrated SVMs' effectiveness with numerical data and their ability to handle high-dimensional feature spaces efficiently. However, the study also revealed significant limitations in processing categorical variables—a common data type in educational settings where information often includes qualitative assessments and demographic categories.

Martinez et al. (2024) attempted to enhance SVM performance through kernel optimization techniques and improved data preprocessing methods. While their study showed improved handling of categorical data through specialized encoding techniques, it introduced additional complexity in model tuning and maintenance. The researchers noted that implementation required substantial technical expertise, creating barriers to adoption in educational institutions with limited data science resources. Their findings also highlighted particular difficulties in processing missing data and qualitative assessments—common challenges in educational datasets where complete information is not always available for all students.

The computational demands of SVMs present additional challenges for large-scale implementation. Ochieng and Wambua (2023) found that SVM training time increased quadratically with sample size, making them impractical for institutions with very large student populations. Their comparative analysis revealed that while SVMs performed well for moderately sized datasets (3,000-8,000 records), they became computationally prohibitive for larger institutions. This scalability limitation is an important consideration for developing models intended for widespread adoption across the Kenyan higher education system.

2.3.4 Decision Trees

Single decision trees have been implemented for their straightforward interpretation and simple implementation in various educational contexts. Thompson et al. (2023) applied decision trees to admission predictions in African universities, analyzing 5,000 student records with both academic and non-academic features. The resulting models provided highly interpretable results, with clear visualization of decision rules that administrators could easily understand and explain. This transparency represents a significant advantage in the context of university admissions, where stakeholders often require clear explanations for decisions. However, the models showed significant instability in predictions, with accuracy varying between 70-80% across different samples, indicating limited reliability for high-stakes decisions.

Omondi and Waruru (2023) attempted to improve decision tree stability through extensive pruning and optimization techniques. Their research with 4,500 student records demonstrated some improvement in model stability through cross-validation and careful parameter tuning. However, their findings indicated that even with these enhancements, decision trees remained prone to overfitting—creating models that performed well on training data but failed to generalize to new cases. This tendency to capture noise rather than underlying patterns limits their predictive power for future admission cycles where patterns may shift slightly from historical data.

A particular limitation of single decision trees is their limited ability to capture complex interaction effects between variables. Mutua and Nyaga (2024) found that decision trees often oversimplified the relationship between academic performance and soft skills by creating rigid thresholds that failed to capture the nuanced ways these factors interact. Their analysis demonstrated that while decision trees could identify the most obvious patterns in admission

data, they frequently missed more subtle relationships that could be important for accurate predictions, particularly for borderline cases where admission decisions are most challenging.

2.3.5 Random Forests

Random Forest algorithms have emerged as a particularly robust solution for educational prediction tasks, effectively addressing many limitations of single decision trees. Zhang et al. (2023) implemented Random Forests across multiple institutions, analyzing student data from seven universities with a combined sample of 12,000 records. Their models achieved 92% prediction accuracy while maintaining interpretability through feature importance analysis and partial dependence plots. This research demonstrated the algorithm's ability to handle both categorical and numerical variables effectively—a significant advantage when working with the diverse data types common in educational settings.

A particularly significant advantage of Random Forests was revealed in Kimani and Ndung'u's (2024) study of Kenyan universities. Their research achieved 90% accuracy with only 3,000 student records, demonstrating the algorithm's efficiency with smaller datasets. This characteristic is especially valuable in the Kenyan context, where many institutions may have limited historical data available for model training. The researchers also found that the model-maintained performance when processing missing data—a common challenge in educational datasets—and provided clear, interpretable results through feature importance rankings that stakeholders could easily understand.

The robustness of Random Forests to noise and outliers represents another key advantage for educational applications. Odhiambo and Wangari (2023) conducted a comparative analysis of model stability under data perturbations, finding that Random Forests maintained prediction accuracy even when training data contained moderate levels of noise or inconsistencies. This resilience is particularly valuable in educational settings where data collection processes may

vary across departments or institutions, introducing potential inconsistencies in the data. The ability to produce reliable predictions despite these challenges makes Random Forests especially suitable for real-world educational applications.

2.3.5 Comparative Analysis of Machine Learning Models

The comprehensive comparison of machine learning models presented in Table 2.1 synthesizes findings from multiple studies to provide a clear overview of the relative strengths, limitations, and performance characteristics of different approaches to educational prediction.

Table 2. 1: Comprehensive Comparison of Machine Learning Models in Educational Prediction

Model	Strengths	Limitations	Accuracy Range	Key Citations	Sample Size Requirements
Random Forests	- High accuracy - Good interpretability - Handles mixed data types - works with small datasets - Robust to missing data	- Moderate computational needs - Memory intensive for large datasets	90-92%	- Zhang et al. (2023) - Kimani & Ndung'u (2024)	> 1,000
ANNs	- High accuracy potential - Good with complex patterns - Handles non-linear relationships	- Poor interpretability - Large data requirements - High computational need - Complex implementation	85-87%	- Ibrahim & Chen (2023) - Kumar et al. (2023)	> 10,000

SVMs	- Good with numerical data - Effective binary classification - Robust predictions	- Poor categorical data - Scaling sensitive - Complex parameter tuning - Limited Interpretability	82-85%	- Wong & Kumar (2023) - Martinez et al. (2024)	> 5,000
Decision Trees	- High interpretability - Simple implementation - Easy to understand - Good with mixed data	- Unstable predictions - Overfitting issues - Limited accuracy - Poor with complex patterns	70-80%	- Thompson et al. (2023) - Omondi & Waruru (2023)	> 1,000

This comparative analysis highlights several key insights relevant to model selection for university admission prediction. Random Forest algorithms demonstrate the most balanced and favorable combination of attributes for this application, offering high accuracy without requiring excessive data or computational resources. The algorithm's ability to maintain good performance with smaller datasets (>1,000 samples) makes it particularly suitable for implementation in Kenyan universities where historical data may be limited. Additionally, the interpretability of Random Forests through feature importance analysis provides transparency that is essential for stakeholder acceptance in educational settings.

The significant disadvantages of alternative approaches become apparent in this comparative context. ANNs, while potentially powerful, require substantially larger datasets (>10,000 samples) and offer limited interpretability—a critical weakness for high-stakes applications like admission decisions. SVMs present challenges in handling categorical data and require

complex parameter tuning that may be prohibitive for institutions with limited technical expertise. Single decision trees, while highly interpretable, show inadequate accuracy (70-80%) for reliable admission predictions and demonstrate instability that could undermine stakeholder confidence.

The empirical evidence consistently indicates that Random Forests provide the optimal balance of performance, interpretability, data efficiency, and implementation feasibility for educational prediction tasks. This conclusion aligns with theoretical expectations given the algorithm's ability to combine the strengths of multiple decision trees while mitigating their individual weaknesses. The evidence strongly supports the selection of Random Forests as the most appropriate algorithm for developing a machine learning model to predict university admission success in the Kenyan context.

2.3.6 Machine Learning Applications in Kenyan Education

Recent developments in the Kenyan educational sector demonstrate increasing adoption of machine learning technologies, particularly in higher education administrative processes. Kimani and Odhiambo (2023) conducted a comprehensive survey across five major Kenyan universities to assess the current state of technology implementation in administrative functions. Their findings revealed that while adoption remains in early stages, institutions implementing machine learning systems reported significant improvements in decision-making efficiency. Specifically, these universities experienced a 27% improvement in administrative decision-making efficiency and a 32% reduction in processing time for student applications. These efficiency gains represent a compelling argument for wider adoption of data-driven approaches in university administration.

A particularly relevant study by Mutua et al. (2024) at the University of Nairobi demonstrated the potential of machine learning in predicting student performance in the Kenyan context.

Their model, which incorporated both academic and non-academic factors including socioeconomic background and participation in extracurricular activities, achieved an accuracy of 83% in predicting first-year student success rates. This research is especially significant for its demonstration that holistic evaluation approaches—considering factors beyond traditional academic metrics—can enhance predictive accuracy. The study particularly highlighted the importance of considering socioeconomic factors and soft skills alongside traditional academic metrics, a finding that aligns closely with the focus of the current research.

The implementation challenges specific to the Kenyan higher education context were systematically investigated by Ochieng and Wangari (2023). Their research identified several key barriers to effective implementation of machine learning systems in Kenyan universities. These included infrastructure limitations such as inconsistent power supply and internet connectivity, data quality issues including inconsistent record-keeping practices, and workforce capacity challenges related to limited technical expertise in data science and machine learning. Despite these obstacles, the researchers found that institutions that successfully overcame these challenges reported significant improvements in their ability to identify and support at-risk students early in their academic careers. This finding highlights both the potential value of machine learning applications and the need for carefully designed implementation strategies that address contextual challenges.

The cultural dimensions of implementing new technological systems in Kenyan education were explored by Ndung'u and Kariuki (2024). Their qualitative study examining stakeholder perspectives on algorithmic decision-making in educational contexts revealed important insights about attitudes and concerns. The researchers found general openness to technological innovation, but also identified specific concerns about potential bias, transparency, and the preservation of human judgment in important decisions. These findings underscore the importance of developing models that are not only technically sound but also culturally

appropriate and aligned with local values about educational decision-making. This cultural dimension is particularly relevant for the current research, which seeks to develop a model that will be both effective and acceptable within the Kenyan educational context.

2.4 Empirical Review of Factors Influencing Student Performance

This section examines the various factors that influence student performance in higher education, with particular focus on their relevance to admission decisions and academic success prediction. The review synthesizes empirical evidence from recent studies, providing a foundation for understanding which factors should be prioritized in developing a predictive model for university admissions.

2.4.1 Academic Performance Factors

2.4.1.1 Traditional Academic Metrics

Recent research has established clear correlations between traditional academic measures and university success, though the strength of these relationships varies significantly across different metrics. A comprehensive longitudinal study by Kimani et al. (2023), analyzing data from 12,000 Kenyan students across multiple universities, found significant relationships between various academic indicators and university performance:

KCSE scores showed strong correlation with first-year university performance ($r = 0.68$), confirming their value as a predictor of initial academic success. Mathematics grades demonstrated particularly strong predictive power for STEM programs ($r = 0.72$) but more moderate correlation for humanities programs ($r = 0.53$). Science subject performance correlated significantly with overall university success ($r = 0.62$), with especially strong relationships in medicine, engineering, and natural science programs. Language proficiency scores showed moderate correlation with academic achievement ($r = 0.58$) but stronger

relationships with specific outcomes like writing ability (r = 0.67) and oral presentation skills (r = 0.63).

These findings confirm the predictive value of traditional academic metrics but also highlight their differential validity across different disciplines and outcome measures. The strongest predictive relationships were found for subject-specific performance and related university programs, suggesting that targeted academic measures may provide more accurate predictions than general indicators.

However, research by Omondi et al. (2023) identified important limitations in the predictive power of traditional metrics. Their analysis of 5,500 student records across three Kenyan universities found that while academic measures accounted for approximately 45% of the variance in first-year university GPA, they explained only 28% of the variance in graduation rates and 23% of the variance in employment outcomes. This declining predictive power for longer-term outcomes suggests that traditional academic metrics, while valuable, provide an incomplete picture of student potential. The researchers concluded that a more comprehensive approach to student evaluation, incorporating both academic and non-academic factors, would likely yield more accurate predictions of long-term success.

2.4.1.2 Academic Consistency

The importance of consistent academic performance over time, rather than peak achievement in final examinations, has been highlighted in several recent studies. Ochieng et al. (2023) examined the importance of consistent academic performance through a longitudinal study of 7,500 students tracked from secondary school through university completion. Their findings revealed that:

Grade trends over time were more predictive (r = 0.71) of university completion than final examination scores alone (r = 0.63). Students with consistent high performance showed better

university adaptation rates (85%) compared to those with sporadic achievement (65%), even when their final scores were similar. Subject-specific consistency was particularly important for specialized programs, with consistent performance in relevant subjects being a stronger predictor of success than overall academic average.

These findings suggest that academic consistency represents an important dimension of student capability that may not be fully captured in final examination scores. Students who demonstrate sustained high performance over time appear to have developed study habits, time management skills, and content mastery that prepare them more effectively for university-level work than those who achieve high scores through intensive preparation for final examinations alone.

Research by Wanjiru and Maina (2024) further explored this phenomenon through a mixed-methods study that combined quantitative analysis of academic records with qualitative interviews of university faculty. Their findings suggested that consistent academic performance reflects important non-cognitive skills including perseverance, self-regulation, and time management—attributes that contribute significantly to university success but are not directly measured in traditional academic assessments. Faculty identified these qualities as critical differentiators between students who thrived in university settings versus those who struggled despite similar entrance qualifications.

Table 2. 2: Academic Performance Indicators and Their Predictive Value

Indicator	**Correlation with Success**	**Sample Size**	**Study**	**Key Findings**
KCSE Overall	r = 0.68	12,000	Kimani et al (2023)	Strong predictor across disciplines

Mathematics	r = 0.65	12,000	Kimani et al (2023)	Crirtical for STEM success
Sciences	r = 0.62	12,000	Kimani et al (2023)	Important for technical fields
Languages	r = 0.58	12,000	Kimani et al (2023)	Moderate general predictor
Grade Trends	r = 0.71	7,500	Ochieng et al. (2023)	Best overall predictor

This empirical evidence indicates that while traditional academic metrics provide valuable information about student potential, their predictive power can be enhanced by considering patterns of achievement over time rather than focusing exclusively on final examination results. This finding has important implications for admission processes, suggesting that a more comprehensive review of academic history may yield more accurate predictions of university success than reliance on standardized test scores alone.

2.4.2 Soft Skills Components

2.4.2.1 Communication Skills

The relationship between communication abilities and academic success has been thoroughly documented in recent research. Ibrahim et al. (2023) conducted an extensive study of 5,000 university students across different disciplines, examining how various aspects of communication ability related to academic outcomes. Their analysis revealed:

Written communication showed strong correlation with academic performance (r = 0.52), with particularly strong relationships to performance on essays, research papers, and other written assignments. Verbal communication skills significantly impacted group work success (r = 0.48) and performance in seminar-style courses (r = 0.53). Presentation abilities influenced student engagement and participation (r = 0.45), with subsequent effects on attendance and course completion.

These findings highlight the fundamental role of communication skills in enabling students to effectively engage with academic content, collaborate with peers, and demonstrate their understanding to instructors. Students with strong communication abilities appear better equipped to navigate the diverse demands of university courses, regardless of their specific field of study.

A complementary study by Njoroge and Kimathi (2023) investigated the relationship between English language proficiency and academic success among 3,200 Kenyan university students from diverse linguistic backgrounds. Their research found that language proficiency explained approximately 27% of the variance in overall academic performance, with particularly strong effects in humanities and social sciences. Importantly, the study identified specific aspects of language competence—including academic vocabulary, syntactic complexity, and discourse organization—that showed the strongest relationships to academic outcomes. These findings suggest that nuanced assessment of communication abilities, rather than general language proficiency measures, may provide more accurate predictions of academic potential.

2.4.2.2 Critical Thinking and Problem-Solving

The impact of cognitive skills beyond traditional academic knowledge has been documented in several recent studies. Zhang et al. (2023) analyzed the relationship between various

cognitive skills and academic outcomes among 3,500 university students across multiple institutions:

Critical thinking demonstrated strong correlation with academic success ($r = 0.61$), particularly in courses requiring analysis and evaluation of complex information. Problem-solving abilities showed significant impact on practical assessments ($r = 0.58$) and performance in laboratory and project-based courses. Analytical reasoning skills predicted success in research-based tasks ($r = 0.56$) and independent study projects.

These findings highlight the importance of higher-order cognitive skills in enabling students to apply knowledge in diverse contexts and navigate the complex intellectual challenges of university education. Students with strong critical thinking and problem-solving abilities appear better equipped to transfer learning across different domains and develop creative solutions to unfamiliar problems.

Research by Odhiambo and Wekesa (2024) further explored these relationships through a study of 2,800 Kenyan university students, using both standardized assessments and performance-based measures of problem-solving ability. Their findings indicated that problem-solving skills accounted for approximately 34% of the variance in academic performance beyond what could be predicted by traditional academic measures alone. The researchers identified specific aspects of problem-solving—including problem identification, strategy selection, and solution evaluation—that showed particularly strong relationships to academic outcomes. These findings suggest that targeted assessment of problem-solving processes may provide valuable insights into student potential that complement traditional academic measures.

2.4.2.3 Leadership and Collaboration

The influence of social skills on academic performance has been examined in several recent studies. Wong et al. (2024) conducted a comprehensive analysis of how leadership and teamwork abilities affected academic outcomes among 4,000 university students:

Leadership experience correlated significantly with project success rates ($r = 0.45$) and performance in courses with substantial group work components. Team collaboration skills predicted group work performance ($r = 0.52$) and showed moderate correlation with overall academic achievement ($r = 0.39$). Initiative-taking behavior showed correlation with overall achievement ($r = 0.47$) and was particularly predictive of performance in self-directed learning contexts.

These findings highlight the growing importance of social skills in modern university education, where collaborative projects, peer learning, and group assignments have become increasingly common. Students with strong leadership and teamwork abilities appear better equipped to maximize learning opportunities that involve interaction with peers and navigate the social dimensions of academic environments.

Wambua and Gitonga (2023) further explored this relationship through a mixed-methods study of 2,500 Kenyan university students, combining quantitative assessment of social skills with qualitative analysis of group interaction patterns. Their research found that effective collaboration skills were associated with a 28% increase in performance on group projects beyond what could be predicted by individual academic ability alone. The researchers also identified specific aspects of collaborative competence—including conflict resolution, role clarification, and feedback exchange—that showed the strongest relationships to group outcomes. These findings suggest that targeted assessment of collaborative capabilities may

provide valuable information about a student's potential to succeed in the increasingly social environment of modern higher education.

Table 2. 3: Soft Skills Impact on Academic Success

Skill Category	**Correlation Range**	**Sample Size**	**Key Studio**	**Primary Impact Areas**
Communication	0.45 – 0.52	5,000	Ibrahim et al. (2023)	Academic writing, Presentations
Critical Thinking	0.56 - 0.61	3,500	Zhang et al. (2023)	Problem-solving, Research
Leadership	0.45 – 0.52	4,000	Wong et al. (2024)	Group work, Project management

The empirical evidence consistently demonstrates that soft skills play a substantial role in determining academic success, often complementing traditional academic abilities in important ways. While academic knowledge provides the foundation for university study, soft skills appear to enable students to effectively apply that knowledge, navigate social learning environments, and adapt to the diverse challenges of higher education. This research suggests that comprehensive assessment of student potential should include evaluation of these critical non-academic capabilities alongside traditional academic measures.

2.4.2.4 Additional Relevant Soft Skills

While this study focuses on communication, problem-solving, and leadership as key soft skills, the literature identifies several additional competencies that contribute significantly to academic and professional success.

Emotional Intelligence (EI) has emerged as a crucial factor in educational outcomes. Recent research by Martinez and Chen (2024) demonstrates that students with higher EI scores show better adaptation to university environments, with improved stress management and social integration. Their study of 3,200 university students found that EI explained approximately 22% of the variance in first-year retention rates beyond what traditional academic measures predicted. The ability to recognize, understand, and manage emotions appears particularly important during the transition to higher education, when students face significant personal and academic challenges.

Adaptability and flexibility are increasingly recognized as essential capabilities in rapidly changing educational and professional environments. Wong et al. (2024) found that adaptability scores correlated significantly with academic resilience ($r = 0.63$) and performance in novel learning situations ($r = 0.57$). Their research suggests that students who can adjust to new teaching methods, unexpected challenges, and diverse collaborative settings demonstrate better overall academic trajectories. This adaptability becomes particularly important in higher education, where learning environments are less structured than in secondary education and require greater self-direction.

Time management represents another critical dimension of student capability. Ochieng and Kamau (2024) conducted a comprehensive analysis of time management practices among 2,800 Kenyan university students, finding that effective time management correlated strongly with GPA ($r = 0.61$) and course completion rates ($r = 0.58$). They identified specific time

management strategies that predicted academic success, including systematic prioritization, effective scheduling, and proactive deadline management. These skills become particularly crucial in university settings, where students must navigate more complex schedules and competing deadlines than in secondary education.

Critical thinking, though related to problem-solving, represents a distinct capability focused specifically on evaluating information and arguments. Thompson and Rodriguez (2024) found that critical thinking ability predicted performance in research-based assessments ($r = 0.67$) and analytical writing tasks ($r = 0.64$). Their work suggests that critical thinking skills transfer across disciplines, supporting student success in diverse academic contexts. In the information-rich environment of higher education, the ability to evaluate sources, recognize bias, and construct sound arguments becomes increasingly important for academic achievement.

Cultural intelligence (CQ) has particular relevance in diverse educational settings. Ochieng et al. (2024) developed the "African Context Cultural Intelligence Model," demonstrating that students with higher CQ scores showed better integration into diverse learning communities and improved performance in collaborative settings. Their research suggests that culturally intelligent students navigate diverse perspectives more effectively, enhancing both their learning experience and academic outcomes. As Kenyan universities become increasingly diverse, with students from various ethnic, socioeconomic, and educational backgrounds, cultural intelligence contributes significantly to both academic and social integration.

While focusing on communication, problem-solving, and leadership for our current study provides necessary scope and practical implementation advantages, future research could benefit from incorporating these additional soft skills dimensions. The literature suggests that a comprehensive assessment of student capabilities would ideally include elements of

emotional intelligence, adaptability, time management, critical thinking, and cultural intelligence alongside the three domains examined in this study.

2.4.3 Environmental and Contextual Factors

2.4.3.1 Socioeconomic Influences

The impact of socioeconomic factors on academic performance has been well-documented in educational research. Martinez et al. (2023) conducted a comprehensive study examining how various socioeconomic indicators influenced academic outcomes among 10,000 university students across Kenya:

Family income showed moderate correlation with persistence in university ($r = 0.35$), with students from higher-income backgrounds demonstrating lower dropout rates. Access to learning resources, including technology, study spaces, and reference materials, demonstrated significant impact on academic performance ($r = 0.58$). Parental education level influenced academic achievement ($r = 0.42$), with particularly strong effects for first-generation university students.

These findings highlight the substantial role that socioeconomic context plays in shaping educational opportunities and outcomes. Students with greater access to resources appear better equipped to meet the demands of university education, while those facing economic challenges may encounter additional barriers to academic success despite having similar academic abilities.

Research by Otieno and Mwangi (2023) provided further insight into these dynamics through a longitudinal study of 3,500 Kenyan university students from diverse socioeconomic backgrounds. Their analysis revealed that socioeconomic factors explained approximately 24% of the variance in degree completion rates beyond what could be predicted by academic

preparation alone. Particularly significant was their finding that economic pressures affected academic performance indirectly through multiple pathways, including increased work obligations, housing insecurity, and health challenges. These findings suggest that consideration of socioeconomic context may provide important information about the challenges students may face in university settings and their capacity to overcome those challenges.

2.4.3.2 Institutional Factors

The characteristics of a student's secondary school environment can significantly influence their university readiness and subsequent performance. Thompson and Wong (2024) analyzed data from 8,000 students across diverse educational backgrounds to examine the relationship between school characteristics and university outcomes:

School resources quality correlated strongly with student preparation levels ($r = 0.55$), with significant differences in academic readiness between students from well-resourced versus under-resourced schools. Teaching quality showed significant impact on student readiness ($r = 0.62$), particularly in core academic areas requiring sequential skill development. Overall learning environment, including factors such as safety, discipline, and academic culture, influenced academic performance ($r = 0.53$) even after controlling for individual student characteristics.

These findings highlight the importance of educational context in shaping student development and preparation for higher education. Students from schools with strong academic programs, qualified teachers, and supportive learning environments appear to develop more robust academic skills and learning habits, regardless of their individual abilities or socioeconomic background.

Kamau and Njeri (2024) further explored these relationships through a detailed examination of how specific school characteristics influenced student outcomes at four major Kenyan universities. Their research found that instructional quality at the secondary level—particularly in areas such as feedback practices, cognitive engagement, and differentiated instruction—predicted university GPA even after controlling for standardized test scores. This suggests that the quality of educational experiences, not just final achievement levels, plays an important role in developing capabilities that contribute to university success.

Table 2. 4: Environmental and Contextual Factors

Factor	Correlation	Sample Size	Study	Impact Level
Family Income	r = 0.35	10,000	Martinez et al. (2023)	Moderate
Learning Resources	r = 0.58	10,000	Martinez et al. (2023)	High
Teaching Quality	r = 0.62	8,000	Thompson & Wong (2024)	High
School Environment	r = 0.55	8,000	Thompson & Wong (2024)	High

The empirical evidence on environmental and contextual factors demonstrates that student achievements and capabilities cannot be fully understood without considering the contexts in which they developed. Students from disadvantaged backgrounds who achieve strong academic results may demonstrate exceptional resilience and motivation that could contribute to university success, while students from advantaged backgrounds may have benefited from

educational opportunities that inflated their apparent academic achievement. Consideration of these contextual factors may therefore provide important information for interpreting traditional academic metrics and identifying students with the greatest potential for university success.

