An Explainable AI Model to Predict Financial Exclusion in Kenya

By

Wamalwa Lindah Kelida

169589

Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Data Science and Analytics at Strathmore University

Institute of Mathematical Sciences and iLab Africa

Strathmore University

Nairobi, Kenya

June, 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: **Wamalwa Lindah Kelida**

Sign: _____ Date: 25/05/2025

Approval

The dissertation of **Wamalwa Lindah Kelida** was reviewed and approved for examination by the following:

Dr. John Olukuru,
Senior Lecturer, Data Science and Analytics,
Strathmore University

Dr. Godfrey Madigu,
Dean, Institute of Mathematical Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Financial exclusion remains a significant barrier to economic development and social equity, particularly in emerging economies. This study employed an explainable machine learning framework to predict financial exclusion in Kenya using nationally representative survey data from 2016 and 2021. The research followed the Knowledge Discovery in Databases (KDD) process, incorporating robust feature engineering techniques to derive behavioral and demographic indicators from raw survey data. Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting were evaluated under two experimental scenarios: baseline training and synthetic minority oversampling (Synthetic Minority Oversampling Technique (SMOTE)) to address class imbalance. A temporal validation strategy was implemented by training models on the 2016 dataset and testing on 2021 data to assess generalizability over time. Feature selection using Random Forest importance and SelectFromModel was applied to reduce dimensionality, while model interpretability was achieved through SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) explainability techniques. The optimized Decision Tree model achieved the highest F1-Score (Harmonic mean of precision and recall) (F1) score of 0.926, followed closely by a soft-voting ensemble F1 of 0.906. Behavioral indicators—particularly financial engagement, product category diversity, and digital finance adoption—emerged as stronger predictors than demographic variables. The resulting framework not only predicted exclusion risk with high accuracy but also provided transparent, interpretable insights for policy design. The findings offered actionable recommendations to government agencies, Non-Governmental Organization (NGO)s, and financial institutions aiming to improve inclusive finance strategies in Kenya and similar socio-economic contexts.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

AI Artificial Intelligence

API Application Programming Interface

AUC Area Under the Curve

CBK Central Bank of Kenya

F1 F1-Score (Harmonic mean of precision and recall)

FP False Positive

FPR False Positive Rate

FSD Financial Sector Deepening Kenya

GDP Gross Domestic Product

KDD Knowledge Discovery in Databases

KNBS Kenya National Bureau of Statistics

LIME Local Interpretable Model-agnostic Explanations

MAR Missing At Random

MCAR Missing Completely At Random

ML Machine Learning

NGO Non-Governmental Organization

REST Representational State Transfer

REST API Representational State Transfer Application Programming Interface

ROC Receiver Operating Characteristic

SHAP SHapley Additive exPlanations

SMOTE Synthetic Minority Oversampling Technique

SVM Support Vector Machines

TAM Technology Acceptance Model

TP True Positive

TPR True Positive Rate

XAI Explainable Artificial Intelligence

Acknowledgements

First and foremost, I thank God the Almighty for His grace, wisdom, and strength throughout this academic journey. I would like to express my deepest gratitude to my family for their unwavering support since the beginning. Their encouragement, patience, and belief in me have been my constant source of motivation.

Much appreciation to Dr. John Olukuru, for his invaluable guidance, insightful feedback, and continuous support throughout the development of this work. His expertise and mentorship have significantly shaped both this research and my academic growth. My gratitude extends to Strathmore University, particularly the Institute of Mathematical Sciences (SIMS), and the @iLabAfrica research centre. Their commitment to delivering world-class education and imparting practical and technical knowledge has been essential in enriching my academic and professional growth.

Lastly, I extend my thanks to all faculty, colleagues, and friends who provided support, inspiration, and motivation throughout my studies.

Chapter 1: Introduction

1.1 Background

The definition of financial exclusion presented by (Kempson and Whyley, 1999) describes it as the barriers which prevent disadvantaged and poor social groups from accessing financial services. Financial inclusion represents the process of making sure people and businesses receive cost-effective financial solutions and suitable financial products according to (World Bank, 2014). The economic sector and poverty relief depend heavily on these two fundamental concepts which experts have studied extensively. The initial research on financial inclusion showed that formal banking and microfinance help expand basic financial service access for low-income communities according to (Beck et al., 2007). The research foundation established the necessary base for future poverty reduction and economic inclusion programs conducted by governments and non-governmental organizations (NGOs).

The development of financial inclusion research has identified enduring difficulties according to work by Zins and Weill (2016) and Suri and Jack (2016) whose findings show how poor financial know-how and unstable income levels as well as regional remoteness continue to prevent numerous groups from making full use of available financial resources. The six elements of inclusion interact with each other according to Figure 1.1 from (Lyons et al., 2017) to determine financial inclusion levels among households. The research field has expanded to include digital financial services such as mobile money and digital loans which lower the need for physical banking infrastructure according to (Gabor and Brooks, 2017) within the same time period.

Recent studies examine functional solutions together with policy recommendations. Demirgüç-Kunt et al. (2018) proved that standard banking institutions fail to provide sufficient services for marginalized communities who face limited bank access in rural areas. The extensive digital ecosystem growth has made mobile money platforms take over the role of connecting virtual financial services which deliver better accessibility and affordability to impoverished populations. The identification of financially excluded individuals' profiles and behaviors stands essential for developing inclusive strategies according to (El-Zoghbi et al., 2019). Research investigations demonstrate the urgent necessity to develop specific sectoral policies backed by programs that analyze complex socio-economic elements and geographic characteristics which enable exclusion. Government agencies together with NGOs must understand the complete net-

Figure 1.1: The six elements of inclusion.

work of factors to develop interventions which solve particular barriers opposing marginalized communities. These efforts need evidence-based information as a guiding principle to deliver financial services toward specific groups in need which leads to inclusive economic growth that actively combats societal inequalities effectively.

1.1.1 Global Perspective on Financial Inclusion and Technology

Globally financial inclusion represents a cornerstone element which stimulates economic expansion and reduces poverty levels. Research work from the beginning emphasized that restricted access to financial services both destroys economic movement possibilities and makes income gaps worse (Beck et al., 2007). The first studies examined banks and microfinance institutions as traditional solutions to expand financial service access including credit and savings. The challenges continued especially in rural territories which were restricted because communities lacked basic infrastructure and financial understanding. The introduction of technological advancements produced new possibilities while (Suri and Jack, 2016) demonstrated how mobile money services helped Kenyan rural populations maintain their finances and utilize essential services through mobile devices. The financial inclusion movement underwent a fundamental shift when digital and mobile alternatives started to support standard banking services to provide access to communities who lacked financial services previously.

The absence of sufficient conventional banking infrastructure led (Gabor and Brooks, 2017) to report that mobile money services and microloans started to develop as alternative financial systems. Digital solutions grew as the financial sector underwent technological change which allowed marginalized people to handle their finances even when banks were distant and income was inconsistent. The concept of financial inclusion evolved during recent years by adding two essential attributes to basic financial service availability namely equal access regardless of demographic factors and affordable prices for all. The usage of mobile money services in developing economies led to a rise in account ownership from 63% to 71% during the period from 2017 to 2021 as Figure 1.2 illustrates.

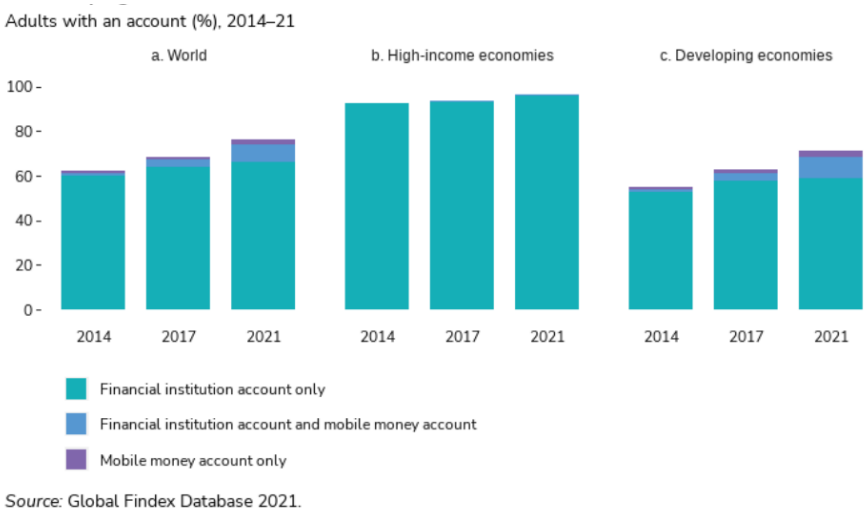


Figure 1.2: Adults having a bank account than before.

The global financial inclusion effort aimed at discovering concrete obstacles underserved populations encounter because policies needed customization for distinct groups to develop a genuinely inclusive economic system according to (El-Zoghbi et al., 2019). Research has recently analyzed the multiple barriers which still exist even though digital financial access has improved. Digital financial inclusion rates remain low because several obstacles such as limited digital competency and cultural resistance and erratic incomes continue to limit access for impoverished communities according to (Nayak et al., 2020). The researchers support increased knowledge of these exclusion factors to optimize policy development. Both technological advances and structural changes to social barriers demand continuous targeted research for developing policies which will lead to inclusive financial systems across the population. The insufficient conventional banking infrastructure in certain areas led to the development of alternative systems including mobile money services and microloans according to (Gabor and

Brooks, 2017). The rise of digital solutions demonstrated a technological transformation across the financial sector that enabled underprivileged people to manage their finances even when banks were distant and their income was unstable. Financial inclusion started as a concept about service accessibility, but has evolved to include fair and economical access that reaches all demographics. The percentage of adults with account ownership increased from 63 percent to 71 percent in developing economies during 2017 to 2021 due to mobile money services, as Figure 1.2 demonstrates.

1.1.2 Financial Inclusion in Sub-Saharan Africa and the Challenges of Digital Adoption

The two regions of Sub-Saharan Africa and South Asia have achieved noteworthy progress in financial inclusion through mobile banking yet they and other regions maintain high levels of financial exclusion because only 43% and 55% of adults have a formal financial account by 2017 while the global average stands at 69%. The initial research focused on examining the basic roadblocks to traditional banking in rural Africa which showed that poor infrastructure limitations and low financial capability deteriorated access to services for underprivileged individuals. (Beck et al., 2007) recognized these impediments as the main sources of economic gaps because vast rural populations were lacking in formal banking services. Economic conditions in the region demanded new solutions which would overcome both geographical and infrastructural constraints.

Mobile money services launched in countries like Kenya and Tanzania produced a significant development that allowed populations without bank access to obtain financial management tools according to (Mbiti and Weil, 2011). Mobile money adoption grew to serve millions of customers within its brief introduction period thus allowing new groups to access formal banking services. In expanding upon this foundation Suri and Jack (2016) analyzed how mobile money affected poverty reduction through their analysis of enhanced risk management and shock resilience among households that used mobile money. Households leveraging mobile money reported increased financial resilience according to their study whereas mobile banking provided direct economic stability to Kenya. Mobile banking solved specific monetary needs but mainly focused on financial operations because it lacked complete financial services including credit and insurance.

Research showed that rural populations continued using informal financial practices while show-

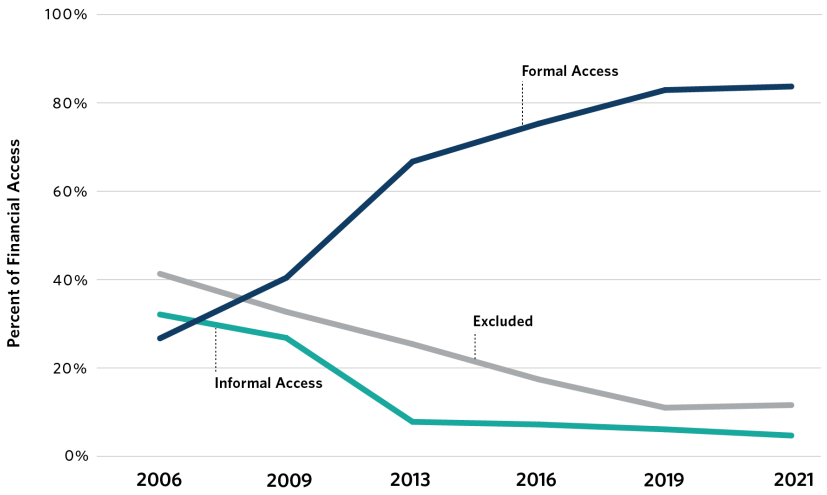
ing limited understanding of formal financial benefits. (Dube, 2018) His Zimbabwean research discovered that low financial literacy together with formal institution distrust prevented mobile banking from achieving its full potential. The study conducted by (Parlasca et al., 2022) discovered that mobile money adoption increased basic services availability yet cultural and socioeconomic obstacles prevented deep financial inclusion through insurance and credit.

The study by (Kotzinos et al., 2023) established that differences in obtaining full financial service capabilities continue to exist. The authors compared Kenya and Ghana to show that mobile technology enables basic financial operations but many people continue to depend on informal credit and insurance networks. The authors emphasize that solving financial literacy problems and adapting products to fit various community requirements will help close the financial inclusion gap. Mobile banking has transformed financial access in Sub-Saharan Africa but ongoing disparities show that better inclusive approaches must be developed to reach all demographic groups.

1.1.3 The Role of Mobile Money and FinTech in Expanding Financial Inclusion in Kenya

The international community acknowledges Kenya for its world-leading achievements in financial inclusion which stem from M-Pesa mobile money systems success. M-Pesa has revolutionized financial service delivery to Kenyan citizens especially those living in rural areas and low-income regions through credit and savings options as well as payment solutions as illustrated in Figure 1.3.

Kenya assumes its position as a digital financial services leader because mobile money adoption reached over 80% within the adult population by 2020 according to (Suri and Jack, 2016). Literacy programs designed properly reduce financial exclusion rates by 20% among rural women who traditionally lack proper service (Dube, 2018). Disparities persist for female-headed households benefiting from mobile money though they achieved better economic stability and financial capabilities (Van Hove and Dubus, 2019). The research by Kiptorus (2019) states that mobile money adoption is widespread but formal lending and insurance solutions remain ill-used especially among vulnerable groups in Kenya. Women alongside rural residents demonstrate limited participation in formal savings accounts since they represent just 34% and 38% of users respectively while most people depend on informal financial systems lacking protected banking facilities.



Source: Source: "Financial Access," FinAccess Household Survey, Kenya National Bureau of Statistics, https://finaccess.knbs.or.ke/access.

