ADDIS ABABA UNIVERSITY
COLLEGE OF TECHNOLOGY AND BUILT ENVIRONMENT
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

PREDICTING CUSTOMER CHURN IN DIGITAL BANKING SERVICES USING MACHINE LEARNING

BY
HLINA NEGUSE

ADVISOR
Dr. BINIAM TADESSE

A thesis submitted to the School of Electrical and Computer Engineering in partial fulfillment of the requirements for the Degree of Master of Science in Computer Engineering

JUNE, 2025

ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY

COLLEGE OF TECHNOLOGY AND BUILT ENVIRONMENT

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

The undersigned have examined the thesis titled:

PREDICTING CUSTOMER CHURN IN DIGITAL BANKING SERVICES USING MACHINE LEARNING

BY

HLINA NEGUSE

Approved by the examining committee members:

	Name	Signature	Date
Advisor:	Dr. Biniam Tadesse	______________	____________
Examiner:	Dr. Sosina Mengistu	______________	____________
Examiner:	Dr. Menore Tekeba	______________	____________

Declaration

I, Hlina Neguse Zerihun, attest that I conducted my own independent research for this thesis.

All information and concepts borrowed from other authors are properly cited and referenced throughout the document. Furthermore, I certify that neither the entirety nor any portion of this thesis has been submitted to fulfill the requirements of any other academic program or institution. The information provided here represents my distinct contribution to the field of study.

Student Name: Hlina Neguse Zerihun

Signature: ____________________

Date: ______________________

JUNE, 2025

Acknowledgments

I would like to express my utmost gratitude to Almighty God for His grace, strength, and guidance at every point of this life.

I am truly grateful to my advisor, Dr. Biniam Tadesse, for his invaluable guidance, support, and effort while conducting this research. His experience and guidance are essential in developing this thesis.

I should thank Mr. Amanuel Kindu, Mr. Yohannes Fenta, and Mr. Habtamu Abera for their assistance, and their positive comments and motivation helped me to conquer many challenges along the way.

I would also like to extend my deep appreciation to my beloved family for their endless love, patience, and encouragement.

Special thanks to my classmates and MSc Computer Engineering students at the college of technology and built environment for their support and shared experiences in this academic journey.

Finally, I need to thank to all those who have supported me throughout this work.

Abstract

In an increasingly competitive digital banking environment, retaining customers is a key consideration in emerging markets like Ethiopia, where the adoption of digital offerings today is increasing rapidly. This study solves the gap of the lack of Ethiopian digital-specific prediction models, and the limited incorporation of sentiment analysis and probabilistic approaches. It proposes a suite of machine learning–based predictions of the risk of customer churn in the Ethiopia digital banking environment utilizing a comprehensive (one year) dataset of behavioral transactions, demographic details, and customer sentiment measures based on surveys. The researchers trained, tested, and compared five supervised learning models: Logistic Regression, Decision Tree, Gradient Boosting, Random Forest, and. Neural Network.

The Random Forest approach performed best overall scoring 95% accuracy based on 0.204 score of log loss and ROC-AUC score of 0.984. The addition of sentiment features significantly improves the model's performance and highlights the potential value of obtaining customer sentiment responses on their likelihood of churn. Feature importance analysis using SHAP revealed that the most influential predictors of churn were EASY_SCORE, REGION, LAST_LOGIN, CONTINUE_SCORE, SECURE_SCORE, BALANCE, and NO_DB_TRN. The study also implemented individualized churn probability predictions. These results affirm the need for customer-centric churn models that account for both behavior and perception.

Overall, this study underscores the importance of blending behavioral analytics with customer feedback to develop proactive, personalized retention strategies in Ethiopia’s growing digital banking ecosystem.

Table of Contents

Declaration I
Acknowledgments II
Abstract III
List of Figures VI
List of Tables VII
List of Acronyms VIII
1. Introduction 1
1.1. Background of the Study 1
1.2. Motivation 6
1.3. Statement of the Problem 8
1.4. Research Questions 9
1.5. Objectives 10
1.5.1. General Objective 10
1.5.2. Specific Objectives 10
1.6. Significance of the Study 10
1.7. Limitation of the Study 12
1.8. Thesis Organization 12
2. Literature Review 14
2.1. Overview of Churn 14
2.2. Machine Learning Churn prediction models 16
2.3. Sentiment Analysis 18
2.4. Existing Studies and Gaps in Research 20
3. Methodology 27
3.1. Research Design 27
3.2. Data Collection 28
3.2.1. Demographic Attributes 28
3.2.2. Banking-Specific Features 28
3.2.3. Behavioral Attributes 29
3.2.4. Transactional Data 29
3.2.5. Churn Indicator 30

3.2.6. Sentiment-Driven Features ... 30
3.3. Data Pre-Processing and Analysis ... 31
3.4. Sentiment Analysis ... 32
3.5. Feature Selection ... 33
3.6. Model Training ... 35
3.7. Model Evaluation ... 39
3.8. Ethical Considerations ... 40
4. Results and Discussion ... 41
4.1. Experimentation Setup ... 41
4.2. Model Performance ... 41
4.2.1. Logistic Regression ... 42
4.2.2. Decision Tree ... 42
4.2.3. Random Forest ... 42
4.2.4. Gradient Boosting ... 43
4.2.5. Neural Network (MLP) ... 43
4.2.6. Model Comparison Summary ... 43
4.3. Performance under Different Feature Sets ... 44
4.4. Impact of Sentiment Analysis on Model Performance ... 45
4.5. Churn Probability Results ... 49
4.5.1. Sample Prediction Results ... 49
4.6. Feature Importance ... 51
4.7. How Results Answer Research Questions ... 53
4.8. Discussion ... 54
5. Conclusion and Recommendations ... 56
5.1. Summary of Findings ... 56
5.2. Contributions ... 57
5.3. Practical Implications ... 58
5.4. Future Work ... 58
References ... 59
Appendices ... 63
Survey Questionnaires ... 63

List of Figures

Figure 1: Number of Digital Banking Users in Ethiopia 3
Figure 2: Transaction Using Digital Banking in Ethiopia 3
Figure 3: Challenges of Digital Banking in Ethiopia... 5
Figure 4: Research Methodology Flowchart... 27
Figure 5: Model Performance Comparison Before vs. After Sentiment Integration.................... 46
Figure 6: ROC-AUC Curve Before vs After Sentiment Analysis 48
Figure 7: Feature Importance Result on SHAP Values 51

List of Tables

Table 1: Comparison of Related Works.. 25
Table 2: Result of Evaluation of Models 41
Table 3: Model Performance under Different Feature Selection 44
Table 4: Churn Probability Result 50

List of Acronyms

ATMs Automated Teller Machines

CRM Customer Relationship Management

CLV Customer Lifetime Value

CAC Customer Acquisition Costs

NLP Natural Language Processing

BERT Bidirectional Encoder Representations from Transformers

LSTM Long Short Term Memory

SMOTE Synthetic Minority Oversampling Technique

MLPClassifier Multi-layer Perceptron Classifier

MLP Multilayer Perceptron

ROC-AUC Area Under the Receiver Operating Characteristic Curve

AUC Area Under the Curve

RFECV Recursive Feature Elimination with Cross-Validation

NO_DB_TRN Number of Digital Banking Transaction

NO_BB_TRN Number of Branch Banking Transaction

SHAP Shapley Additive Explanations

1. Introduction

1.1. Background of the Study

Digital banking describes how traditional banking services have been converted into digital formats, enabling users to access various banking services, manage their accounts, and make transactions online[1]. It is the digitization of all traditional banking products, systems, and processes intended for the purpose of providing services to customers through Internet-based channels.