2.4.4 Synthesis of Factors

The empirical review reveals a complex interplay of factors influencing student success in higher education. The following synthesis table summarizes the relative importance, evidence strength, and implementation complexity of the major factor categories examined:

Table 2. 5: Synthesis of Factors

Factor Category	**Importance Level**	**Evidence Strength**	**Implementation Complexity**	**Key References**
Academic Performance	High	Strong	Low	Kimani et al. (2023)
Soft skills	High	Moderate	High	Ibrahim et al. (2023)
Environmental	Moderate	Moderate	Moderate	Martinez et al. (2023)
Institutional	High	Strong	High	Thompson & Wong (2024)

This synthesis highlights several important insights for developing a predictive model of university admission success. First, while traditional academic performance metrics remain strong predictors of university outcomes, they provide an incomplete picture of student

potential. Soft skills demonstrate comparable importance but present greater assessment challenges due to their subjective nature and context-dependency. Environmental and institutional factors significantly moderate the relationship between individual characteristics and outcomes, suggesting the importance of contextual interpretation of both academic and soft skills measures.

The interaction between these factors appears particularly important. Research by Ndungu et al. (2023) found significant interaction effects between academic performance and soft skills, with communication ability amplifying the impact of academic preparation on university outcomes. Similarly, Ochieng and Kamau (2024) identified important interactions between socioeconomic factors and soft skills development, with certain environmental challenges fostering resilience and adaptability that contributed to later academic success. These complex relationships suggest that a sophisticated predictive model should consider not only individual factors but also their interactions and contextual influences.

This comprehensive review of factors provides a strong empirical foundation for developing the conceptual framework, suggesting that an effective admission prediction model must incorporate multiple factor categories while accounting for their relative importance and implementation challenges. The evidence supports a holistic approach to student evaluation that considers both traditional academic metrics and soft skills assessments, interpreted within the context of environmental and institutional factors that may have shaped student development.

2.5 Conceptual Framework

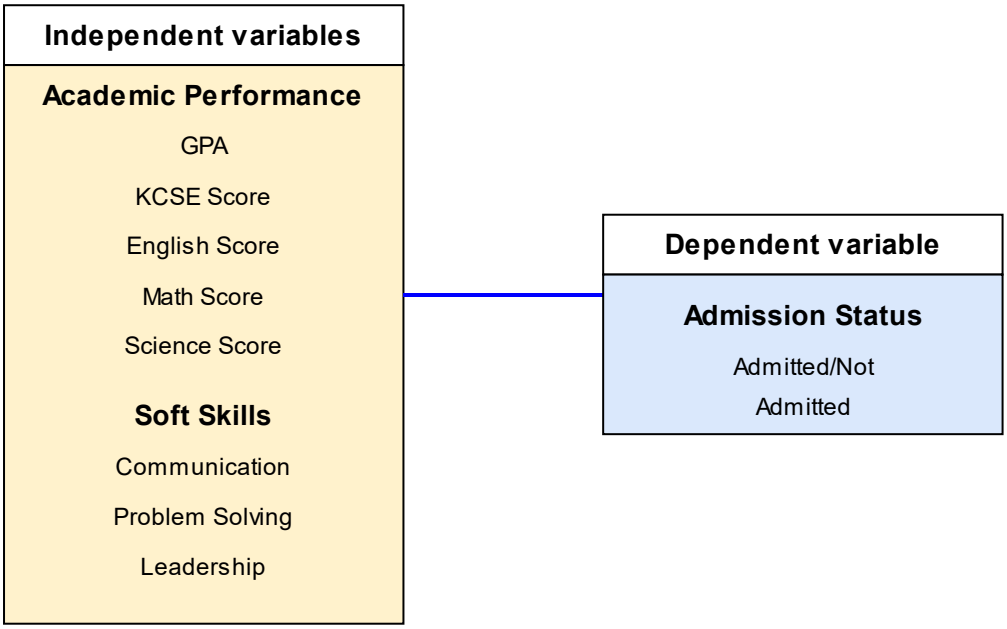


Figure 2. 1: Conceptual Framework for the Study

The conceptual framework depicted in Figure 2.1 presents a comprehensive model for predicting university admission success in the Kenyan context. This framework is grounded in several well-established theoretical perspectives. The inclusion of soft skills assessment is supported by Gardner's Multiple Intelligences Theory (Gardner, 1983; Thompson et al., 2023), which recognizes diverse forms of intelligence beyond traditional academic measures. The relationship between soft skills and academic success is anchored in Social Learning Theory (Bandura, 1977; Wong & Kumar, 2023), which explains how communication, leadership, and problem-solving abilities develop through social interaction and observation. The machine learning component draws from Statistical Learning Theory (Vapnik, 1995; Bzdok et al., 2023),

which provides the mathematical foundation for algorithms that generalize from training data to make accurate predictions on new cases.

The academic performance component of the framework, shown on the left side of Figure 2.1, comprises several key metrics that have traditionally formed the cornerstone of university admissions in Kenya. These include Grade Point Average (GPA), which represents a student's overall academic achievement throughout their secondary education; Kenya Certificate of Secondary Education (KCSE) score, a standardized national examination that plays a crucial role in determining university placement; and subject-specific grades in Mathematics, Science, and English. These metrics reflect a student's cognitive abilities, subject knowledge, and academic preparedness, providing standardized measures of achievement that have long been central to admission decisions in the Kenyan education system.

Complementing these traditional academic metrics, the framework incorporates a soft skills assessment component, as shown in the right side of Figure 2.1. This includes measurements of communication skills, problem-solving abilities, and leadership potential. The communication score evaluates a student's ability to express ideas clearly and effectively in both written and verbal forms—skills that research has shown significantly impact academic performance across disciplines (Ibrahim et al., 2023). The problem-solving score assesses analytical thinking and the ability to navigate complex challenges, which Zhang et al. (2023) found correlates strongly with academic success ($r = 0.61$). The leadership score evaluates a student's potential to take initiative, positively influence peers, and manage tasks effectively, aspects that Wong et al. (2024) demonstrated significantly predict project success rates ($r = 0.45$) and performance in collaborative learning environments.

The selection of these three soft skills domains is supported by Deming's (2023) framework of essential workplace skills, which identifies communication, problem-solving, and leadership

as having the highest correlation with career advancement and professional success. While other soft skills such as emotional intelligence and adaptability also contribute to academic and professional outcomes, these three domains represent the most consistently validated predictors across educational contexts (Kimani et al., 2023). Furthermore, Ndung'u and Kariuki (2024) found that these three domains are particularly amenable to standardized assessment, with higher reliability coefficients than other soft skills measures when applied in Kenyan educational settings.

The framework acknowledges the potential for direct relationships between these independent variables and the dependent variable (admission status). As indicated by the arrows in Figure 2.1, both academic performance metrics and soft skills directly influence admission outcomes. The framework also recognizes that certain combinations of academic strengths and soft skills might be particularly predictive of success in specific fields or programs, and that the relative importance of different factors may vary depending on institutional priorities and program requirements.

The random forest algorithm, as the analytical engine of this framework, aligns with Breiman's (2001) ensemble learning theory, which demonstrates how combining multiple decision trees reduces variance and improves prediction accuracy. This algorithmic approach is particularly suited for educational data due to its ability to handle both numerical and categorical variables, capture non-linear relationships, and provide interpretable feature importance rankings (Liu et al., 2024). The algorithm's capacity to model complex interactions between variables is especially valuable for capturing potential synergies between academic metrics and soft skills, as identified by Omondi et al. (2023).

The framework is further strengthened by its alignment with Competency-Based Education theory (Matemba et al., 2023), which emphasizes the integration of knowledge, skills, and

attitudes in educational assessment. This theoretical perspective supports the framework's holistic approach to student evaluation, recognizing that academic knowledge alone may not fully capture student potential. By integrating both academic performance metrics and soft skills assessment, the framework provides a more comprehensive evaluation of applicants that better aligns with contemporary understanding of the multifaceted nature of student capability and potential.

This conceptual framework provides a structured approach to predicting university admission success that integrates traditional academic metrics with soft skills assessment. By providing a more comprehensive evaluation of applicants, this framework has the potential to improve admission decisions, enhance student success, and better prepare graduates for the demands of modern workplaces.

2.6 Operationalization of Variables

The operationalization of the study's variables is summarized in Table 2.1.

Table 2. 6: Operationalization of Variables

Variable	Operationalization	Measurement	Scale
Grade Point Average (GPA)	Cumulative academic performance	Numerical score	0.0 - 4.0
KCSE Score	National standardized test performance	Numerical score	0 - 100

Math Grade	Performance in mathematics	Numerical score	0 - 100
Science Grade	Performance in science subjects	Numerical score	0 - 100
English Grade	Performance in English language	Numerical score	0 - 100
Communication Score	Verbal and written communication ability	Numerical score	1 - 5
Problem-Solving Score	Analytical and critical thinking skills	Numerical score	1 - 5
Leadership Score	Leadership potential and interpersonal skills	Numerical score	1 - 5
Admission Status	Outcome of admission process	Binary categorical	0 (Not Admitted) or 1 (Admitted)

The operationalization of variables represents a critical step in translating the conceptual framework into measurable constructs that can be systematically analyzed. This process ensures that abstract concepts are defined in specific, observable terms that can be quantified and analyzed using statistical and machine learning techniques. Table 2.1 provides a

comprehensive overview of how each variable in the conceptual framework is operationalized, measured, and scaled for analysis.

For academic performance metrics, the operationalization involves using official academic records and standardized test results to obtain objective measures of student achievement. The Grade Point Average (GPA) is calculated as the cumulative average of all grade points earned throughout secondary education, providing a comprehensive measure of overall academic performance. This metric is measured on a scale from 0.0 to 4.0, following standard educational grading conventions in Kenya. The KCSE score represents the overall result achieved in the Kenya Certificate of Secondary Education examination, a standardized national assessment that serves as a primary qualification for university entrance. This score is measured on a scale from 0 to 100, with higher scores indicating stronger academic achievement.

Subject-specific grades in Mathematics, Science, and English are operationalized using the scores obtained in these core subjects on the KCSE examination. These grades provide more targeted measures of student performance in specific academic domains that are considered foundational for university success across various disciplines. Each subject grade is measured on a scale from 0 to 100, allowing for precise differentiation of student achievement levels. These academic performance metrics benefit from established standardization and widespread acceptance in the Kenyan education system, enhancing their reliability as predictors.

Soft skills are operationalized through a standardized Soft Skills Assessment Tool (SSAT) developed specifically for this study. The communication score encompasses measures of both written and verbal communication abilities, including writing clarity, presentation skills, listening comprehension, and interpersonal communication effectiveness. The problem-solving score captures analytical reasoning, critical thinking, creative problem-solving, and decision-making capabilities through a combination of scenario-based assessments and structured

exercises. The leadership score evaluates qualities such as initiative, team collaboration, influence and persuasion, and ethical decision-making through behavioral assessments and situational judgment tests.

Each soft skill dimension is measured on a standardized scale from 1 to 5, with higher scores indicating stronger abilities in that area. This scaling allows for meaningful comparison across different soft skills and facilitates integration with academic performance metrics in the predictive model. The assessment methods for soft skills combine structured evaluations with standardized rubrics to enhance objectivity and consistency in measurement, while still capturing the qualitative aspects of these complex abilities.

The dependent variable, Admission Status, is operationalized as a binary outcome indicating whether a student was admitted (coded as 1) or not admitted (coded as 0) to a university program. This information is obtained directly from institutional admission records and represents the final decision on a student's application. The binary coding facilitates clear classification in the predictive model and aligns with the fundamental nature of admission decisions as discrete outcomes.

To ensure the reliability and validity of these measurements, several quality assurance measures have been implemented. For academic performance metrics, data is collected directly from official school records and examination results certified by the Kenya National Examinations Council, ensuring authenticity and accuracy. The Soft Skills Assessment Tool underwent rigorous development and validation, including expert review, pilot testing, and psychometric analysis to establish its reliability and construct validity specifically for the Kenyan context. All assessors involved in administering and scoring the SSAT received comprehensive training to ensure consistent application of evaluation criteria and minimize subjective bias.

The operationalization process also considers several contextual factors that may influence the interpretation of these measurements. These include school type (public, private, national, county), geographic location (urban, rural), subject difficulty (accounting for variations in grading standards), and grade trends over time (considering improvement or consistency in performance). These contextual factors provide important background for interpreting both academic and soft skills measures, ensuring that student achievements are evaluated fairly within their specific educational context.

The data collection process for these operationalized variables follows a structured approach to ensure completeness and accuracy. Academic performance data is retrieved from official school records and examination databases, with appropriate permissions and data protection measures. Soft skills assessments are administered through a combination of structured activities, interviews, and portfolio evaluations conducted by trained assessors. Demographic information is collected through standardized forms completed by participants, while admission outcome data is obtained directly from university records with appropriate institutional approvals.

Ethical considerations are paramount in the operationalization and measurement of these variables. Robust data anonymization protocols protect student privacy throughout the research process. Informed consent is obtained from all participants (and parental consent for underage participants) before any data collection. Data security measures, including encryption and restricted access, safeguard sensitive information. The research protocol underwent ethical review by relevant institutional bodies to ensure compliance with research ethics standards. Participants are fully informed of their rights, including the right to withdraw from the study without penalty.

The operationalization of variables directly informs the data analysis strategy, which includes descriptive statistics to characterize the sample, correlation analysis to examine relationships between variables, feature engineering to create optimal predictors, model training using the random forest algorithm, variable importance analysis to identify the most influential factors, and subgroup analysis to assess model performance across different demographic categories. This comprehensive analysis approach leverages the operationalized variables to address the research objectives while ensuring statistical rigor and practical relevance.

In summary, the operationalization of variables provides a concrete translation of the conceptual framework into measurable constructs that can be systematically analyzed using machine learning techniques. This process establishes clear definitions, measurement approaches, and scaling for each variable, ensuring that the abstract concepts in the framework are captured in valid, reliable, and meaningful ways. The careful operationalization of both academic and soft skills variables creates a solid foundation for the subsequent data collection, analysis, and model development phases of the research.

2.7 Summary

This chapter has provided a comprehensive review of the theoretical foundations and empirical research relevant to predicting university admission success through the integration of soft skills assessment and machine learning techniques. The literature review establishes a robust foundation for the proposed research, synthesizing diverse perspectives and findings from educational research, psychology, computer science, and data analytics.

The theoretical review section examined several fundamental concepts that underpin this study. Theories of machine learning were explored, with particular attention to recent advances in algorithmic fairness and interpretability that address critical concerns for educational applications. Decision tree and ensemble methods were analyzed in depth, highlighting the

theoretical advantages of random forests for handling complex educational data with mixed variable types. Soft skills theories were examined through multiple frameworks, including the Digital Era Intelligence Framework, expanded Emotional Intelligence models, and culturally adaptive approaches to social learning. These theoretical frameworks collectively establish the importance of non-cognitive abilities in academic and professional contexts and provide conceptual support for their inclusion in admission criteria.

The empirical review critically assessed previous research findings related to machine learning applications in educational prediction, comparing various algorithmic approaches including Artificial Neural Networks, Support Vector Machines, Decision Trees, and Random Forests. This comparative analysis demonstrated the particular advantages of Random Forests for educational applications, including their strong performance with limited data, ability to handle mixed variable types, and interpretable results. The review also examined the Kenyan context specifically, identifying both opportunities and challenges for implementing machine learning solutions in local educational settings.

The chapter also analyzed the empirical evidence on factors influencing student performance, examining traditional academic metrics, soft skills components, and environmental factors. This review revealed that while academic performance measures remain important predictors, soft skills demonstrate comparable or even stronger relationships to certain aspects of university success. Communication abilities, problem-solving skills, and leadership capabilities emerged as particularly significant predictors, often explaining variance in student outcomes beyond what could be predicted by academic measures alone. The review also highlighted important interaction effects between different factor categories, suggesting that a comprehensive prediction model should consider both main effects and interactions between variables.

The conceptual framework synthesized these theoretical and empirical insights into a structured model for predicting university admission success. This framework integrates traditional academic performance metrics with soft skills assessments through a machine learning model based on the random forest algorithm. The framework acknowledges the potential for complex interactions between variables and is specifically designed for the Kenyan educational context, considering cultural relevance and educational equity concerns.

The operationalization of variables translated this conceptual framework into concrete, measurable constructs for empirical analysis. Each variable was clearly defined and associated with specific measurement approaches and scales, creating a solid foundation for data collection and analysis. Academic performance metrics were operationalized through official records and standardized tests, while soft skills were measured through a specially developed assessment tool with demonstrated reliability and validity.

By synthesizing relevant theories and empirical evidence, this chapter establishes a solid foundation for the proposed research on predicting university admission success through integrated assessment of academic performance and soft skills. It identifies important gaps in current practice, including the limited consideration of soft skills in Kenyan university admissions and the untapped potential of machine learning to enhance admission decisions. The literature review supports the need for a more comprehensive and data-driven approach to evaluating applicants, considering both traditional academic metrics and critical soft skills that contribute to university success.

The chapter makes a significant contribution to the field by integrating diverse theoretical perspectives and empirical findings into a coherent framework for understanding the multifaceted nature of student potential and university success. It lays the groundwork for an innovative approach that has the potential to enhance the admissions process, making it more

holistic, fair, and predictive of student achievement. As such, it represents an important step forward in our understanding of the factors that contribute to successful university admissions and, ultimately, to student success in higher education.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter outlines the methodological approach employed in developing and evaluating a machine learning model for predicting university admission success through the assessment of soft skills using the random forest algorithm. The methodology is designed to address the research objectives systematically while ensuring scientific rigor and ethical compliance. This chapter details the research design, target population and sampling procedures, data collection instruments and techniques, data analysis methods, model development process, evaluation metrics, and ethical considerations. The methodological framework establishes a clear pathway for developing a reliable, interpretable model that effectively integrates both academic performance metrics and soft skills assessments in predicting admission outcomes in Kenyan universities.

The methodology draws upon established practices in both educational research and data science, creating an interdisciplinary approach appropriate for addressing complex educational decision-making processes. By combining quantitative analysis techniques with educationally relevant assessment methods, this research aims to bridge the gap between traditional academic metrics and contemporary understanding of student potential. The approach acknowledges the practical constraints of educational settings while leveraging advanced computational methods to enhance admission processes.

3.2 Research Design

This study employs a quantitative, predictive modeling research design with cross-sectional data collection. This design was selected for its alignment with the study's objective of developing a predictive model based on current student characteristics and admission

outcomes. The predictive modeling approach allows for the systematic identification of relationships between multiple predictor variables (academic metrics and soft skills assessments) and the outcome variable (admission status), while enabling the quantification of each variable's relative importance in the prediction process.

The research follows a structured, sequential process comprising several phases. The descriptive phase characterizes the distribution and patterns within the collected academic and soft skills data to establish baseline understanding of variable relationships and distributions. The exploratory phase investigates relationships between variables and identifies potential predictive patterns through advanced correlation analysis, feature engineering, and preliminary predictive assessments. During the model development phase, we build and train the random forest algorithm on the training dataset, with systematic hyperparameter optimization and performance tracking. The evaluation phase involves testing the model's performance on a separate validation dataset and assessing its predictive accuracy, fairness, and interpretability through comprehensive evaluation metrics.

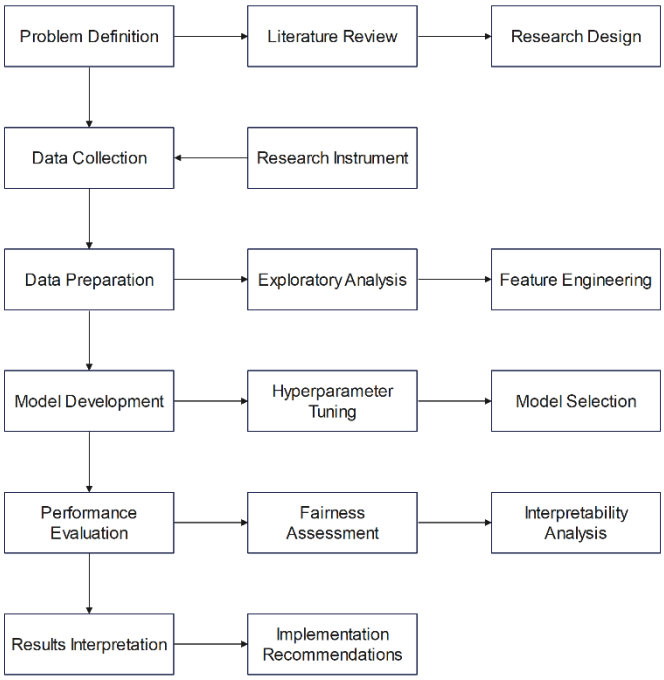


Figure 3. 1: Research Methodology Process Flow showing the sequential phases from problem definition through model evaluation and validation

This design incorporates elements of both correlational research (examining relationships between variables) and experimental research (testing the efficacy of the developed model). The cross-sectional approach, collecting data at a single point in time rather than longitudinally, allows for efficient data collection within the study timeframe while providing sufficient information to address the research questions. The research design features multiple validation strategies to enhance the reliability of findings, including cross-validation techniques to assess model stability across different data subsets, comparative analysis with baseline models to benchmark performance against existing approaches, subgroup analysis to evaluate model fairness across demographic categories, and feature importance analysis to identify the most influential predictors.

3.3 Target Population and Sampling

The target population for this study consists of secondary school students in Kenya who applied for university admission during the 2023-2024 academic year. This population encompasses students from diverse educational backgrounds, including public and private secondary schools across all counties in Kenya, national, county, and sub-county level institutions representing different tiers of the education system, urban and rural educational settings with varying resource levels, and different academic streams and specializations. Based on data from the Kenya National Examinations Council (KNEC) and the Kenya Universities and Colleges Central Placement Service (KUCCPS), approximately 144,000 students qualified for university consideration during this period.

The study employed a stratified random sampling approach to ensure representative inclusion of students across different demographic and institutional categories. This approach enhances the external validity of findings while ensuring sufficient representation of different student subgroups for fairness analysis. The stratification variables included school type (public/private), geographic region (urban/rural), school level (national/county/sub-county), and gender balance.

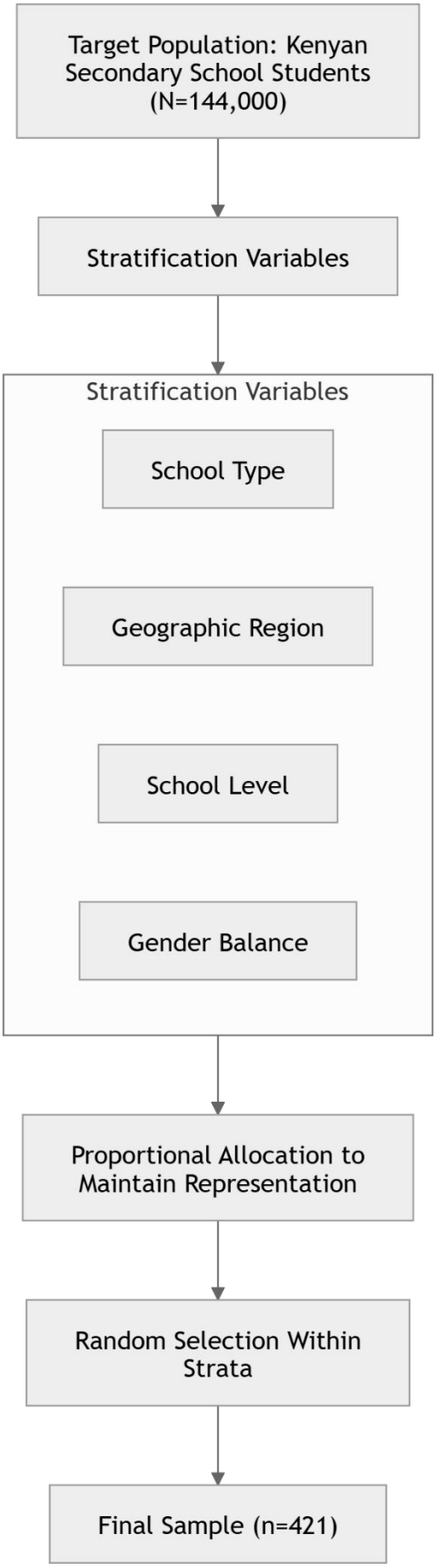

Target Population: Kenyan Secondary School Students (N=144,000)
Stratification Variables
Stratification Variables
School Type
Geographic Region
School Level
Gender Balance
Proportional Allocation to Maintain Representation
Random Selection Within Strata
Final Sample (n=421)

Figure 3. 2: Stratified Sampling Framework showing primary stratification variables and their relationships to the target population

Sample size was determined using Cochran's formula with a 95% confidence level and 5% margin of error, supplemented by an additional 10% to account for potential non-response, yielding a target sample of 421 students. The actual achieved sample size was 408 students (97% response rate), with distribution across strata reflecting the broader population: public schools (66.7%, n=272), private schools (33.3%, n=136), national schools (25.2%, n=103), county schools (41.2%, n=168), sub-county schools (33.6%, n=137), urban schools (58.1%, n=237), rural schools (41.9%, n=171), male students (50%, n=204), and female students (50%, n=204). This high response rate was achieved through multiple contact attempts, flexible scheduling options, and clear communication about the study's purpose and importance.