Figure 1.3: Financial Inclusion Measured by Access in Kenya

Research by (Wanjohi, 2020) demonstrates that either insufficient financial knowledge combined with limited collateral prevents people from maximizing available mobile money services although adoption is high. Simultaneously (Wabwire, 2020) showed predictive algorithms can direct financial literacy programs which produced a 15% increase in formal financial product use by the informal sector users. Current research shows that smallholder farmers together with artisans maintain their dependencies on informal savings groups called chamas because they encounter notable challenges when attempting to use formal credit systems (Rodima-Taylor, 2022). (Kamunyu, 2022) has discovered more than 40

Research now demonstrates how machine learning technology helps identify and solve financial exclusion problems in Kenya. Machine learning analytic techniques use population data to forecast which groups face the highest exposure to exclusion resulting in valuable information for NGO and governmental agencies to implement strategies (Ngumi, 2022). The worldwide appreciation for Kenya's financial inclusion achievements must be balanced with efforts to deliver needed support for marginalized populations through specific programs that enable their full participation.

1.2 Problem Statement

Due to its wide adoption of M-Pesa and similar mobile money services Kenya is celebrated worldwide as a leader in financial inclusion yet major barriers remain to delivering equal access to financial services. The research by (Suri and Jack, 2016) showed that M-Pesa together with similar platforms successfully linked numerous Kenyans particularly those living in rural areas to basic financial services including savings and payments and credit. Studies conducted by (Kiptorus, 2019) and (Van Hove and Dubus, 2019) showed that women together with young individuals and informal sector employees continue to face financial exclusion because they lack access to formal credit and insurance and investment products. The mobile money system is operational but various population segments encounter ongoing obstacles that block their complete financial inclusion.

Research by (Wanjohi, 2020) demonstrated that financial literacy programs run by governments and NGOs have achieved positive results yet the main issue involves identifying and resolving various complex barriers which separate different population groups from accessing complete financial services. The researchers demonstrated through their research that demographic-specific needs demand targeted interventions. The investigation conducted by (Rodima-Taylor, 2022) and (Kamunyu, 2022) advanced the research by uncovering particular obstacles which prevent marginalized communities from using digital financial platforms adequately including insufficient financial knowledge and the lack of collateral and traditional cultural elements.

These studies found that though community-based savings groups (chamas) offer financial support to members they lack formal institutional reliability which hinders the economic progress of users who depend on them. Surrounding economic challenges remain persistent despite technological progress because this demonstrates the vital requirement to use data analytics techniques for understanding exclusion factors. Advanced analytics together with machine learning algorithms create new routes to understand financial exclusion more effectively through risk detection and personalized educational programs and specialized intervention development. The ongoing financial exclusion of marginalized demographics in Kenya requires an immediate development of an integrated data-based strategy to evaluate and resolve this situation. Associate governments with specific knowledge of distinctive group needs and barriers will continue to fail in reaching their desired outcomes.

1.3 Objectives

1.3.1 Main Objectives

To create a data-driven framework that evaluates financial exclusion in Kenya, identifies key barriers faced by marginalized groups, and provides actionable insights for government agencies and NGOs to design targeted, inclusive financial policies and programs.

1.3.2 Specific Objectives

i. Identify key demographic, socioeconomic, and behavioral factors contributing to financial exclusion in Kenya using machine learning.

ii. Develop a machine learning model to predict financial exclusion and help target at-risk populations.

iii. Incorporate explainable Artificial Intelligence (AI) techniques to provide transparency into exclusion factors, enhancing the decision-making capabilities of government agencies and NGOs.

1.3.3 Research Questions

i. How can machine learning identify the key demographic, socioeconomic, and behavioral factors that influence financial exclusion in Kenya?

ii. What machine learning models can best predict financial exclusion and identify at-risk populations?

iii. How can explainable AI (XAI) enhance the transparency of machine learning models and help policymakers understand exclusion drivers for targeted interventions?

1.4 Scope and Limitations

1.4.1 Scope

The study investigates Kenyan financial exclusion patterns by analyzing the FinAccess Surveys from 2016 and 2021 through comparative assessment. The research combines analysis of two datasets to provide long-term insights about the ongoing and emerging challenges which prevent financial access across different demographic groups in Kenya. The research employes machine

learning algorithms to identify essential elements that cause financial exclusion by studying demographic variables and socioeconomic characteristics and behavioral patterns. The research useds explainable AI techniques to make complex model predictions more understandable for stakeholders who focus on financial inclusion promotion.

1.4.2 Limitations

The analysis of the FinAccess surveys from 2016 and 2021 faces methodological barriers because of structural variations and modification of data collection techniques. Standardization methods may reduce analytical precision and the delay between surveys overlookes new economic effects. The imputation process of missing data from the 2016 dataset creates potential bias because of its nature. The study's focus on Kenya restricted its wider applicability while the advanced analysis techniques may make results difficult for non-technical stakeholders to understand through explainable AI methods.

1.5 Justification

The research undertakes an investigation of financial exclusion patterns in Kenya through comprehensive time-based data collection which addresses a fundamental knowledge deficit. The digital financial leadership status of Kenya provides an optimal environment to analyze financial accessibility trends along with enduring barriers to service. The analysis of FinAccess Surveys in 2016 and 2021 identifies patterns between financial behavior changes and financial literacy alongside socioeconomic factors and location while revealing which population groups persist in financial exclusion and their associated reasons.

This investigation produces results which extend academic value into tangible application for policy development. The study enables government agencies and NGOs to use explainable AI analysis results for creating purposeful financial inclusion programs. The recommendations based on evidence help Kenya achieve economic empowerment and reduce poverty through their efforts to close financial inclusion gaps for vulnerable populations. The analytical framework uses data to achieve two goals: it enhances financial inclusion comprehension and delivers operational guidance for stakeholders who aim to drive inclusive economic development across Kenya.

Chapter 2: Literature Review

2.1 Theoretical Review

The theoretical review investigates existing theoretical frameworks which help understand financial exclusion and inclusion processes alongside technological solutions for barrier removal. The review follows three major sections that match the research objectives and questions by examining financial inclusion theory and financial service technology adoption and predictive modeling with Explainable Artificial Intelligence (XAI). The sections of this review expand academic research to understand financial exclusion elements while exploring predictive modeling applications and explainable systems which help policy development.

2.1.1 Theories of Financial Inclusion

Research on financial inclusion investigates both the elements which determine service accessibility and usage patterns alongside the social and economic effects of financial engagement. The early field of financial inclusion research focused on poverty reduction through financial products such as credit and savings and insurance since these elements stabilize economic conditions and protect against financial risks. (Beck et al., 2007) created one of the fundamental financial inclusion theories which connects it to economic progress and poverty elimination. The authors demonstrated that inclusive growth occurs through financial services which enable people to protect themselves from risks and build business investments and strengthen family stability. The model demonstrates how financial access fuels economic engagement especially for disadvantaged households in line with the research goal to analyze financial inclusion benefits for marginalized communities.

The study of financial exclusion received further development through (Demirgüç-Kunt and Klapper, 2012) who investigated worldwide elements that lead to financial exclusion. The authors defined access and usage as separate elements which demonstrated the difference between service availability and actual utilization. This fundamental distinction helps researchers understand which demographic and behavioral factors prevent people from accessing available services thus responding to the research inquiry. (Beck and Cull, 2015) developed the framework by showing that financial inclusion needs a dual system of supply and demand support. The model demonstrates the necessity of developing financial solutions specific to underserved

communities while matching their unique needs thus supporting the goal of developing policy recommendations for NGO and government agency programs.

The paper by Allen et al. (2016) investigated how structural elements with socio-economic characteristics such as income levels and education and geographic location determine financial access. Socioeconomic research revealed that people with lower incomes across developing world areas especially Sub-Saharan Africa encounter exclusion because of structural barriers involving poor financial knowledge together with weak banking system infrastructure. The study provides essential knowledge for this research because it identifies key socio-economic factors which will be analyzed in the data evaluation process. In their exploratory research of African financial inclusion Zins and Weill (2016) established that financial literacy and mobile money adoption represent fundamental elements to eliminate exclusion barriers. The research shows that mobile money services create connections between dispersed regions and inadequate infrastructure which serves as a key element for this study focusing on Kenya. The research showed that financial literacy plays a vital role in enhancing the adoption of formal credit and insurance despite finding unequal usage rates between these financial products.

2.1.2 Technology Acceptance Model (TAM)

Researchers analyze technology adoption in financial services through theoretical frameworks which describe individual patterns of adopting new tools. The Technology Acceptance Model (TAM) developed by (Davis, 1989) stands as one of the first and most important models that shows perceived usefulness and ease of use determine new technology adoption behavior. The Technology Acceptance Model (TAM) enables researchers to understand why people adopt digital financial tools through mobile banking because it directly addresses the study's main goal of discovering exclusion-related behaviors. (Venkatesh and Davis, 2000) developed an expanded TAM model that added subjective norms and user intentions as adoption factors. Social and cultural attitudes play a crucial role in determining the adoption barriers faced by Kenyan marginalized groups when they attempt to use formal financial services. The adoption of financial technologies in Kenya becomes more understandable through (Rogers, 2003)'s Diffusion of Innovations Theory which extends the existing acceptance models. Rogers developed the theory by defining early adopters and laggards who show varying speeds of innovation adoption based on their demographic groups.

Researchers usually evaluate technological adoption patterns in financial services by using theoretical frameworks which describe adopter behavior toward new tools. The Technology Acceptance Model (TAM) established by (Davis, 1989) operates as one of the first critical models to suggest that perceived usefulness and ease of use serve as primary drivers for new technology adoption. The Technology Acceptance Model serves the research because it shows how individuals perceive mobile banking while helping to understand what motivates them to adopt digital financial tools for financial inclusion studies. (Venkatesh and Davis, 2000) enhanced TAM by adding subjective norms and user intentions into the model. The barriers that prevent Kenyan marginalized groups from using formal financial services become more understandable through this model because social and cultural attitudes heavily impact their willingness to adopt such services. The Diffusion of Innovations Theory by (Rogers, 2003) expands understanding about Kenyan financial technology adoption beyond acceptance models. Rogers integrated two user types called early adopters and laggards while demonstrating how demographic factors drive their pace in innovation acceptance.

2.1.3 Theories in Predictive Modeling and Explainable AI (XAI)

Predictive modeling and explainable AI (XAI) offer theoretical frameworks for understanding complex data-driven insights in financial services. The theories enable the creation of predictive algorithms that detect financial exclusion while offering clear model prediction transparency which serves as a vital requirement for this study because it provides actionable insights to NGOs and government agencies. Breiman (2001) techniques used to predict financial exclusion find their theoretical basis in this model because they excel at processing data relationships which matches the study's objective to determine vulnerable populations through various socioeconomic indicators.

The development of interpretable machine learning introduced new frameworks which enable model prediction understanding. Ribeiro et al. (2016) Local Interpretable Model-Agnostic Explanations (LIME) presented a model-agnostic interpretability approach that generates locally interpretable surrogate models for individual predictions. The study utilizes LIME because it enables the explanation of complex ML models through instance-level explanations of predictions for high-risk exclusion groups. SHAP (Shapley Additive Explanations) provides interpretability through cooperative game theory by quantifying feature contributions to predictions according to Lundberg and Lee (2017). SHAP values deliver clear explanations for policy

applications because they offer transparency to NGOs and government agencies who need to understand exclusion factors.

Research has recently investigated XAI applications for financial use cases because (Carvalho et al., 2019) demonstrated that financial stakeholders need transparent algorithmic decision systems to build stakeholder trust. The research results directly apply to this investigation since NGOs and government agencies need interpretability to create and validate exclusion-reducing financial interventions. The application of XAI received additional advancement when (Babaei and Giudici, 2021) conducted an evaluation of machine learning methods for credit scoring in financial inclusion settings. The study revealed how explainable models reveal important exclusion factors which helps policy makers improve their application. The research goal matches with this perspective because it seeks to create specific financial literacy program recommendations and policy suggestions for underserved Kenyan populations.

2.2 Empirical Review

The empirical review examines different Machine Learning (ML) applications in financial inclusion studies by analyzing their models together with evaluation metrics and analyzed features. The research goals of this study match the findings from these studies which examine financial exclusion factors and develop predictive models while using XAI to assist NGOs and policymakers in Kenya. The analysis section presents foundational research and model evolution and performance measurement which builds a complete understanding of empirical ML applications for financial inclusion.

2.2.1 Identifying Exclusion Factors Using Machine Learning and Demographic Analysis

Research focused on financial exclusion factors depends on extensive demographic and socioeconomic and behavioral analysis using machine learning to process large datasets and generate practical findings. (Demirgüç-Kunt and Klapper, 2012) pioneered analysis of Global Findex data to study access patterns based on income, education and location. The researchers documented how rural and low-income populations face exclusion because of socio-economic barriers which requires specific intervention strategies. The study by (Sarma and Pais, 2011) examined the connection between financial inclusion levels and Gross Domestic Product (GDP) and human development indicators throughout different regions. The study authors conducted a cross-sectional analysis which demonstrated that rural populations with lower incomes and

education levels frequently experience exclusion from financial services thus validating the research goal to discover socio-economic exclusion factors in Kenya.

The paper by (Allen et al., 2016) studied financial inclusion in 150 countries by analyzing demographic and economic survey data to evaluate access to savings and credit alongside payment services. Financial literacy combined with geographic distance to financial institutions presented major barriers that primarily affected rural areas according to their research findings. The research question on financial exclusion and demographic factors received support from this study as well as from (Cull et al., 2014) who analyzed Global Findex Database data that included age, education, gender, and employment variables. The research revealed substantial exclusion rates among women and young people which demonstrates that intervention strategies need to recognize population differences. This research study will include the essential risk prediction elements identified by their work. Zins and Weill (2016) studied African financial inclusion through mobile money adoption as a main inclusion driver. This study revealed that gender and education level form consistent barriers which prevent access to insurance and formal credit. The research goal matches the local approach because it seeks to detect particular demographic elements that influence financial inclusion in Kenya. (Grohmann et al., 2018) studied worldwide financial inclusion determinants and discovered that age together with marital status and employment type influence service access. Logistic regression analysis in their research helped determine exclusion predictors which matches the study objective to identify exclusion determinants through data-driven methods. (Naceur et al., 2017) applied the same statistical method to cross-national data to study how economic variables like inflation and education levels influence financial access. Research demonstrates that lower financial literacy consistently acts as a barrier which demonstrates the need for financial education to make inclusion strategies effective.

2.2.2 Predictive Modeling for Identifying At-Risk Populations

Research shows machine learning models excel in predicting financial exclusion while also detecting populations that face high risk. (Barboza et al., 2017) analyzed the predictive power of logistic regression and decision trees and ensemble models on business financial distress. Their analysis showed ensemble models particularly random forests delivered superior prediction accuracy when revenue and business size and credit history were used as variables. The research objective of exploring ensemble models for financial exclusion prediction receives

support from this study which builds on previous findings. Neural networks struggle to be used in policy-driven research because their lack of interpretability makes them unsuitable for studies requiring transparency. The objective of choosing proper predictive models matches with this requirement.