Digital banking is an important tool for increasing financial inclusion for marginalized populations. Digital wallets and mobile banking are some creative services making it easier than ever for people to access financial services[2]. Digital banking has altered the global financial landscape in recent years because it provides consumers with speed, convenience, and accessibility while improving the operating efficiencies of financial institutions[3]. Account-based digital banking can help organizations reduce costs, increase efficiencies, and automate some processes, which can greatly increase operational efficiency through digital transformation.

All the traditional banking services offered by a bank are now digitized and are offered 24/7 on computers, smartphones, and smart devices, giving customers complete independence from physical visits to a bank location[1]. Customers can also access statements from their bank account, withdraw cash, transfer funds, manage their checking/savings accounts, open a digital bank account, manage loans, pay bills, write checks, and aggregate their transactions. Besides reducing operational costs, banks can use mobile banking technology to communicate with customers, engage customers, and provide better service delivery processes. While it has been reported that studies show banks can save from 20 – 40% of operational costs by digitizing their services, by accessing credit products and investment options it appears small cash savings that customers get from banks at improved customer service levels[4].

Mobile and internet banking services, which can be categorized as digital banking, have experienced massive growth worldwide. Mobile banking alone has been growing exponentially across the globe with billions of users using banks' mobile apps for savings, transactions, or money management according to the most recent estimates it is reported that there are over 2 billion active users of mobile banking worldwide[7]. Through increased accessibility and client reach, this change has allowed banks to increase profitability, improve customer engagement, and reduce operating expenses. Through shifting user needs and boosting access to clients, banks are able to increase profits, decrease operational expenses, and improve customer engagement.

The rise of smartphones and internet connectivity has led to greater adoption of mobile banking in Ethiopia. The rapid growth in mobile bank accounts shows how digital banking can help change, and perhaps democratize financial access, especially in areas with poor traditional banking infrastructure[9]. For example, data from the National Bank of Ethiopia shows that the number of mobile money accounts has dramatically increased in Ethiopia, nearing 60 million users.

This data represents a double increase in December 2022 compared to December 2021, showing a broad trend in the nation's banking [10]. An overview of Ethiopia's Financial Sector Indicators as of March 2023 revealed the significant expansion. According to the snapshot, there are 35 million debit card holders, 4.8 million online banking users, and 22 million mobile banking users in Ethiopia [10].

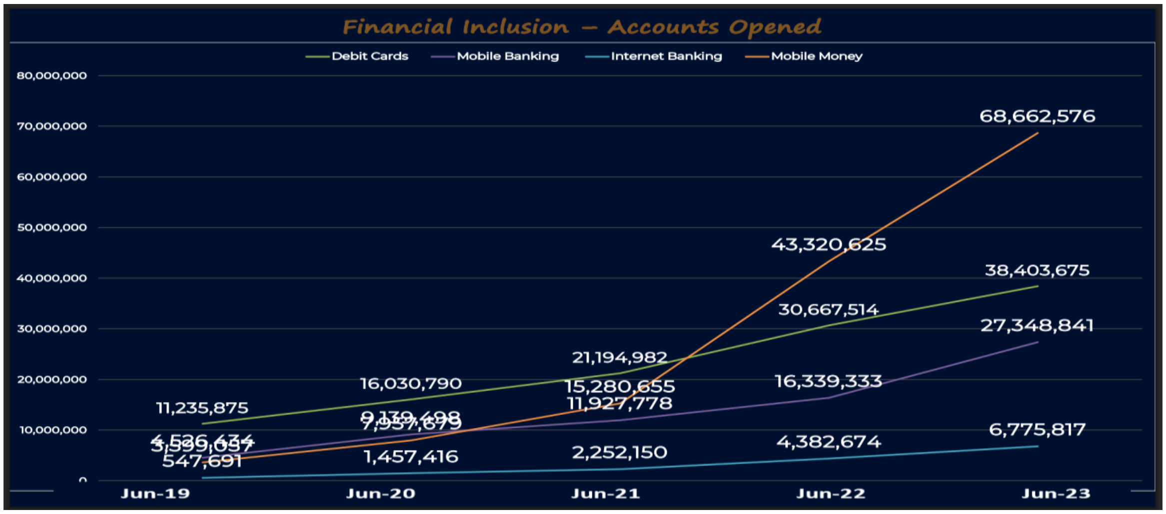

Figure 1: Number of Digital Banking Users in Ethiopia [11]

Figure1, shows the fast expansion of financial service channels in Ethiopia between June 2019 and June 2023. The most notable increase was in Mobile Money, which soared from just over 547,691 accounts in 2019 to over 68.6 million in 2023, thanks to outreach to underserved areas and telecom based services. With an increase from 11.2 million to 38.4 million accounts, debit cards continued to be widely utilized. Additionally, mobile banking grew rapidly, with 27.3 million users by 2023. The number of accounts in internet banking increased more slowly, from roughly 540,000 to 6.8 million [11].