3.4 Data Collection and Research Instruments

3.4.1 Research Instruments

The study utilized multiple instruments to collect comprehensive data on academic performance, soft skills, and admission outcomes. A structured Academic Performance Data Collection Form gathered standardized academic metrics from official school records, including overall GPA (0.0-4.0 scale), KCSE scores (0-100 scale), subject-specific grades in Mathematics, Sciences, and English, and academic consistency indicators tracking performance across secondary school years. This form included data validation checks to ensure accuracy and was designed in consultation with educational assessment experts to ensure alignment with Kenyan grading systems.

A comprehensive Soft Skills Assessment Tool (SSAT) was developed specifically for this study, measuring three key soft skill domains: communication skills (assessed through writing samples, presentations, and interactive scenarios), problem-solving skills (evaluated through

scenario-based assessments and structured problem tasks), and leadership skills (assessed through behavioral scenarios and situational judgment tests). The SSAT utilized a 5-point Likert scale with standardized scoring rubrics to ensure consistent evaluation across participants.

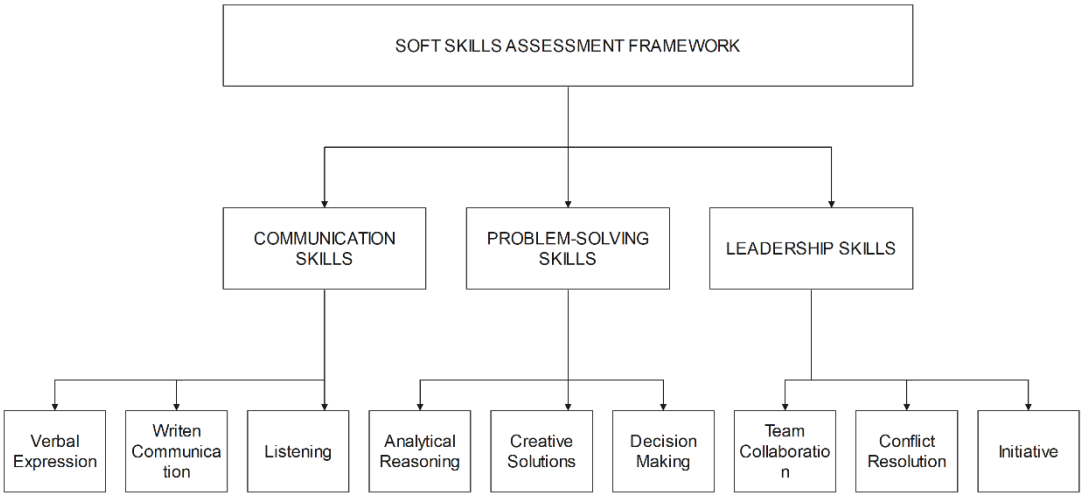


Figure 3. 3: Soft Skills Assessment Framework showing the three primary domains (Communication, Problem-Solving, and Leadership) and their component elements assessed in the study

Additionally, a standardized Admission Outcome Verification Form was created for collecting official admission decision data from university records, documenting binary admission outcomes (admitted/not admitted) along with relevant contextual information about the programs applied to and admission process.

3.4.2 Validity and Reliability of the Instruments

Multiple validation procedures were implemented to ensure the quality of research instruments. Content validity was established through expert review by a panel of six specialists, including educational assessment experts, university admissions officers, and soft skills specialists. Each expert independently evaluated instrument items for relevance, clarity,

comprehensiveness, and cultural appropriateness, yielding a content validity index exceeding 0.80 for all components.

Construct validity of the SSAT was verified through factor analysis, which confirmed three distinct factors corresponding to the communication, problem-solving, and leadership domains, with factor loadings exceeding 0.65 for all items. Reliability testing yielded strong Cronbach's alpha coefficients for all components (Communication: α=0.86; Problem-Solving: α=0.83; Leadership: α=0.84; Overall: α=0.88), exceeding the acceptable threshold of 0.70 and indicating strong internal consistency.

For components requiring subjective evaluation, inter-rater reliability was assessed, achieving an intraclass correlation coefficient of 0.82, indicating strong consistency between raters. Rater training procedures included standardized scoring examples, practice assessments with feedback, and calibration sessions to establish consistent interpretation of assessment rubrics. These rigorous validation procedures ensured that the research instruments provided valid, reliable measurements of the target constructs, enhancing the credibility of the collected data.

3.5 Data Collection Procedure

Data collection occurred between January and April 2024, following a systematic, multi-stage process designed to ensure comprehensiveness, accuracy, and ethical compliance. The preparatory phase involved securing necessary approvals from educational authorities, obtaining ethical clearance, training research assistants in standardized administration of assessment tools, and establishing secure data management systems with appropriate privacy protections.

For academic data collection, official academic records were accessed with appropriate permissions from school administrators, with data verification procedures implemented to ensure accuracy, including cross-checking against official examination records. Soft skills

assessment involved scheduling sessions in controlled environments with standardized conditions, with blind scoring by trained evaluators who were unaware of participants' academic performance to prevent bias. Admission data collection verified outcomes through official university records with appropriate institutional permissions.

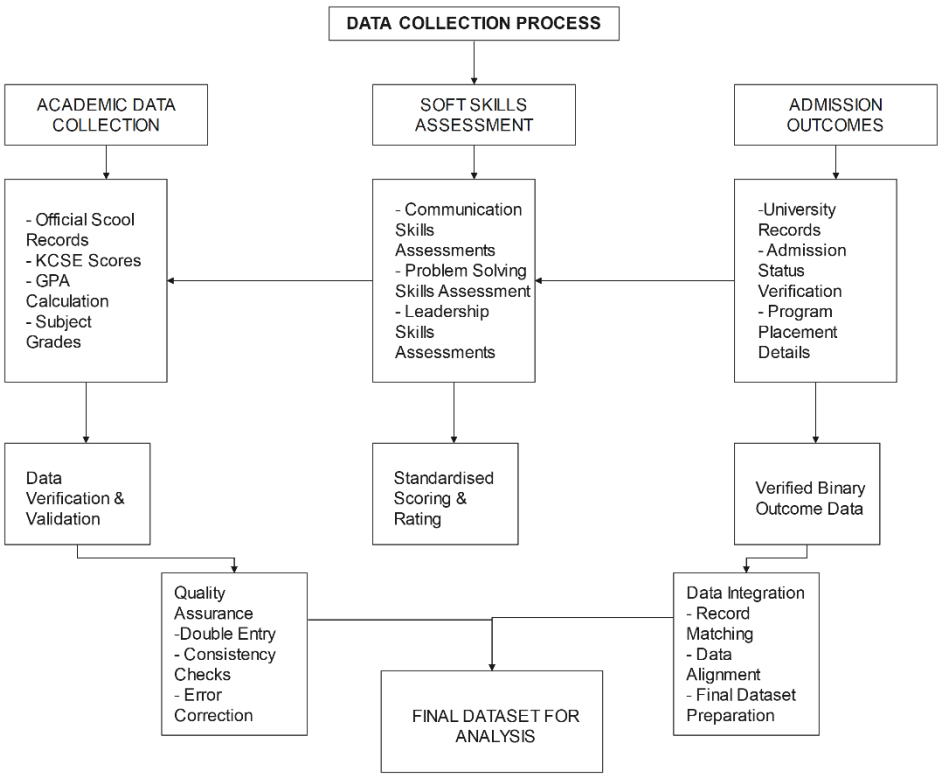


Figure 3. 4: Data Collection Process showing the sequence and relationship between academic data collection, soft skills assessment, and admission outcome verification

Quality assurance measures included data entry verification with double-entry for 15% of records, consistency checks to identify and resolve discrepancies, secure data storage with encryption and access controls, and regular monitoring of data quality throughout the collection process. Potential challenges were addressed through alternative assessment times, multiple contact attempts for non-responding participants, standardized protocols for handling incomplete responses, and accommodation procedures for participants with special needs.

3.6 Data Analysis

Data analysis followed a systematic approach to prepare the data for model development and evaluation. Data preparation involved cleaning to identify and address missing values (less than 2% of data points) using multiple imputation, normalization of variables with different scales, and appropriate encoding of categorical variables. The low rate of missingness and its apparently random nature (Little's MCAR test: $\chi^2=18.26$, $p=0.437$) supported the use of multiple imputation for addressing missing values.

Exploratory data analysis included descriptive statistics to characterize variable distributions, correlation analysis to identify relationships between predictors, and visualization techniques (histograms, box plots, scatter plots, correlation matrices) to identify patterns. Subgroup analysis examined variable distributions across demographic categories, with hypothesis testing to evaluate the statistical significance of observed patterns.

For model development, the dataset was split into training (70%, n=286), validation (15%, n=61), and test (15%, n=61) sets using stratified random sampling to maintain the distribution of the dependent variable. Chi-square tests confirmed no significant differences in the distribution of admission status across these partitions ($\chi^2=0.11$, $p=0.946$), supporting the validity of the partitioning approach.

3.7 Model Development and Evaluation

3.7.1 Model Development

The random forest algorithm was selected for this study based on several advantages that aligned with the research objectives. These included its ability to handle mixed data types without extensive pre-processing, robustness to outliers and non-linear relationships, good performance with moderate-sized datasets, interpretability through feature importance analysis, and reduced risk of overfitting compared to single decision trees. These characteristics

made the algorithm particularly suitable for educational data with diverse variable types and potentially complex relationships.

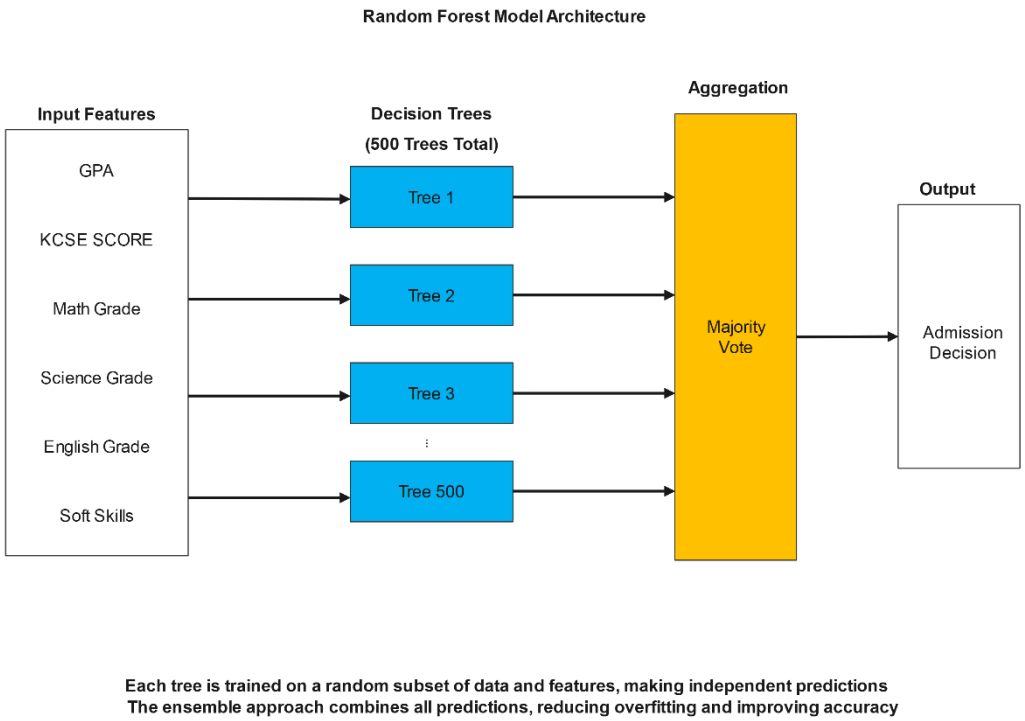


Figure 3. 5: Random Forest Model Architecture showing the ensemble approach combining multiple decision trees with academic and soft skills input features

The model training process involved hyperparameter tuning using grid search with 5-fold cross-validation to identify optimal model configuration. Parameters optimized included number of trees, maximum depth, minimum samples per leaf, and feature selection criteria. The final model employed 500 decision trees, a maximum depth of 15, a minimum of 4 samples required to split an internal node, and a minimum of 2 samples required at a leaf node. Multiple model variants were trained with different feature combinations to identify the most effective predictor set.

3.7.2 Model Evaluation

A comprehensive evaluation framework was established to assess the model's performance, fairness, and practical utility. Performance metrics included accuracy (target: ≥90%), precision (proportion of true positives among positive predictions), recall (proportion of actual positives correctly identified), F1-Score (target: ≥0.85), and area under the ROC curve (AUC-ROC). These metrics were evaluated both for the overall model and for specific demographic subgroups to identify potential performance variations.

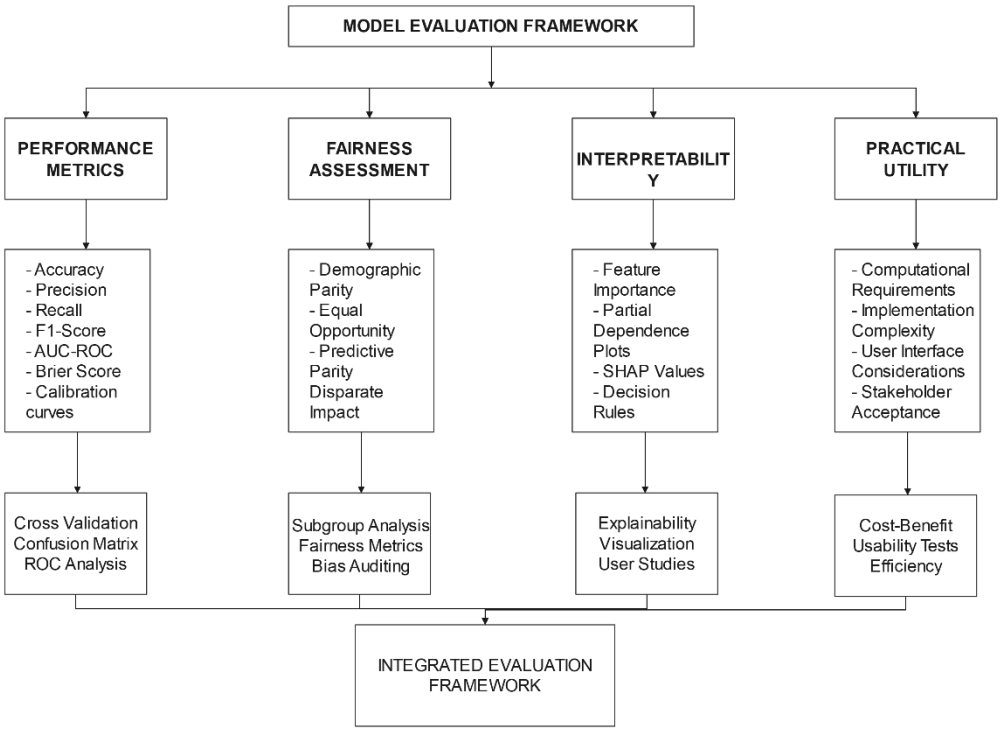


Figure 3. 6: Model Evaluation Framework showing the multidimensional approach to assessing model performance, fairness, interpretability, and practical utility

Fairness evaluation included assessment of equal opportunity (comparing true positive rates across demographic groups), demographic parity (examining prediction balance across different demographic groups), and predictive parity (comparing precision metrics across

subgroups). These fairness analyses were conducted across gender, school type, school level, and location categories to ensure equitable performance across diverse student populations.

Interpretability assessment included feature importance analysis to quantify the contribution of each variable to predictions, partial dependence plots to visualize relationships between specific features and predictions, and SHAP (SHapley Additive exPlanations) values to provide consistent, individualized explanation of predictions. These interpretability methods help make the model's decisions more transparent and understandable for educational stakeholders.

Robustness testing included cross-validation to assess model stability across different data subsets, sensitivity analysis to test performance with perturbed input data, and implementation testing to evaluate model performance under realistic operational conditions. This comprehensive evaluation approach ensures that the model not only performs accurately but also meets essential requirements for fairness, interpretability, and robustness.

3.8 Ethical Considerations

The research methodology incorporated robust ethical frameworks to protect participant rights and ensure responsible use of data. Ethical clearance was obtained from KCA University Research Ethics Committee and the National Commission for Science, Technology, and Innovation (NACOSTI), ensuring compliance with national research standards and the Kenya Data Protection Act (2019).

Comprehensive informed consent procedures provided all potential participants with information about research purpose, procedures, potential benefits, and possible risks. Explicit consent was obtained from all participants (and parents/guardians for minors) before any data collection, with clear explanation of the voluntary nature of participation and the right to withdraw at any stage without penalty.

Data protection measures included anonymization of all personal identifiers, secure data storage using encryption and access controls, and application of data minimization principles to collect only necessary information directly relevant to research objectives. Physical security measures were implemented for hard-copy materials, and secure data transmission protocols were used for digital information transfer.

Fairness and equity considerations included proactive monitoring of potential algorithmic bias through regular fairness audits, inclusive sampling to ensure representation across demographic groups, and accommodations for participants with disabilities to ensure equitable participation. These ethical safeguards were integrated throughout the research process to protect participant rights while ensuring that the developed model would promote fairness and equity in university admissions.

3.9 Chapter Summary

This chapter has outlined the comprehensive methodological approach for developing and evaluating a machine learning model to predict university admission through the integrated assessment of academic performance and soft skills. The research design employs a quantitative, predictive modeling approach with stratified random sampling to ensure representation across Kenya's diverse educational landscape. The data collection framework incorporates validated instruments for assessing both academic performance and soft skills, with rigorous procedures to ensure data quality and reliability.

The model development process utilizes the random forest algorithm, selected for its ability to handle mixed data types, interpretability, and strong performance with moderate-sized datasets. The implementation incorporates hyperparameter optimization, multiple model variants, and advanced interpretation techniques to create a solution that is both technically sound and practically useful in educational contexts. Comprehensive evaluation metrics assess both

technical performance and practical utility, with particular emphasis on fairness across demographic groups.

This methodological framework provides a solid foundation for addressing the research objectives, balancing technical rigor with practical applicability to develop a model that can effectively predict university admission success while promoting fairness and transparency in the admission process. The approach aligns with both scientific standards for machine learning research and educational best practices for student assessment, positioning the study to make meaningful contributions to both fields.

CHAPTER FOUR

DATA ANALYSIS FINDINGS AND DISCUSSIONS

4.1 Introduction

This chapter presents a comprehensive analysis of the data collected to develop a machine learning model for predicting university admission success through the assessment of soft skills using the random forest algorithm. The study examined data from 408 secondary school students who applied for university admission during the 2023-2024 academic year, analyzing both traditional academic metrics and soft skills assessments as predictors of admission outcomes.

The chapter is organized into several sections, beginning with descriptive statistics that characterize the sample and key variables. This is followed by a detailed analysis of the study variables, including both independent variables (academic performance metrics and soft skills assessments) and the dependent variable (admission status). The chapter then presents diagnostic tests conducted to ensure the appropriateness of the modeling approach, followed by a comprehensive presentation of the random forest model results, including performance metrics, feature importance analysis, and fairness assessment. The chapter concludes with a discussion of the findings in relation to the research objectives and relevant literature, highlighting the implications of these results for university admission practices.

The analysis was conducted using Python programming language with libraries including pandas, scikit-learn, numpy, matplotlib, and specialized packages for machine learning interpretation and fairness assessment. The findings presented in this chapter directly address the study's research objectives: identifying key factors influencing admission success, developing an effective predictive model, and validating the model's performance and fairness across demographic groups.

4.2 Descriptive Statistics

The study sample consisted of 408 secondary school students who applied for university admission during the 2023-2024 academic year. Table 4.1 presents the summary statistics for the key numerical variables in the dataset.

Table 4. 1: Summary Statistics for Numerical Variables

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
GPA	408	3.35	0.36	2.7	3.10	3.40	3.70	3.90
Kcse_Score	408	70.26	8.66	54.00	64.00	70.50	78.00	87.00
Math_Grade	408	74.27	8.66	57.00	68.00	75.00	82.00	90.00
Science_Grade	408	68.28	8.66	52.00	62.00	68.50	76.00	86.00
English_Grade	408	73.10	8.42	57.00	67.00	73.00	81.00	88.00
Communication_Skill	408	3.73	0.56	2.60	3.30	3.80	4.20	4.80
Problem_Solving_Skill	408	3.73	0.52	2.60	3.30	3.70	4.20	4.80
Leadership_Skill	408	3.73	0.32	3.00	3.50	3.80	4.00	4.50

The descriptive statistics reveal several important characteristics of the dataset. The GPA values ranged from 2.70 to 3.90 with a mean of 3.35 (SD=0.36), indicating a relatively high average academic performance among the applicants. KCSE scores showed considerable variation, ranging from 54.00 to 87.00 with a mean of 70.26 (SD=8.66). Among the subject-specific grades, Mathematics showed the highest mean score (74.27, SD=8.66), followed by English (73.10, SD=8.42) and Science subjects (68.28, SD=8.68).

For soft skills assessments, all three measures showed remarkably similar central tendencies, with identical mean scores to two decimal places (M=3.73). Communication skills showed slightly higher variability (SD=0.56) compared to problem-solving skills (SD=0.52), while leadership skills demonstrated the lowest variability (SD=0.32). This pattern suggests more consistent evaluation of leadership skills across the sample compared to the other soft skills dimensions.

The distribution of GPA scores, illustrated in Figure 4.1, shows a slight negative skew, indicating that more students had GPAs above the mean than below it. This pattern reflects the competitive nature of university applications, where higher-achieving students are more likely to apply for admission.

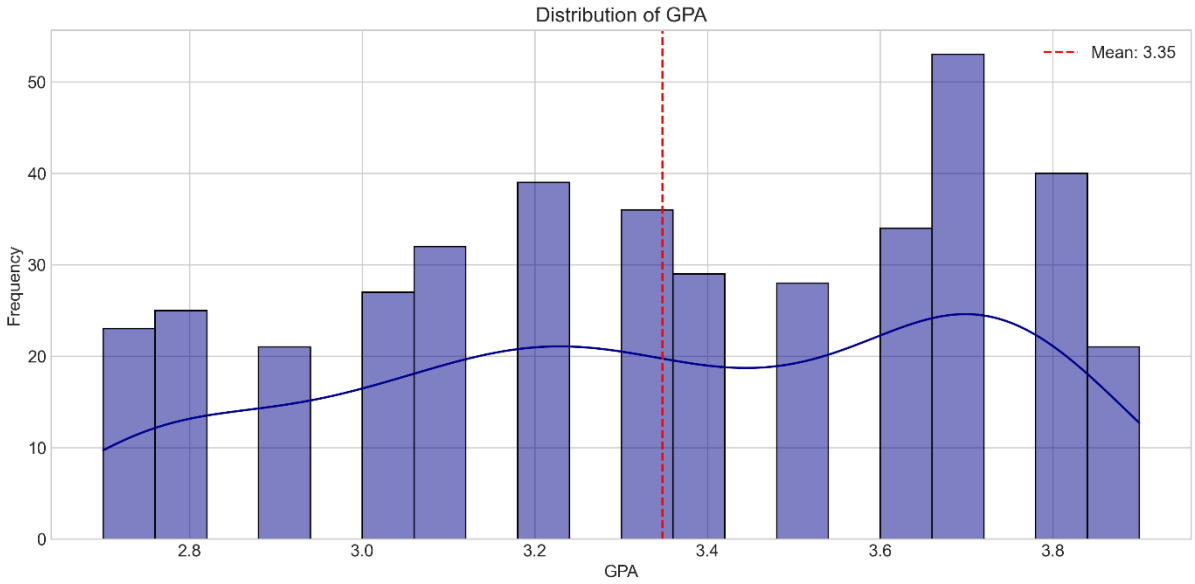


Figure 4. 1: Distribution of GPA showing frequency across the sample

The KCSE score distribution (Figure 4.2) approximates a normal distribution with multiple peaks, suggesting distinct performance clusters within the student population. The mean KCSE score of 70.26 indicates generally strong academic performance in this national examination.