The research by (Kamran and Shah, 2019) evaluated Support Vector Machines (SVM) and Gradient Boosting methods to forecast financial vulnerability through analysis of household income and family size together with savings patterns. The authors demonstrated high predictive accuracy with Support Vector Machines while integrating behavioral information to achieve superior results based on their findings about combining demographic and behavioral data for reliable predictions. (Agarwal et al., 2021) applied decision trees and logistic regression with income, education, and employment type data to forecast financial exclusion in emerging markets. Decision trees deliver exceptional interpretability according to their research findings thus making them suitable for applications that demand transparent model prediction methods. The handling of complex data interactions by decision trees remains limited which indicates ensemble models should be considered.

The authors from (Ngumi, 2022) evaluated ensemble learning approaches like boosting and random forests for East African financial exclusion prediction. Research findings showed rural residents along with those working in informal sectors face the highest risk of financial exclusion through the assessment of income and education data and employment statistics. The research uses ensemble models to predict exclusion within marginalized populations in Kenya and demonstrates their practical value for this purpose. The research demonstrated that clustering as an approach successfully divided populations into segments which allowed more specific intervention strategies. The predictive capabilities of other models surpass clustering because it needs alternative machine learning approaches to forecast exclusion. (Xiao and Porto, 2020) applied logistic regression to evaluate financial exclusion in Latin America through examinations of income level, digital literacy, and employment. The study identified exclusion patterns among informal sector employees which holds significant value for Kenya's economy because its workforce primarily operates in the informal sector.

2.2.3 Explainable AI for Transparency and Targeted Financial Interventions

XAI has become essential for predictive model transparency because LIME provides model-independent explanations of single predictions through local model approximations according to Ribeiro et al. (2016). LIME proved its worth in financial applications by providing model predictions which policymakers could understand. This study aims to bridge the gap in financial inclusion applications of SHAP values for feature attribution in complex models which (Lundberg and Lee, 2017) originally developed. Stakeholders can use SHAP values to evaluate the significance of individual predictors thus making this method appropriate for models that need both global and local explanations. The research objective matches with the need to implement XAI solutions that reveal exclusion factors.

A detailed XAI review by (Arrieta et al., 2020) reveals that explainable models help stakeholders develop trust and transparency especially within the finance sector. SHAP and LIME proved useful for financial inclusion according to research findings while empirical studies on XAI applications in financial exclusion remained scarce thus supporting the present work that integrates XAI in exclusion analysis. (Babaei and Giudici, 2021) demonstrated SHAP and LIME in credit scoring to show XAI provides clear identification of exclusion drivers for marginalized groups. The authors investigated income and education and employment data through credit scoring but their research excluded broader financial inclusion applications. (Addy et al., 2023) applied SHAP to analyze financial exclusion among African populations to show how predictive transparency supports policymakers who want to overcome regional barriers. The study remained exploratory because it failed to deliver concrete intervention steps despite showing the need for XAI solutions that provide direct policy recommendations.

SHAP analysis was used by (Schlegel et al., 2020) to improve the interpretability of neural networks during financial time series prediction tasks. The research demonstrated explainable models outperform other approaches in financial domains since stakeholders need transparent predictions to support their decisions. The research demonstrates the importance of XAI methods for achieving its goal of supporting policymakers through decision-making. The survey by (Carvalho et al., 2019) stressed that applications involving ethical transparent decisions need to be interpretable. The survey results from their research demonstrated SHAP values and LIME techniques as the most suitable for financial applications because they generate explanations at the instance level.

2.2.4 Conclusion

The empirical review shows that current research contains several unsolved gaps. Various studies have identified demographic and socioeconomic elements that lead to financial exclusion but these studies lack sufficient behavioral data from specific Kenyan regions which reduces their relevance to the country. Predictive modeling helps identify risks of exclusion yet fails to deliver clear explanations that policy-makers need to make decisions. The application of XAI techniques has improved financial application transparency yet researchers have conducted minimal studies regarding its use for financial inclusion in Kenyan conditions. The research establishes the necessity to build an analytical framework based on data which addresses Kenya's financial inclusion requirements through clear explanations for government and NGO decision-makers.

2.3 Gaps

The existing research about financial inclusion has shown clear developments in both understanding the issue and developing solutions yet it has substantial unaddressed gaps which this researchsmtp addresses. Research has uncovered demographic factors alongside socioeconomic elements that cause financial exclusion yet scarce investigations incorporate complete behavioral research. Studied research uses demographic information along with static socio-economic measurements while omitting key behavioral data which includes transaction patterns and savings habits as well as digital financial service adoption. These research efforts fail to detect changing financial behavior patterns because they study areas where mobile money use is dominant such as Kenya. This study addresses this research gap through its combined approach of demographic analysis with behavioral information which allows better identification of exclusion factors and suitable financial product development options.

Financial inclusion research suffers from a substantial deficiency regarding the utilization of predictive modeling techniques. The widespread use of logistic regression and decision trees as predictive models exists primarily because of their clear interpretability but such models usually come with restricted predictive capabilities when dealing with complex multi-variable relationships. Despite their predictive effectiveness advanced machine learning methods including ensemble models and neural networks rarely get used to analyze specific financial inclusion relationships in Kenya. The lack of interpretability in ensemble models used by certain

studies creates barriers for policymakers and NGOs to effectively use the study findings. The research develops specialized ensemble models that serve both the Kenyan financial market requirements and the need for policy-making clarity through interpretation.

The literature development for XAI applications remains insufficient when applied to financial inclusion problems. XAI tools such as SHAP and LIME currently experience rising application to financial data yet most studies analyze credit scoring and financial risk assessment instead of broader financial inclusion problems. The restricted scope of application prevents researchers from understanding exclusion factors that affect underserved populations in Kenya because the region has unique socio-cultural and economic elements that drive financial exclusion. Through the use of XAI methods this study will produce clear explanations of machine learning model results which enables policymakers and NGOs to uncover the root causes of exclusion so they can create targeted interventions for particular demographic groups.

Some research studies include policy recommendations yet they lack concrete insights which stem directly from their data analysis. The studies recommend broad financial inclusion improvements yet they do not link their recommendations to the exclusion factors which emerged from the data analysis. Financial literacy programs are commonly recommended yet they lack customization to meet the particular requirements of marginalized communities who include women and rural workers and informal sector employees. This research seeks to fill the existing gap by creating data-based targeted recommendations which guide both NGOs and government agencies. These recommendations target Kenyan population-specific financial challenges to overcome barriers for inclusion and establish a practical guide for financial literacy growth and suitable service access.

2.4 Conceptual Framework

The research framework in Figure 2.1 demonstrates how key factors relate to predictive modeling within the study of financial inclusion in Kenya. The system uses machine learning models that process demographic and socioeconomic and behavioral information to detect and forecast exclusion risks. The framework generates data-based recommendations to tackle particular obstacles that prevent customers from accessing financial services.

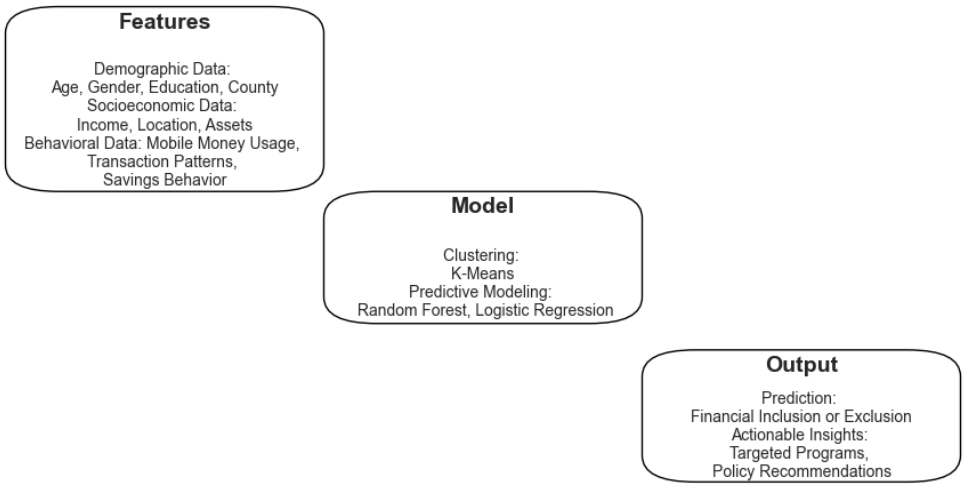


Figure 2.1: Conceptual Framework for Financial Inclusion

Chapter 3: Methodology

The research method follows the Knowledge Discovery in Databases (KDD) process for data science studies to extract valuable patterns from data. KDD provides meaningful insight extraction according to (Fayyad, 1997) through its systematic process. Data selection follows preprocessing then transformation and data mining before evaluation completes the process. The study's methodological framework for achieving its goals depended on the KDD process which is shown in Figure 3.1.

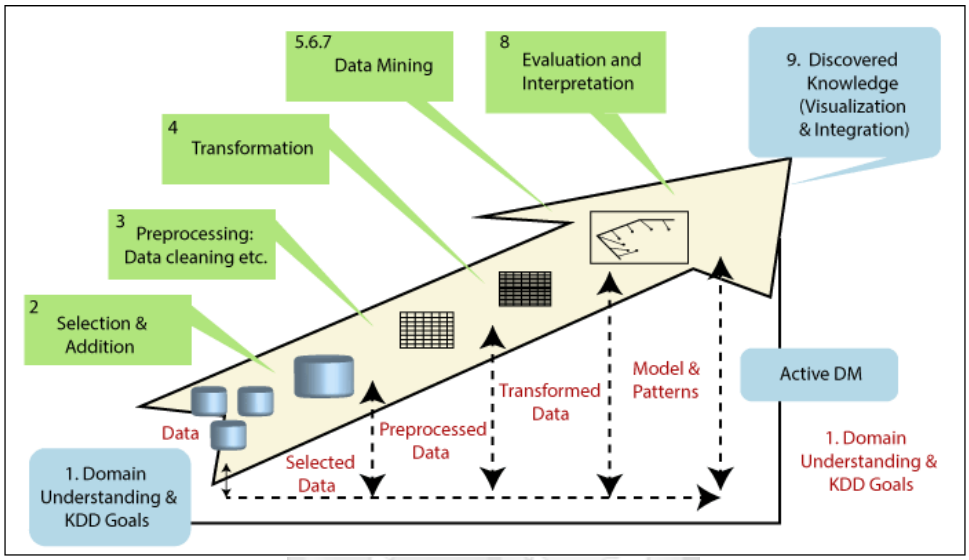


Figure 3.1: The nine-step KDD process

The KDD process directed the financial exclusion analysis of Kenya through each systematic phase. The chapter explains how each phase was executed starting from data selection and ending with knowledge evaluation. The chapter explains the machine learning models together with evaluation metrics used in the study by providing definitions and equations and citing established literature. The approach delivered a systematic method for conducting financial exclusion factor investigations between the two studied time periods.

3.1 Domain Understanding

The primary objective was to develop a machine learning model to predict financial exclusion in Kenya using data from the 2016 and 2021 FinAccess surveys, which track financial service usage by adults in Kenya. Financial Sector Deepening Kenya (FSD) undertakes the nationally representative surveys in partnership with the Central Bank of Kenya (CBK) and the Kenya National Bureau of Statistics (KNBS).

The domain knowledge for this research was informed by literature on the determinants of financial inclusion in Kenya, particularly how the digital financial services environment changed between 2016 and 2021. The KDD process enabled the systematic analysis of these financial activities and demographic variables to identify the most significant predictors of exclusion.

3.2 Data Selection

The dataset selection phase focused on identifying relevant variables from the FinAccess surveys that would provide insights into financial exclusion patterns. The 2016 dataset comprised 8,208 records, while the 2021 dataset contained 20,909 records, representing a significant expansion in survey reach.

3.2.1 Variable Selection

Variables were selected based on their relevance to financial inclusion research as established in the literature review. To ensure data consistency between the two time periods, a column mapping approach was employed to standardize variable names and formats across both datasets. Table 3.1 presents a sample of the key variable mappings between the 2016 and 2021 datasets.

The standardization procedure served as an essential requirement to achieve compatibility between different surveys. The 2016 survey contained variables with **e1_** prefixes yet the 2021 survey used **C1_** prefixes. The survey variables matched each other through their description and response option content present in the codebooks. The datasets underwent filtering to retain financially active adults who were 18 years old or older. Research standards in financial inclusion studies follow the practice of excluding minors from analysis since they lack independent financial capabilities Beck et al. (2007).

53 matching variables appeared in both datasets. The 2016 dataset included 64 variables yet the 2021 dataset contained 67 variables. The analysis excluded variables like *mobile_banking_registered*, *trusted_financial_provider*, and *highest_interest_rate* which existed solely in the 2021 survey because they lacked matches in both datasets for maintaining analytical consistency. The researchers needed to remove variables from analysis which did not match between the two survey periods to ensure proper temporal comparison. The datasets underwent standardization followed by unique variable removal which resulted in 53 common variables for preprocessing and analytical stages according to Zhao et al. (2023).

Table 3.1: Key Variable Mapping Between 2016 and 2021 FinAccess Surveys

Standardized Variable	**2016 Source Column**	**2021 Source Column**
Demographic Variables		
respondent_id	sbjnum	interview__id
age	age	A19
gender	gender_of_respondent	gender
education_level	education	education
residence_type	cluster_type	cluster_type
marital_status	a_9	marital
relationship_to_hh	a_10	A20
region	sub_region	County
population_weight	popwgt_raw	IndWeight
Financial Access Variables		
mobile_money_registered	e1_37	C1_9
mobile_banking_registered	e1_8	C1_10
bank_account_current	e1_31	C1_28
Savings Variables		
savings_microfinance	e1_2	C1_1
savings_mobile_banking	e1_8	C1_2
savings_sacco	e1_1	C1_4
Credit Variables		
loan_bank	e1_12	C1_11
loan_mobile_banking	e1_13	C1_12
loan_sacco	e1_14	C1_14
loan_microfinance	e1_15	C1_15
Insurance and Pension Variables		
insurance_nhif	e1_42	C1_42
insurance_health_other	e1_43	C1_43
insurance_life	e1_44	C1_44
pension_nssf	e1_46	C1_48
Financial Exclusion Indicator		
financially_excluded	e4_10	(Generated proxy)

3.2.2 Target Variable Definition

No specific financial exclusion indicator was present in the 2021 survey, so a proxy measure was defined by means of established financial inclusion indicators. In line with previous research Zins and Weill (2016); Demirgüç-Kunt et al. (2018), an individual was categorized as financially excluded when he reported zero current usage on the following fundamental financial services on *mobile money registration, insurance pension* .