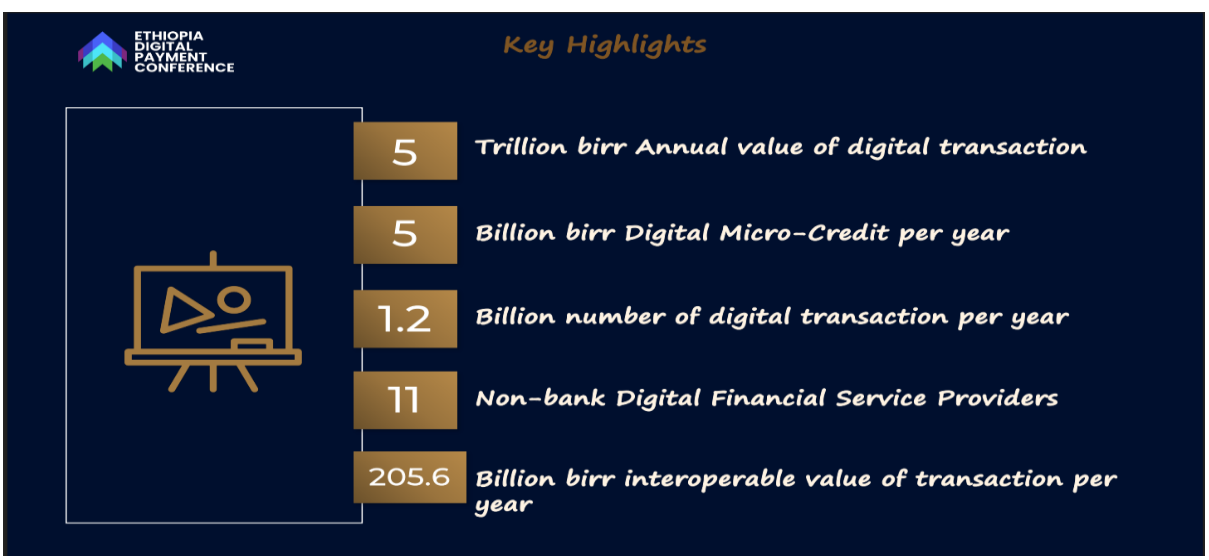

Figure 2: Transaction Using Digital Banking in Ethiopia [11]

Figure 2, showcasing the Ethiopia's progress in digital finance. As a sign of its increasing accessibility and uptake, it claims 5 billion birr in digital microcredit and 5 trillion birr in digital transactions annually. The ecosystem is growing quickly, with 11 non-bank digital service providers and 1.2 billion transactions annually. Furthermore, interoperable transactions totaling 205.6 billion birr show a high level of system integration. All things considered, these figures demonstrate Ethiopia's growing financial inclusion and strong digital payment environment[11].

This change is not without difficulties, though. Banks must reconsider how they engage with their clients and how they will satisfy their expanding needs in light of the digital revolution. Therefore, it is critical to ask: What models and techniques are available for measuring shifting customer satisfaction[12]? To be competitive and meet client needs, the banking sector must improve and guarantee the provision of high-quality web-based technical services[13]. Many clients experience problems like unwieldy user interfaces, poor customer service, security challenges, and a dearth of customized services. These factors can lead to dissatisfaction and, consequently, increased churn rates[14]. As a result, banks must proactively identify at-risk clients and put in place efficient retention plans that will be customized to meet their requirements.

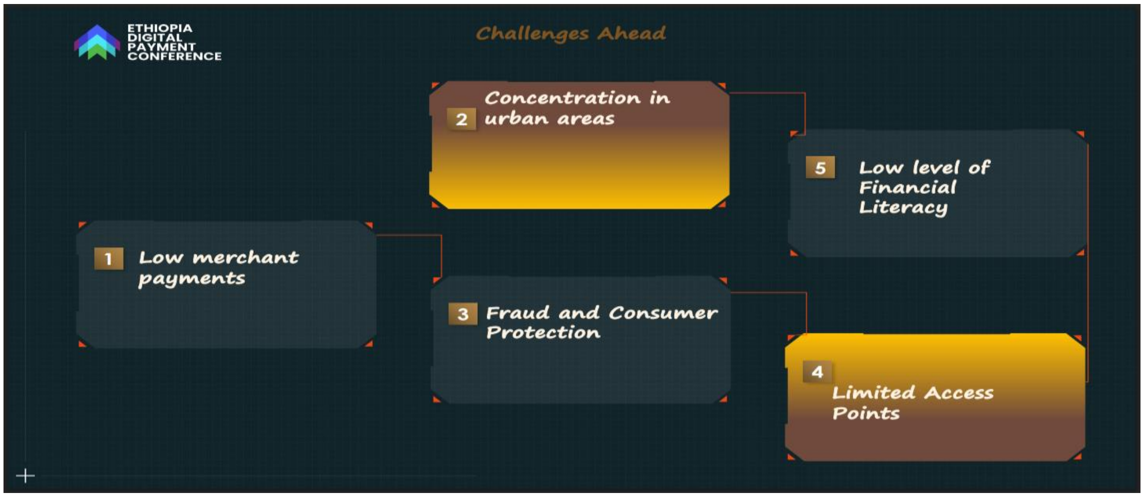

Figure 3: Challenges of Digital Banking in Ethiopia [11]

Figure 3, outlines five main challenges hindering digital payment adoption. These include low merchant payments, limiting where consumers can use digital services, and urban concentration, which excludes rural areas from access. Fraud and poor consumer protection raise issues associated with trust, as well as limited access points such as agents and ATMs meaning less access to services especially in rural areas. Finally, low financial literacy stops many people from even knowing how to use the digital tools available. Taken together, these challenges highlight the necessity for heightened access, improved protections, and public education to ensure Digi-Financial inclusion in Ethiopia[11].

As the banking industry grows rapidly, particularly through digital services, financial institutions are faced with a critical gap in customer retention[15]. The developments in technology have allowed customers to access an abundance of banking choices, making it easier to switch than before. This has resulted in increasing rates of customer churn; where customers close their relationships with the bank, which is a reality in a rapidly changing landscape that signifies great threats to banks' growth and profitability[16]. Understanding the reasons for customer attrition is important for banks wanting to better service delivery and develop sustainable loyalty as large digital banking is emerging in a competitive financial environment in Ethiopia[17].

In digital banking, churn prediction is an identification of customers likely to stop using a digital product or service of the bank. This is an important strategy for financial institutions that are looking to improve their customer retention. Machine learning models use large amounts of transaction and demographic information to evaluate churn prediction by uncovering trends and predicting future behavior [16]. The study assesses machine learning techniques for predicting customer churn in Ethiopia's banking sector. The research addresses trends in consumer behavior, preferences, and customer comments with financial services through new data analytics and sentiment analysis. Some of the machine learning techniques that provide reliable frameworks for analyzing large datasets[25] are Random Forest, Gradient Boosting, decision trees, neural networks (MLP), and Logistic Regression as well as techniques that provide banks the opportunity to learn more about the reasons for customer churn. This study uses customer sentiment analysis, gathered from survey responses, and transactional, and demographic information to evaluate the churn prediction model and to obtain an increase in accuracy and interpretation. Customers' emotions can be precursors of dissatisfaction or an intention to leave[28].

1.2. Motivation

As digital banking usage continues to grow across the country, addressing customer churn is now mission-critical[9]. However, even as the opportunity of digital banking grows, for many financial institutions in Ethiopia, disengagement from digital services is happening across the board and there are no systems or metrics in place to measure or stem this loss[1].