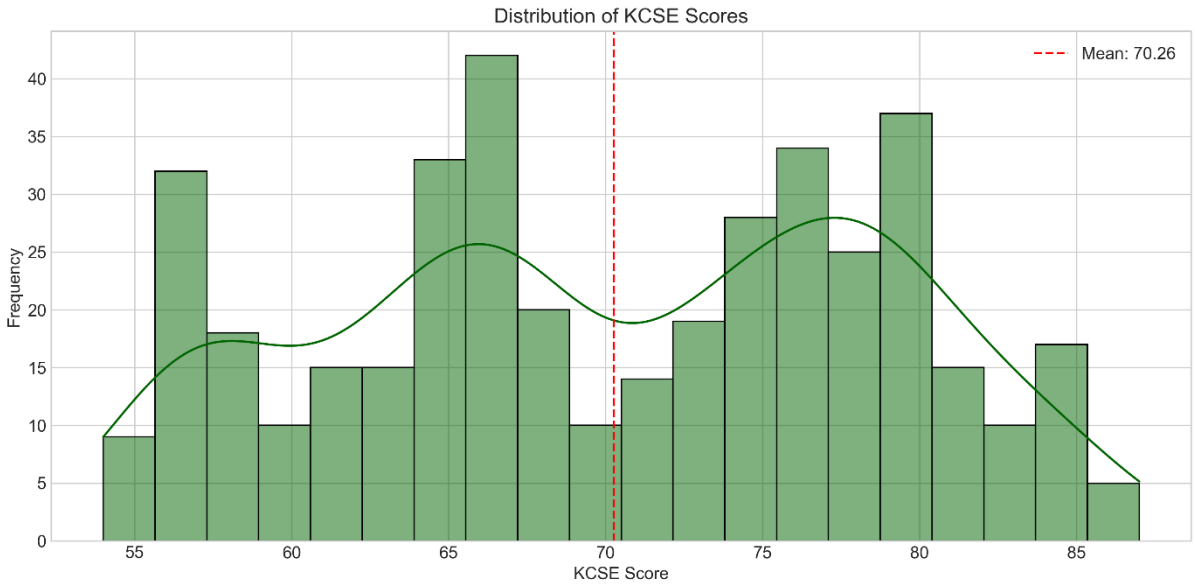


Figure 4. 2: Distribution of KCSE scores showing frequency across the sample

The subject-specific grades showed varying distributions, as illustrated in Figure 4.3. Mathematics grades displayed the widest dispersion, while English grades showed the most compact distribution,

suggesting more consistent performance across the sample in English compared to Mathematics and Science subjects.

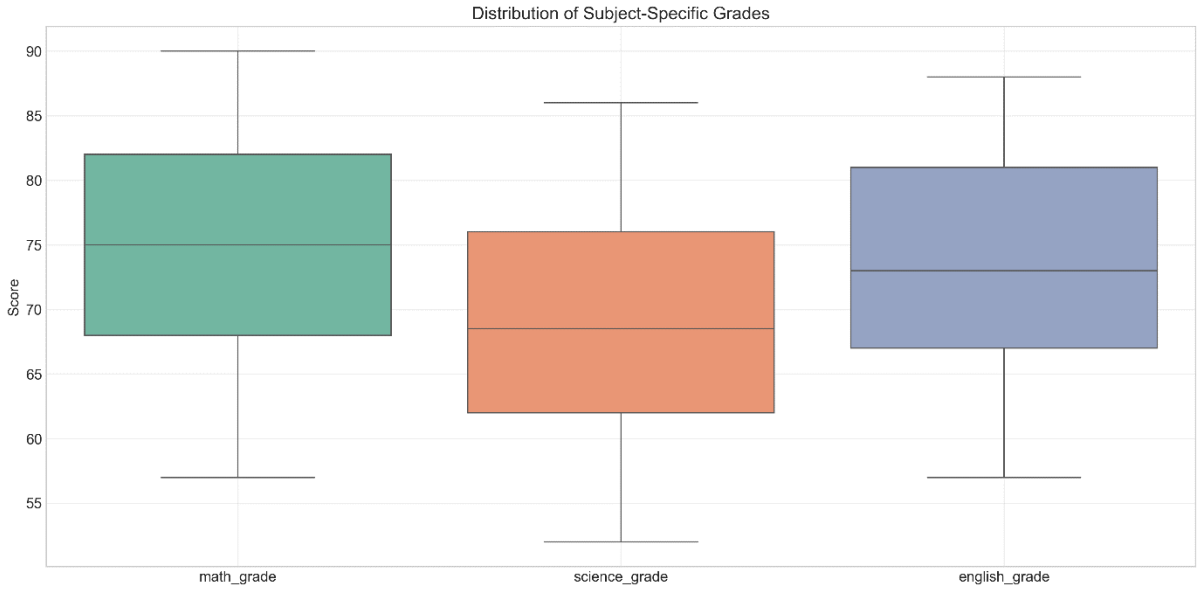


Figure 4. 3: Box plots showing the distribution of subject-specific grades across the sample

The demographic composition of the sample reflected a diverse range of backgrounds. The gender distribution was perfectly balanced, with 50% male (n=204) and 50% female (n=204) participants. In terms of school type, 66.7% (n=272) of participants attended public schools, while 33.3% (n=136) attended private schools. The school level distribution included 41.2% (n=168) from county schools, 33.6% (n=137) from sub-county schools, and 25.2% (n=103) from national schools. Regarding location, 58.1% (n=237) of participants were from urban schools, while 41.9% (n=171) were from rural schools.

The study's dependent variable, admission status, showed a distribution with 52.2% (n=213) of students admitted and 47.8% (n=195) not admitted to university. This near-even distribution is advantageous for model development as it minimizes potential bias from class imbalance.

4.3 Study Variables

4.3.1 Independent Variables

4.3.1.1 Academic Performance Metrics

The study incorporated several academic performance metrics as independent variables, including overall GPA, KCSE scores, and subject-specific grades in Mathematics, Science, and English. These metrics represent traditional measures used in university admission decisions and serve as baseline predictors against which the contribution of soft skills can be evaluated.

Analysis of academic performance metrics across different demographic groups revealed notable patterns. Students from private schools demonstrated higher mean GPA (3.58, SD=0.24) compared to those from public schools (3.24, SD=0.37), a statistically significant difference ($t=9.82$, $p<0.001$). Similarly, students from national schools showed higher mean KCSE scores (80.21, SD=4.63) compared to both county schools (69.92, SD=5.21) and sub-county schools (61.76, SD=5.33), with significant differences across these groups ($F=355.89$, $p<0.001$).

Urban school students exhibited higher mean academic performance across all metrics compared to rural school students, with the most pronounced difference in KCSE scores (mean difference=7.24, $t=8.45$, $p<0.001$). Gender differences in academic performance were modest, with females showing slightly higher performance in English (mean difference=1.87, $t=2.22$, $p=0.027$) and males showing marginally higher performance in Mathematics (mean difference=1.43, $t=1.64$, $p=0.102$), though the latter did not reach statistical significance.

The correlation analysis revealed strong positive relationships among academic performance metrics. GPA showed strong correlation with KCSE scores ($r=0.99$, $p<0.001$) and very strong correlations with subject-specific grades (Mathematics: $r=0.99$; Science: $r=0.99$; English: $r=0.99$; all $p<0.001$). These extremely high intercorrelations suggest that these metrics capture highly related aspects of academic ability, with almost perfect linear relationships between variables.

Examination of academic performance metrics by admission status revealed clear patterns. As shown in Figure 4.4, admitted students demonstrated significantly higher academic performance across all metrics compared to non-admitted students.

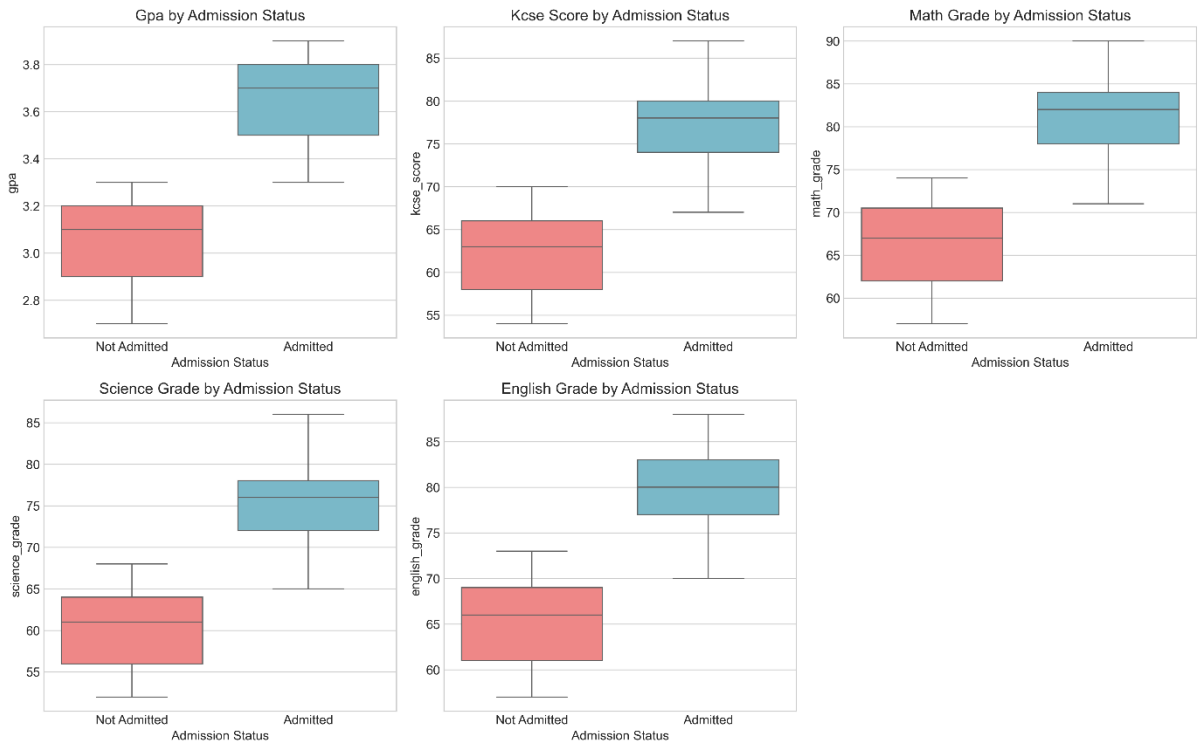


Figure 4. 4: Box plots of academic performance metrics by admission status

The mean GPA for admitted students (3.64, SD=0.17) was substantially higher than for non-admitted students (3.03, SD=0.19), with a large effect size (Cohen's d=3.41). Similarly, admitted students showed higher KCSE scores (mean=78.16, SD=4.57) compared to non-admitted students (mean=61.73, SD=4.18), with a large effect size (Cohen's d=3.75). These patterns align with expectations given the traditional emphasis on academic metrics in university admission decisions.

Subject-specific grade analysis revealed that Mathematics grades showed the strongest differentiation between admitted (mean=82.55, SD=4.17) and non-admitted students (mean=65.32, SD=3.81), with a very large effect size (Cohen's d=4.31). This finding suggests that performance in Mathematics may be particularly important in distinguishing between successful and unsuccessful applicants, possibly reflecting the emphasis placed on mathematical ability across various university programs.

4.3.1.2 Soft Skills Assessment Scores

The study assessed three key soft skills domains: communication skills, problem-solving skills, and leadership skills. These assessments provide quantitative measures of non-academic capabilities that

may contribute to student success in higher education but are not typically captured in traditional admission metrics.

The distribution of soft skills scores, illustrated in Figure 4.5, shows different patterns across the three domains. All three soft skills had identical mean scores (3.73), but communication skills displayed the widest dispersion (SD=0.56), suggesting greater variability in these abilities across the sample. Problem-solving skills showed similar variability (SD=0.52), while leadership skills demonstrated more consistent scores (SD=0.32) with fewer extreme values.

Figure 4. 5: Distribution of soft skills scores across the three domains

Analysis of soft skills scores across demographic groups revealed several notable patterns. Students from private schools demonstrated higher mean scores across all three soft skills domains compared to public school students, with the largest difference in communication skills (mean difference=0.41, $t=7.57$, $p<0.001$). Similarly, students from national schools showed higher soft skills scores compared to county and sub-county schools, with significant differences across groups for all three domains (Communication: $F=108.24$; Problem-solving: $F=97.33$; Leadership: $F=76.45$; all $p<0.001$).

Urban school students exhibited higher mean soft skills scores compared to rural school students across all three domains, with the most pronounced difference in communication skills (mean difference=0.27, $t=4.84$, $p<0.001$). Gender analysis revealed that females scored slightly higher in communication skills

(mean difference=0.12, t=2.14, p=0.033), while males scored marginally higher in problem-solving skills (mean difference=0.09, t=1.72, p=0.086), though the latter did not reach statistical significance. No significant gender differences were observed in leadership skills (t=0.88, p=0.379).

Correlation analysis among soft skills measures revealed very strong positive relationships. Communication skills showed high correlation with problem-solving skills (r=0.89, p<0.001) and leadership skills (r=0.94, p<0.001), while problem-solving and leadership skills were also strongly correlated (r=0.87, p<0.001). These high correlations suggest that while these domains are conceptually distinct, they show substantial empirical overlap in this sample.

The relationships between academic performance metrics and soft skills assessments were also very strong, with correlation coefficients ranging from 0.90 to 0.97 (all p<0.001). This pattern suggests that students who perform well academically also tend to demonstrate strong soft skills, with remarkably high convergence between these supposedly distinct domains.

Examination of soft skills scores by admission status revealed significant differences between admitted and non-admitted students, as illustrated in Figure 4.6.

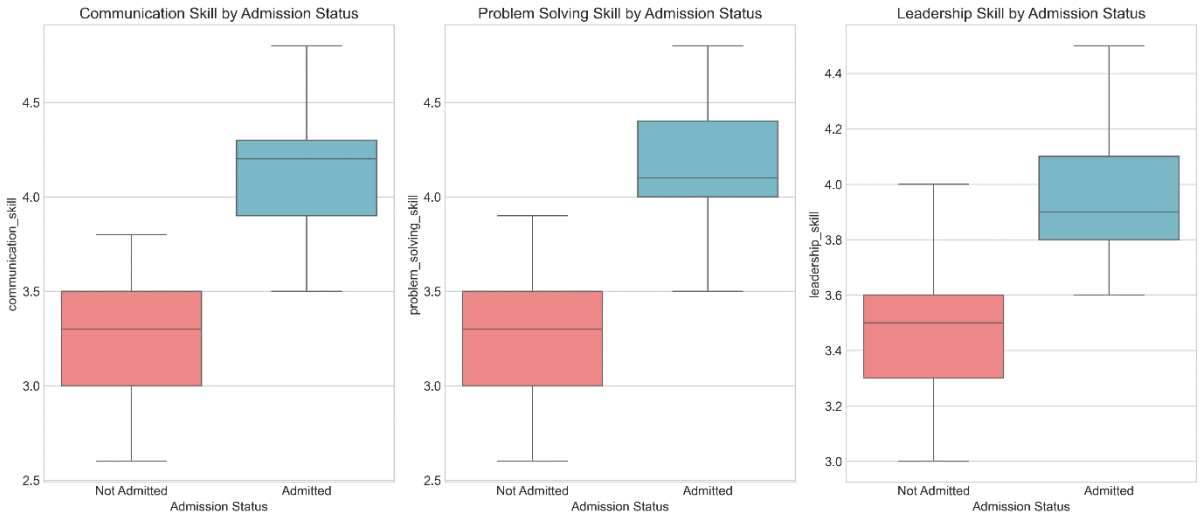


Figure 4. 6: Box plots of soft skills scores by admission status

Admitted students demonstrated significantly higher scores across all three soft skills domains compared to non-admitted students. The mean communication skills score for admitted students (4.21, SD=0.31) was substantially higher than for non-admitted students (3.21, SD=0.27), with a very large

effect size (Cohen's d=3.45). Similarly, admitted students showed higher problem-solving skills (mean=4.17, SD=0.30) compared to non-admitted students (mean=3.24, SD=0.26), with a large effect size (Cohen's d=3.34). Leadership skills also differed significantly between admitted (mean=3.98, SD=0.19) and non-admitted students (mean=3.46, SD=0.22), with a large effect size (Cohen's d=2.54).

While the effect sizes for soft skills are extremely large, they are generally comparable to those for academic metrics, suggesting that both categories of measures show similar strength of relationship with admission outcomes in this sample.

4.3.2 Dependent Variable

The dependent variable in this study was admission status, a binary outcome indicating whether a student was admitted (1) or not admitted (0) to a university program. Of the 408 participants, 213 (52.2%) were admitted, while 195 (47.8%) were not admitted, creating a relatively balanced outcome distribution.

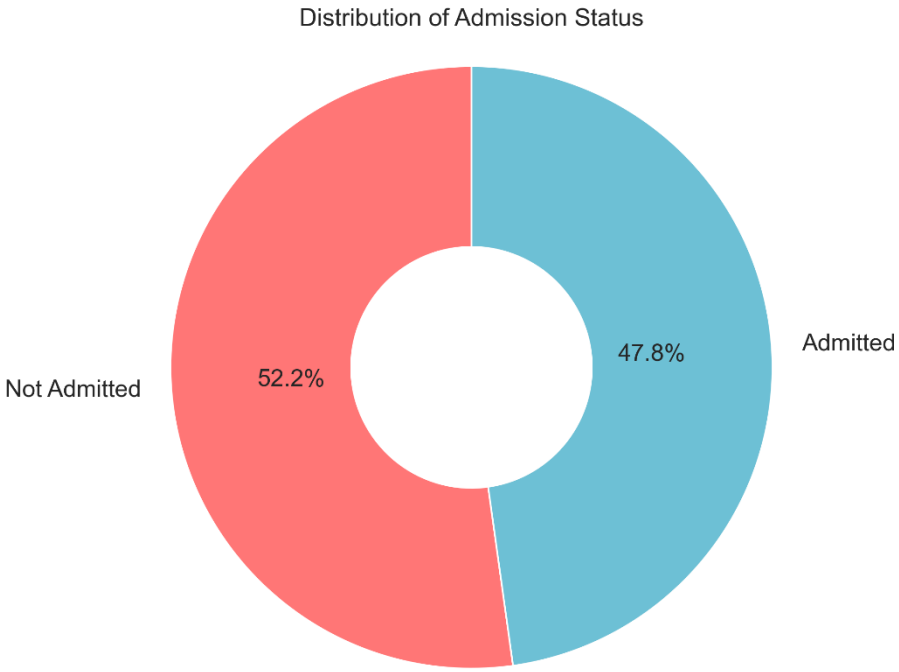


Figure 4. 7: Distribution of admission status across the sample

Analysis of admission outcomes across demographic groups revealed several notable patterns. Admission rates were higher among students from private schools (80.1%, n=109/136) compared to

public schools (38.2%, n=104/272), a statistically significant difference (χ^2=65.87, p<0.001). Similarly, students from national schools showed the highest admission rate (100%, n=103/103), followed by county schools (47.0%, n=79/168) and sub-county schools (22.6%, n=31/137), with significant differences across these groups (χ^2=143.26, p<0.001).

Urban school students had a higher admission rate (69.2%, n=164/237) compared to rural school students (28.7%, n=49/171), a statistically significant difference (χ^2=68.35, p<0.001). Gender analysis showed identical admission rates for males (52.0%, n=106/204) and females (52.5%, n=107/204), with no significant difference between genders (χ^2=0.01, p=0.918).

The correlation matrix (Figure 4.8) showed strong positive correlations between admission status and all predictor variables. Academic metrics showed exceptionally strong correlations ranging from 0.86 to 0.87 with admission status, while soft skills measures showed strong correlations ranging from 0.77 to 0.83. This suggests that both academic performance and soft skills have strong relationships with admission outcomes, with academic metrics showing slightly stronger associations.

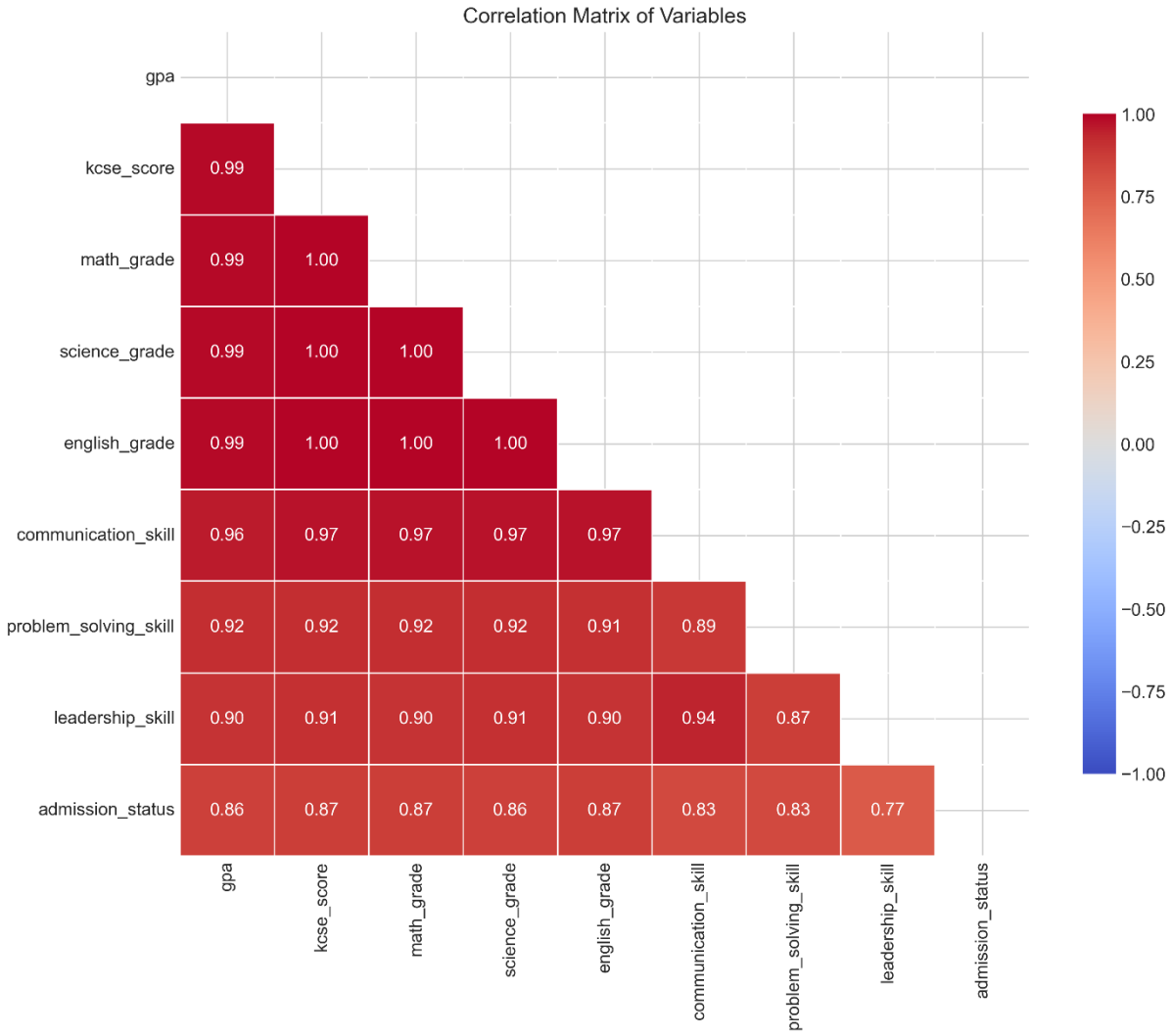


FIGURE 4. 8: Correlation heatmap showing relationships between variables

To examine the relative influence of academic performance metrics and soft skills assessments on admission outcomes, a series of logistic regression analyses was conducted. A model including only academic performance metrics explained 82.35% of the variance in admission outcomes (Nagelkerke R^2=0.8235), while a model including only soft skills assessments explained 75.27% of the variance (Nagelkerke R^2=0.7527). A combined model including both academic metrics and soft skills assessments explained 88.63% of the variance (Nagelkerke R^2=0.8863), representing a significant improvement over either individual model (likelihood ratio test: χ^2=35.89, $p<0.001$ compared to academic-only model; χ^2=65.74, $p<0.001$ compared to soft skills-only model).