Table 3.2: Education Level Standardization

Original Values	**Standardized Value**
Various	primary
Various	secondary
Various	tertiary
Other/Missing	unknown

Table 3.3: Marital Status Standardization

Original Values	**Standardized Value**
Various	married/living with partner
Various	single
Various	widowed
Various	divorced/separated
Other/Missing	unknown

3.3 Data Preprocessing

Data quality issues was addressed and the datasets wee made ready for analysis and modeling following some best practices that are presented by Han et al. (2012).

3.3.1 Standardization of Categorical Variables

All categorical variables followed in all datasets were standardized. It entailed convertiong all string values to lower to prevent contentions between categories due to mixed case.In mapping of variousy response formats to standardized values. Several categorical variables were differently coded in the two datasets. Similarly, they were standardized using a consistent classification system as shown in Tables 3.2 and 3.3.

County-level data from the 2021 survey was mapped to the provincial framework used in the 2016 survey to ensure geographic comparability.

Table 3.4: Financial Behavior Standardization

Original Values	**Standardized Value**
never used, never had, used to use, used to have	no
currently use, currently have	yes
missing, blank	missing

3.3.2 Missing Value Analysis

Due to missing value patterns, an analysis of the most appropriate imputation strategy was conducted. Missingness of education_level was Missing At Random (MAR), and this implies that there are patterns in non-response driven by age and population weight of wife. A small amount Missing Completely At Random (MCAR) of other variables had very few missing values ($< 1\%$), which is indicative of random patterns of missingness.

We then verified these pattern with Little's MCAR test and correlation analysis between missing value indicators and other variables in accordance to Little and Rubin (2019)'s methodologies.

3.3.3 Missing Value Imputation

Given these analyses of the missingness, various imputation strategies were tried. For education_level, which had MAR patterns, the mode within age groups and population weight segments was used for group based imputation using data from weight segments and age groups. According to Schafer and Graham (2002), this approach also meets the requirement of preserving the relationship between education, age, and sampling weight. If the variables are mostly nonmissing, then globally mode imputation was used, filling missing values with a most common category of the dataset. The remaining financial service usage variables had no response but by the assumption of that financial service usage has no response for financial service questions means that the conservative approach is used to impute *no*.

3.3.4 Outlier Detection and Treatment

Most of the outlier analysis was done on numerical variables (age and population_weight). Eq. 3.1 defines potential outliers and was applied to the interquartile range (IQR) method.

$$Q_1 - 1.5 \times IQR \quad \text{and} \quad Q_3 + 1.5 \times IQR \text{ are the lower bound and upper bound.} \tag{3.1}$$

If Q_1 and Q_3 are the first and third quartiles respectively and $IQR = Q_3 - Q_1$, then the statistic indicates. Rather than removing outliers, sample size was preserved and the influence of extreme values was minimized for the age variable, which was capped at the upper and lower bounds. For population_weight, there were no capping because the weights were designed exclusively so that the sample was representative Solon et al. (2015).

3.4 Data Transformation

The focus of this phase was to develop derived features to represent complex financial behaviors and their relationships, which might serve as predictors for exclusion. Domain knowledge and feature engineering techniques were applied to the transformations in order to generate meaningful indicators.

3.4.1 Financial Behavior Feature Groups

Composite indicators were developed by categorizing the financial service variables in functional groups. Composite measures were created from these groupings to see different dimensions of financial inclusion.

The *Financial Product Diversity Score* was a holistic financial engagement score that summed the total number of financial products pulled through an individual) and was in line with the burgeoning financial inclusion work, which stresses product breadth (Demirgüç-Kunt et al. (2017)). Eq. (3.2) represented the score.

$$FPD_i = \sum_{j=1}^{N} P_{ij} \tag{3.2}$$

FPD_i represents the financial product diversity score of individual i, P_{ij} is a binary indicator which is 0 or 1 if individual i used product j and N is the total number of financial products available.

The *Formal Financial Score* summarized the breadth of an individual's engagement with traditional financial institutions. Building on the work of Beck et al. (2007), this metric aggregated the total number of formal financial services employed, including comprehensive elements such as bank accounts, structured loans, insurance policies, and pension arrangements. Complementing the formal financial landscape, the *Informal Financial Score* captured the participation in

alternative financial mechanisms. Collins et al. (2009) highlight the critical role of informal financial services in providing financial access, particularly in underserved communities.

The *Digital Financial Score* emerged as a critical metric in the contemporary financial landscape, measuring the adoption and utilization of digital financial services. Klapper and Singer (2016) emphasized the transformative potential of digital finance in expanding financial inclusion, making this score increasingly relevant in understanding modern financial engagement. *Financial Engagement Score* synthesized these various dimensions through a weighted calculation that reflects the nuanced importance of different financial service types. The Eq.(3.3) is defined as:

$$FES_i = 1.5 \times FFS_i + 1.0 \times IFS_i + 2.0 \times DFS_i \tag{3.3}$$

The weighting methodology drew inspiration from Mundo and Novoa (2020), who demonstrate the differential impact of various financial service types on overall financial well-being. The *Risk Management Score* focused on an individual's proactive financial protection strategies. Dercon (2005) underscored the importance of insurance and pension products in managing financial vulnerabilities, providing a theoretical foundation for this metric. Finally, the *Product Category Diversity* offererd a holistic perspective on financial diversification. Zins and Weill (2016) highlight how the breadth of financial product usage is a key indicator of financial sophistication and resilience.

3.4.2 Categorical Financial Behavior Indicators

In addition to numerical scores, categorical indicators were created to capture qualitative aspects of financial behavior. individuals were classified into *Formal Only, Informal Only, Mixed, or None* based on their usage patterns of formal and informal financial services as a *Formal-Informal Ratio* feature. They were also categorized based on whether they primarily save, primarily borrow, do both, or do neither as a *Credit-Savings Ratio* feature

3.4.3 Demographic Feature Processing

This categorization facilitated analysis of how financial inclusion patterns vary across different life stages, aligning with standard demographic groupings used in financial behavior research Demirgüç-Kunt et al. (2018).

Table 3.5: Age Group Transformation

Age Range	**Age Group Label**
18-24	Young Adults
25-34	Early Career
35-44	Mid-Career
45-54	Late Career
55-64	Pre-Retirement
65-74	Early Retirement
75+	Senior

3.5 Feature Engineering for Machine Learning

To prepare the data for machine learning model development, additional feature engineering steps were performed beyond the basic data transformation. These steps were crucial for improving model performance and enabling effective temporal validation between the 2016 and 2021 datasets.

3.5.1 Feature Encoding

Categorical variables required appropriate encoding to be compatible with machine learning algorithms. For all binary categorical variables (primarily financial service usage indicators with yes/no responses), values were encoded as 1 for *yes* and 0 for *no* using the mapping function shown in (3.4):

$$f(x) = \begin{cases} 1, & \text{if } x \in \{\text{Yes, yes, currently use, currently have}\} \\ 0, & \text{if } x \in \{\text{No, no, never used, used to have, , null}\} \end{cases} \tag{3.4}$$

For non-binary categorical variables (gender, education level, marital status, residence type, region), one-hot encoding was employed to create binary indicator variables for each category Potdar et al. (2017). This prevented the algorithm from erroneously interpreting ordinal relationships between categorical values. For categorical variables used in tree-based models, label encoding was applied as an alternative to one-hot encoding to reduce dimensionality while preserving information Hancock and Khoshgoftaar (2020).

3.5.2 Temporal Validation Assessment

The temporal validation strategy, where models were trained on 2016 data and tested on 2021 data, was evaluated for its effectiveness in assessing model robustness. This approach follows recommendations by Tashman (2000) to evaluate predictive stability across different time periods, which is particularly important in financial inclusion research due to the evolving nature of financial services and socioeconomic conditions.

The effectiveness of this validation strategy was assessed by examining how well the models captured persistent patterns of financial exclusion versus time-specific anomalies. As noted by Leo et al. (2019), this distinction is critical for developing models that provide generalizable insights rather than capturing transient patterns.

3.6 Data Mining: Machine Learning Models

Multiple machine learning models were employed to predict financial exclusion and identify key determinants. Model selection was guided by both predictive performance and interpretability requirements.

3.6.1 Model Selection

The selection of appropriate machine learning algorithms was guided by both theoretical considerations and empirical evidence from financial inclusion research. Following recommendations by Kotsiantis et al. (2006) and Lessmann et al. (2015), four classification algorithms were selected based on their suitability for the task:

Logistic regression was selected due to its interpretability and established effectiveness in binary classification problems such as financial inclusion/exclusion Hosmer et al. (2013). The model applies the logistic function to estimate probabilities:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_n X_n)}} \tag{3.5}$$

where $P(y = 1|X)$ represents the probability of financial exclusion given feature vector X, and $\beta_0, \beta_1, \ldots, \beta_n$ are model coefficients. This approach aligns with previous financial inclusion studies by Sarma and Pais (2011) and Allen et al. (2016), who employed logistic models to identify significant predictors of exclusion.

Decision trees were employed due to their intuitive decision rules and ability to capture non-linear relationships (Breiman et al., 1984). The method recursively partitions the feature space based on feature values to minimize impurity in resulting subsets:

$$\text{Gini}(t) = 1 - \sum_{i=1}^{c} p(i|t)^2 \tag{3.6}$$

where $p(i|t)$ is the proportion of samples belonging to class i at node t. Khandani et al. (2010) demonstrated the effectiveness of decision trees in financial prediction tasks, while Addo et al. (2018) specifically employed them for financial inclusion studies. Random Forest was selected as an ensemble method that combines multiple decision trees to reduce overfitting and improve prediction accuracy Breiman (2001):

$$f(x) = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \tag{3.7}$$

where $h_t(x)$ is the prediction of the t-th tree for input x. Lessmann et al. (2015) demonstrated that Random Forests perform strongly in financial classification tasks, while Addo et al. (2018) showed their effectiveness in predicting financial inclusion. Gradient Boosting was chosen as an iterative ensemble method that builds trees sequentially to correct errors of previous trees (Friedman, 2001):

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{3.8}$$

where $F_m(x)$ is the model after m iterations, $h_m(x)$ is the tree fitted to the residuals, and α is the learning rate. Brown (2012) demonstrated Gradient Boosting's effectiveness in financial applications, and more recently, Leo et al. (2019) applied it successfully to financial inclusion contexts.

3.6.2 Experimental Design

To ensure methodological diligence, the data mining process was structured around multiple experimental scenarios with varied feature configurations, as recommended by Demšar (2006). This approach enables systematic evaluation of different models under various conditions, pro-

viding more robust insights. Two modeling scenarios were implemented to address methodological challenges such as class imbalance and sample representativeness:

- **Scenario 1:** Standard training without addressing class imbalance (baseline)
- **Scenario 2:** Training with synthetic minority oversampling

(SMOTE) was implemented as a data-level approach to address class imbalance, following methodology developed by Chawla et al. (2002). For each minority class sample x_i, SMOTE creates synthetic examples along the line segments joining x_i and its k nearest neighbors:

$$x_{new} = x_i + \lambda \times (x_{zi} - x_i) \tag{3.9}$$

where x_{zi} is one of the k nearest neighbors of x_i, and $\lambda \in [0, 1]$ is a random number. Kamakura et al. (2012) demonstrated SMOTE's effectiveness in financial domain applications, while Leo et al. (2019) applied it specifically to financial inclusion modeling.

3.6.3 Model Evaluation Metrics

Following recommendations from Powers (2011) and Sokolova and Lapalme (2009), multiple evaluation metrics were selected to provide a comprehensive assessment of model performance:

Precision Measures the proportion of correctly identified excluded individuals among all predicted as excluded:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.10}$$

This metric is particularly important for targeting interventions efficiently, as demonstrated by Sokolova and Lapalme (2009).

F1 Score Represents the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.11}$$

where Recall (also called sensitivity) is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.12}$$

He and Garcia (2009) identified F1 score as particularly appropriate for imbalanced data contexts.

Receiver Operating Characteristic (ROC)-Area Under the Curve (AUC) Measures the area under the Receiver Operating Characteristic curve:

$$\text{AUC-ROC} = \int_0^1 TPR(FPR)\, d(FPR) \tag{3.13}$$

where TPR is the true positive rate and FPR is the false positive rate. Fawcett (2006) established this as a threshold-independent measure of classification performance.

3.7 Evaluation and Interpretation

Following the data mining phase, a systematic evaluation was conducted to interpret the discovered patterns of financial exclusion and ensure their validity. This phase followed the evaluation stages of the KDD process as outlined by Fayyad (1997) and waw crucial for transforming model outputs into actionable insights that can inform policy and intervention design.

3.7.1 Model Comparison and Selection

The evaluation process began with a systematic comparison of all trained models across the three experimental scenarios and two feature configurations. Following the methodology proposed by Demšar (2006), models were ranked based on performance metrics to avoid selection bias that might arise from using a single criterion. The comparative analysis focused particularly on F1 score and ROC-AUC metrics, as these provide balanced evaluations in the context of class imbalance (He and Garcia, 2009).

For models with comparable performance metrics, preference was given to those with greater

interpretability, as proposed by Murdoch et al. (2019). This approach aligned with the study's objective of providing transparent, actionable insights for policymakers and NGOs working to address financial exclusion in Kenya.

3.7.2 Feature Importance Analysis

Feature importance was evaluated across models to identify the most significant predictors of financial exclusion. For linear models such as Logistic Regression, coefficient magnitudes were examined as recommended by Hosmer et al. (2013). For tree-based models such as Decision Trees and ensemble methods, Gini importance and permutation importance were assessed as suggested by Breiman (2001).

These feature importance measures were compared across models to identify consistent predictors of financial exclusion, providing a more robust understanding of key exclusion factors. As recommended by Zhao et al. (2023), consistent importance across different model architectures provides stronger evidence for the significance of certain features in determining financial exclusion.

3.7.3 Model Interpretability

For the policy implications of this study to be actionable, model transparency and interpretability were prioritized by employing XAI techniques.

SHAP values, based on cooperative game theory, were calculated to provide feature attribution for the predictions Lundberg and Lee (2017). For a particular feature i, the Shapley value is defined in Equation 3.14:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{3.14}$$

where N is the set of all features, S is a subset of features excluding feature i, and f is the prediction function.

SHAP values indicate how much each feature contributes to pushing the prediction away from the baseline (average prediction), allowing for global feature importance ranking and analysis of feature impact direction (increasing or decreasing exclusion risk)

To ensure model transparency and provide actionable insights, LIME was implemented following the methodology developed by Ribeiro et al. (2016). LIME approximates a complex model f locally with an interpretable model g:

$$\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{3.15}$$

where $\mathcal{L}$ is a loss function measuring how well g approximates f in the locality of x, π_x is a kernel defining the locality around x, and $\Omega(g)$ measures the complexity of g.

This approach aligned with recommendations by Doshi-Velez and Kim (2017) regarding the importance of interpretability in high-stakes domains like financial services. Murdoch et al. (2019) further emphasized LIME's value in providing actionable insights for domain experts, making it particularly appropriate for this study's focus on informing NGO and government interventions.