The idea of addressing customer churn becomes particularly urgent when considered in the wider context of Ethiopia's national priorities; achieving financial inclusion, digitizing public services, and unlocking economic growth through digital transformation. High levels of customer churn directly put these objectives at risk, eroding customer confidence and trust in digital services; stalling the adoption of new and innovative platforms; and reducing the return on investments made into financial technology[12].

The imperative to retain existing customers is better than acquiring new customers, especially given that switching on the customer's end is often[20]. Financial institutions still often wait to try and retain a customer until there's been some form of contract breach and even those approaches remain mostly information-less, generalized, and reliant on assumptions. Understanding why customers decide to leave or not to return, many with a local decision-maker's knowledge; reflect the challenges of too large of a market lumped into the same customer buckets. The distances at which consumers are adopting digital for the above reasons such as those driven by the national agenda for greater financial inclusion and digital transformation[1] is all the more concerning. Customers may not have the luxury of waiting for banks to do or demonstrate their efforts around retention, as they will quickly move on to competing banks with comparable digital products, experiences, and opportunities at which point predictive tools are not local and understood down to the behavioral shifts that are relevant in Ethiopia. With a gap like this, it is impossible for local banks to react accordingly and retain their customers, and only exacerbates the limited information retained on customer behavior.

In the new digital banking era, understanding customer emotion is not only a necessity, it is a critical issue. On digital platforms, the amount of human interaction is severely limited, and emotional feedback is one of the only direct indicators available for an assessment of customer experience [28]. A bad customer emotional experience, say frustration, confusion, or distrust (or worse, a combination of all three when it comes to their financial institution), will usually lead them to halt their relationship (churn) with your institution without any indication that they did so, actively and silently [29]. Unfortunately, traditional churn prediction methods that require transactional or demographic data alone cannot capture these more subtle but powerful indications.

This project is motivated by the potential of filling that gap by developing a machine learning-based [16] churn prediction model for the digital banking ecosystem in Ethiopia. The proposed model provides empirically validated insights gleaned from behavioral data, sentiment data, and banking transactional data. Models such as this one provide banks with actionable insights that will allow them to build better retention programs for their customers while facilitating a more resilient, inclusive, and efficient financial system in the process[3].

1.3. Statement of the Problem

Customer churn continues to pose challenges for Ethiopia's burgeoning digital banking scene. As more banks move to digital and even online to deliver banking services, they find themselves in a highly competitive and evolving customer expectation space [17]. However, many seem to struggle with identifying their churn probability and do not have the analytical capability to take meaningful and proactive client retention actions [18].

Existing churn prediction models in Ethiopia generally focus on overall customer churn of the bank and fail to capture digital-specific churn drivers such as app usability issues, login frequency, transaction behavior, unique cultural, technological, and behavioral dynamics of local digital banking users, and service personalization[17]. These models typically overlook key factors tied to digital user experience, limiting their effectiveness in addressing churn within mobile and online platforms.

Moreover, there is a lack of localized, probabilistic churn prediction models that can provide actionable, risk-based forecasts. The absence of such models weakens the ability of financial institutions to implement timely, proactive retention strategies and accurately allocate intervention resources[5].

Another critical gap is the minimal use of sentiment analysis, which offers valuable insights into customer satisfaction, frustration, and loyalty[21,23]. Without capturing customer sentiment from feedback and surveys, banks miss subtle but important indicators of dissatisfaction that transactional data alone cannot reveal.

Without robust, locally adapted predictive tools, Ethiopian banks risk continued customer attrition[19], diminished profitability, and setbacks to national financial inclusion and digital transformation efforts.

Accordingly, this study addresses these gaps by developing a machine learning–based churn prediction model tailored to Ethiopia's digital banking environment. By incorporating digital-specific behavioral patterns, sentiment analysis, and a probabilistic modeling approach, the research aims enhances customer retention.

1.4. Research Questions

Based on the problems identified above, this study seeks to address the following main research questions:

1. Which machine learning model most effectively predicts customer churn in Ethiopia's digital banking sector?
2. What is the impact of sentiment analysis on improving churn prediction accuracy?
3. Which features in Ethiopia's digital banking sector most influence customer churn?

1.5. Objectives

1.5.1. General Objective

The general objective of this research is to develop and evaluate machine learning models enhanced with sentiment analysis for predicting customer churn in Ethiopian digital banking, enabling banks to implement targeted retention strategies.

1.5.2. Specific Objectives

The specific objectives are as follows:

1. To analyze and select relevant features for churn prediction by incorporating behavioral, technological, socioeconomic, and cultural variables specific to the Ethiopian digital banking context.
2. To implement and compare multiple machine learning algorithms for predicting customer churn and suitability for the Ethiopian market.
3. To integrate sentiment analysis to enrich the dataset and enhance model performance.
4. To develop a model interpretation framework that allows stakeholders to understand key predictors of churn.

1.6. Significance of the Study

The paper holds diverse implications and significance across multiple domains. Here's a breakdown by stakeholder group:

Banks: By understanding what drives customer churn, banks can enhance their service set and alleviate pain points specific to digital banking [15]. The findings of this study can be used to guide customer retention practices. By understanding the churn factors, banks can take action to remove customer churn risk before it has occurred. By avoiding churn, banks can keep more customers and maximize customer lifetime value [14].

Data Scientists and Analytics Professionals: This paper provided data scientists with a case example of the complete demonstration of applying machine learning models [24] specific to customer churn in digital banking, particularly in a context such as Ethiopia, in a developing market. Analysts were able to reflect on features and variables specific to the Ethiopian banking sector, developing a more informed approach to model building.

Marketing Team: Marketing teams can use specific information regarding at-risk customers for focused retention campaigns. Knowing who the at-risk customers are, ensures that marketing can focus their efforts on the right customers and get the right message out to the customers [2]. This would be useful for marketing to better understand some pain points and possible gaps in service better produce and shape the overall customer experience and customer service and understand precisely how to reduce churn [10].

Policy Makers: Policymakers can leverage findings to address barriers to digital banking adoption, with the aim of retaining customers while supporting financial inclusion across the country [7]. A reduction in churn across banks helps to solidify the banking sector and contributes to greater economic stability and growth [9].

Academics and Researchers: The paper contributes to the limited research/narrative regarding machine learning applications in emerging markets, and more so in sub-Saharan Africa [15]. The work provides a reference point for future research on customer churn in similar banking markets, which can be developed with different approaches [3].

Technology Providers: Providers of digital banking solutions can use these insights to develop features that directly address churn issues, creating more value for banking clients. It also helps them understand customer behavior in digital banking, improving the efficacy of customer relationship management (CRM) solutions they offer[7].

In summary, this study provides a critical foundation for reducing churn in digital banking, advancing machine learning in the finance industry, and supporting customer-centric decision-making in Ethiopian and similar emerging markets.