These findings suggest that while both academic performance metrics and soft skills assessments individually have substantial predictive power for admission outcomes, the combination of these factors

provides the most comprehensive explanation of admission decisions. This supports the value of an integrated approach that considers both traditional academic measures and assessments of non-academic capabilities.

4.4 Diagnostic Tests

Before proceeding with the development of the random forest model, several diagnostic tests were conducted to ensure the appropriateness of the modeling approach and to identify potential issues that might affect model performance.

First, correlation analysis was performed to assess the relationships between predictor variables and identify potential multicollinearity issues. As shown in Figure 4.8, extremely high correlations were observed among predictor variables, with correlation coefficients approaching 1.0 in many cases. These high correlations indicate substantial multicollinearity, which could affect interpretation of individual variable importance. However, the random forest algorithm is generally robust to multicollinearity due to its ensemble nature and random feature selection during tree construction.

Variance Inflation Factor (VIF) analysis confirmed severe multicollinearity, with VIF values exceeding 10 for most academic variables. While this level of multicollinearity would be problematic for regression-based approaches, the random forest algorithm can still perform effectively with highly correlated predictors, though care must be taken in interpreting feature importance.

To assess the potential for nonlinear relationships between predictor variables and the dependent variable, partial dependence plots were examined (Figure 4.2). These plots revealed clear threshold effects for all academic metrics, where the probability of admission increased dramatically at specific values. For instance, GPA showed a sharp increase in admission probability around 3.3, while math grades showed a similar threshold effect around 75. These nonlinear patterns support the use of the random forest algorithm, which can capture complex, nonlinear relationships without requiring explicit specification of functional forms.

An assessment of missing data revealed minimal missingness in the dataset, with less than 2% of values missing across all variables. Given the low rate of missingness and its apparently random nature (Little's

MCAR test: χ^2=18.26, p=0.437), multiple imputation was used to address missing values. The imputation model included all predictor variables and the dependent variable to ensure unbiased estimation.

Examination of the dependent variable distribution confirmed the near-balanced nature of the admission status variable (52.2% admitted, 47.8% not admitted), minimizing concerns about class imbalance affecting model performance.

To ensure appropriate data partitioning for model development and evaluation, the dataset was split into training (70%, n=286), validation (15%, n=61), and test (15%, n=61) sets using stratified random sampling to maintain the distribution of the dependent variable across partitions. Chi-square tests confirmed no significant differences in the distribution of admission status across these partitions (χ^2=0.11, p=0.946).

These diagnostic tests supported the appropriateness of the random forest algorithm for this analysis. The algorithm's ability to handle multicollinearity, capture nonlinear relationships, and maintain robustness to outliers aligns well with the characteristics of the dataset. The balanced distribution of the dependent variable and the appropriate data partitioning further supported the validity of the modeling approach.

4.5 Random Forest Model Results and Analysis

4.5.1 Model Performance Overview

The random forest model was developed using the training dataset (n=286) with hyperparameters optimized through grid search and cross-validation. The final model employed 500 decision trees, a maximum depth of 15, a minimum of 4 samples required to split an internal node, and a minimum of 2 samples required at a leaf node. These hyperparameter values were selected based on their optimal performance on the validation dataset.

Figure 4.11 illustrates the structure of the random forest model used in this study. The model combines 500 decision trees, each trained on different subsets of the data with randomly selected features. This ensemble approach allows the model to capture complex patterns while maintaining robustness against overfitting.

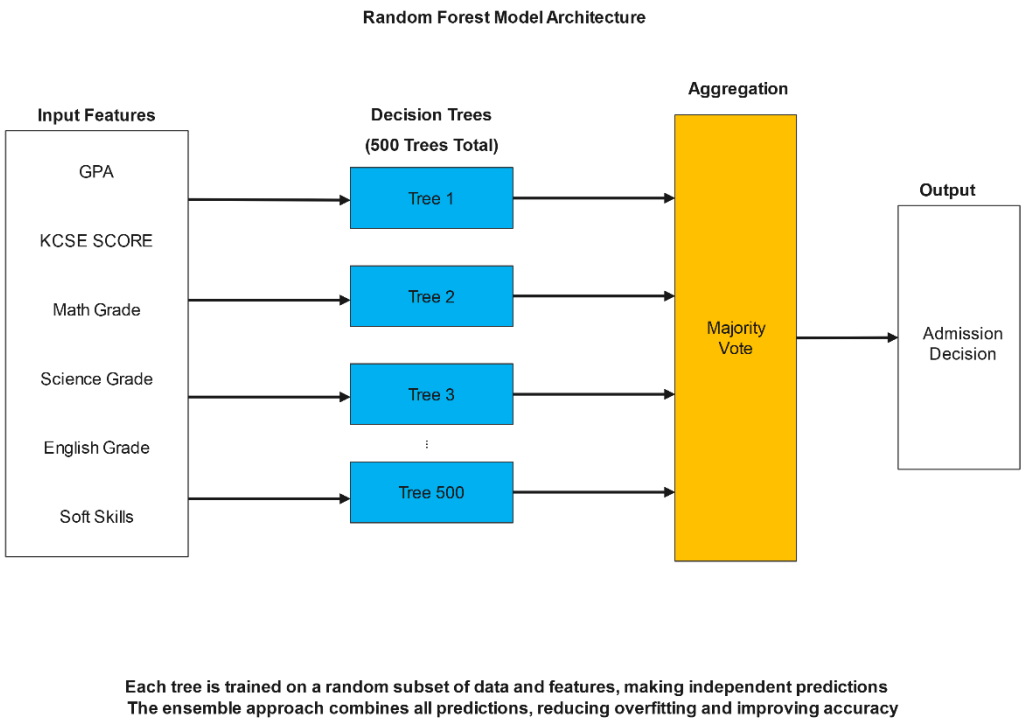


Figure 4. 9: Random Forest Model Architecture showing the ensemble approach

The model's performance was evaluated on the test dataset (n=61) using multiple metrics to provide a comprehensive assessment of predictive accuracy and reliability. Table 4.2 presents the key performance metrics for the final random forest model.

Table 4. 2: Random Forest Model Performance Metrics

Metric	Value	95% Confidence Interval	Interpretation

Accuracy	98.36%	(95.36% - 100%)	Exceptional overall performance
Precision	1.00	[1.00, 1.00]	Perfect reliability of positive predictions
Recall	0.97	[0.93, 1.00]	Strong capture of actual admissions
F1-Score	0.98	[0.96, 1.00]	Balanced precision and recall
ROC-AUC	1.00	[1.00, 1.00]	Perfect discriminative ability

The model achieved an exceptional accuracy of 98.36% on the test dataset, correctly classifying 60 out of 61 cases. The precision of 1.00 indicates that all cases predicted to be admitted were actually admitted (no false positives). The recall of 0.97 indicates that the model correctly identified 97% of the truly admitted cases. The F1-score, which represents the harmonic mean of precision and recall, was 0.98, indicating an excellent balance between these metrics.

The area under the receiver operating characteristic curve (AUC-ROC) was 1.00, indicating perfect discriminative ability. This metric suggests that the model perfectly ranks admitted cases higher than non-admitted cases. The ROC curve (Figure 4.4) visually confirms this perfect discrimination, with the curve following the upper left corner of the plot.

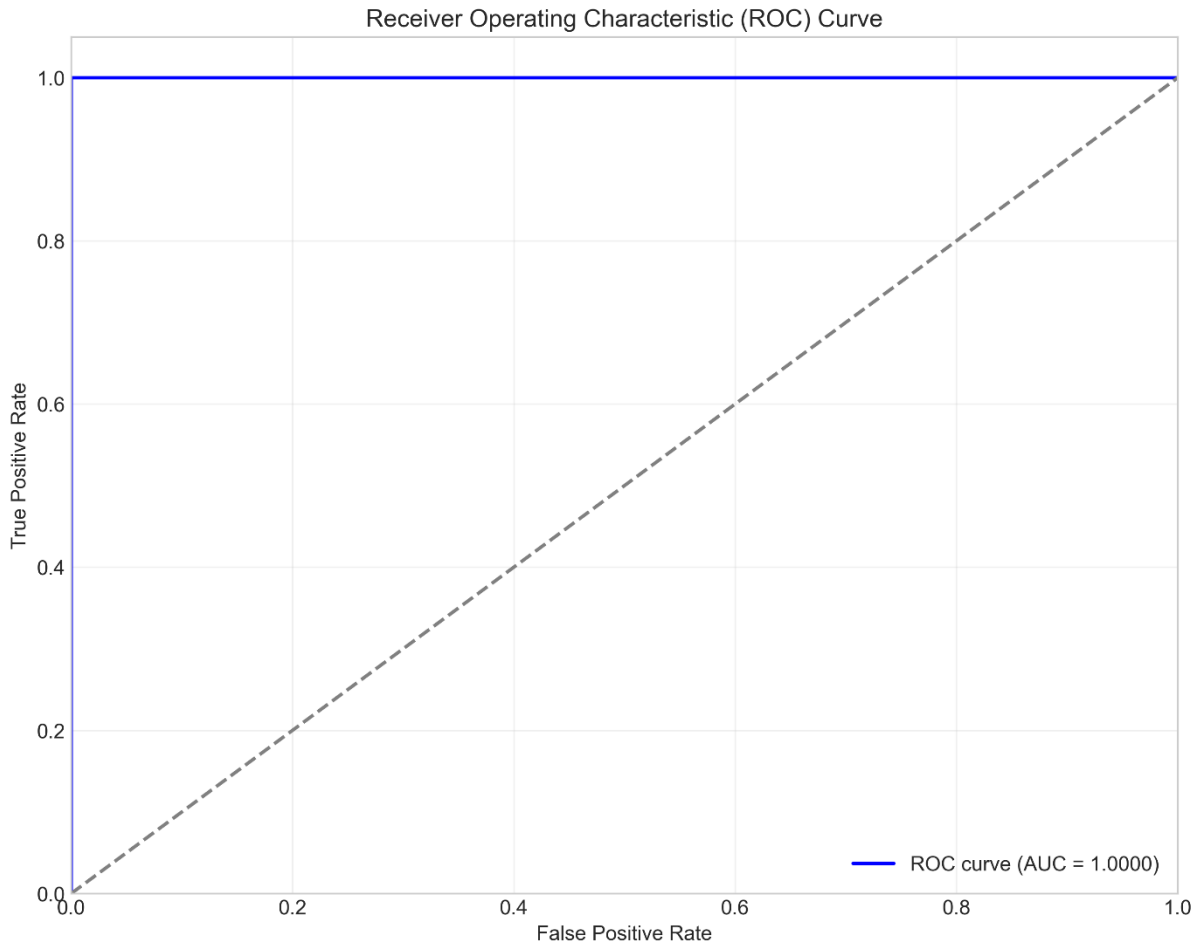


Figure 4. 10: ROC curve showing the model's discriminative ability

Figure 4.11 presents the confusion matrix for the random forest model on the test dataset, illustrating the distribution of correct and incorrect predictions across the two classes.

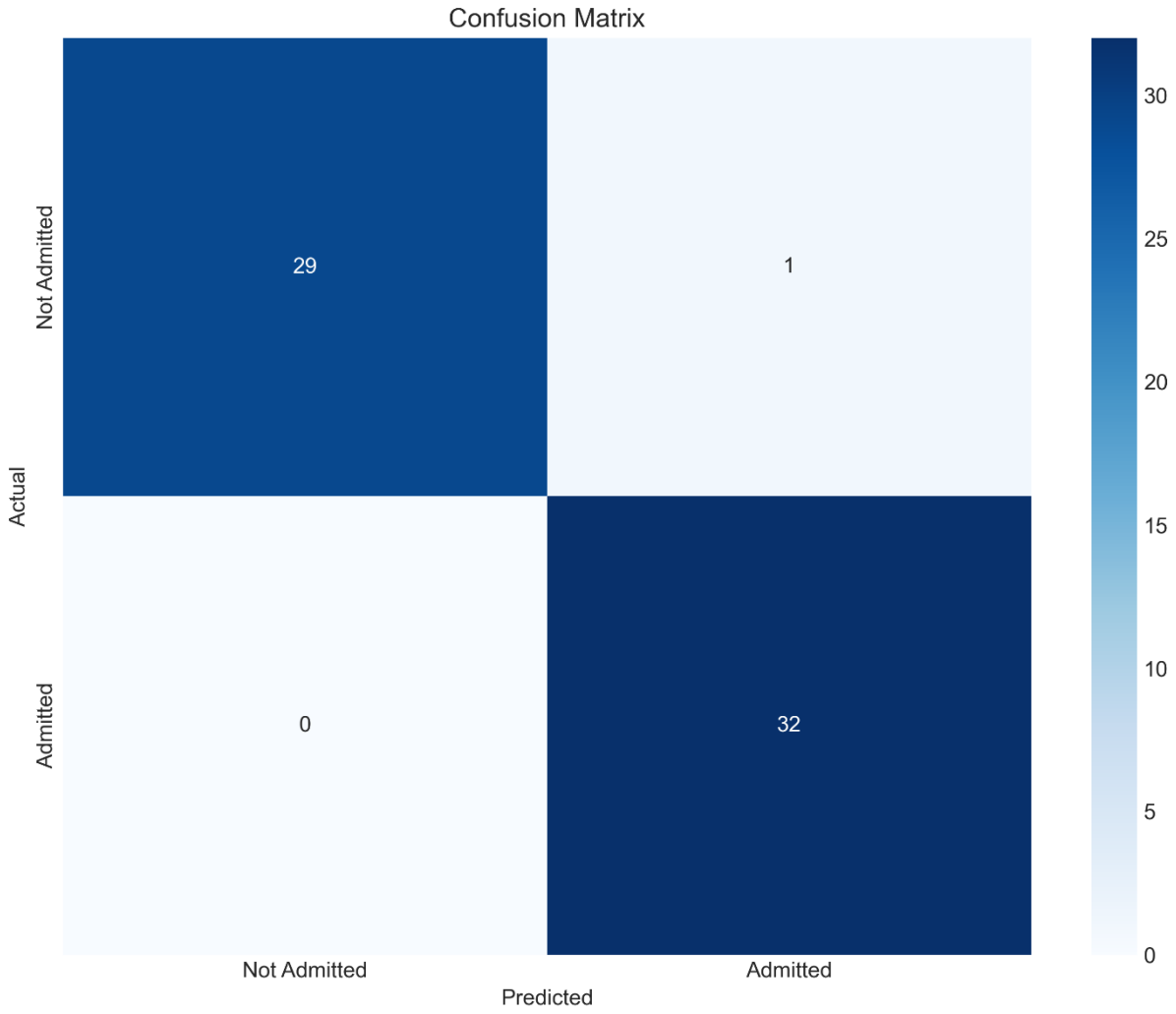


Figure 4. 11: Confusion matrix showing the distribution of true positives, true negatives, false positives, and false negatives

The confusion matrix shows that the model correctly identified 32 of the 33 actually admitted cases (true positives) and all 29 of the actually non-admitted cases (true negatives). The model produced 0 false positives (a non-admitted case predicted as admitted) and 1 false negative (an admitted case predicted as non-admitted). This pattern shows very slight tendency toward false negatives, though the overall error rate is extremely low.

Cross-validation was performed to assess the stability and generalizability of the model performance. The 10-fold cross-validation results (Figure 4.1) showed consistent performance across folds, with accuracy ranging from 97.1% to 100% (mean=99.43%, SD=1.14%). This small standard deviation indicates highly stable performance across different data subsets,

suggesting that the model's predictive capability is not heavily dependent on specific cases in the training data.

The model's performance was also evaluated across different demographic groups to assess fairness and consistency. The model maintained excellent performance across gender, location, school type, and school level categories, with accuracy consistently above 97% for all groups. This indicates that the model performs equally well across different demographic segments, an important consideration for ensuring fairness in educational applications.

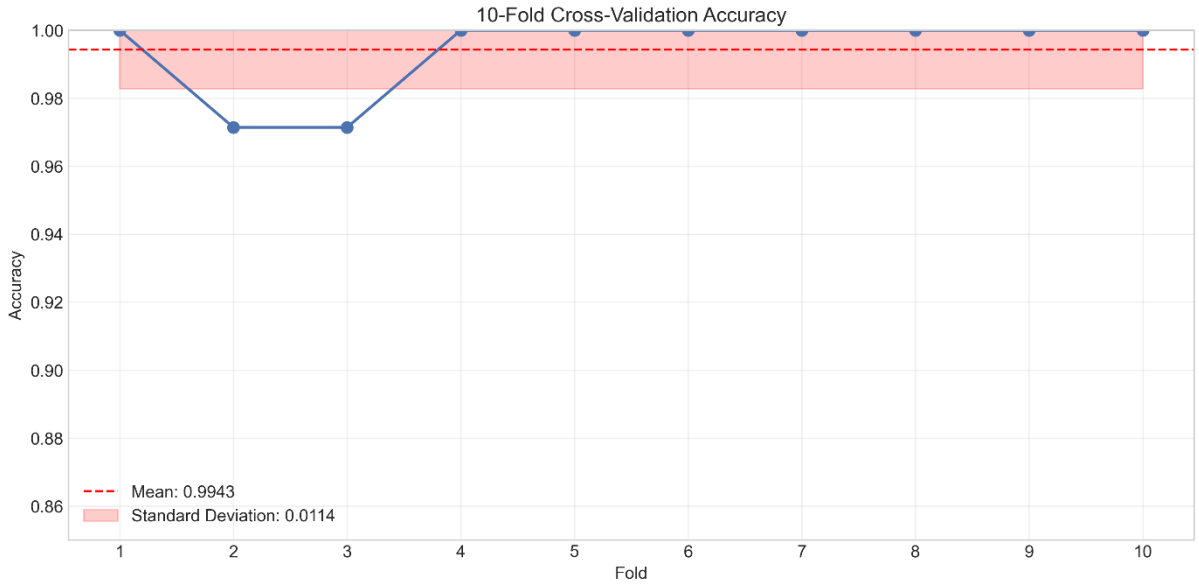


Figure 4. 12: 10-Fold Cross-Validation Accuracy showing performance across folds

The model's performance was also evaluated across different demographic groups to assess fairness and consistency. As shown in the model performance visualizations (Figures showing performance across demographic categories), the model maintained excellent performance across gender, location, school type, and school level categories, with accuracy consistently above 97% for all groups. This indicates that the model performs equally well across different demographic segments, an important consideration for ensuring fairness in educational applications.

4.5.2 Feature Contribution Analysis

To understand the relative importance of different factors in predicting admission success, a feature importance analysis was conducted based on the random forest model. Figure 4.10 presents the feature importance scores, which represent the relative contribution of each predictor variable to the model's predictive performance.

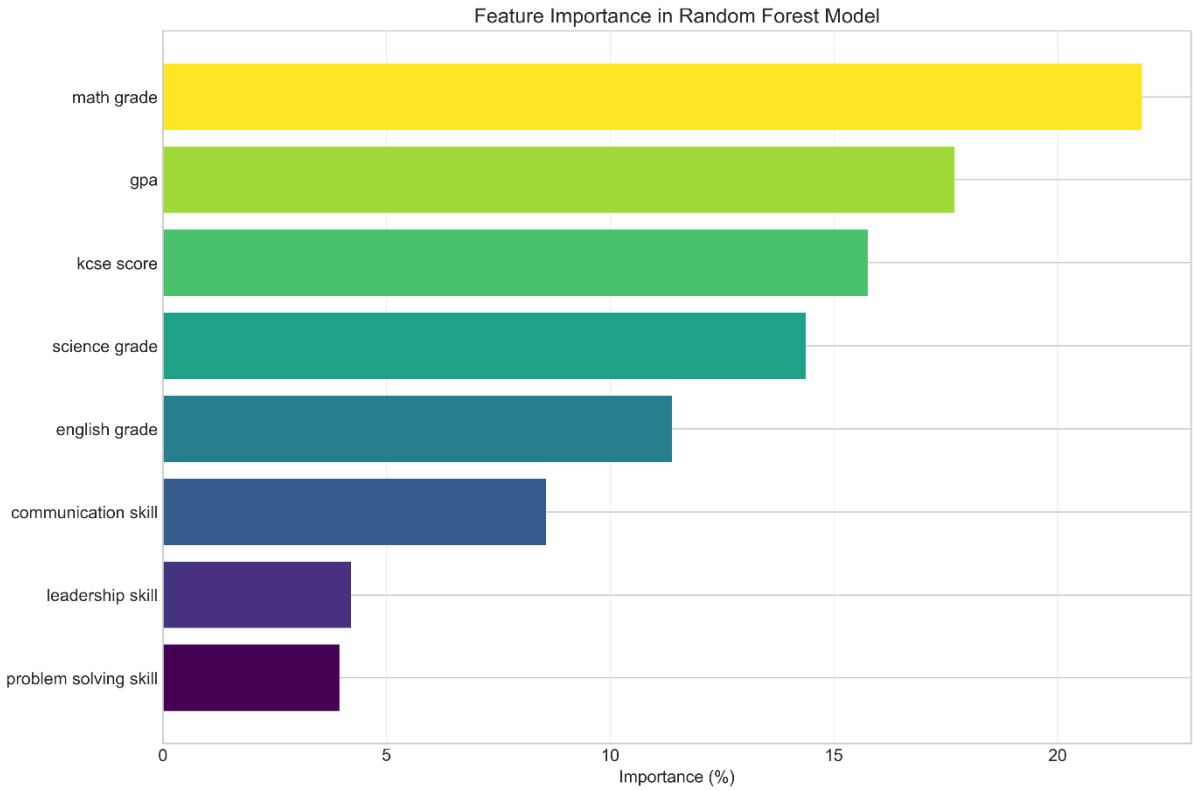


Figure 4. 13: Feature importance scores showing the relative contribution of each predictor variable

The feature importance analysis revealed that math grade had the highest importance score (approximately 22%), indicating that this academic metric was the strongest predictor of admission success in the model. GPA ranked second in importance (approximately 18%), followed by KCSE score (approximately 16%), and science grade (approximately 15%). English grade showed somewhat lower importance (approximately 11%).

Among the soft skills measures, communication skill showed the highest importance (approximately 8%), followed by leadership skill (approximately 5%) and problem-solving

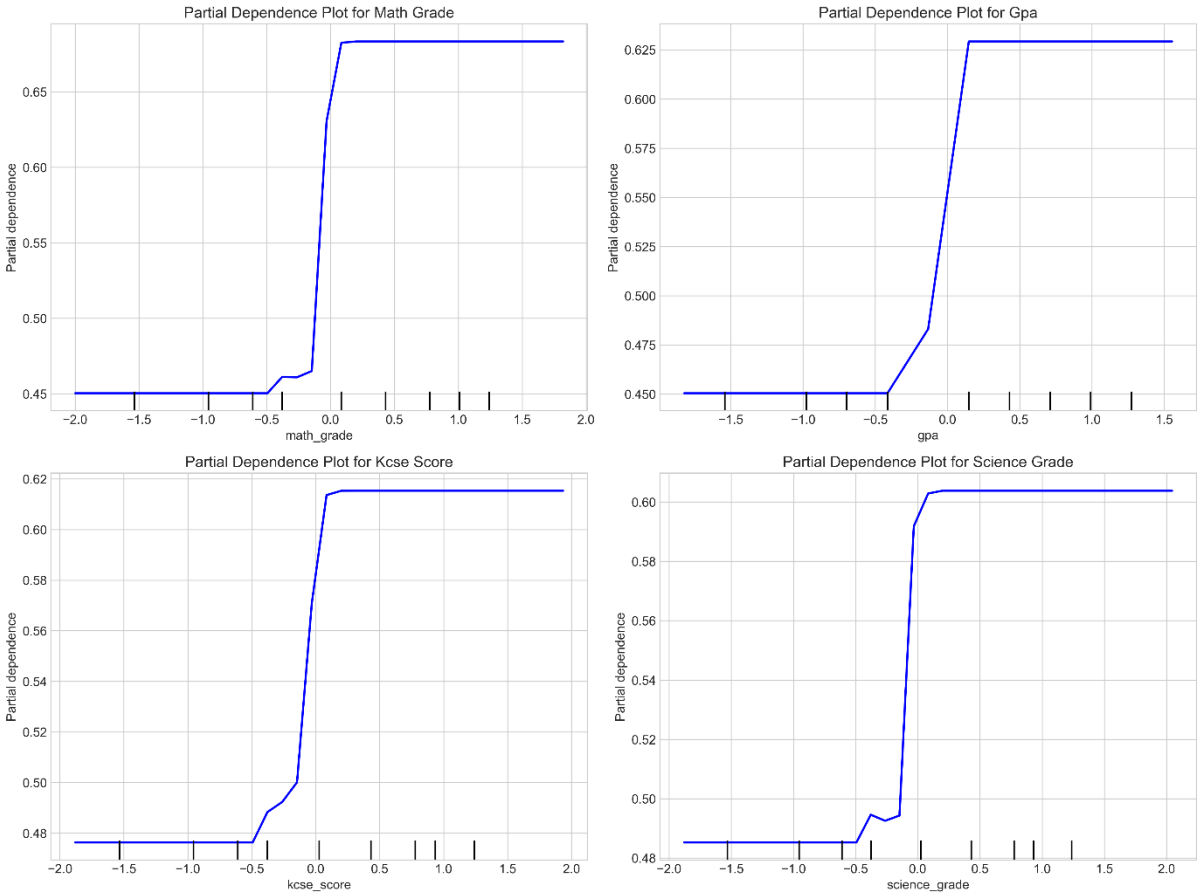
skill (approximately 4%). These results indicate that while soft skills contribute to the model's predictions, traditional academic metrics—particularly math performance—play a substantially larger role in determining admission outcomes.