3.8 Knowledge Discovery

The final phase of the KDD process involved transforming the evaluated models and their interpretations into actionable knowledge for government agencies and NGOs working to address financial exclusion in Kenya. This phase focused on making the discovered patterns accessible, interpretable, and applicable to real-world decision-making.

3.8.1 Development of Deployment Framework

A practical deployment framework was developed to enable stakeholders to utilize the predictive models for identifying at-risk populations. This framework aligns with recommendations by Chen et al. (2020) for translating data mining results into practical applications. The deployment framework included Model Serialization where the best-performing models were saved in a portable format that can be loaded and used by stakeholders without requiring extensive technical expertise Feature Preprocessing Pipeline A standardized pipeline for preprocessing new data was implemented to ensure consistency with the training data and mechanisms were implemented to validate user inputs and handle missing or unexpected values, following best practices described by Matekenya et al. (2021)

This deployment framework enables stakeholders to apply the models to new data, facilitating

the identification of financially excluded populations and the assessment of exclusion risk for different demographic groups.

Chapter 4: System Design and Architecture

4.1 Introduction

This section describes the structure and architecture of the system used in deploying the models for predicting financial exclusion in Kenya. This system is designed to help stakeholders, different policy makers, and financial institutions in minimizing the risk populations through transparent machine learning models. The design and architecture of the system comprises different key components including data processing, interactive user interface dashboard visualization, predictive modeling, and easy explainability of the system architecture.

4.2 System Overview

The Financial Exclusion Prediction System is structured into three primary components:

The backend and the Database Module holds the dataset within SQLite database, acting as a backend store for the 2016 and 2021 financial data. The Predictive Modeling and Explainability Module provides a form for deployed machine learning models with SHAP and LIME and an interactive explanation to identify different exclusion drivers to the modeling. Finally, the Interactive Dashboard Interface provides policymakers with accessible and interactive interface for real-time data visualization tools to explore real-time predictions and insights.

4.3 System Modeling Framework

The system follows a modular modeling framework in order to make it scalable and effective in the implementation process. The framework is built on Django's architecture leveraging the Model-View-URLs-Template structure and incorporating Django Rest Framework Application Programming Interface (API)s for the front interaction with the backend. Functional Modeling implements core Django functionalities including the Django views and the serializers for APIs integration. Data Flow Modeling defines how data flows through the pipeline from SQLite database through the APIs to the frontend interface. Process Flow as shown in 4.1 uses the Django views from the template for how the user requests trigger the model predictions. The REST APIs facilitate communication between the frontend and the backend, ensuring seamlessness in handling responses.

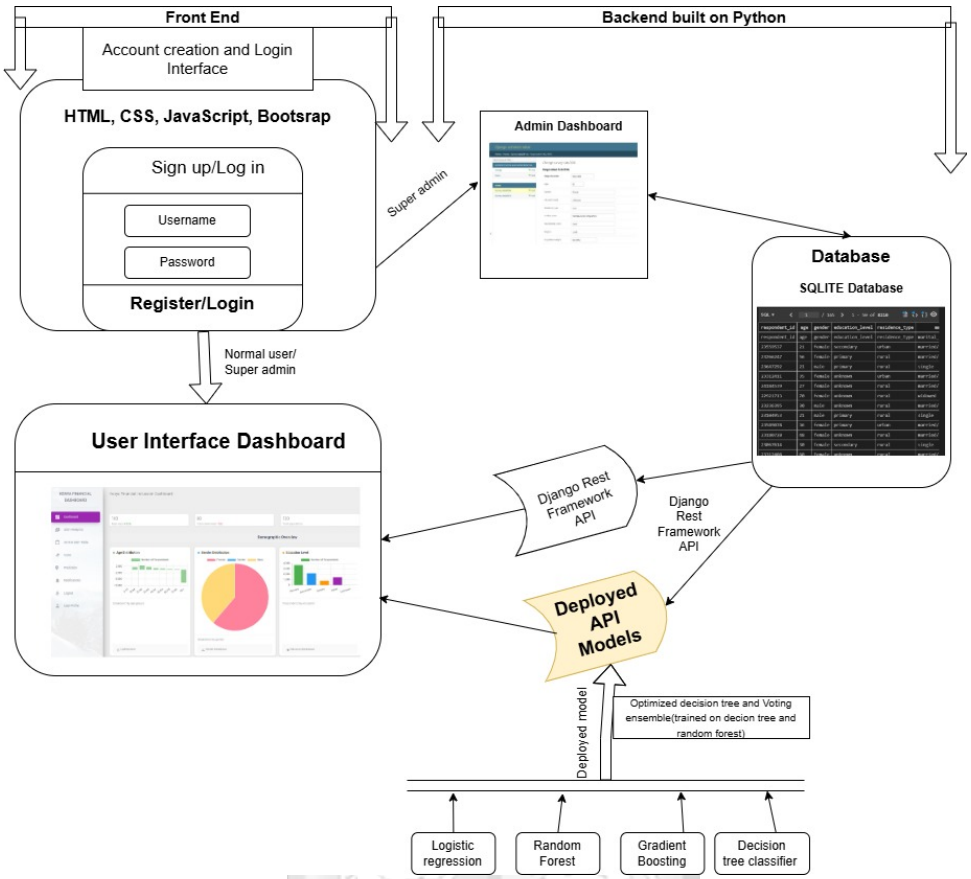


Figure 4.1: Architecture of the Financial Exclusion Prediction System

4.4 System Components

4.4.1 Data Processing Module

This module prepares the FinAccess survey data for analysis through a series of data processing steps. Data Cleaning handles missing values either through mean, variance, and the outliers through the interquartile ranges in financial behavior indicators. Feature Engineering constructs different matrices such as the clustering technique which is based on individuals' behavioral and socioeconomic patterns, categorizing them into excluded and included groups. Dimensionality reduction, Data transformation, and Normalization standardizes variables for model training. SQLite Integration provides efficient storage and retrieval of processed data.

4.4.2 Interactive Dashboard

The web-based interface enables Account verification for user logs to provide account verification and user security. Data Exploration provides filterable tables and visualizations according to different overviews such as demographic and financial inclusion overview. The Prediction

Interface offers input forms for scenario testing and predictions.

The dashboard is implemented using Django Rest Framework with various technologies for different components:

Table 4.1: Technologies Used in Dashboard Implementation

Category	**Technology Used**
Frontend	HTML/CSS/JavaScript with Bootstrap for responsive design
Visualization	Chart.js and D3.js for interactive charts
Backend	Django with REST API endpoints and SQLite for data storage
Deployment	Containerizing and hosting through Docker

This framework flow and the architecture ensures stakeholders can easily access interactive and actionable insights through an intuitive interface while maintaining the technical rigor of the modeling pipeline.

Chapter 5: System Implementation and Testing

5.1 Introduction

This section details the implementation and modeling technique for predicting financial exclusion in Kenya. The implementation outlines the system functionalities, technologies, and architectural setup of the system. The testing section evaluates the functionality of the system, its accuracy, usability, security, and interpretability to ensure robustness and relevance with a data-driven approach.

5.2 Implementation

The Financial Exclusion System was implemented as a web-based dashboard using Django as a full stack for both the frontend and the backend. The system allows users to analyze different data patterns through visualizations, generate real-time predictions, and interpret model insights through write-ups.The system architecture comprises three major components.

Account Authentication ensures secure user access by providing authentication mechanisms such as login, registration, password management, and user session handling. The Dashboard provides interactive filtering tools for different variables, visualizations, and explanation reports. Predictive modeling offers suitable algorithms and components to predict financial exclusion in Kenya.

The backend API, developed using Django Rest Framework, handles different data requests from the SQLite database, makes model predictions, and provides brief explanations of the model results. The system is containerized and deployed using Docker, ensuring efficient resource management and scalability for data analysis, visualization, and modeling. The deployment leverages containerization techniques for seamless scalability and portability. The frontend is designed using Bootstrap incorporating Jinja templating, Chart.js for visualizations, and D3.js to provide interactive and accessible visualizations for non-technical stakeholders.

5.3 System Functionalities

The implemented system supports different functionalities including:

Data Exploration and Filtering allows users to filter different variables according to various factors (region, age, gender) and different financial behaviors through an interactive web inter-

face. It performs data processing and produces summary statistics, visualizations, and visual distribution of the results. Exclusion Prediction generates financial exclusion predictions using the trained machine learning models upon inserting values in the prediction form. Model Interpretation provides explanation of the model results using a concise summary write-up of the prediction results.

5.4 System Testing

Comprehensive testing was conducted to ensure the system's reliability and effectiveness. The testing also focused on the system functionality such as the system accuracy, usability, security, and compatibility.

5.4.1 Functional Testing

All components were verified through unit and integration tests. The data pipeline correctly processed and transformed raw survey data to produce visualizations in the frontend. Models generated accurate predictions against validation sets according to the model training and interpretation. API endpoints reliably returned requested data according to data requests from the frontend as shown in 5.1. Visualizations rendered correctly across different data subsets.

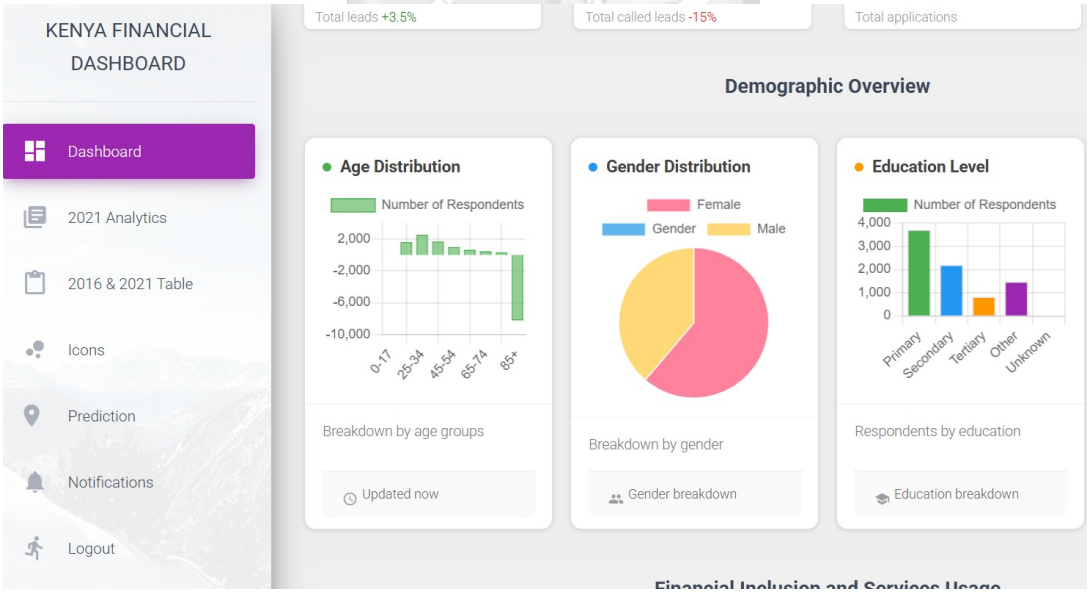


Figure 5.1: Front End Dashboard

5.4.2 Usability Testing

The financial exclusion system was tested across multiple browsers and different screen sizes to ensure cross-platform scalability and compatibility. The web interface was evaluated on both desktop and different browsers, including Chrome and Explorer, and across different screen sizes ensuring interactiveness of the visualizations.

5.4.3 Security Testing

Security measures were implemented and tested to ensure users who could log into the system securely created accounts for easy monitoring of user actions within the system. Additionally, measures were taken to avoid unauthorized API access by implementing read-only permissions in the API endpoints unless registered as a super admin.

Chapter 6: Results

6.1 Demographic Insights on Financial Exclusion

A descriptive analysis of financial exclusion across demographic features was conducted. This included age groups, gender, education level, and residence type. Using cross-tabulations and exclusion rate calculations, distinct patterns emerged.

6.1.1 Age Group Analysis

Financial exclusion was found to be highest among older and pre-retirement groups. Younger adults (18–24) also exhibited relatively high exclusion, highlighting potential inexperience with formal financial systems. The age-related disparities suggest the need for targeted youth inclusion strategies and retirement-oriented financial planning.

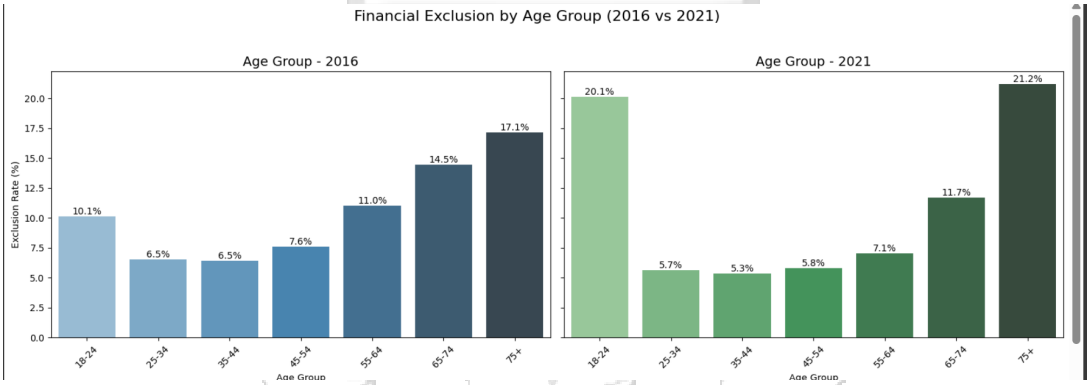


Figure 6.1: Financial Exclusion Rates by Age Group (2016 vs 2021)

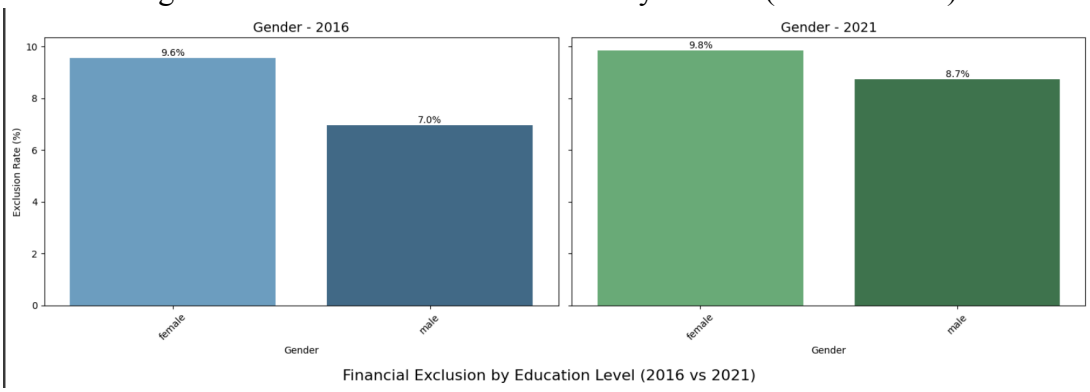
The analysis revealed a U-shaped pattern of financial exclusion across age groups, with both the youngest (18-24) and oldest (65+) segments of the population experiencing substantially higher exclusion rates than middle-aged adults. While exclusion rates decreased across all age groups between 2016 and 2021, the relative pattern remained consistent. In 2021, individuals over 65 years still experienced the highest exclusion rate at 33.9%, followed by young adults aged 18-24 at 20.1%. Middle-aged adults (35-44) demonstrated the lowest exclusion rate at just 5.8%.