1.7. Limitation of the Study

This study is limited by its reliance on data from a single Ethiopian bank, which may affect the generalizability of the findings across the broader banking sector. The model also lacks real-time adaptability and individual-level explanation of churn drivers, which future research should address for more personalized and dynamic churn management.

1.8. Thesis Organization

This thesis is organized into five chapters, each contributing to a comprehensive understanding of customer churn prediction in the Ethiopian digital banking sector.

- **Chapter 1: Introduction:** Introduces the background, motivation, problem statement, research questions, objectives, and significance of the study.
- **Chapter 2: Literature Review:** Explains the churn, reviews existing studies related to customer churn, machine learning in financial services, and the role of sentiment analysis. This chapter identifies research gaps and justifies the proposed approach.
- **Chapter 3: Methodology:** Describes the research design, data sources, preprocessing techniques, feature selection, and the machine learning models used. It also explains the evaluation metrics.
- **Chapter 4: Results and Discussion:** Presents the results of model training and evaluation. It interprets key findings, including feature importance, the impact of

sentiment analysis, and implications for Ethiopian digital banking. It also presents the experimental setup.

- **Chapter 5: Conclusion and Recommendations**: Summarizes the major findings, discusses their practical significance, and offers recommendations for banks and future researchers. It also highlights the study's limitations and areas for future work.

2. Literature Review

2.1. Overview of Churn

The process by which people or organizations stop doing business with a company or stop using its goods or services is known as customer churn, or customer attrition[25]. For businesses across various sectors, it is a crucial performance metric, as it has a direct impact on sustainability, growth, and revenue[16]. Two primary categories of churn exist: involuntary churn, which happens for reasons like expired payment methods, legal problems, or unsuccessful service migrations, and voluntary churn, where clients voluntarily decide to terminate their relationship with a service provider, frequently as a result of dissatisfaction, better alternatives, or changing needs[26]. Churn can be quantified using several indicators, including revenue churn, customer lifetime value (CLV) depreciation, and the churn rate, which is the percentage of customers lost during a specific period[26]. In order to keep their client base, organizations must closely monitor and solve the underlying causes of high churn rates, which indicate problems with customer satisfaction, service quality, or competitiveness[15].

Digital banking churn, or the loss of clients who stop using digital banking channels (such as online banking portals, mobile banking applications, or other digital financial services) is concerning for a number of reasons[14]. First, churn is now top of mind for banks and financial institutions, due to the rise in financial technology advancements and increasing reliance on digital banking channels for financial transactions. Second, digital banking churn can stem from a number of influences, such as poor user experience or usability, complex systems, outages or security issues, a lack of customer support, lower capabilities than competitors, a lack of personalized services[2]. Customers have more opportunity than ever to switch to other providers that have a comparatively more favorable digital performance, with less of a hassle, in a highly competitive, technology-enabled marketplace[17]. Furthermore, younger clients, especially digital ones, expect around-the-clock access to their money and services, nimble transaction timeframes, immediate customer support, and superior levels of security[18]. If those expectations aren't met, clients remain unhappier, more dissatisfied, and possibly leading to attrition. Perhaps most concerning about digital banking churn is that digital banking clients tend to be younger, more tech-savvy than traditional clients, and more likely to diversify their financial service providers/offerings over time, making them highly durable to organizations that are ultimately best served to keep them.

Churn has a serious and varied impact on the banking industry. First, there is an immediate revenue drop, especially if a high-value or long-time customer churns[24]. Banks spend a considerable amount on acquiring customers via marketing, onboarding, or promotions; losing the customer before realizing any value affects return on investment. Also, when a customer churns, the customer lifetime value (CLV) decreases, and the number of customers who can be cross-sold or upsold other products (credit cards, loans, insurance, investments) [27] also decreases. This stifles growth and negatively affects strategic revenue diversification. More churn also increases customer acquisition costs (CAC). When a bank has high churn, they must continuously replace customers to keep their books in balance, which means an allocation of resources or funds could be used to innovate or something else, [27] rather than replace customers.

Additionally, churn also impacts brand image and trust. Disgruntled customers can sway the opinions of many potential customers on social media or in reviews, and the damage can quickly multiply through bad word-of-mouth. Churn also affects all customer engagement data and analytics that prevent banks from building predictive models or adequately creating the personalized layers of service they all aim for[15]. Operationally, regular churn undermines long-term planning, affects workforce allocation and can monopolize compliance, support, and IT resources.

In an environment like Ethiopia and other similar emerging markets where digital banking is growing [10], managing churn effectively is even more important. Banks need to balance their interests in improving digital transformation with their attempts to understand what cultural, socioeconomic, and technological factors might determine customer behavior [18]. If these banks do nothing proactively to premeditate churn in such environments, they may hinder their financial inclusion strategy as well as exacerbate any reduction in digital trust and create further separation from customer needs to banks' products. Therefore, engaging with potential poor-performing customers with churn predictive models, assessing the pain-points of these customers, and executing targeted retention strategies is critical for Banks to be competitive, customer-centric, and profitable in an increasingly digital financial environment [19].

2.2. Machine Learning Churn prediction models

Churn prediction models are critical, especially in emerging markets, for understanding and combating customer attrition within a rapidly evolving digital ecosystem[27]. Emerging markets have unique considerations related to infrastructure limitations, digital literacy differences, and cash-based economic norms that shape the ways customers behave relative to developed markets [15]. Machine learning models are well-suited to churn prediction in this space because they can process complex and high-dimensional data from many differing sources, such as customers' transactional histories, demographics, and behavioral data [14].

Typically churn prediction models include logistic regression that provides interpretability and serves as a baseline model; decision trees and random forest that can capture non-linear relationships, while also ranking the importance of predictive features; and gradient boosting models such as XGBoost which offers good accuracy and efficiency given large datasets[14, 16]. Deep learning models including neural networks are also capable of processing time series or sequential data making them relevant for longitudinal churn analysis when enough data is available [16]. Clustering as a technique, such as k-means, can also be used to inform customer segments at increased risk of churn which can improve model accuracy and retention strategies.

Localized information of these models are very necessary in developing countries to resolve issues of data availability, quality, and socioeconomic factors[22]. Sociodemographic characteristics would need to be modified in certain ways to align with cultural norms and economic conditions, but sentiment research on social media or customer reviews might help provide some insight into the drivers of consumer happiness and discontent[22]. The contextualization and external data, such as local income levels, and cell penetration levels, would be useful to gain insight into the digital banking context of customer churn reasons.