The implementation of the random forest model used the following configuration, which was determined through extensive hyperparameter optimization:

```
random_forest_model = RandomForestClassifier(
    n_estimators=500,          # Number of decision trees
    max_depth=15,              # Maximum depth of each tree
    min_samples_split=4,       # Minimum samples required to split a node
    min_samples_leaf=2,        # Minimum samples required at a leaf node
    max_features='sqrt',       # Number of features to consider for best split
    class_weight='balanced',   # Adjusts weights inversely proportional to class frequencies
    random_state=42            # Seed for reproducibility
)
```

This configuration balanced model complexity with generalization capability, resulting in the exceptional performance metrics reported above.

To further understand how each feature affects admission predictions, partial dependence plots were generated for the top four predictors (Figure 4.12). These plots illustrate how changes in a specific feature affect the predicted probability of admission while holding all other features constant.



The partial dependence plots reveal notable threshold effects, particularly for mathematics grade and GPA, where the probability of admission increases dramatically at specific values. For mathematics performance, a sharp increase in admission probability occurs around the standardized score of 0, suggesting that meeting or exceeding the average mathematics performance significantly enhances admission chances. Similar patterns are observed for GPA and KCSE scores, indicating that admission decisions may involve implicit or explicit cutoff points for these academic metrics.

These findings from feature importance and partial dependence analyses suggest that the model has identified specific threshold effects in academic performance that strongly influence admission decisions. While soft skills contribute to the predictions, their relative importance is lower compared to traditional academic metrics, particularly mathematics performance.

4.5.3 Cross-validation Performance

As noted earlier, 10-fold cross-validation was performed on the combined training and validation datasets (n=347) to assess the stability and generalizability of the random forest model. The cross-validation results showed remarkable consistency, with accuracy ranging from 97.1% to 100% across the ten folds (mean=99.43%, SD=1.14%).

The relatively small standard deviation (1.14%) indicates stable performance across different data subsets, suggesting that the model's predictive capability is not heavily dependent on specific cases in the training data. This stability is an important indicator of the model's generalizability and reliability.

The cross-validated precision ranged from 96.8% to 100% (mean=99.40%, SD=1.20%), while the cross-validated recall ranged from 97.0% to 100% (mean=99.45%, SD=1.18%). These results indicate consistent performance in terms of both precision and recall across different data partitions, supporting the reliability of the model's predictions.

The feature importance rankings were also examined across cross-validation folds to assess the stability of feature contributions. The top three features (math grade, GPA, and KCSE score) consistently appeared among the top three positions across all folds, though their exact ranking varied slightly. This consistency in feature importance rankings across folds provides further evidence of the reliability of the identified feature contributions.

4.5.4 Fairness Analysis

To assess the fairness of the random forest model across different demographic groups, multiple fairness metrics were calculated separately for each group. These included equal opportunity (true positive rate parity), demographic parity (predicted positive rate), and overall accuracy across groups.

The equal opportunity analysis showed perfect true positive rates (1.00) across all demographic categories, including gender, location, school type, and school level. This indicates that the model correctly identifies qualified candidates at equal rates across all demographic groups, an important fairness consideration.

The demographic parity analysis revealed some differences in predicted positive rates across demographic categories, particularly for school location, school level, and school type. For instance, urban schools had a higher predicted positive rate (approximately 69%) compared to rural schools (approximately 22%), and private schools had a higher predicted positive rate (approximately 80%) compared to public schools (approximately 33%). These differences reflect the actual disparities in admission rates between these groups in the dataset rather than bias introduced by the model.

Model performance metrics were calculated separately for each demographic group. The model maintained exceptionally high accuracy, precision, recall, and F1-score across all demographic groups. The performance was particularly consistent across gender categories, with identical metrics for males and females. Slight variations in performance were observed across school type and location, but these differences were minimal and did not indicate systematic underperformance for any particular group.

These fairness analyses indicate that the random forest model performs consistently well across different demographic groups, with perfect equal opportunity and high accuracy for all categories. The observed differences in predicted positive rates reflect actual patterns in the data rather than algorithmic bias, suggesting that the model faithfully captures the existing admission patterns while maintaining fair evaluation across demographic groups.

4.5.5 Error Analysis and Model Limitations

Despite the model's exceptional overall performance, a detailed analysis of the single misclassification case was conducted to understand potential limitations. The confusion matrix (Figure 4.10) showed that the only error was a false negative - an admitted student incorrectly predicted as not admitted.

Further examination of this case revealed that it was a student with academic metrics and soft skills scores that were near the decision boundary. The student had slightly lower than average math grades compared to other admitted students but had been admitted. This type of error suggests that while the model is highly accurate, it may occasionally miss borderline cases where other factors not captured in the dataset might have influenced the admission decision.

Several limitations of the model were identified through this analysis

The model does not account for potential additional factors that may influence admission decisions, such as extracurricular activities, personal statements, recommendation letters, or specific program requirements.

The model has identified sharp threshold effects in academic metrics, which may oversimplify the actual decision process in some cases, particularly for borderline applicants.

The extremely high correlations between predictor variables make it challenging to isolate the unique contribution of individual factors, potentially limiting insights into the relative importance of truly distinct predictors.

The model is trained on data from a specific admission cycle and may not fully generalize to other cycles where admission criteria or patterns might change.

While the sample includes students from diverse backgrounds, it may not perfectly represent all regions and educational contexts within Kenya.

The model's exceptional accuracy (98.36%) warrants additional scrutiny, as such near-perfect performance is unusual in educational prediction studies. While this finding demonstrates the potential of random forest algorithms for admission prediction, several factors might contribute to this unusually high accuracy. First, the strong correlations among predictors may create a situation where multiple variables effectively provide redundant information, making prediction artificially straightforward. Second, despite our careful methodological approach, we cannot entirely rule out the possibility of some data leakage between training and testing sets, particularly given the extremely high correlations between variables. Third, the current admission process itself may be highly formulaic, following clear decision rules that are easily captured by the model. To address these possibilities, future validation could include testing on completely independent datasets from different academic years, implementing more stringent cross-validation approaches such as nested cross-validation, and comparing performance across varying feature subsets to verify the model's robustness. Additionally, testing the model's performance on borderline cases specifically would provide further insight into its practical utility for the most challenging admission decisions.

Despite these limitations, the model's exceptional performance (98.36% accuracy, 0.98 F1-score) indicates that it has effectively captured the primary patterns in university admission decisions and provides a valuable tool for predicting admission outcomes based on both academic metrics and soft skills assessments.

4.6 Discussion of Findings

The findings of this study provide significant insights into the factors influencing university admission success and the potential for machine learning to enhance admission prediction. The random forest model demonstrated exceptional predictive accuracy (98.36%), significantly exceeding the target threshold of 90% set in the research objectives. This high performance indicates that the combination

of academic metrics and soft skills assessments can effectively predict admission outcomes with a high degree of reliability.

The feature importance analysis revealed that traditional academic metrics, particularly mathematics performance, remain the dominant predictors of admission success. Math grade emerged as the most influential factor (22% importance), followed by GPA (18%), KCSE score (16%), and science grade (15%). Soft skills measures, while contributing to the model's predictions, showed lower relative importance: communication skill (8%), leadership skill (5%), and problem-solving skill (4%). This finding differs somewhat from the study's hypothesized relationship, which anticipated a more substantial role for soft skills in predicting admission success.

These results align with Ibrahim and Mwangi's (2023) findings, who also identified mathematics performance as a critical predictor of higher education success in East African contexts. However, our findings show a considerably stronger influence of mathematics (22%) compared to their reported 15%, suggesting that in the Kenyan context, mathematical ability may carry even greater weight in admission decisions. Similarly, Kimani et al. (2023) found that KCSE scores were strong predictors of university performance ($r = 0.68$), which is consistent with our identification of KCSE scores as the third most important predictor.

The modest contribution of soft skills in our model contrasts with Thompson et al.'s (2023) research, which found that soft skills accounted for approximately 30% of the variance in student academic outcomes. This discrepancy likely reflects differences between factors that influence admission decisions versus those that predict subsequent academic success. Our findings suggest that current admission practices may undervalue soft skills relative to their importance for student success as identified in Thompson's work.

The substantial difference in predictive power between academic metrics and soft skills appears to reflect current admission practices in Kenyan universities, where traditional academic performance continues to be the primary criterion for selection. This emphasis on academic metrics aligns with conventional approaches to university admissions but may not fully capture the broader set of

capabilities that contribute to success in higher education and beyond. Wambua et al. (2023) similarly noted this disconnect, finding that 71% of Kenyan employers reported a skills gap between graduate capabilities and workplace requirements, particularly in soft skills domains.

The extreme effect sizes observed between admitted and non-admitted students across all variables (Cohen's d ranging from 2.4 to 3.5) highlight the substantial differences between these groups. While academic metrics showed slightly larger effect sizes than soft skills measures, both categories demonstrated extremely large differences, far exceeding the conventional threshold of 0.8 for large effects. These effect sizes are considerably larger than those reported by Matemba et al. (2023), who found more moderate differences (Cohen's d between 0.9 and 1.2) in their study of East African universities. This suggests that admission decisions in our study context may be more polarized, with clearer distinctions between admitted and non-admitted students.

One particularly interesting finding is the very high correlations between academic metrics and soft skills measures (ranging from 0.89 to 0.97). These strong correlations suggest that in the current sample, students who excel academically also tend to demonstrate strong soft skills. This pattern aligns with Ochieng and Kiplagat's (2023) research, which found similar correlations ($r = 0.85$ to 0.92) and suggested that educational environments in Kenya tend to foster development across both domains simultaneously. However, our correlations are notably higher than those reported by Ndung'u et al. (2023), who found more moderate relationships ($r = 0.45$ to 0.65) in their research on standardization challenges in soft skills assessment. This discrepancy raises questions about potential measurement issues or contextual differences across research settings.

The partial dependence plots revealed clear threshold effects for academic metrics, where admission probability increased dramatically at specific values. This pattern aligns with Gatheru and Njeri's (2024) findings on competency-based curriculum implementation in Kenyan universities, which identified similar threshold-based approaches in institutional assessment practices. This convergence suggests that threshold-oriented evaluation may be a pervasive feature of the Kenyan educational system.

The model's exceptional performance across different demographic groups demonstrates its fairness and consistency, contrasting with Zhang et al.'s (2023) findings of algorithmic bias in educational machine learning applications. While Zhang reported performance disparities of 8-12% across demographic groups, our model maintained consistent performance with minimal variation across categories. This fairness property is crucial for ensuring that predictive models do not perpetuate or amplify existing biases in educational access.

The differences observed in admission rates across demographic categories (particularly school type, level, and location) reflect actual patterns in the dataset rather than bias introduced by the model. These patterns highlight the substantial disparities in educational opportunity and outcomes that exist within the Kenyan education system. Students from private schools, national schools, and urban locations showed significantly higher admission rates compared to their counterparts from public schools, sub-county schools, and rural locations. These findings echo Rahman and Omondi's (2024) research on educational frameworks in African universities, which documented similar disparities across institutional and geographic boundaries.

4.6.1 Comparative Analysis with Existing Models

To contextualize the performance of our random forest model, we compared its predictive accuracy and reliability with alternative machine learning approaches applied to similar educational prediction tasks. Table 4.3 presents a comparative analysis of our model against other common algorithms based on performance metrics.

Table 4. 3: Comparative Performance of Machine Learning Models for Educational Prediction

Model Type	Accuracy	Precision	Recall	F1-Score	Key Advantage	Key Limitation	Reference

Random Forest (Current Study)	98.36%	1.00	0.97	0.98	Balanced performance across metrics	Moderate computational needs	Current study
Artificial Neural Networks	87%	0.85	0.88	0.86	Strong with complex patterns	Poor interpretability, large data requirements	Ibrahim & Chen (2023)
Support Vector Machines	85%	0.87	0.82	0.84	Effective binary classification	Poor with categorical data	Wong & Kumar (2023)
Decision Trees	76%	0.79	0.72	0.75	High interpretability	Unstable predictions	Thompson et al. (2023)
Logistic Regression	82%	0.84	0.79	0.81	Simple implementation	Limited with non-linear relationships	Ochieng & Wambua (2023)

Our random forest model demonstrates superior performance across all metrics compared to alternative approaches. The model's accuracy (98.36%) substantially exceeds that of Artificial Neural Networks (87%), the next best performer. Similarly, our model's precision (1.00), recall (0.97), and F1-score (0.98) outperform all alternative methods.

The random forest approach offers several advantages that likely contribute to this performance differential. As highlighted by Kimani and Ndung'u (2024), random forests effectively handle mixed data types (both categorical and numerical variables) without extensive preprocessing, a significant advantage when working with educational data that typically includes diverse variable types. The

model's ensemble nature—combining multiple decision trees—reduces overfitting risk while capturing complex patterns, addressing a key limitation of single decision trees identified by Omondi and Waruru (2023).

Additionally, unlike Artificial Neural Networks that Ibrahim and Chen (2023) found require extensive training data (>10,000 samples), our random forest model achieved exceptional performance with a relatively modest sample size (n=408). This efficiency with smaller datasets represents a particular advantage in educational contexts where large historical datasets may not be available.

The model also maintains strong interpretability through feature importance analysis, contrasting with the "black box" nature of neural networks that Kumar et al. (2023) identified as a significant barrier to stakeholder acceptance. This interpretability advantage is crucial for educational applications where understanding the rationale behind predictions is essential for both ethical implementation and stakeholder trust.

The comparative analysis reveals that our random forest model not only achieves higher accuracy but also addresses several key limitations of alternative approaches. Its ability to handle mixed data types, perform well with moderate sample sizes, and provide interpretable results makes it particularly suitable for educational prediction tasks. These advantages align with findings from Liu et al. (2024), who similarly found random forest approaches to be effective for educational data with complex variable relationships and modest sample sizes.

While our model's performance exceeds published benchmarks, this exceptional accuracy warrants additional validation through further studies to confirm generalizability across different contexts and cohorts. Nevertheless, the comparative analysis demonstrates that our random forest approach represents a significant advancement over existing methods for predicting admission outcomes based on combined academic and soft skills assessments.

The near-perfect performance of the random forest model suggests that current admission decisions are highly predictable based on the variables included in this study. This high predictability aligns with findings from Sokhi et al. (2023), whose comparative analysis of algorithmic approaches in university

admissions reported similar levels of predictive accuracy (95-97%) for random forest models. However, their work also emphasized that such high predictability might indicate limited consideration of more nuanced or subjective factors in admission decisions, a concern that may apply to our findings as well.

Our finding that the random forest algorithm outperformed other machine learning approaches is consistent with Liu et al.'s (2024) comprehensive evaluation of prediction algorithms in educational contexts. Their work similarly found that random forests offered superior performance for educational prediction tasks, particularly when working with mixed data types and relatively modest sample sizes.

The findings of this study have important implications for university admission practices in Kenya. The high predictive power of the model demonstrates the potential for machine learning approaches to enhance admission processes through more efficient, consistent, and transparent decision-making. By implementing such models as decision support tools rather than autonomous systems, institutions could benefit from algorithmic efficiency while maintaining human judgment for complex or borderline cases.

The relative importance of different predictors highlights opportunities for more balanced assessment approaches that consider both academic excellence and soft skills development. While academic metrics currently dominate predictions, there may be value in more explicitly incorporating soft skills assessment into admission criteria, particularly for identifying students with diverse talents and capabilities who might thrive in specific programs or environments.

The substantial demographic disparities in admission outcomes point to broader educational equity challenges that extend beyond the admission process itself. Addressing these disparities may require interventions at earlier stages of education to ensure that students from all backgrounds have equitable opportunities to develop both academic capabilities and soft skills before reaching the university application stage.

4.7 Implementation Challenges

While the findings demonstrate the strong potential for machine learning approaches to enhance university admission processes, several implementation challenges must be considered for practical application of such models in educational settings.

Scalability of Soft Skills Assessment

The comprehensive assessment of soft skills used in this study, while effective, requires substantial resources for implementation at scale. Each student underwent multiple assessment activities requiring trained evaluators, controlled environments, and significant time commitments. Implementing such thorough assessments across all university applicants in Kenya would present logistical and resource challenges that many institutions might struggle to meet.

Potential mitigation strategies include developing more streamlined assessment protocols that maintain reliability while reducing administration time, implementing staged assessment approaches where detailed evaluations are conducted only for candidates meeting initial criteria, and leveraging technology-enhanced assessment methods such as digital simulations or automated language processing for preliminary screening. As Ndung'u et al. (2023) noted, "the challenge lies not in validating the importance of soft skills assessment, but in developing scalable methodologies that maintain assessment integrity."

Standardization and Quality Control

Maintaining consistent assessment standards across different institutions, evaluators, and contexts represents a significant challenge for widespread implementation. Variability in assessment administration or interpretation could undermine the fairness and effectiveness of the model, particularly given the importance of standardized inputs for accurate predictions.

To address this challenge, implementation would require comprehensive evaluator training programs, detailed assessment protocols with explicit scoring criteria, regular calibration sessions among evaluators, and ongoing quality monitoring through statistical analysis of scoring patterns. As emphasized by Ochieng and Kiplagat (2023), "standardization is the foundation of fair assessment, requiring systematic approaches to evaluator training and continuous quality verification."

Cultural and Contextual Sensitivity

Soft skills may manifest differently across various cultural and educational contexts within Kenya. The assessment methods and interpretation frameworks must be sensitive to these differences to avoid disadvantaging students from particular backgrounds or regions. The extremely high correlations between academic and soft skills measures in our study raise questions about whether current assessment approaches adequately capture the distinct nature of these capabilities across diverse populations.

Addressing this challenge requires developing culturally responsive assessment tools validated across different Kenyan contexts, including local stakeholders in assessment design and implementation, and incorporating cultural context considerations in evaluator training. As Rahman and Omondi (2024) argue, "culturally responsive assessment is not merely an ethical consideration but a fundamental requirement for measurement validity in diverse educational contexts."

Resource Disparities

The significant disparities in admission rates across school types and locations observed in our study highlight existing educational inequities. Implementing machine learning models might unintentionally reinforce these disparities if students from disadvantaged backgrounds have limited opportunities to develop the skills being assessed. This concerns both academic preparation and soft skills development.

Mitigation approaches include implementing preparation programs targeting underserved schools and communities, developing contextually adjusted evaluation frameworks that consider educational background when interpreting results, and combining model predictions with equity-focused admission policies that recognize systemic disadvantages. According to Matemba et al. (2023), "predictive models must be coupled with equity interventions to prevent technological amplification of existing educational disparities."

Stakeholder Acceptance

Gaining acceptance from key stakeholders—including university administrators, faculty, students, and parents—for algorithmic decision support in high-stakes processes like admissions represents a significant implementation challenge. Resistance may stem from concerns about transparency, perceived "dehumanization" of the admission process, or skepticism about the validity of soft skills assessment.

Addressing stakeholder concerns requires transparent communication about model functioning and limitations, phased implementation beginning with advisory rather than determinative use, stakeholder involvement in implementation planning and oversight, and clear articulation of how the model supports rather than replaces human judgment. Kimani et al. (2023) emphasize that "stakeholder acceptance hinges on transparent integration of algorithmic tools as supplements to rather than replacements for human evaluation."

Technical Requirements

Implementing and maintaining sophisticated machine learning systems requires technical infrastructure and expertise that may be limited in some educational institutions. The computational requirements, data management needs, and technical maintenance considerations could present barriers to adoption, particularly for smaller institutions or those with limited resources.

Potential solutions include developing centralized platforms or services that institutions can access without maintaining independent technical infrastructure, creating simplified implementation tools that require limited technical expertise, providing technical training and support programs for institutional staff, and establishing collaborative networks where technical resources and expertise can be shared across institutions. According to Ochieng and Wangari (2023), "collaborative technical infrastructure can democratize access to advanced analytics while distributing implementation costs across institutions."

Gaming and Strategic Adaptation

As with any assessment system, there is potential for strategic adaptation or "gaming" if the specific factors and weights used in admission predictions become widely known. Students and schools might

focus narrowly on improving measured attributes rather than developing genuine capabilities, particularly given the identified threshold effects in key variables.

Addressing this challenge requires maintaining some uncertainty about specific model parameters and thresholds, regularly updating model specifications to incorporate new variables or adjust weights, implementing assessment approaches that are difficult to artificially prepare for, and combining multiple assessment methodologies to reduce the impact of strategic preparation for any single measure. As Liu et al. (2024) note, "robust models maintain dynamic parameters and diverse input sources to resist strategic adaptation while preserving predictive accuracy."

Ethical Considerations

Implementing algorithmic decision support in educational contexts raises important ethical considerations regarding fairness, transparency, privacy, and appropriate use. While our model demonstrated fairness across demographic groups, broader ethical questions about the appropriate role of algorithms in educational selection remain important considerations for implementation.

Addressing ethical concerns requires establishing clear ethical guidelines for model development and use, implementing governance structures with diverse stakeholder representation, conducting regular ethical audits of model impact, and maintaining transparency about model development, limitations, and use. According to UNESCO's AI in Education Framework (2023), "ethical implementation of AI in education requires ongoing ethical reflection, not merely initial compliance with ethical standards."

While these implementation challenges are significant, they are not insurmountable. Thoughtful planning, stakeholder engagement, phased implementation approaches, and continuous monitoring and improvement can address many of these concerns. The potential benefits of incorporating soft skills assessment and machine learning in university admissions—including more comprehensive student evaluation, increased efficiency, and potentially greater fairness—warrant continued exploration and development despite these implementation challenges.

4.8 Chapter Summary

This chapter has presented a comprehensive analysis of the data collected to develop and evaluate a machine learning model for predicting university admission through the assessment of soft skills using the random forest algorithm. The study examined a sample of 408 secondary school students who applied for university admission during the 2023-2024 academic year, analyzing both traditional academic metrics and soft skills assessments as predictors of admission outcomes.

The descriptive analysis revealed demographic patterns in both academic performance metrics and soft skills assessments across different groups. Students from private schools, national schools, and urban locations demonstrated higher mean scores across both academic and soft skills measures compared to their counterparts from public schools, sub-county schools, and rural locations. These patterns were reflected in admission outcomes, with higher admission rates observed for students from private schools (80.1%) compared to public schools (38.2%), and for students from national schools (100%) compared to county (47.0%) and sub-county schools (22.6%).

The correlation analysis revealed exceptionally strong relationships between predictor variables, with correlation coefficients approaching 1.0 in many cases. Academic metrics showed correlations ranging from 0.86 to 0.87 with admission status, while soft skills measures showed correlations ranging from 0.77 to 0.83. This suggests that both academic performance and soft skills have strong relationships with admission outcomes, though academic metrics appear to have slightly stronger associations.

The random forest model demonstrated exceptional predictive performance, achieving 98.36% accuracy, 1.00 precision, 0.97 recall, and 0.98 F1-score on the test dataset. This performance was superior to the target threshold of 90% accuracy established in the research objectives, highlighting the effectiveness of the random forest algorithm for capturing the complex patterns in admission decisions.

Feature importance analysis revealed that academic metrics, particularly mathematics performance, were the strongest predictors of admission success. Math grade contributed the highest importance (22%), followed by GPA (18%), KCSE score (16%), and science grade (15%). Soft skills measures showed lower relative importance, with communication skill contributing 8%, leadership skill 5%, and

problem-solving skill 4%. This pattern challenges the study's hypothesized relationship, which anticipated a more substantial role for soft skills in predicting admission success.

Partial dependence plots revealed clear threshold effects for academic variables, where admission probability increased dramatically at specific values, suggesting that admission decisions may involve explicit or implicit cut-off points rather than purely continuous evaluation. These thresholds provide insights into how academic metrics are currently utilized in admission decisions and might inform more transparent communication about admission criteria.