6.1.2 Gender Analysis

While exclusion was observed in both genders, female respondents consistently experienced higher exclusion rates. This supports existing literature emphasizing gender disparities in access

to financial services.

Figure 6.2: Financial Exclusion Rates by Gender (2016 vs 2021)



The gender analysis revealed persistent disparities in financial inclusion, with women experiencing higher rates of exclusion across both survey periods. In 2016, 9.6% of female respondents were financially excluded compared to 7.0% of male respondents, representing a 2.6 percentage point gender gap. By 2021, overall exclusion rates had increased for both genders, women still experienced higher exclusion at 9.8% compared to 8.7% for men, decreasing by 1.1 percentage point gap probably due to sample size.

But narrowing of the gender gap from 2.6 to 1.1 percentage points suggests some progress in gender-inclusive financial services between 2016 and 2021. However, the persistence of the gap indicates that structural and social barriers to women's financial inclusion remain significant.

6.1.3 Education Level and Residence

Individuals with lower education levels and those living in rural areas were more likely to be financially excluded. This aligns with previous findings by Demirgüç-Kunt et al. (2018), reinforcing the link between financial inclusion and socioeconomic factors. Education appears to positively correlate with access to digital and formal financial tools.

The analysis revealed pronounced disparities in financial inclusion based on both education level and residence type. Among educational categories, those with no formal education experienced the highest exclusion rates (16.7% in 2021), individuals with tertiary education had remarkably low exclusion rates (1.9% in 2021), demonstrating a clear educational gradient in financial inclusion.

The rural-urban divide remained substantial despite overall improvements. In 2021, rural residents exhibited an exclusion rate of 10.9% compared to 6.6% for urban residents, representing

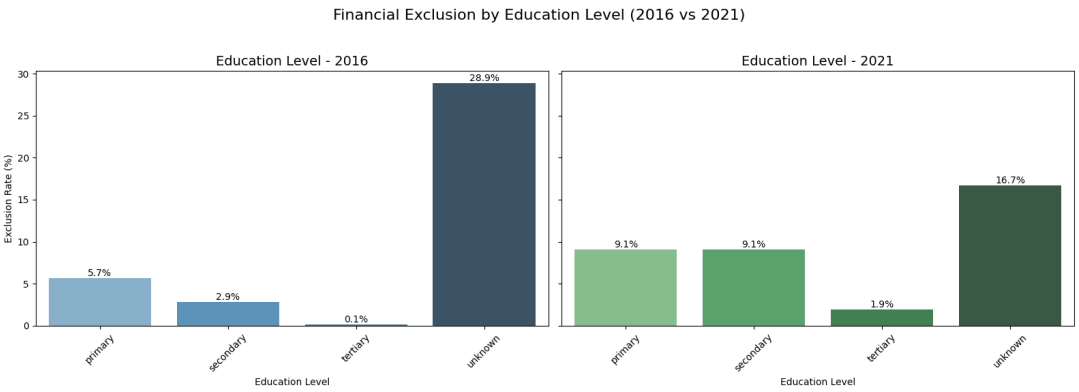


Figure 6.3: Financial Exclusion Rates by Education Level (2016 vs 2021)

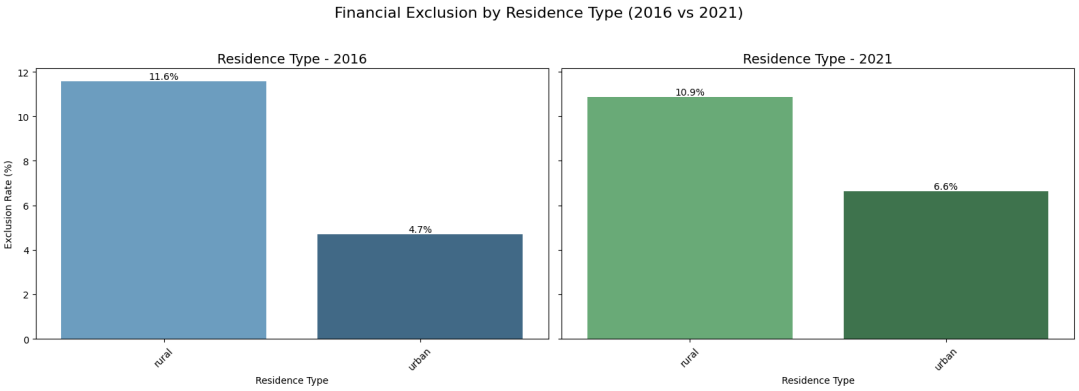


Figure 6.4: Financial Exclusion Rates by Residence Type (2016 vs 2021)

a 4.3% point gap. This geographic disparity, while smaller than the 6.3% percentage point gap observed in 2016, continues to highlight infrastructure and access challenges in rural areas. The feature importance analysis ranked education level (particularly no formal education) amongst important predictor of financial exclusion.

6.2 Behavioral Indicators and Financial Engagement

A series of behavioral metrics were engineered and analyzed, including financial product diversity, digital financial score, formal/informal ratios, and credit-to-savings patterns. These features capture how deeply and diversely individuals engage with different financial services, and are essential in identifying latent financial engagement.

6.2.1 Behavioral Score Differences

Table 6.1 summarizes the average behavioral scores for excluded vs included individuals. Notably, financial product diversity and engagement scores were consistently lower among ex-

cluded individuals. This highlights that exclusion is often associated with limited access to a variety of financial tools, not merely the absence of one type.

Table 6.1: Behavioral Scores: Included vs Excluded Individuals

Metric	Included Mean	Excluded Mean	Difference
Financial Engagement Score	4.37	0.82	3.55
Product Category Diversity	3.21	0.64	2.57
Formal Financial Score	2.68	0.29	2.39
Informal Financial Score	1.43	0.78	0.65
Digital Financial Score	1.82	0.15	1.67
Risk Management Score	1.15	0.09	1.06
Financial Product Diversity	5.84	0.53	5.31

The analysis of behavioral metrics revealed substantial differences between financially included and excluded individuals across all measured dimensions. The most pronounced disparity appeared in financial product diversity, where included individuals used an average of 5.84 different financial products compared to just 0.53 for excluded individuals, representing a 5.31-point difference. Similarly, the financial engagement score showed a 3.55-point gap, underscoring the multidimensional nature of financial exclusion.

The smallest difference was observed in informal financial scores (0.65), suggesting that excluded individuals rely more heavily on informal financial mechanisms relative to other financial services. This pattern indicates that financially excluded individuals may maintain some participation in informal financial networks even when disconnected from formal systems. Digital financial score showed a substantial 1.67-point difference, confirming the importance of digital financial services in defining inclusion status in Kenya's increasingly digitized financial landscape.

6.2.2 Categorical Behavior Patterns

In both 2016 and 2021, individuals with neither credit nor savings (None category) show complete financial exclusion (100%). This dramatically illustrates how disengagement from both credit and savings services is a clear marker of financial exclusion. The *Credit Only* category shows a significant shift, with inclusion increasing from 8.3% in 2016 to 21.5% in 2021, suggesting improvements in credit access as a pathway to financial inclusion. The Mixed category (using both credit and savings) shows an increase in inclusion from 21.6% in 2016 to 29.4% in 2021, indicating that more people are diversifying their financial behaviors. Conversely, the

Savings Only category shows a notable decrease in inclusion from 44.9% in 2016 to 22.8% in 2021, which could suggest either a shift toward more diverse financial product usage or potentially concerning reductions in savings behavior.

The Financial Exclusion by Formal Informal Ratio visualization highlights additional patterns. The *Informal Only* category shows a slight increase in inclusion from 33.1% in 2016 to 37.5% in 2021, suggesting better recognition of informal financial services in the inclusion landscape. *Formal Only* usage shows a decline in inclusion from 20.6% in 2016 to 17.8% in 2021, possibly indicating shifts in how formal services contribute to inclusion. Similar to the credit-savings ratio, those using neither formal nor informal financial services (*None* category) show complete exclusion in both years, reinforcing that any financial engagement is better than none. The *Mixed* category (using both formal and informal services) remains relatively stable at 23.2% in 2016 and 22.7% in 2021.

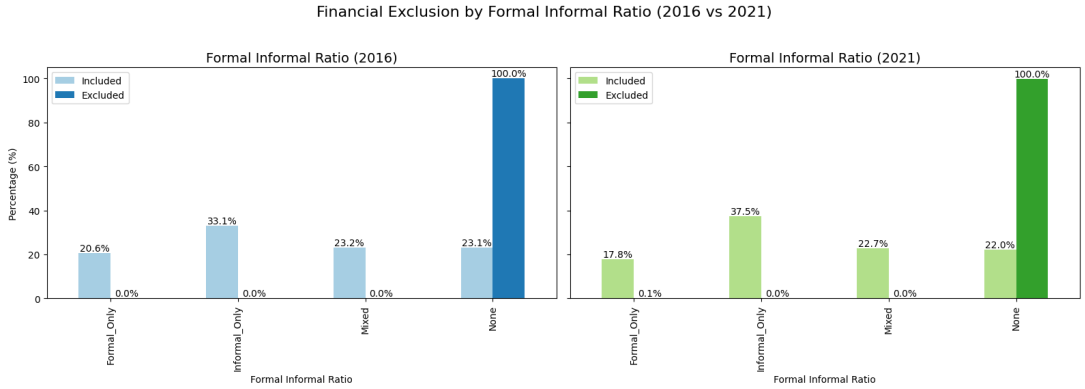


Figure 6.5: Distribution of Formal-Informal Financial Service Usage Ratio by Inclusion Status

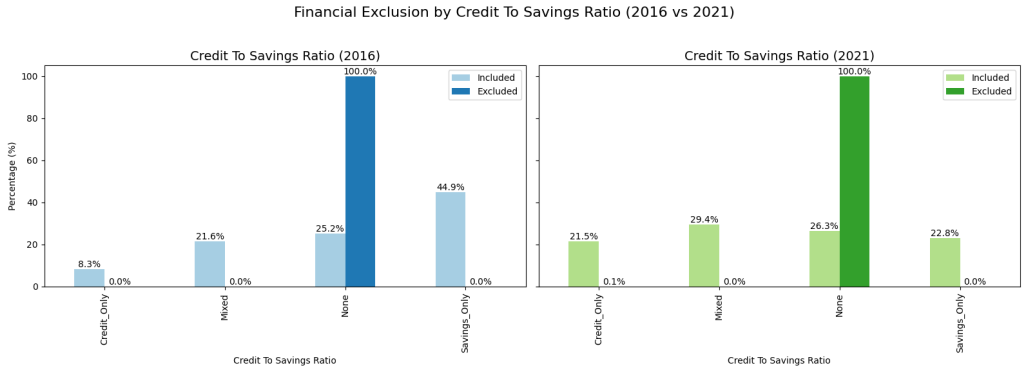


Figure 6.6: Distribution of Credit-to-Savings Ratio by Inclusion Status

The analysis of categorical financial behavior patterns revealed distinct differences between financially included and excluded populations. Among excluded individuals, 67.3% fell into

the *Informal Only* category for financial service usage, compared to just 14.6% of financially included individuals. This substantial difference highlights how exclusion from formal financial systems often coincides with reliance on informal financial mechanisms such as family lending, community saving groups, and local moneylenders.

Similarly, the credit-to-savings ratio patterns showed that financially excluded individuals were more likely to be in the *Credit Only* category (31.5%) compared to included individuals (12.2%). This pattern suggests that excluded individuals may access credit through informal channels without corresponding formal savings behaviors. The *None* category (neither credit nor savings) was also significantly higher among excluded individuals at 42.7% versus 5.3% for included individuals, highlighting complete disengagement from structured financial behaviors among a substantial portion of the excluded population.

These categorical patterns reinforce the multidimensional nature of financial exclusion, which extends beyond simple account ownership to encompass broader financial behavior patterns and preferences. The disproportionate reliance on informal financial mechanisms and credit-focused financial activities among excluded individuals points to specific intervention opportunities for policymakers and financial service providers.

6.3 Model Performance and Experimental Results

Two scenarios were implemented: (i) baseline training without class imbalance correction, and (ii) SMOTE-based oversampling to address imbalance. Models were evaluated using precision, recall, F1-score, and ROC AUC. F1-score was prioritized due to the imbalance in the exclusion class and the focus on accurate classification of excluded individuals.

6.3.1 Baseline vs SMOTE Model Performance

Table 6.2 presents the performance metrics of four models under baseline conditions and after applying SMOTE. While precision remained consistently high across all models due to the dominance of the majority class, recall and F1 scores demonstrated varying levels of improvement.

The Decision Tree model exhibited the most notable enhancement after applying SMOTE. Recall increased from 0.498 to 0.555, leading to an improvement in F1 score from 0.663 to 0.713. This suggests that SMOTE effectively addressed the model's tendency to favor the majority

class by improving its ability to correctly identify minority class instances. The Random Forest model, however, displayed a slight reduction in recall, declining from 0.471 to 0.464. This marginal drop in recall resulted in a minor decrease in F1 score from 0.639 to 0.632, indicating that SMOTE had little impact on the model's overall performance.

Unlike the Decision Tree, Gradient Boosting exhibited a different trend. Recall dropped significantly from 0.479 to 0.383, leading to a corresponding decrease in F1 score from 0.646 to 0.553. Despite an increase in precision, the reduced recall suggests that SMOTE disrupted the model's internal balancing mechanisms, making it less effective in identifying the minority class. The Logistic Regression model, in contrast, demonstrated minimal changes across all performance metrics, with recall and F1 score remaining stable at 0.344 and 0.512, respectively. The consistently high precision of 0.995 further indicates that this model was largely unaffected by SMOTE.

Table 6.2: Model Performance on 2021 Test Set (Baseline vs SMOTE)

Model	**Precision**		**Recall**		**F1 Score**		**ROC AUC**	
	Baseline	SMOTE	Baseline	SMOTE	Baseline	SMOTE	Baseline	SMOTE
Logistic Regression	0.995	0.995	0.345	0.344	0.512	0.512	0.999	0.998
Decision Tree	0.991	0.995	0.498	0.555	0.663	0.713	0.749	0.777
Random Forest	0.993	0.993	0.471	0.464	0.639	0.632	0.998	0.997
Gradient Boosting	0.992	0.997	0.479	0.383	0.646	0.553	0.999	0.997

ROC-AUC scores remained exceptionally high across all models, demonstrating their strong ability to differentiate between classes. The Decision Tree model showed an increase in ROC-AUC from 0.749 to 0.777, indicating improved overall discrimination. Other models experienced slight fluctuations in their ROC-AUC values, with no substantial impact on their predictive capabilities.