Specifying an appropriate activation logic for churn prediction models used in developing economies includes considerations for transparency and explainability to customers who may be getting an understanding of digital banking products[26]. Ethical issues, such as uneven data bias, and data privacy standards could be relevant to make sure there are no unintended consequences especially when taking into consideration regulations and technology restrictions[26]. Modified and optimized, machine learning models as churn prediction have huge potential to help digital banks in developing economies proactively see customer needs, mitigate churn, and create financial inclusion, and stability in the market[3].

2.3. Sentiment Analysis

Sentiment analysis (also widely referred to as opinion mining) is a natural language processing (NLP) technique, used for the identification and extraction of subjective information from text[28]. Its use is typically to assess whether the sentiment being expressed in the text (usually referred to as a review) is either positive, negative, or some form of neutral sentiment. Through this technique, organizations can gain a better understanding of the feelings, attitudes, and opinions of their customers by analyzing text and various text data (including but not limited to: reviews, responses to survey questions, social media posts, emails, transcripts of support chat) [28]. Sentiment analysis can be assessed at the document level (with the analysis of the whole text), sentence level (with the assessment of individual sentences), or aspect level (looking at the sentiment regarding particular aspects or features) [29]. Techniques vary from rule-based systems that rely on pre-defined dictionaries, to machine learning and deep learning models (BERT or LSTM) that can learn contextual and semantic nuances in the language [30]. In customer experience management, sentiment analysis has emerged as a vital tool to help convert unstructured feedback into structured feedback. This helps customer experience management professionals extract actionable insights that inform decisions taken by a business to improve upon, or tailor services [28].

In churn prediction, sentiment analysis acts as an engine of creative insight for the traditional model where emotional and behavioral information is left out. By experience, churn models typically incorporate numerical and categorical data from structured data such as transaction data, demographic information, and account usage historical analysis. However, sentiment analysis allows an organization to assess customers' emotional state, which can be early indicators of discontent and the likelihood of switching[23]. For example, analyzing customer feedback, from sources like surveys, complaint tools, or app stores, to assess negative sentiments, such as feelings of disappointment or frustration, may demonstrate that a customer may be on the verge of leaving. Similarly, assessing customer expressions of dissatisfaction on social media can allow businesses to capture the feeling of dissatisfaction in the heat of the moment without having it formally expressed, but nonetheless have an effect on churn behavior [22]. Customer care contacts, such as surveys, email communications, chat discussion transcripts with bots, and all call transcripts, can undergo sentiment analysis to know if customers are contacting the business, not for frustrations, but maybe persistently negative emotions or other unresolved feelings[29]. By including sentiment ratings as features of input data for modeling churn, businesses can improve the accuracy of prediction in churn prediction machine learning models, while finding high-risk clients more effectively.

Sentiment studies or sentiment analysis are particularly beneficial in fields like banking where customer retention is paramount [30]. It is also exceptionally relevant given the customer interaction conversion from physical interactions to digital services in the form of apps, online banking portals, and digital customer service through chatbot and platforms due to business digitizing more of their banking services. These developed systems also produce massive amounts of unstructured data, which could be analyzed for sentiment. Take the example of a customer expressing dissatisfaction in their review, or in a feedback form, due to regular crashes of their banking app, hidden fees from their monthly statement, or bad customer service experiences from agents. This behavioral dissatisfaction could be an early warning indicator of drop-off or churn[21]. The integration of sentiment research into churn prediction models enables banks to evaluate predictively, using emotional churn triggers, to take proactive measures to retain customers. Banks could deal with emotional drivers of churn in a variety of ways, such as fee waivers, and improved level of service through better support or recommendations for products that suit them. Applying best practices through such recommendations or dealings could reduce attrition and enhance customer loyalty [16].

The influence of sentiment analysis when using churn prediction in digital banking is heightened in developing countries like Ethiopia where digital financial services are new, and customer expectations and trust are still evolving [7]. With respect to these contexts, it is important to leverage sentiment analysis to consider the meaning of local language, cultural expressions, and feedback that is relevant to a community and develop reliable churn models, along with culturally sensitive retention strategies. By running analysis on sentiment data along with other components such as transaction frequency, digital customer engagement, and demographics, banks can build more reliable predictors that account for rational and emotional aspects of customer behavior [22]. This ultimately enhances the performance of churn prediction systems but also enhances an organization's ability to develop meaningful long-term relationships plagued by short-lived customer loyalty in an increasingly competitive digital marketplace [16].

2.4. Existing Studies and Gaps in Research

The paper [14] covers the critical issue of customer churn in the banking sector and highlights the major role that predictive modeling can play in retaining customers. The dataset was used from Kaggle and trained different statistical and machine-learning techniques, such as k-means clustering, logistic regression, decision trees, random forests, and support vector machines, to explore customer churn. One of the findings from the study was that while the random forest was the best performer (around 97% accuracy), customer segmentation neither improved the models to a significant degree nor resulted in significantly improved prediction accuracy for any of the models considered. The authors recognized the necessity for competent data preprocessing and balancing treatments such as SMOTE in addressing class imbalance in their dataset. The authors also pointed out that their methodology could be applied to other sectors outside of banking. The paper ended with future directions for research that involved developing real-time applications or further consideration of the effects of features on churn prediction on outcomes. Overall, the work contributes valuable insights into the application of machine learning in customer churn analysis, though it acknowledges limitations related to dataset scope and the need for broader validation.

The paper [16] investigates the critical issue of customer churn in the banking sector, utilizing real data from a community bank in the Southern US. The research constructs and compares ten machine learning classification models alongside various sampling techniques to enhance predictive accuracy. Key findings reveal that models such as Random Forest, XGBoost, AdaBoost, and Bagging Meta classifiers achieve high overall accuracy (ranging from 87% to 96%) and strong AUC values (between 0.9 and 0.93). The study highlights the importance of addressing data imbalance, which causes high false negative rates in predictions. Through comprehensive feature selection and diagnostic analysis, the paper identifies significant characteristics influencing customer churn behavior, providing valuable insights for banks aiming to retain customers. This work contributes to a relatively underexplored area in financial services, offering robust methodologies and practical implications for improving customer retention strategies. However, this paper did not include sentiment analysis for understanding customer emotions to get better predictions as well as for retention strategies.

Targeting the Commercial Bank of Ethiopia, the study [17] examines the crucial problem of customer attrition in the banking industry. Using a dataset of 204,161 cases with eleven attributes, the authors create a machine-learning model to forecast client attrition. Logistic

Regression, Random Forest, Support Vector Machine, K-nearest neighbor, and Deep Neural Network are among the supervised learning algorithms they use, and they use strategies like SMOTE for class imbalance. According to the findings, the Deep Neural Network performs better than other models, attaining 79.32% accuracy as well as excellent precision and recall levels. It is important to grasp customer behavior in a competitive banking market, and this study highlights an important contribution of using modern machine learning techniques to close the gap in the effectiveness of customer retention and loss management tools. The study does not use more features for model enhancement as well as confirm more accuracy.