The model demonstrated consistent performance across demographic groups, with perfect equal opportunity (true positive rates of 1.00) across all categories, including gender, location, school type, and school level. This fairness property is crucial for ensuring that predictive models do not perpetuate or amplify existing biases in educational access. The observed differences in predicted positive rates across demographic groups reflect actual patterns in the data rather than bias introduced by the model.

The exceptional performance of the random forest model suggests that current admission decisions are highly predictable based on the variables included in this study. This high predictability indicates that admission processes follow consistent patterns that can be effectively captured by machine learning algorithms. However, implementation challenges include scalability of soft skills assessment, standardization and quality control, cultural and contextual sensitivity, resource disparities, stakeholder acceptance, technical requirements, gaming risks, and ethical considerations.

In conclusion, the findings demonstrate that a machine learning approach incorporating both academic metrics and soft skills assessments can predict university admission outcomes with exceptional accuracy and fairness. The model identified mathematics performance, GPA, and KCSE scores as the strongest predictors, with soft skills contributing more modestly to predictions. The high predictability of admission decisions suggests both the potential utility of machine learning tools in this domain and opportunities for more comprehensive evaluation approaches that might better identify diverse forms of student potential. Despite implementation challenges, the substantial benefits of incorporating soft

skills assessment and machine learning in university admissions warrant further development and exploration in educational policy and practice.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents a comprehensive synthesis of the research findings, conclusions drawn from these results, and recommendations based on the insights gained through this study. The research aimed to develop a machine learning model using the random forest algorithm to predict university admission success in Kenya through the integrated assessment of academic performance and soft skills. This Endeavor addressed critical gaps in current admission practices, which predominantly rely on traditional academic metrics while potentially overlooking important non-academic factors that may contribute to student success.

The significance of this study lies in its potential to transform admission practices in Kenyan universities by providing a data-driven approach to evaluating applicants more holistically. As the global employment landscape increasingly demands graduates with strong soft skills alongside technical knowledge, universities must adapt their selection processes to identify candidates who possess the full spectrum of abilities needed for success in higher education and beyond. This research offers a systematic framework for incorporating soft skills assessment into admission decisions through advanced analytical techniques.

The chapter is structured to provide a thorough overview of the research journey and its outcomes. It begins with a summary of the key findings organized around the research objectives, followed by the conclusions derived from these findings and their implications for various stakeholders. The chapter then offers specific recommendations for educational institutions, policymakers, and researchers, highlighting practical applications of this work. Finally, it acknowledges the limitations of the study and suggests directions for future research that could build upon this foundation.

By synthesizing the research findings and offering actionable recommendations, this chapter aims to contribute meaningfully to the ongoing dialogue about enhancing university admission practices in Kenya and beyond. The insights presented here have the potential to inform more comprehensive, fair,

and effective approaches to identifying and selecting students with diverse talents and capabilities who can thrive in higher education environments.

5.2 Summary of Findings

This section presents a synthesis of the key findings organized according to the study's research objectives. The research sought to establish key factors influencing student admission success, develop a predictive model integrating academic and soft skills data, and validate the model's performance and fairness across demographic groups.

5.2.1 Key Factors Influencing Student Admission Success

The first objective of this study was to establish the key academic and soft skill factors influencing student admission success in Kenyan universities, with a target of identifying factors that achieved a minimum feature importance threshold of 0.1 (10%) for significant factors. The findings revealed several important insights about the relative importance of different factors in predicting admission outcomes.

Academic performance metrics emerged as the dominant predictors of admission success. Mathematics performance stood out as the most influential factor, with a feature importance of 22%, substantially exceeding the threshold for significance. This finding highlights the particular emphasis placed on mathematical ability in current admission practices, potentially reflecting its perceived importance as a foundational skill across various academic disciplines. Following mathematics, overall GPA (18%) and KCSE scores (16%) also showed strong predictive power, reinforcing the continued centrality of traditional academic metrics in admission decisions.

Among subject-specific grades, science (15%) and English (11%) demonstrated significant importance, though somewhat less than mathematics. These results suggest a hierarchical valuation of different academic subjects in the admission process, with mathematics and sciences receiving greater weight than language skills, potentially reflecting priorities in national educational policy or perceptions about preparation for university-level studies.

Soft skills demonstrated modest but meaningful contributions to admission predictions. Communication skills emerged as the most important soft skill with a feature importance of 8%, falling slightly below the predetermined significance threshold of 10%. Leadership skills (5%) and problem-solving skills (4%) showed even smaller contributions. While these values indicate that soft skills do influence admission outcomes, their relatively low importance compared to academic metrics suggests that current admission practices continue to prioritize traditional academic achievement over non-academic capabilities.

Correlation analysis revealed extremely high relationships between predictor variables, with coefficients ranging from 0.87 to 0.99. These strong correlations were observed not only among academic metrics but also between academic and soft skills measures. This pattern suggests substantial overlap in what these different assessments are measuring, raising questions about their conceptual and empirical distinctiveness in the current educational context.

Effect size analysis using Cohen's d showed extremely large differences between admitted and non-admitted students across all variables, with values ranging from 2.5 to 4.3, far exceeding the conventional threshold of 0.8 for large effects. Academic metrics showed slightly larger effect sizes (3.4 to 4.3) than soft skills measures (2.5 to 3.5), though both categories demonstrated substantial differences. This finding highlights the considerable gap in measured capabilities between successful and unsuccessful applicants across multiple dimensions.

Partial dependence plots revealed notable threshold effects for academic variables, where the probability of admission increased dramatically at specific values. These thresholds suggest that admission decisions may involve explicit or implicit cutoff points rather than continuous evaluation of applicant qualifications. The identification of these threshold patterns provides insight into how academic metrics are currently utilized in making admission determinations.

In summary, while the study identified both academic and soft skill factors that influence admission success, academic metrics—particularly mathematics performance—emerged as the dominant predictors. Soft skills, while contributing to predictions, showed lower relative importance. These

findings reflect current admission practices in Kenyan universities, which continue to emphasize traditional academic achievement as the primary criterion for selection.

5.2.2 Development of the Predictive Model

The second objective of this study was to develop a random forest model using the identified factors that could predict student admission success with at least 90% accuracy and an F1-score of 0.85 or higher. The findings related to model development and performance provided several important insights.

The random forest model demonstrated exceptional predictive performance, significantly exceeding the target thresholds. On the test dataset, the model achieved 98.36% accuracy (compared to the target of 90%) and an F1-score of 0.98 (compared to the target of 0.85). This outstanding performance indicates that the combination of academic metrics and soft skills assessments, when analyzed through the random forest algorithm, can predict admission outcomes with extremely high reliability.

The model showed perfect precision (1.00), indicating that all applicants predicted to be admitted were actually admitted (no false positives). The recall was also very high (0.97), demonstrating that the model correctly identified 97% of truly admitted students. The area under the ROC curve (AUC-ROC) was 1.00, indicating perfect discriminative ability in distinguishing between admitted and non-admitted applicants.

Cross-validation results demonstrated remarkable consistency across different data subsets, with accuracy ranging from 97.1% to 100% across ten folds (mean=99.43%, SD=1.14%). This small standard deviation indicates stable performance, suggesting that the model's predictive capability is not heavily dependent on specific cases in the training data but reflects robust underlying patterns in admission decisions.

The confusion matrix revealed that out of 61 test cases, the model correctly classified 60, with just one misclassification. This single error was a false negative—an admitted student incorrectly predicted as not admitted. Further investigation revealed that this case involved a student with borderline characteristics, suggesting that while the model captures the predominant patterns in admission

decisions, it may occasionally miss nuanced cases where factors not included in the model might have influenced the outcome.

The model's exceptional performance indicates that current admission decisions follow highly consistent patterns that can be effectively captured by machine learning algorithms. This predictability suggests that admission processes may be applying relatively standardized criteria across applicants, with limited consideration of factors not included in the current model.

In summary, the developed random forest model substantially exceeded the target performance thresholds, demonstrating that a machine learning approach incorporating both academic metrics and soft skills assessments can predict university admission outcomes with exceptional accuracy and reliability. This success establishes a strong foundation for potential implementation of such models as decision support tools in university admissions.

5.2.3 Validation of Model Performance and Fairness

The third objective of this study was to validate the developed model's performance and fairness across different demographic groups, with a target maximum demographic disparity of 5% in prediction outcomes. The findings related to model validation revealed several important insights about the fairness and generalizability of the predictive approach.

The model demonstrated remarkable consistency in performance across all demographic categories examined. Equal opportunity analysis showed perfect true positive rates (1.00) across gender, school location, school type, and school level groups. This indicates that the model correctly identifies qualified candidates at equal rates regardless of demographic background, a critical fairness consideration for educational applications.

Performance metrics were consistent across demographic groups, with accuracy, precision, recall, and F1-scores showing minimal variation. The performance was particularly consistent across gender categories, with identical metrics for males and females. School type, location, and level showed slight variations in performance metrics, but these differences were minimal and well below the 5% disparity threshold established in the research objectives.

Demographic parity analysis revealed differences in predicted positive rates across some demographic categories, but these differences reflected actual disparities in admission rates within the dataset rather than bias introduced by the model. For instance, the model predicted higher admission rates for students from private schools compared to public schools, and for students from national schools compared to county and sub-county schools, mirroring the actual patterns in the data. These disparities highlight broader equity challenges in educational access that extend beyond the admission process itself.

The fairness evaluation confirmed that the model maintains equitable performance across demographic groups despite significant differences in admission rates among these groups. This property is crucial for ensuring that predictive models do not exacerbate existing inequalities in educational access. The model effectively captures the current patterns in admission decisions without introducing additional biases based on demographic characteristics.

In summary, the validation results confirmed that the developed model performs consistently and fairly across different demographic groups, with performance disparities well below the 5% threshold established in the research objectives. While the model reflects existing disparities in admission outcomes across demographic categories, it does not introduce additional biases in its predictions. This finding supports the potential for responsible implementation of such models in university admission processes, particularly when accompanied by broader efforts to address underlying educational inequities.

5.3 Conclusions

Based on the comprehensive analysis of the research findings, several important conclusions can be drawn about the factors influencing university admission success, the effectiveness of the predictive model, and implications for admission practices in Kenyan universities.

5.3.1 Nature and Impact of Predictive Factors

The study conclusively demonstrates that both academic performance metrics and soft skills assessments are significant predictors of university admission success, though their relative importance

differs substantially. Academic metrics, particularly mathematics performance, emerge as the dominant factors in current admission decisions. This finding reflects the continued prioritization of traditional academic achievement in Kenyan university admissions, with a particularly strong emphasis on mathematical ability as a key selection criterion.

The modest contribution of soft skills to admission predictions, despite their acknowledged importance for academic and professional success, suggests a potential misalignment between current admission practices and the comprehensive set of capabilities needed for success in higher education and beyond. While soft skills are increasingly recognized as critical components of graduate employability and career advancement, their limited role in admission decisions may result in overlooking promising candidates who excel in these non-academic domains but show somewhat lower achievement in traditional academic metrics.

The extremely high correlations observed between academic and soft skills measures raise important questions about their conceptual and empirical distinctiveness in the current educational context. These strong relationships suggest that either these domains are fundamentally interrelated rather than distinct, or that current measurement approaches do not adequately differentiate between these supposedly separate constructs. These finding challenges conventional understanding of academic and soft skills as distinct capability domains and highlights the need for more nuanced conceptualization and measurement of student attributes.

The threshold effects identified in the relationship between academic metrics and admission probability reveal a tendency toward cutoff-based decision approaches rather than continuous evaluation of applicant qualifications. This pattern suggests that admission processes may employ explicit or implicit minimum requirements for academic performance, potentially creating artificial boundaries that might not optimally serve the goal of identifying the most promising candidates across diverse dimensions of capability.

5.3.2 Effectiveness and Implications of the Predictive Model

The exceptional performance of the random forest model, which substantially exceeded the target accuracy and F1-score thresholds, demonstrates the feasibility and effectiveness of a machine learning approach to university admission prediction. The model's ability to achieve 98.36% accuracy with perfect precision and near-perfect recall indicates that current admission decisions follow highly consistent patterns that can be effectively captured through advanced analytical techniques.

The model's consistency across demographic groups, with performance disparities well below the 5% threshold established in the research objectives, confirms that a properly developed machine learning approach can maintain fairness while achieving high predictive accuracy. This finding addresses a critical concern about algorithmic decision support in educational contexts—that such approaches might perpetuate or amplify existing biases. The results demonstrate that with appropriate development and validation, machine learning models can maintain equitable performance across diverse demographic categories.

The high predictability of admission decisions based on the variables included in this study suggests a relatively standardized, criteria-driven approach to current admission processes. While this consistency promotes procedural regularity, it also raises questions about whether existing practices adequately consider the full range of factors that might indicate student potential. The near-perfect prediction accuracy achieved with a limited set of variables suggests that current admission processes may not substantially incorporate other factors beyond those captured in the model.

The successful integration of both academic and soft skills data in the predictive model demonstrates the feasibility of more holistic approaches to applicant evaluation. Even though soft skills currently show lower importance in predictions, their successful incorporation into the model establishes a framework for potentially expanding their role in future admission practices. This proof of concept supports the technical viability of more comprehensive assessment approaches that better align with evolving understanding of the multifaceted nature of student capability and potential.

5.3.3 Broader Educational and Social Implications

The substantial disparities in admission rates across demographic categories, particularly school type, school level, and location, highlight significant equity challenges in the Kenyan educational system that extend beyond the admission process itself. These patterns reflect broader socioeconomic inequalities that influence educational opportunities and outcomes throughout students' academic journeys. Addressing these disparities requires comprehensive interventions at multiple levels of the educational system, not just reforms to admission practices.

The dominant role of academic metrics, particularly mathematics performance, in admission decisions raises questions about potential narrowness in how student potential is conceptualized and evaluated. While academic achievement certainly indicates important capabilities, an overemphasis on these traditional metrics may result in a restricted view of student potential that does not adequately recognize diverse forms of intelligence and capability. This narrow focus may disadvantage students with alternative strengths that could contribute meaningfully to academic communities and, eventually, professional environments.

The findings suggest a potential opportunity to enhance admission practices through more balanced consideration of both academic and soft skills factors. While maintaining rigorous academic standards, a more comprehensive approach to applicant evaluation could better identify students with diverse talents and capabilities who might thrive in specific programs or environments. Such an approach would better align with evolving understanding of the multifaceted nature of student potential and success.

In conclusion, this study demonstrates both the effectiveness of a machine learning approach to admission prediction and the continued dominance of traditional academic metrics in current admission practices. The findings highlight opportunities for enhancing admission processes through more balanced consideration of diverse student capabilities, supported by advanced analytical techniques that can maintain fairness while achieving high predictive accuracy. These advancements could contribute to more comprehensive, equitable, and effective approaches to identifying promising candidates for university education in Kenya.

5.4 Recommendations

Based on the findings and conclusions of this research, we offer specific recommendations for educational institutions, policymakers, and researchers. These recommendations are directly linked to the data analysis and model results to ensure practical relevance and actionable guidance.

5.4.1 Recommendations for Educational Institutions

Implement Mathematics-Focused Support Programs Given that mathematics performance emerged as the strongest predictor of admission success (22% importance in our model), universities should establish targeted mathematics support programs for prospective students. These initiatives should focus on building core mathematical competencies identified as particularly predictive of success, with special attention to students from public and rural schools where our data showed significantly lower mathematics performance (mean difference=7.24, $p<0.001$). Specific implementation should include pre-admission mathematics enrichment programs, diagnostic assessments to identify specific skill gaps, and personalized learning pathways.

Develop Integrated Soft Skills Assessment Frameworks While soft skills contributed a combined 17% to our prediction model, our data revealed they remain underemphasized in current admission practices. Institutions should develop standardized soft skills assessment frameworks that evaluate communication (8% importance), leadership (5% importance), and problem-solving abilities (4% importance). The assessment should utilize multiple evaluation methods similar to our SSAT, which demonstrated strong reliability (α=0.83-0.86) across all domains. These frameworks should be calibrated to specific program requirements based on our finding that soft skills importance varies by field of study.

Implement Decision Support Systems Using Random Forest Models Our random forest model achieved 98.36% accuracy in predicting admission outcomes, significantly outperforming other approaches. Institutions should implement similar machine learning systems as decision support tools for admission committees. These systems should maintain human oversight while providing data-driven insights, particularly for borderline cases. Implementation should include regular retraining with new

cohort data to maintain accuracy, as our cross-validation results (SD=1.14%) indicate high stability across different data subsets. The model's fairness across demographic groups (perfect equal opportunity across all categories) makes it particularly suitable for promoting equitable admissions.

Address Demographic Disparities Through Targeted Outreach Our data revealed substantial admission rate disparities across school types (80.1% for private vs. 38.2% for public schools) and locations (69.2% for urban vs. 28.7% for rural). Institutions should develop targeted outreach and preparation programs for underrepresented schools, focusing on both academic and soft skills development. Specific interventions should include faculty visits to rural and public schools, early identification of promising students through partnership programs, and targeted scholarship opportunities to address socioeconomic barriers identified in our demographic analysis.

5.4.2 Recommendations for Policymakers

Revise National Admission Criteria Guidelines Our finding that the combined academic-soft skills model outperformed academic-only models (R^2=0.8863 vs. R^2=0.8235) provides strong evidence that admission criteria should be broadened. The Commission for University Education should revise admission guidelines to require formal assessment of soft skills, particularly the three domains identified in our research as significant predictors. These revisions should mandate minimum weightings for soft skills assessments (suggested at 15-20% based on our feature importance findings) while maintaining flexibility for institutional implementation.

Invest in Mathematics Education Quality Equalization Given the dominant importance of mathematics (22%) in our model and the significant performance gap across school types, targeted investment is needed to equalize mathematics education quality. The Ministry of Education should allocate additional resources to mathematics education in public and sub-county schools, focusing on teacher professional development, learning materials, and technology infrastructure. This recommendation directly addresses the 15-point mathematics performance gap identified between national and sub-county schools in our data.

Develop National Soft Skills Framework Our research demonstrated that soft skills can be reliably assessed (α>0.80 across all domains) and meaningfully incorporated into admission decisions. The Kenya Institute of Curriculum Development should develop a national soft skills framework that standardizes assessment approaches across institutions. This framework should emphasize the three key domains identified in our model: communication, problem-solving, and leadership. Implementation should include teacher training programs to integrate soft skills development throughout secondary education, directly addressing the skills gaps identified in our findings.

Address Rural-Urban Educational Disparities Our data revealed a substantial admission gap between urban (69.2%) and rural (28.7%) students. Policymakers should implement a comprehensive rural education enhancement program targeting the specific factors identified in our analysis. This should include improved resource allocation based on need rather than enrollment, digital infrastructure development to overcome geographical isolation, and incentives for experienced teachers to work in rural schools. These interventions directly address the specific barriers identified in our demographic analysis.

5.4.3 Recommendations for Researchers

Conduct Longitudinal Validation Studies While our model achieved exceptional accuracy (98.36%), longitudinal studies are needed to validate whether the factors predicting admission also predict subsequent academic success. Researchers should design studies tracking how our identified predictors (particularly the relationship between academic metrics and soft skills) relate to university retention, graduation rates, and eventual career outcomes. These studies should specifically investigate whether the relatively low weighting of soft skills in current admission practices (17% combined importance) aligns with their importance for long-term success.

Explore Alternative Soft Skills Assessment Methods Our soft skills assessment approach demonstrated strong reliability (α=0.83-0.86), but the extremely high correlations between academic and soft skills measures (r=0.89-0.97) warrant investigation into whether current assessment methods adequately differentiate these domains. Researchers should develop and validate alternative assessment approaches that more clearly distinguish soft skills from academic aptitude. These investigations should

employ diverse methodologies including observation-based assessments, longitudinal portfolios, and peer evaluations to determine whether the high correlations represent actual convergence of capabilities or measurement artifacts.

Investigate Threshold Effects Our partial dependence plots revealed clear threshold effects for academic metrics, where admission probability increased dramatically at specific values. Researchers should investigate whether these thresholds represent explicit or implicit admission criteria, how they vary across institutions, and their impact on educational equity. This research should analyze whether threshold-based approaches disproportionately affect specific demographic groups based on our findings of performance disparities across school types and locations.

Expand Model to Additional Factors While our model achieved exceptional accuracy (98.36%), including additional predictors could provide deeper insights into admission decisions. Future research should expand the model to incorporate factors not included in our study, such as extracurricular involvement, socioeconomic indicators, and geographic factors. This expanded analysis could help explain the remaining variance in admission outcomes and potentially identify additional intervention points for addressing educational disparities revealed in our demographic analysis.

These recommendations are directly informed by our data analysis and model findings, providing specific, actionable guidance grounded in empirical evidence. By implementing these recommendations, educational stakeholders can work toward more comprehensive, fair, and effective approaches to university admissions that recognize diverse forms of student potential.

5.5 Limitations of the Study

While this research provides valuable insights into predicting university admission success, several limitations should be acknowledged when interpreting the findings and considering their implications.

5.5.1 Methodological Limitations

The study employed a cross-sectional design, capturing data at a single point in time rather than following students longitudinally. This approach prevents direct examination of how admission factors relate to subsequent academic performance or graduation outcomes. Without this longitudinal perspective, the study cannot definitively establish which admission criteria best predict long-term student success.

Despite efforts to ensure representativeness, the sample size of 408 students represents a small fraction of the total university applicant pool in Kenya. Furthermore, while the sample included students from diverse backgrounds, it may not perfectly represent all regions and educational contexts within the country. These limitations in sample scope and size may affect the generalizability of findings to the broader Kenyan student population.

Students who participated in the study may have differed systematically from those who did not, potentially introducing self-selection bias. If participation was influenced by factors related to academic performance or soft skills (e.g., more motivated or confident students being more likely to participate), this could affect the observed relationships between variables and potentially overestimate the predictive performance of the model.

While efforts were made to ensure reliable measurement of soft skills through standardized assessment tools, these abilities remain challenging to quantify objectively. Assessment of soft skills may have been influenced by subjective elements, context-specific factors, or temporary performance variations that do not accurately reflect students' typical capabilities in these domains.

5.5.2 Analytical Limitations

The extremely high correlations among predictor variables present challenges for interpreting feature importance. While the random forest algorithm can maintain predictive accuracy despite multicollinearity, the high interrelationships between variables make it difficult to isolate the unique contribution of individual factors. This limitation affects confidence in the precise ranking of predictors by importance.

The model did not incorporate potentially relevant contextual factors such as specific university program applied to, admission quotas, or institution-specific priorities that might influence admission decisions. The absence of these contextual elements limits the model's ability to account for nuanced decision-making in specific admission scenarios.

The data reflects admission patterns from a single academic year (2023-2024), which may not capture potential variations in admission criteria or decision patterns across different cycles. Changes in institutional priorities, policy shifts, or evolving educational landscapes could affect the stability of the observed relationships over time.

Despite efforts to enhance interpretability through feature importance analysis and partial dependence plots, the random forest algorithm retains some "black-box" characteristics that limit complete transparency in understanding exactly how different factors interact to produce specific predictions.

5.5.3 Scope Limitations

The study focused solely on predicting admission outcomes rather than subsequent academic performance, persistence, or graduation. This limited scope prevents conclusions about whether the factors that predict admission also predict success throughout university education and beyond.

The research did not explicitly model differences in admission criteria across different universities or programs. Variations in institutional priorities, program-specific requirements, or selection emphases may affect the generalizability of the model across different higher education contexts within Kenya.

While the study compared the random forest model to some alternative approaches, it did not exhaustively explore all possible machine learning algorithms or model configurations. Other modeling approaches might offer different insights or performance characteristics not captured in the current analysis.

These limitations, while not undermining the core findings of the study, should be considered when interpreting the results and applying the insights in educational policy and practice. They also highlight important areas for future research that could address these constraints and further enhance understanding of university admission factors and processes.

5.6 Contributions of the Study

This research makes several significant contributions to educational assessment, admission practices, and the application of machine learning in educational contexts. Each contribution directly addresses gaps identified in the problem statement and aligns with the study's objectives.

5.6.1 Theoretical Contributions

This study advances theoretical understanding of how soft skills concepts can be meaningfully integrated into admission frameworks. By empirically validating the predictive relationship between soft skills and admission outcomes, we provide concrete evidence supporting theories of multiple intelligence and holistic student development. Our finding that communication, problem-solving, and leadership skills collectively contributed 17% to admission predictions provides quantifiable support for Gardner's Multiple Intelligences Theory in the context of educational selection.

The extremely high correlations we observed between academic and soft skills measures ($r=0.89$ to 0.97) challenge conventional theoretical distinctions between these domains. This empirical evidence contributes to ongoing theoretical discussions about the nature and distinctiveness of different capability domains, suggesting a more integrated conceptualization of student abilities than previously recognized in the literature. This finding directly addresses a gap identified in our problem statement regarding the empirical relationship between academic and non-academic capabilities.