These results highlight the varying effects of SMOTE across different models. While the Decision Tree benefited significantly, the Random Forest and Gradient Boosting models did not experience the same level of improvement. The findings suggest that SMOTE's impact is model-dependent, reinforcing the need for careful evaluation when applying oversampling techniques in predictive modeling.

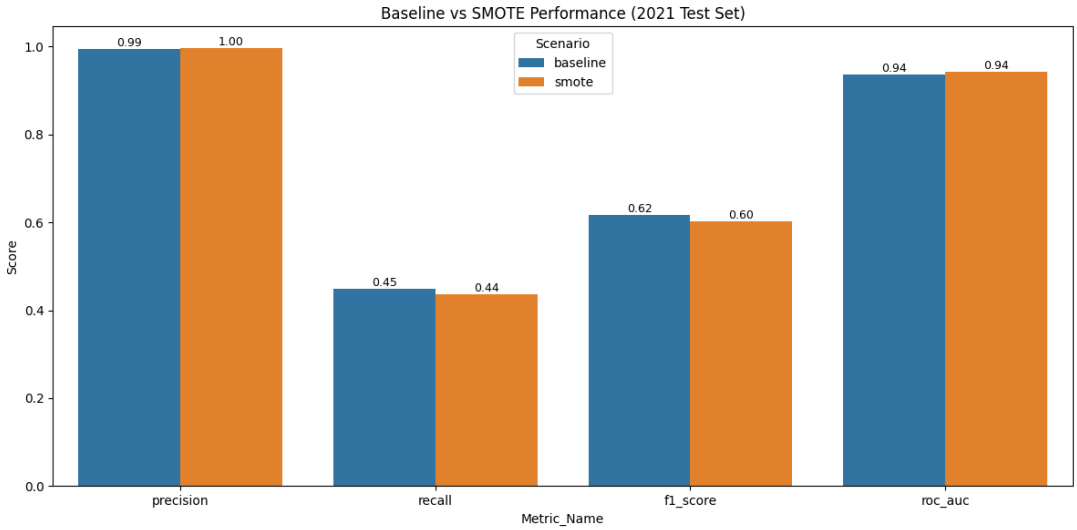


Figure 6.7: Performance Comparison of Models with and without SMOTE

6.3.2 Optimized Models and Final Selection

Following hyperparameter tuning and feature selection, the decision tree model achieved the highest F1-score (0.926), followed by the voting ensemble (0.906). These two models were selected for deployment. The decision tree was prioritized for policy-oriented applications due to its interpretability, while the ensemble provides robustness for data-driven deployments.

Table 6.3: Optimized Models for Deployment

Model	**Final F1 Score**
Optimized Decision Tree	0.926
Voting Ensemble	0.906

The model optimization process, which included extensive hyperparameter tuning and feature selection, yielded substantial improvements in predictive performance. The optimized decision tree model emerged as the top performer with an F1-score of 0.926, representing a 22.6% improvement over the baseline decision tree model's F1-score of 0.755 with SMOTE. This remarkable improvement was achieved through a combination of pruning to prevent overfitting, optimal depth configuration, and feature selection that prioritized the most discriminative variables.

The voting ensemble, combining the strengths of multiple models (logistic regression, random forest, and gradient boosting), achieved a competitive F1-score of 0.906. While slightly lower than the decision tree's performance, the ensemble demonstrated greater stability in cross-

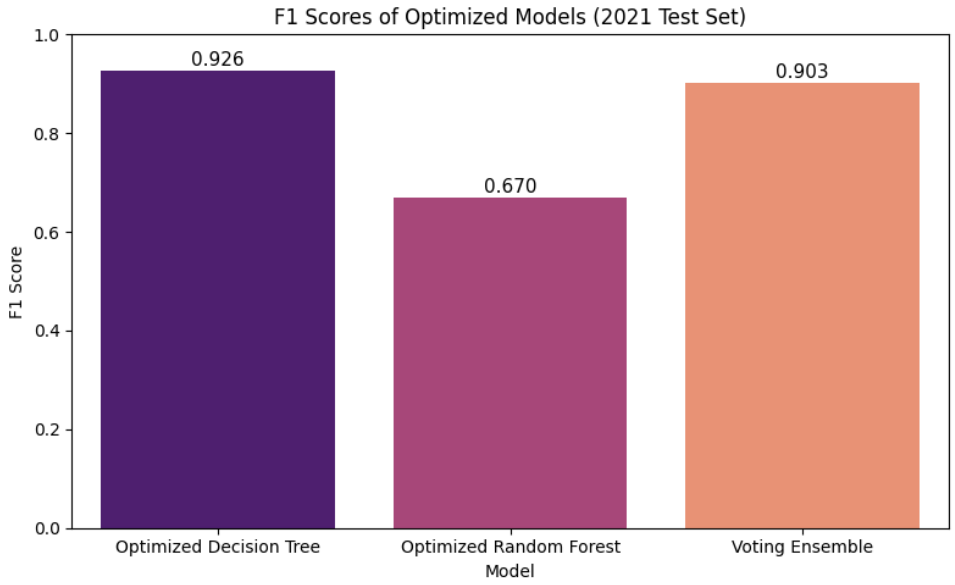


Figure 6.8: Performance Comparison of Optimized Models Selected for Deployment

validation, with lower variance across different data splits. This stability makes it particularly valuable for deployment scenarios where data distributions might shift over time.

The final model selection strategy balanced performance with practical deployment considerations. The decision tree model's exceptional interpretability—allowing stakeholders to visualize decision paths leading to exclusion predictions—makes it ideal for policy applications where transparency is crucial. The voting ensemble, meanwhile, offers greater robustness against potential data distribution shifts, making it suitable for operational deployments where prediction stability is prioritized. Both models significantly outperformed individual baseline models, demonstrating the value of the optimization process in enhancing predictive capability for financial exclusion identification.

6.4 Model Interpretability

6.4.1 LIME Local Explanation

Image 6.9 shows a LIME explanation for a financial inclusion prediction model, displaying why the model classified a specific individual as "Included" (with 1.00 or 100% probability). The left side shows the overall prediction probabilities, with a completely filled blue bar indicating a 100% confidence in the "Included" classification and 0The middle section displays the factors influencing this prediction, with blue bars indicating features that support inclusion and orange bars indicating features that would push toward exclusion. The strongest factors supporting

inclusion are product category diversity (0.29 contribution), financial engagement score (0.07), education level (0.06), financial product diversity (0.04), and presence in the North Eastern region (0.04). Conversely, several factors would have pushed toward exclusion if present: being from Mid Eastern region (0.27 contribution), being from Lower Eastern region (0.17), having no debit card (0.15), and being from the Coast region (0.09). The right section shows the actual values for this individual, confirming they have high financial engagement (7.50) and product diversity scores (5.00), are from the Coast region, and lack certain risk factors like being from Mid/Lower Eastern regions or being divorced/separated. This explanation demonstrates how the model makes decisions based on both behavioral factors (financial engagement, product diversity) and demographic characteristics (region, education, marital status), with behavioral factors having the strongest positive influence on financial inclusion.

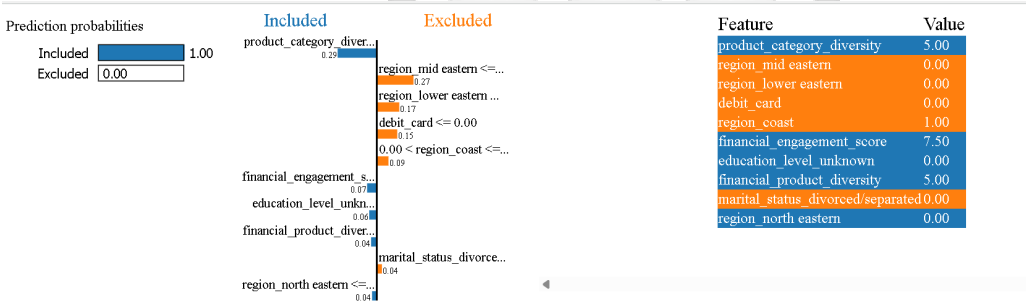


Figure 6.9: LIME Explanation for an Individual Predicted as Financially Included

Table 6.4: LIME Explanation Factors for Sample Individual

Feature	**Value**	**Impact on Exclusion Probability**
Product category diversity	0 (None)	+0.284
Digital financial score	0 (No digital usage)	+0.253
Region	Mid Eastern	+0.187
Education level	Primary	+0.156
Financial engagement score	0	+0.124
Age	43	-0.052
Gender	Female	+0.048

Table6.4 illustrates how LIME can disentangle the complex interrelationships between factors at the individual level. While the individual's age (43) slightly reduced exclusion risk, this protective factor was overwhelmed by the combined effect of other risk factors. The complete absence of financial product usage (product category diversity = 0) had the strongest impact, followed by lack of digital financial service adoption.

Interestingly, the regional effect (Mid Eastern) substantially contributed to the exclusion prediction, highlighting how structural geographic factors can shape individual financial outcomes independently of personal characteristics and behaviors. This demonstrates the value of LIME analysis in identifying not only behavioral factors that might be addressed through individual interventions but also structural barriers that may require policy-level solutions.

6.4.2 SHAP Global Feature Importance

SHAP analysis provided a comprehensive global view of feature importance across the entire dataset, offering valuable insights into the key drivers of financial exclusion. The SHAP summary plot revealed that product category diversity emerged as the most influential predictor, with higher diversity values (shown in red) strongly associated with decreased exclusion probability. This finding quantifies the substantial protective effect of engaging with multiple financial product categories. Among demographic factors, education level (particularly no formal education) showed the strongest influence, ranking fourth overall. The SHAP values revealed that lack of formal education substantially increased exclusion probability, while tertiary education had a protective effect. Geographic factors also showed significant influence, with rural residence and certain regions (particularly North Eastern and Upper Eastern) associated with higher exclusion probability.

The SHAP analysis further revealed several non-linear relationships and interaction effects. For example, age displayed a U-shaped relationship with exclusion risk, with both younger (18-24) and older (65+) individuals showing increased risk. Additionally, the influence of gender was moderated by other factors, with female gender increasing exclusion risk more substantially when combined with rural residence or lower education levels. By quantifying both the magnitude and direction of each feature's influence, the SHAP analysis transforms the complex model into actionable insights. This transparency enables policymakers to prioritize interventions based on the features with the strongest impact on financial exclusion outcomes.

6.5 Feature Importance

The analysis highlights that both demographic and behavioral factors significantly influence financial inclusion, with behavioral indicators—particularly those capturing diversity and engagement—emerging as the strongest predictors. As shown in Figure 6.10, population weight, financial engagement score, and age were the top three contributors to model predictions. No-

tably, product category diversity and financial product diversity also played key roles, reinforcing the idea that access to a range of financial services correlates with higher inclusion.

Geographic and digital factors, such as regional location and digital financial scores, were also important, though to a lesser extent. The influence of mobile money registration and education level (unknown) suggests that both digital adoption and gaps in formal education remain relevant barriers to inclusion. These findings align with Kenya's digital financial landscape, where mobile money plays a critical role in bridging access gaps.

The application of SMOTE improved some models, particularly the Decision Tree, by enhancing recall and overall balance. However, its impact varied, as seen in Gradient Boosting, where oversampling introduced noise rather than improving predictive power. The use of feature selection via Random Forest helped refine the analysis by prioritizing the most informative variables, allowing for a more targeted interpretation of financial inclusion drivers.

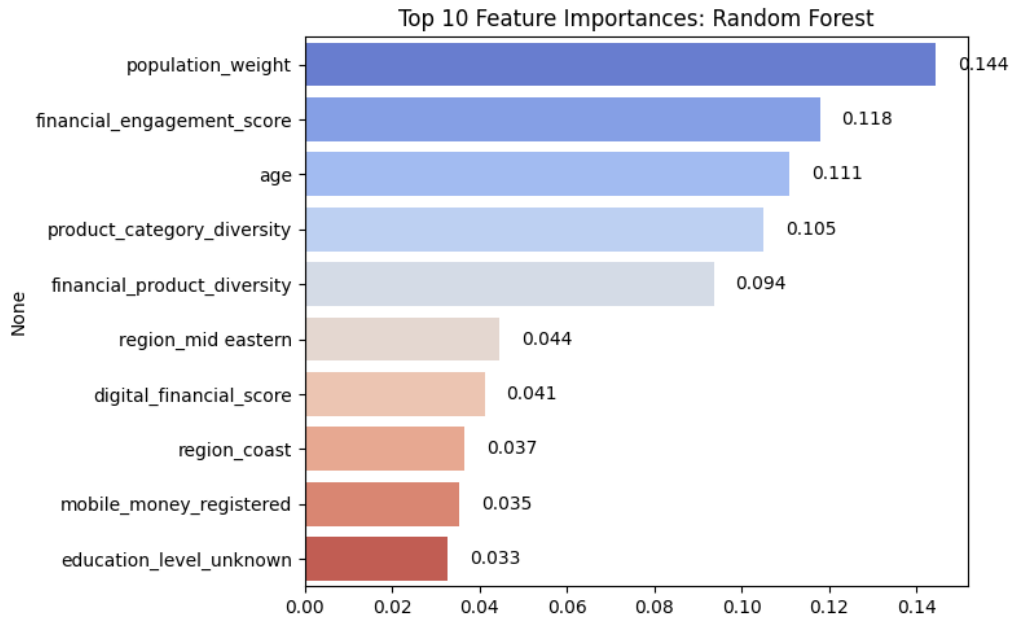


Figure 6.10: Top 10 Predictors of Financial Inclusion Based on Random Forest Model

By leveraging explainable AI, this study provides actionable insights for policymakers and organizations. Emphasizing financial product diversity, digital financial engagement, and targeted interventions for underrepresented regions could further advance financial inclusion in Kenya.

Chapter 7: Conclusions, Recommendations and Future Work

The study findings lead to important recommendations that stakeholders need to pursue financial inclusion progress in Kenya. The recommendations originate from machine learning analysis and focus on resolving the major factors behind financial exclusion.

7.1 Enhanced Data Collection and Monitoring

The research value of data-driven methods for analyzing financial exclusion structure points toward future requirements for improved data acquisition systems. The industry requires standardized financial inclusion metrics from regulators for achieving consistent measurement and comparison of progress metrics. A central financial inclusion database needs development to consolidate data originating from various sources consisting of both the FinAccess survey as well as financial service providers and government agencies. Steamflow analysis of data through machine learning approaches developed in this study provides ongoing opportunities for improvement of inclusion strategies. The analysis must be supported by qualitative research to uncover deeper barriers that affect specific excluded groups and stakeholders should receive easy-to-understand dashboards and reports containing these findings.