The paper [24] on customer churn prediction in telecommunications covers many machine learning algorithms, and emphasizes the necessity of correctly predicting customer retention to better ensure business sustainability. The authors have weaved their way through the difficulties of churn, such as imbalanced data and the complicated nature of customer behavior and presented a thorough analysis of various algorithms, including logistic regression, decision trees, K-nearest neighbors, random forests, Gaussian naive Bayes, XGBoost, gradient boosting, and neural networks. Through experimental comparisons using a Kaggle dataset of 7043 customers, the study evaluates model performance based on accuracy, AUC, and KS statistics, finding that gradient boosting methods, particularly XGBoost, outperform others in terms of predictive accuracy while also noting the efficiency of logistic regression for practical applications. The paper concludes by emphasizing the need for models that balance complexity and interpretability, as well as the importance of adapting to evolving customer behaviors, suggesting future research directions in dynamic modeling and real-time learning. Overall, it offers valuable insights into leveraging machine learning for effective customer retention strategies in the telecom sector, but this paper would be better if it used a real data set, which results in a more accurate value.

The paper [15] explores the critical issue of customer churn in Zimbabwe's banking sector, highlighting the implications of losing clients on revenue and reputation. The authors, Zvinodashe Revesai et al., employ various classical machine learning algorithms such as logistic regression, decision trees, random forests, and XGBoost alongside a basic deep learning model to analyze customer data. Their results reveal that while all models demonstrate very good levels of accuracy when predicting churn, the deep learning model performed better than classical algorithms both in accuracy and recall. There are limitations to the study - such as limited features on the data and a small sample size - that could by nature limit the generalizability of findings. The authors contend that banks should employ some form of a good churn prediction model to identify at-risk customers who they can take preventative measures to retain. They outline several avenues for future studies to improve churn analytics, in particular diversifying and experimenting with the datasets, and looking at the effect of specific retention actions. Overall, the paper provides insights into customer retention strategies in a difficult economic climate and has practical implications for banks in Zimbabwe.

The research article [19] emphasizes the need for churn prediction techniques that are able to predict customer attrition, which is of utmost concern in the banking industry. They used several machine learning techniques to examine customer behavior and churn prediction which include Random Forest, Logistic Regression, Gradient Boosting Classifier, Extreme Gradient Boosting Classifier, and Light Gradient Boosting Machine Classifier. In the study, a dataset consisting of 49,707 data entries and 11 features from the Commercial Bank of Ethiopia was implemented to identify some of the important aspects of data preparation including data normalization and feature engineering. The Light Gradient Boosting Machine classifier gets the maximum performance measures, including an accuracy of 98%, precision of 97%, and recall of 100%, when the author uses the Synthetic Minority Oversampling Technique (SMOTE) to solve the class imbalance. This comprehensive study provides information about the effectiveness of a number of predictive models in a banking context and highlights the importance of machine learning for enhancing customer retention campaigns. The report also urges more research to improve these models even more and modify them to reflect changing consumer behavior. It would be better for this paper, though, if sentiment analysis was incorporated into both the retention and prediction algorithms.

The below table illustrates the comparison of different studies on their model selection, data source, and limitation.

Unlike these studies, which use generic, non-localized data, this project employs a real dataset sourced directly from an Ethiopian bank, allowing for a highly context-specific model that accurately reflects local customer behavior and banking patterns. Additionally, while existing studies often apply generic churn prediction models, they typically lack focus on digital banking, where customer engagement patterns, profitability, and retention dynamics differ significantly from traditional banking. Moreover, most studies rely solely on transaction and demographic data without considering customer sentiment that are particularly influential in emerging markets. This research addresses these gaps by leveraging sentiment analysis from customer feedback, which provides deeper insights into customer satisfaction and the external pressures shaping churn behavior. Importantly, the study also incorporates individualized churn probability predictions that allow for a customer-centric approach to retention, enabling banks to identify and prioritize at-risk customers with greater precision. By integrating these local and sentiment-based features into the predictive model, the proposed approach offers a more nuanced and context-specific understanding of churn in digital banking, setting it apart from traditional models that do not account for the unique cultural and economic factors relevant to Ethiopia.

The combination of the random forest model, SHAP, BERT-based sentiment analysis is optimal for this study because it addresses important gaps in churn prediction for Ethiopian digital banking by handling complex data and capture important patterns in customer behavior, adds transparency by explaining which features influence the predictions, and extracts valuable insights from customer feedback, capturing emotions that traditional data miss but strongly affect churn.

Table 1: Comparison of Related Works

Ref.	Title	Model Used	Data Source	Limitation
[14]	Customer Churn Prediction in the Banking Sector Using Machine Learning-based Classification Models	K-means clustering, logistic regression, decision trees, random forests	Kaggle	Reliance on an external dataset and a lack of specific features
[16]	Bank Customer Churn Prediction Using Machine Learning Framework	Random Forest, XGBoost, AdaBoost, and Bagging Meta	Community bank in the Southern US	Lack of sentiment analysis
[17]	Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia	Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor, and Deep Neural Network	CBE	Less features (attributes)
[24]	Customer Churn Prediction in Telecom Based on Machine Learning	Logistic regression, decision trees, K-nearest neighbors, random forests, Gaussian naive Bayes, XGBoost, gradient boosting, and neural networks	Kaggle	Lack of real data (used synthetic data from Kaggle not from bank)

[15]	Customer Churn Analytics using Classical Machine Learning Algorithms and Deep Neural Networks: A Case of Zimbabwe Banks	Logistic regression, decision trees, random forests, and XGBoost	Zimbabwean banks	Small sample size of data and feature set
[19]	Customer churn prediction using machine-learning techniques in the case of commercial bank of Ethiopia	Random Forest, Logistic Regression, Gradient Boosting Classifier, Extreme Gradient Boosting Classifier, and Light Gradient Boosting Machine Classifier	CBE	Lack of sentiment analysis

3. Methodology

3.1. Research Design

Customer churn prediction is the method of identifying customers more likely to discontinue using the bank's digital banking service. This work develops several different machine learning models to predict customer churn in an Ethiopian digital banking context. By employing sentiment, transactional, and behavioral data, this analysis provides a very accurate, data-driven method that can ensure strategic proactive retention strategies. As shown in the Figure 4, the process is broken down into multiple crucial stages, each of which plays an essential part in achieving the main goal of the study.