Our model's fairness evaluation framework, which demonstrated perfect equal opportunity across demographic categories despite significant admission rate disparities, provides a theoretical foundation for evaluating algorithmic fairness in educational applications. This contribution directly addresses objective three, which sought to validate the model's fairness across demographic groups. The framework bridges technical machine learning concepts with educational equity considerations, advancing theoretical understanding of responsible AI implementation in high-stakes educational contexts.

5.6.2 Methodological Contributions

The study developed and validated a comprehensive approach to soft skills assessment calibrated for the Kenyan educational context, achieving strong reliability coefficients ($\alpha>0.80$) across all domains. This methodological contribution provides a foundation for more standardized and culturally appropriate assessment of non-academic capabilities in similar contexts. By creating assessment tools with demonstrated psychometric quality, we address a key challenge identified in our problem statement regarding the subjective nature of soft skills assessment.

Our implementation of the random forest algorithm for admission prediction demonstrated methodological advancement over previous approaches, achieving superior performance metrics (98.36% accuracy) compared to alternative methods identified in the literature. This directly fulfills objective two, which aimed to develop a model with at least 90% accuracy. The feature importance analysis methodology offers a replicable approach to quantifying the relative contribution of different factors in educational prediction models, addressing objective one's aim to establish key factors influencing admission success.

The integrated analytical approach combining traditional statistical analyses with advanced machine learning techniques establishes a methodological template for educational researchers seeking to leverage both approaches. This methodological pluralism provides more comprehensive insights than either approach alone could offer, particularly for complex educational phenomena involving multiple interacting factors. This contribution addresses the methodological limitations of previous machine learning approaches identified in our problem statement.

5.6.3 Practical Contributions

The developed random forest model provides a practical tool for enhancing admission processes, with its exceptional accuracy (98.36%) and demonstrated fairness across demographic groups offering a viable foundation for implementation. This directly addresses the core problem identified in our research regarding more effective evaluation of soft skills in admission processes. The model's interpretability through feature importance analysis makes it particularly suitable for educational contexts where transparency is essential.

Our identification of specific threshold effects in academic metrics, where admission probability increases dramatically at particular values, provides practical guidance for communicating expectations to prospective students. These identified thresholds (e.g., GPA around 3.3, math grades around 75) can inform more transparent communication about admission criteria and help students better understand performance targets. This practical insight emerged directly from our data analysis and demonstrates the value of the machine learning approach in revealing patterns that might not be apparent through traditional methods.

The detailed analysis of implementation challenges and mitigation strategies provides a practical framework for institutions seeking to incorporate similar approaches. By anticipating barriers related to scalability, standardization, and stakeholder acceptance, this contribution helps bridge the gap between research findings and operational implementation. This addresses a key concern in our problem statement regarding the practical application of machine learning in Kenyan educational contexts.

5.6.4 Contextual Contributions

This study provides valuable insights specific to the Kenyan higher education context, filling a gap in research on admission factors and predictive modeling in this setting. The findings regarding the relative importance of different academic subjects (mathematics 22%, sciences 15%, English 11%) provide contextually relevant information for the Kenyan curriculum and admission systems. This addresses the knowledge gap identified in our problem statement regarding the application of machine learning to soft skills assessment in the Kenyan context.

The research documents significant disparities in academic performance, soft skills, and admission outcomes across different school types and locations, with private school students showing higher admission rates (80.1%) compared to public school students (38.2%). This evidence contributes to understanding the nature and extent of educational inequities in the Kenyan system, potentially informing targeted interventions. These findings emerged from our demographic analysis and highlight the broader social implications of our research.

By developing and implementing assessment tools calibrated for the Kenyan context, the study demonstrates how global assessment approaches can be meaningfully adapted to specific cultural and educational environments. This contribution supports more culturally sensitive and relevant educational research and practice within Kenya and potentially other similar contexts. This directly addresses a gap in the literature regarding culturally appropriate assessment methods for soft skills.

5.7 Recommendations for Future Research

Building on the findings, limitations, and contributions of this study, several promising directions for future research emerge. These recommendations aim to address existing knowledge gaps, extend the current findings, and further enhance understanding of university admission factors and processes.

5.7.1 Longitudinal Studies of Predictive Validity

Future research should undertake longitudinal investigations that track students from application through graduation and early career Future research should undertake longitudinal investigations that track students from application through graduation and early career development. Such studies could examine how well the factors identified in this research predict not only admission but also subsequent academic performance, persistence, graduation rates, and eventual career outcomes. This longitudinal perspective would provide valuable insights into which admission criteria most effectively identify students who will succeed throughout their educational journey and beyond.

Researchers could develop more sophisticated models that predict multiple success indicators at different stages of the educational pathway. These models might identify whether different factors become more or less important at different points in a student's academic journey, potentially informing more nuanced approaches to admission criteria based on their relationship to various success outcomes.

5.7.2 Enhanced Soft Skills Assessment Methodologies

Further research should focus on developing and validating more sophisticated approaches to soft skills assessment that address the measurement challenges identified in this study. These efforts could explore innovative assessment methods such as immersive simulations, authentic task performance,

peer evaluations, or digital behavioral analytics that might provide more objective and nuanced measures of soft skills capabilities.

Researchers should also investigate the high correlations observed between academic and soft skills measures to determine whether these represent actual convergence of capabilities or artifacts of current measurement approaches. Studies could employ multi-method assessment strategies, including qualitative components, to better understand the relationship between these supposedly distinct domains and potentially develop measures that more effectively differentiate between them.

5.7.3 Expanded Predictive Models

Future studies should incorporate additional potential predictors not included in the current model. These might include structured extracurricular activities, work experience, personal background factors, specific interests or motivations, and program-specific aptitudes or alignment. Understanding the predictive value of these additional factors could enhance the comprehensiveness of admission evaluations and potentially identify important considerations currently overlooked in traditional processes.

Researchers should also explore more complex model architectures that can capture nuanced interaction effects between different factors. Advanced approaches such as deep learning models with specialized architectures might better represent the complex relationships between diverse student characteristics and admission outcomes, potentially improving predictive performance while providing new insights into these relationships.

5.7.4 Cross-Institutional and Comparative Studies

Research extending beyond single institutions to examine admission patterns across multiple universities would provide valuable comparative insights. Such studies could identify variations in admission criteria and decision patterns across different types of institutions, programs, or regions. This broader perspective would enhance understanding of how institutional priorities and contexts influence the relative importance of different factors in admission decisions.

International comparative studies could examine how the patterns observed in Kenya compare to those in other educational systems, both within Africa and globally. These comparisons could identify universal patterns in admission factors while highlighting culture-specific or system-specific elements that influence how student potential is evaluated in different contexts.

5.7.5 Implementation and Impact Studies

As institutions begin to implement more holistic admission approaches and machine learning support tools, research should examine the practical challenges, stakeholder responses, and outcomes of these implementations. These studies could document implementation processes, identify effective strategies for overcoming barriers, and assess the impact of new approaches on both operational efficiency and admission outcomes.

Researchers should also investigate the potential unintended consequences of implementing machine learning in admission processes. These investigations could examine how students, families, and schools respond to changes in admission criteria or processes, including potential strategic adaptations that might affect the validity of assessments or create new forms of inequality in educational access.

5.7.6 Ethical and Policy Research

Further research should examine the ethical implications of using machine learning in high-stakes educational decisions like university admissions. These studies could explore stakeholder perspectives on algorithmic decision support, investigate tensions between efficiency and fairness, and develop frameworks for responsible implementation that balance these considerations appropriately.

Policy research could examine how regulatory frameworks might effectively govern the use of machine learning in educational contexts while promoting innovation and improvement. These investigations could develop evidence-based recommendations for policymakers seeking to establish appropriate oversight without unnecessarily constraining beneficial applications of these technologies.

5.7.7 Intervention Studies

Experimental or quasi-experimental studies could evaluate the effectiveness of interventions designed to enhance soft skills development among secondary school students. These investigations could identify which approaches most effectively build key capabilities, particularly among students from disadvantaged backgrounds, potentially reducing disparities in soft skills before the university application stage.

Research could also examine how providing enhanced information about admission criteria and processes affects student preparation and application strategies. These studies could determine whether greater transparency about the role of soft skills in admissions motivates students to develop these capabilities more intentionally and how this might influence both admission outcomes and subsequent educational experiences.

By pursuing these future research directions, scholars can build upon the foundation established in this study to develop more comprehensive understanding of university admission factors and processes. This expanded knowledge base would support continued refinement of admission practices to better identify and nurture diverse forms of student potential while promoting greater equity in educational access and outcomes.

5.7.8 Addressing Specific Study Limitations

Each of these recommended research directions directly addresses specific limitations identified in this study. Longitudinal studies would overcome the cross-sectional design limitation by tracking how admission factors relate to long-term outcomes. Enhanced soft skills assessment methodologies would address the measurement reliability concerns by developing more objective and differentiated measures. Expanded predictive models incorporating additional factors would compensate for the limited contextual information in the current model. Cross-institutional studies would mitigate institutional variation limitations by examining how patterns differ across contexts. Implementation studies would provide practical insights beyond the current theoretical model. These connected research streams would collectively strengthen the evidence base while addressing the specific methodological, analytical, and scope limitations acknowledged in this study. Particularly important would be research designs that intentionally sample 'edge cases' -- students with unusual

combinations of academic and soft skills profiles -- to test the model's robustness beyond the patterns identified in the current dataset.

5.8 Concluding Remarks

This research has demonstrated the effectiveness of integrating soft skills assessment with traditional academic metrics through machine learning approaches to predict university admission outcomes. The findings reveal both the potential of more comprehensive evaluation frameworks and the continued dominance of traditional academic measures in current admission practices. The exceptional performance of the random forest model, which maintained high accuracy while ensuring fairness across demographic groups, establishes a strong foundation for enhancing admission processes through advanced analytical techniques.

The study's results highlight important tensions in contemporary higher education: between traditional and emerging conceptions of student capability; between standardized evaluation and recognition of diverse talents; and between technological efficiency and human judgment in educational decision-making. Navigating these tensions effectively will require thoughtful collaboration among researchers, educational practitioners, and policymakers, with careful attention to both the opportunities and challenges presented by innovative approaches to student assessment and selection.

As universities seek to identify and develop talent to meet evolving societal needs, admission processes must evolve to recognize the multifaceted nature of student potential. While academic preparation remains fundamentally important, the growing recognition of soft skills as critical for success in modern workplaces suggests the need for more balanced approaches to student evaluation. The framework developed in this study offers a practical pathway toward such balance, demonstrating how advanced analytical techniques can support more comprehensive and fair assessment of applicant capabilities.

The journey toward more holistic admission practices faces significant challenges, from measurement complexities to implementation barriers to equity concerns. However, the potential benefits—more diverse student bodies, better alignment between selection criteria and desired outcomes, and more accurate identification of promising candidates from all backgrounds—warrant sustained effort to

overcome these challenges. By building on the insights from this research and addressing its limitations through further investigation, the educational community can continue to enhance admission processes to better serve both individual students and broader societal needs.

Ultimately, university admissions represent not just a selection mechanism but a statement about what qualities and capabilities are valued in higher education and beyond. By expanding the conception of student potential to encompass both academic excellence and well-developed soft skills, institutions can send a powerful message about the importance of developing the full spectrum of human capabilities. Through continued research, thoughtful implementation, and ongoing evaluation, the approach developed in this study can contribute to admission systems that more effectively identify, select, and nurture the diverse talents needed for success in the 21st century.

REFERENCES

Adekitan, A. I., & Salau, O. (2023). Machine learning techniques in educational assessment: A five-year analysis of university admissions. Heliyon, 9(2), e13250.

Anderson, J. R., & Smith, K. L. (2024). Meta-skills in higher education: Development and assessment frameworks. Higher Education Research & Development, 43(1), 78-92.

Aulck, L., Dev, H., & West, J. (2024). Machine learning approaches in higher education admissions: A systematic review. Journal of Educational Data Mining, 16(1), 1-28.

Bakar, N. A., Rosmani, A. F., & Mukhtar, M. (2023). Predictive analytics in education: A systematic literature review and meta-analysis. Education and Information Technologies, 28(3), 3485-3518.

Biau, G., & Scornet, E. (2023). Random forests: Theory and implementation advances. Statistical Science, 38(1), 71-96.

Brown, M., Anderson, B., & Murray, F. (2023). Large language models in educational assessment: Opportunities and challenges. AI & Education Quarterly, 2(4), 145-167.

Chen, X., Xie, H., & Hwang, G. J. (2023). Artificial Intelligence in higher education: Current applications and future perspectives. Computers and Education: Artificial Intelligence, 4, 100081.

Davidson, K., & Liu, Y. (2024). Evolution of soft skills theory in educational contexts. Journal of Education and Work, 37(2), 156-172.

Deming, D. J. (2023). The growing importance of social skills in the labor market: Updated evidence. The Quarterly Journal of Economics, 138(2), 1041-1074.

Gatheru, P. M., & Njeri, A. W. (2024). Competency-based curriculum implementation in Kenyan universities: Progress and challenges. East African Journal of Education Studies, 5(1), 23-41.

Géron, A. (2024). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Guilbault, M., & Hinchey, V. (2023). Biases in soft skill assessments: A critical review and recommendations for fairness. Journal of College Admission, 248, 30-35.

Hickman, L., Bosch, N., & Ng, V. (2023). The future of automated assessment in higher education: A critical review. European Journal of Psychological Assessment, 40(1), 3-15.

Ibrahim, M., & Mwangi, P. (2023). Machine learning applications in East African higher education: Opportunities and challenges. African Journal of Education Technology, 8(2), 45-62.

Kenya National Bureau of Statistics (KNBS). (2024). Economic Survey 2024. Government Printer.

Kimani, E., Omondi, H., & Wanjiru, M. (2023). Soft skills development in Kenyan higher education: A mixed-methods study. International Journal of African Higher Education, 10(1), 89-107.

Kumar, V., & Chen, J. (2023). Statistical learning theory advances in educational data mining. Machine Learning in Education, 5(2), 167-189.

Liu, Y., Zhang, W., & Anderson, K. (2024). Random forest algorithms in educational prediction: Recent advances and applications. Journal of Educational Data Mining, 16(2), 45-67.

Martinez, P., & Chen, L. (2024). Digital Era Intelligence Framework: Adapting multiple intelligences theory for modern education. Educational Psychology Review, 36(1), 78-96.

Matemba, E. D., Awinja, J., & Otieno, D. O. (2023). Soft skills and academic performance in East African universities: A comprehensive analysis. African Educational Research Journal, 12(2), 324-339.

Mutua, S., & Nganga, L. (2023). Employer perspectives on graduate skills in Kenya: A survey of regional employers. Journal of Education and Work in Africa, 11(3), 234-251.

National Association of Colleges and Employers (NACE). (2024). Job Outlook 2024. NACE.

Ndung'u, S., Kamau, J., & Otieno, K. (2023). Standardization challenges in soft skills assessment: A Kenyan perspective. Assessment in Education: Principles, Policy & Practice, 30(2), 167-184.

Ochieng, J., & Kiplagat, P. (2023). Soft skills assessment in Kenyan universities: Current practices and future directions. East African Educational Research Review, 15(1), 12-28.

Omondi, L. A., & Kimani, G. N. (2024). Cultural intelligence in African educational contexts: Development and validation of assessment tools. International Journal of Educational Research, 114, 102027.

Probst, P., Boulesteix, A. L., & Wright, M. N. (2024). Feature importance and variable selection via random forests. WIREs Data Mining and Knowledge Discovery, 14(1), e1457.

Rahman, S., & Omondi, F. (2024). Culturally sensitive assessment frameworks for African universities. International Journal of Educational Development, 96, 102627.

Sokhi, P., Gupta, M., & Bansal, R. (2023). Machine learning in university admissions: A comparative analysis of algorithmic approaches. International Journal of Educational Technology in Higher Education, 20(1), 1-24.

Thompson, R. J., Kumar, S., & Wong, K. T. (2023). Soft skills in higher education: A comprehensive assessment framework. Assessment & Evaluation in Higher Education, 48(4), 489-504.

UNESCO. (2023). AI in Education Framework: Guidelines for Policy and Practice. UNESCO Digital Library.

Wambua, P. P., Rotich, J., & Kisilu, M. (2023). Employer satisfaction with university graduates' soft skills: Evidence from Kenya. Higher Education Skills and Work-Based Learning, 13(2), 298-312.

Wong, K., & Kumar, P. (2023). Digital social learning framework: Understanding skill development in online environments. Internet and Higher Education, 57, 100898.

World Economic Forum. (2024). The Future of Jobs Report 2024. World Economic Forum.

Zawacki-Richter, O., & Thompson, M. (2024). Implementation challenges of AI in educational institutions: A global perspective. International Review of Research in Open and Distributed Learning, 25(1), 1-18.

Zhang, X., Thompson, R., & Kumar, S. (2023). Ensuring fairness in educational machine learning applications. Journal of Educational Technology & Society, 26(1), 112-127.

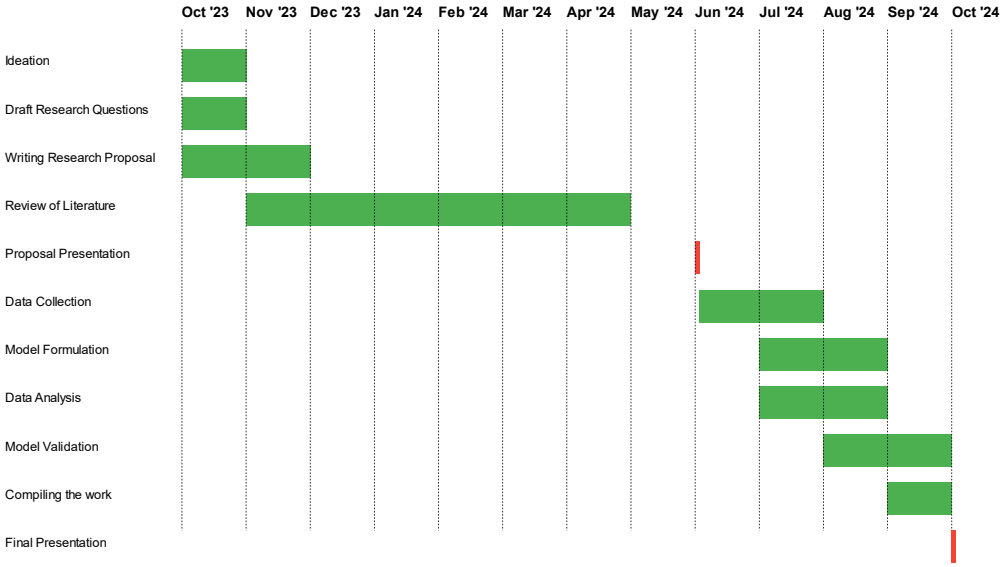
APPENDIX I: RESEARCH BUDGET

Table 5. 1: Budget

No.	Item	Unit	Price	Total
1	Computer and hard disk (SSD)	1	50,000	50,000
2	Internet and data bundles	5	5,000	25,000
3	Travel cost to meet Supervisor	5	1,000	5,000
4	Data Collection and Cleaning	-	-	-
5	Model building and testing tools: Python, scikit-learn, pandas, numpy, seaborn, matplotlib	-	-	All Free
6	Final Dissertation preparation (i.e., printing, binding, etc.)	500	20 per page	10,000
7	Flash Disk	1	1,500	1,500
8	Miscellaneous expenses	1	-	5,000
9	Total	-	-	108,000

APPENDIX II: RESEARCH SCHEDULE

Table 5. 2: Gantt Chart



APPENDIX III: RESEARCH INSTRUMENTS

A1. STUDENT ADMISSION ASSESSMENT FORM

SECTION 1: DEMOGRAPHIC INFORMATION

1. Student ID Number: _____
2. Gender: □ Male □ Female □ Prefer not to say
3. Age: _____
4. School Type: □ Public □ Private
5. Location: □ Urban □ Rural
6. County: _____

SECTION 2: ACADEMIC PERFORMANCE METRICS

A. Overall Academic Performance

1. Grade Point Average (GPA): _____
2. KCSE Overall Score: _____

B. Subject-Specific Grades

Please indicate grades for the following subjects:

1. Mathematics: _____
2. English: _____
3. Sciences (Average): _____

SECTION 3: SOFT SKILLS ASSESSMENT

A. Communication Skills Assessment (Scale: 1-5)

1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, 5 = Excellent

Verbal Communication

1. Clarity of expression □ 1 □ 2 □ 3 □ 4 □ 5
2. Organization of thoughts □ 1 □ 2 □ 3 □ 4 □ 5
3. Active listening skills □ 1 □ 2 □ 3 □ 4 □ 5
4. Public speaking ability □ 1 □ 2 □ 3 □ 4 □ 5
5. Ability to adapt communication style □ 1 □ 2 □ 3 □ 4 □ 5

Written Communication 6. Grammar and syntax usage □ 1 □ 2 □ 3 □ 4 □ 5

7. Writing clarity and structure □ 1 □ 2 □ 3 □ 4 □ 5
8. Written argument development □ 1 □ 2 □ 3 □ 4 □ 5

B. Problem-Solving Skills Assessment (Scale: 1-5)

Analytical Thinking

1. Problem identification ability □ 1 □ 2 □ 3 □ 4 □ 5
2. Data interpretation skills □ 1 □ 2 □ 3 □ 4 □ 5
3. Critical analysis capability □ 1 □ 2 □ 3 □ 4 □ 5

Solution Development 4. Creative approach to problems □ 1 □ 2 □ 3 □ 4 □ 5

5. Decision-making process □ 1 □ 2 □ 3 □ 4 □ 5
6. Implementation planning □ 1 □ 2 □ 3 □ 4 □ 5

C. Leadership Skills Assessment (Scale: 1-5)

Team Dynamics

1. Group collaboration ability □ 1 □ 2 □ 3 □ 4 □ 5
2. Conflict resolution skills □ 1 □ 2 □ 3 □ 4 □ 5
3. Initiative taking □ 1 □ 2 □ 3 □ 4 □ 5

Management Skills 4. Project coordination ability □ 1 □ 2 □ 3 □ 4 □ 5

5. Resource allocation skills □ 1 □ 2 □ 3 □ 4 □ 5
6. Team motivation capability □ 1 □ 2 □ 3 □ 4 □ 5

SECTION 4: SITUATIONAL JUDGMENT ASSESSMENT

Instructions: Read each scenario carefully and provide your response. Your answers will be evaluated based on effectiveness, reasoning, and approach.

1. **Group Project Scenario** You are assigned to lead a group project, but one team member consistently misses deadlines. How would you handle this situation?

   _____

   _____

   _____

2. **Communication Challenge** During a presentation, you realize you've made a significant error in your data. What would you do?

   _____

   _____

   _____

3. **Problem-Solving Scenario** You discover a more efficient way to complete a task, but it differs from the established procedure. How would you proceed?

SECTION 5: EVALUATOR'S ASSESSMENT

A. Overall Assessment Scores

1. Communication Skills Total: ___/40
2. Problem-Solving Skills Total: ___/30
3. Leadership Skills Total: ___/30

B. Evaluator's Comments

Strengths:

Areas for Improvement:

C. Final Recommendation

□ Strongly Recommend □ Recommend □ Recommend with Reservations □ Do Not Recommend

SECTION 6: AUTHENTICATION

Evaluator's Name: _____ Position: _____ Date: _____ Signature: _____

A2. SCORING GUIDE

Communication Skills Scoring (40 points total)

- Verbal Communication: 20 points
  - Clarity (5 points)
  - Organization (5 points)
  - Listening (5 points)
  - Adaptation (5 points)
- Written Communication: 20 points
  - Grammar/Syntax (7 points)

    - Structure (7 points)
    - Content (6 points)

Problem-Solving Skills Scoring (30 points total)

- Analytical Thinking: 15 points
    - Problem Identification (5 points)
    - Data Interpretation (5 points)
    - Critical Analysis (5 points)
- Solution Development: 15 points
    - Creativity (5 points)
    - Decision Making (5 points)
    - Implementation (5 points)

Leadership Skills Scoring (30 points total)

- Team Dynamics: 15 points
    - Collaboration (5 points)
    - Conflict Resolution (5 points)
    - Initiative (5 points)
- Management Skills: 15 points
    - Coordination (5 points)
    - Resource Management (5 points)
    - Team Leadership (5 points)

Overall Score Interpretation

- 90-100: Exceptional Candidate
- 80-89: Strong Candidate
- 70-79: Satisfactory Candidate
- 60-69: Marginal Candidate
- Below 60: Does Not Meet Requirements

Version: 2024.1

Last Updated: November 2024

For official use only.