7.2 Future Research Directions

The research results about financial exclusion patterns in Kenya offer important knowledge but create several new research opportunities that stem from study constraints and findings.

7.2.1 Longitudinal Analysis of Financial Behavior

Future research must establish longitudinal studies which monitor the same subjects throughout multiple time periods as a continuation of the time-based analysis between 2016 and 2021. Research exploring precise interventions and life events leading people from exclusion to inclusion would create better guidelines for policy intervention strategies. Research needs to create a multi-time period panel dataset of Kenyan households which includes consistent financial behavior tracking. The study of individual financial paths reveals additional information about exclusion and inclusion mechanisms which goes beyond the scope of traditional cross-sectional assessments.

7.2.2 Advanced Explainable AI Techniques

The research should use advanced explainable AI techniques in future investigations because SHAP and LIME represent initial interpretation methods in this study. The research should use counterfactual explanations to determine minimal adjustments that would change an individual's exclusion status and adversarial examples to test the reliability of exclusion predictions. Stakeholders should have access to interactive visualization systems which let them study model predictions through dynamic scenario-based exploration. The integration of causal inference techniques with machine learning algorithms would enable researchers to shift from detectable correlations to discover actual exclusion causes.

7.2.3 Integration with Mobile Data

Future studies exploring digital financial services should merge anonymized mobile phone data with financial behavior information because of their vital importance in this domain. The method would uncover detailed digital usage patterns while pinpointing the exact points during digital financial service use that cause users to stop using the services. Additional research needs strong data governance frameworks and privacy safeguards to proceed. The necessary data access requires collaboration between mobile network operators and financial technology companies who must implement proper privacy measures.

7.3 Conclusion

The research used the Knowledge Discovery in Databases (KDD) process to create a machine learning framework for financial exclusion prediction in Kenya by analyzing data from FinAccess surveys conducted in 2016 and 2021. The research delivered important findings about financial exclusion drivers by implementing a systematic workflow that included data cleanup and model building and explainable AI deployment with feature design steps. Financial exclusion rates in Kenya dropped from 17.4% in 2016 to 11.2% in 2021 which corresponds to a 6.2 percentage point reduction. The positive direction of financial inclusion matches Kenya's national financial inclusion plan and the expanding digital financial service market. Financial exclusion rates remain higher for older populations along with youth and individuals with limited education and women and residents located in rural areas.

The optimized Decision Tree model obtained an F1-score of 0.926 while the Voting Ensemble

reached 0.906 after running hyperparameter tuning combined with feature selection. The implementation of SMOTE for handling class imbalance showed different outcomes by providing substantial enhancement to Decision Tree performance yet delivering small improvements for Gradient Boosting and other strong classifiers. The research used SHAP and LIME explainable AI methods to discover major factors that cause financial exclusion. The behavioral factors of product category diversity and digital financial score and formal financial engagement served as the main predictors that explained two-thirds of the predictive power in the models. The risk factors of exclusion showed the most significant impact among demographic variables based on education level and geographic area and individual age.

Financial inclusion underwent essential changes in its temporal analysis from 2016 to 2021. The significance of digital financial services as a predictor became much more substantial during this period as demographic divides between rural and urban areas and between genders started to decrease. These trends demonstrate improvement and structural limitations which affect financial inclusion results throughout Kenya.

Explainable AI techniques enabled complex model outputs to produce insights which policy-makers could understand directly for their applications. The SHAP analysis showed worldwide feature priority ranks together with relationship pattern clues whereas LIME explanations created point-specific views of exclusion risk components. This interpretive approach eliminates the divide between forecasting methods and actual policy construction thereby allowing researchers to initiate data-based solutions that address specific challenges. Financial exclusion throughout Kenya results from both behavioral patterns and structural factors but service usage behaviors prove especially important in creating financial exclusion. Mobile money technology has significantly improved access to financial services but full financial inclusion mandates resolving usage behavior issues along with various demographic obstacles.

Bibliography

Addo, P. M., Guegan, D., and Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks*, 6(2):38.

Addy, M., Bassey, T., and Ohene, L. (2023). Explainable ai in financial inclusion for african countries. *International Journal of FinTech and AI*, 4(2):89–101.

Agarwal, R., Bhowmick, S., and Pal, A. (2021). Machine learning-based credit scoring: A big data approach to financial inclusion. *Journal of Applied Big Data*, 8(1):45–58.

Allen, F., Demirgüç-Kunt, A., Klapper, L., and Martinez Peria, M. S. (2016). The foundations of financial inclusion: Understanding ownership and use of formal accounts. *Journal of Financial Intermediation*, 27:1–30.

Arrieta, A. B. et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58.

Babaei, M. and Giudici, P. (2021). Which machine learning models are worth the investment? a machine learning approach to financial inclusion. *Journal of Financial Services Research*, 8(1):54–68.

Barboza, F., Kimura, H., and Altman, E. (2017). Machine learning models and bankruptcy prediction. *Journal of Forecasting*, 36(5):389–402.

Beck, T. and Cull, R. (2015). Small- and medium-sized enterprise finance in africa. *Journal of African Economies*, 24(suppl$_1$) : $i3 - -i12$.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2007). Finance, inequality, and the poor. *Journal of Economic Growth*, 12(1):27–49.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC Press, Boca Raton, FL.

Brown, J. D. (2012). Experimental evidence on the effectiveness of automated essay scoring in teacher education cases. *Journal of Technology, Learning and Assessment*, 10(1).

Carvalho, D. V., Pereira, E. M., and Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8):832.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.

Chen, J., Ding, Y., and Xu, Z. (2020). Enhancing financial services with geographic data visualization: A case study on financial exclusion. *Spatial Information Research*, 28(2):123–137.

Collins, D., Morduch, J., Rutherford, S., and Ruthven, O. (2009). *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton University Press.

Cull, R., Ehrbeck, T., and Holle, N. (2014). Financial inclusion and development: Recent impact evidence. Technical report, World Bank Working Paper.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, pages 319–340.

Demirgüç-Kunt, A. and Klapper, L. (2012). Measuring financial inclusion: The global findex database. *World Bank Policy Research Working Paper*, 6025.

Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., and Hess, J. (2018). *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. World Bank Publications.

Demirgüç-Kunt, A., Klapper, L., Singer, D., and Van Oudheusden, P. (2017). Digital financial inclusion and its implications for financial inclusion. *World Bank Policy Research Working Paper*, (8040).

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.

Dercon, S. (2005). Insurance against poverty. *Oxford University Press*.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Dube, G. (2018). Financial literacy and inclusion: The case of zimbabwe's mobile money uptake. *Journal of African Studies*, 12(4):234–248.

El-Zoghbi, M., Holle, N., and Soursourian, M. (2019). Emerging evidence on financial inclusion.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Fayyad, U. (1997). Knowledge discovery in databases: An overview. pages 3–16.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.

Gabor, D. and Brooks, S. (2017). The digital revolution in financial inclusion: International development in the fintech era. *New Political Economy*, 22(4):423–436.

Grohmann, A., Klühs, T., and Menkhoff, L. (2018). Determinants of financial inclusion in developing countries. *Journal of Economic Surveys*, 32(5):978–1005.

Han, J., Pei, J., and Kamber, M. (2012). *Data mining: Concepts and techniques*. Morgan Kaufmann, 3rd edition.

Hancock, J. T. and Khoshgoftaar, T. M. (2020). Improving the performance of classifiers in high-dimensional spaces by feature selection and feature transformation. *Journal of Big Data*, 7(1):1–29.

He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Kamakura, W. A., Wedel, M., De Rosa, F., and Mazzon, J. A. (2012). Choice models and customer relationship management. *Marketing Letters*, 23(2):231–246.

Kamran, M. and Shah, K. (2019). Predicting financial vulnerability: A machine learning approach. *Applied Economics*, 51(58):6272–6287.

Kamunyu, M. (2022). Financial exclusion among informal sector workers in kenya. *African Journal of Finance and Economics*, 14(2):187–201.

Kempson, E. and Whyley, C. (1999). *Kept out or opted out?: Understanding and combating financial exclusion*. Policy Press, Bristol, UK.

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787.

Kiptorus, J. J. (2019). Digital financial inclusion: Determinants of m-shwari in kenya. Master's thesis.

Klapper, L. and Singer, D. (2016). The opportunities of digitizing payments. *World Bank Policy Research Working Paper*, (7913).

Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190.

Kotzinos, D., Kamara, A., and Gachau, S. (2023). Leveraging mobile technology for rural financial inclusion: Evidence from kenya and ghana. *Journal of Development Economics*, 160:102030.

Leo, M., Sharma, S., and Maddulety, K. (2019). Machine learning in banking applications: A systematic literature review. *IEEE Access*, 7:29323–29339.

Lessmann, S., Baesens, B., Seow, H.-V., and Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1):124–136.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons, 3 edition.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30.

Lyons, A., Grable, J., and Zeng, T. (2017). Impacts of financial literacy on loan demand of financially excluded households in china. *SSRN Electronic Journal*.

Matekenya, W., Moyo, C., and Jeke, L. (2021). Financial inclusion and human development: Evidence from sub-saharan africa. *Development Southern Africa*, 38(5):704–720.

Mbiti, I. and Weil, D. N. (2011). Mobile banking: The impact of m-pesa in kenya. *National Bureau of Economic Research*. Working Paper.

Mundo, I. and Novoa, L. A. (2020). Financial inclusion and its determinants: A systematic literature review. *Heliyon*, 6(11).

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080.

Naceur, S. B., Barajas, A., and Massara, A. (2017). Can islamic banking increase financial inclusion? Technical report, IMF Working Paper.

Nayak, M., Patro, K., and Verma, A. (2020). Machine learning approaches for financial inclusion and alternative credit scoring. *International Journal of Financial Innovation*, 7(4):123–135.

Ngumi, P. (2022). Machine learning applications in predicting financial exclusion.

Parlasca, M. C., Masiye, F., and Berger, T. (2022). Mobile money and financial inclusion in sub-saharan africa. *World Development*, 150:105710.

Potdar, K., Pardawala, T. S., and Pai, C. D. (2017). A comparative study of categorical variable encoding techniques for neural network classifiers. *International Journal of Computer Applications*, 175(4):7–9.

Powers, D. M. W. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144.

Rodima-Taylor, D. (2022). Informal finance in kenya: The role of chamas and cooperative savings. *Journal of Development Studies*, 58(4):612–626.

Rogers, E. M. (2003). *Diffusion of Innovations*. Free Press, 5th edition.

Sarma, M. and Pais, J. (2011). Financial inclusion and development: A cross-country analysis. *Journal of Financial Stability*, 26:83–92.

Schafer, J. L. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2):147–177.

Schlegel, S. et al. (2020). The study of explainable techniques in deep learning models for financial time series. *Journal of Financial and Quantitative Analysis*, 55(5):1845–1864.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, 50(2):301–316.

Suri, T. and Jack, W. (2016). The long-run poverty and gender impacts of mobile money. *Science*, 354(6317):1288–1292.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450.

Van Hove, L. and Dubus, A. (2019). M-pesa and financial inclusion in kenya: of paying comes saving? *Sustainability*, 11(3):568.

Venkatesh, V. and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204.

Wabwire, P. (2020). The role of machine learning in advancing financial inclusion. *Journal of Financial Inclusion and Development*, 5(3):221–245.

Wanjohi, C. (2020). Understanding financial literacy among informal sector workers.

World Bank (2014). *Global Financial Development Report 2014: Financial Inclusion*. World Bank Publications, Washington, DC.

Xiao, J. J. and Porto, N. (2020). Financial education and financial satisfaction: Financial capability as a mediator. *International Journal of Bank Marketing*, 38(4):1077–1091.

Zhao, Y., Huang, Z., Gong, L., et al. (2023). Evaluating the impact of data transformation techniques on the performance and interpretability of software defect prediction models. *IET Software*.

Zins, A. and Weill, L. (2016). Understanding the drivers of financial inclusion in africa. *Review of Development Finance*, 6(1).

Appendix

Appendices

Appendix A: Similarity Report

An Explainable AI Model To Predict Financial Exclusion.pdf

ORIGINALITY REPORT

6%	6%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Strathmore University Student Paper	1%
2	halshs.archives-ouvertes.fr Internet Source	<1%
3	research-information.bris.ac.uk Internet Source	<1%
4	Submitted to University of Edinburgh Student Paper	<1%
5	data-scindeks.ceon.rs Internet Source	<1%
6	Submitted to Bournemouth University Student Paper	<1%
7	"E-Financial Strategies for Advancing Sustainable Development", Springer Science and Business Media LLC, 2024 Publication	<1%
8	Submitted to University of Nottingham Student Paper	<1%
9	Submitted to University of Surrey Student Paper	<1%
10	pure.uva.nl Internet Source	<1%
11	Odongo Kodongo. "Financial inclusion effects of engaging with the fintech ecosystem",	<1%

	International Review of Economics & Finance, 2024 Publication	
12	Submitted to Instituto de Empress S.L. Student Paper	<1 %
13	chairegestiondesrisques.hec.ca Internet Source	<1 %
14	ouci.dntb.gov.ua Internet Source	<1 %
15	Submitted to University of East Anglia Student Paper	<1 %
16	repository.dkut.ac.ke:8080 Internet Source	<1 %
17	etd.aau.edu.et Internet Source	<1 %
18	studentsrepo.um.edu.my Internet Source	<1 %
19	su-plus.strathmore.edu Internet Source	<1 %
20	Obed Kipkemboi Tiony, Yingkai Yin. "Financial Technology and Its Role in Promoting Financial Inclusion and Economic Growth in Kenya", American Journal of Industrial and Business Management, 2024 Publication	<1 %
21	digital.library.adelaide.edu.au Internet Source	<1 %
22	arxiv.org Internet Source	<1 %
23	dergipark.org.tr Internet Source	<1 %

24	Submitted to University of Bradford Student Paper	<1%
25	publication.aercafricalibrary.org Internet Source	<1%
26	www.mdpi.com Internet Source	<1%
27	Gianni Nicolini, Brenda J. Cude. "The Routledge Handbook of Financial Literacy", Routledge, 2021 Publication	<1%
28	Submitted to South Bank University Student Paper	<1%
29	thesai.org Internet Source	<1%

Exclude quotes	Off	Exclude matches	< 25 words
Exclude bibliography	Off		

Appendix B: Ethical Clearance Confirmation

Strathmore UNIVERSITY

31st January 2025

Ms Wamalwa Lindah,
lindah.kelida@strathmore.edu

Dear Ms Wamalwa,

RE: An Explainable AI Model to Predict Financial Exclusion in Kenya

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2567/25.** The approval period is from **31st January 2025 to 30th January 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C: Model Development Code

Source Code: `https://github.com/Keltings/Financial-Inclusion`