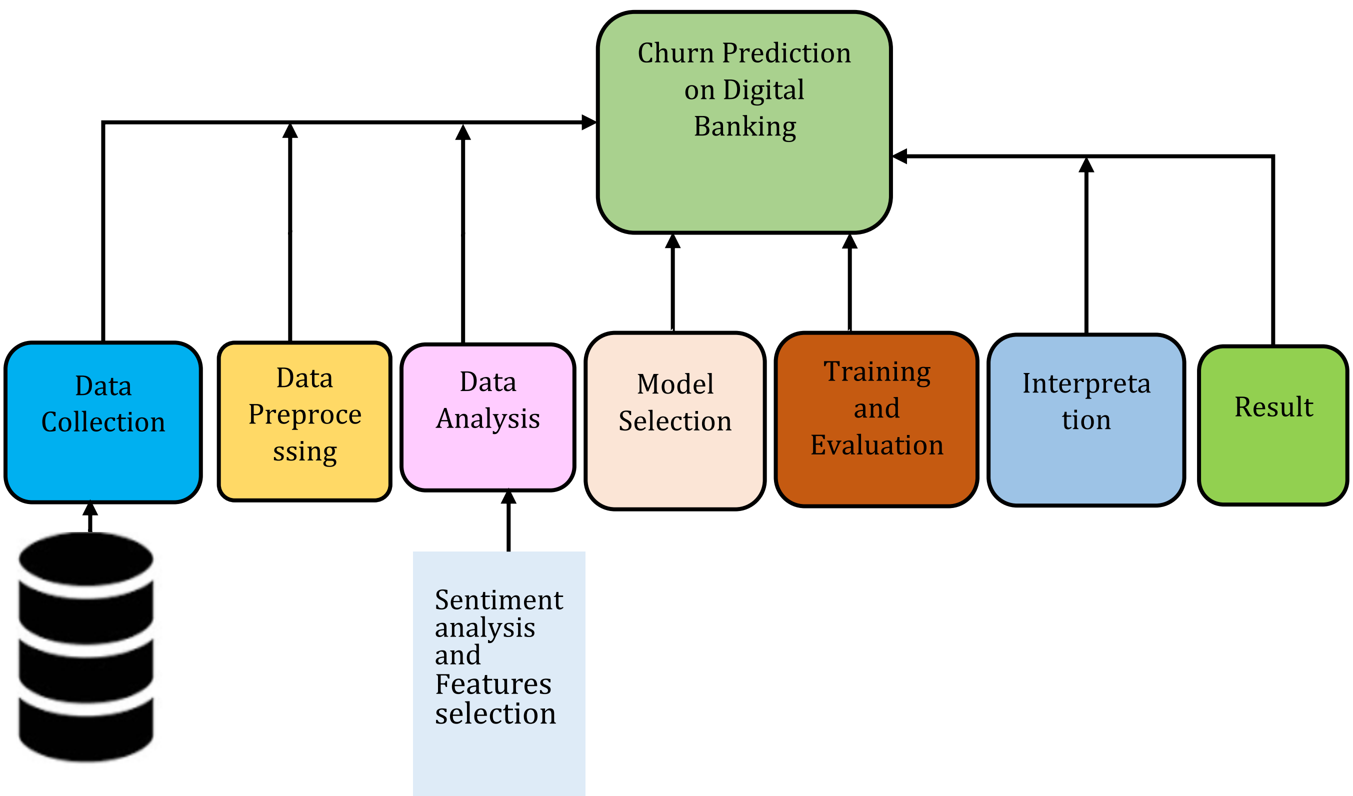

Figure 4: Research Methodology Flowchart

3.2. Data Collection

This research collected data over a year period within an Ethiopian bank, collecting a rich and varied array of customer-related attributes required to explore digital banking usage and assess customer churn. The dataset consisted of 783,466 rows and 23 columns, giving a total of 18,019,718 data entries and representing 35,293 unique customers. The resulting dataset ultimately included 23 attributes in total, including demographic, behavioral, transactional, and sentiment variables to represent customer engagement.

3.2.1. Demographic Attributes

Fundamental information about the customers is provided by the demographic characteristics. Important qualities include:

- REGION: Determines the customers' geographic distribution, allowing for the examination of regional consumption trends.
- BRANCH_NAME: The customer's specific branch is an indicator, which may affect service accessibility and preference.
- SEX: Gender information, which is helpful when examining trends in behavior and satisfaction among various gender groupings.
- CUSTOMER_ID: An individual customer's unique identification that makes it easier to accurately merge and track data from several sources.

3.2.2. Banking-Specific Features

To understand the financial relationship and account characteristics of customers, several banking-specific variables were included:

- ACCOUNT_TYPE: Indicates the type of account (e.g., savings, current) held by the customer.
- BALANCE: Reflects the current financial standing of the customer.
- ACCOUNT_STATUS: Categorizes accounts as active, dormant, or closed, which directly correlates with usage behavior and potential churn.

3.2.3. Behavioral Attributes

These features focus on the customer's relationship duration and interaction frequency:

- AGE: Denotes the customer's age, potentially influencing technology adoption and digital preferences.
- TENURE: Represents the length of time a customer has maintained a banking relationship, which can be a predictor of loyalty or risk of churn.
- LATEST_LOGIN: Captures the most recent digital login activity, acting as a proxy for digital engagement.

3.2.4. Transactional Data

To assess banking channel preferences and usage intensity, the dataset includes both aggregate and granular transaction data:

NO_DB_TRN: Number of digital banking transactions within a defined time window.

NO_BB_TRN: Number of branch banking transactions within the same period, allowing for comparative analysis of digital versus traditional banking.

- TOTAL_CHARGE: The total cost incurred from digital banking activities, which can impact satisfaction and retention.

- TRANSACTIONS and TRANSACTIONAMOUNT: Provide detailed logs of individual transactions and their respective values, offering deep insight into transaction behavior.

3.2.5. Churn Indicator

Importantly, the dataset has a clearly defined and labeled target variable which is CHURN and it tells whether the client stopped using the digital banking services during the time of the data sample. This target variable makes it possible to look at supervised machine-learning models to predict customer churn.

- CHURN: A binary variable that identifies whether a customer has ceased using digital banking services during the observation period. This Churn variable was labeled as 1 (churned) if the customer had stopped using digital banking for more than one year, while customers who were still actively using digital banking were labeled as 0 (non-churned), based on the National Bank of Ethiopia's directive.

3.2.6. Sentiment-Driven Features

Additionally, survey data was collected to help better understand consumer behavior and emotions. To sift out how clients felt about this bank's services, sentiment data was processed and converted into sentiment scores using BERT and standard scoring. After processing different sentiment results from the survey, we have made averages and variances for the sentiment texts and use the categorical responses as separate features, leading to the creation of the following derived features:

- AVERAGE_SENTIMENT: The overall sentiment score derived from customer feedback.

- SENTIMENT_VARIANCE: Measures the consistency or fluctuation in sentiment responses.
- CONTINUE_SCORE: Reflects customer inclination to continue using the service.
- SECURE_SCORE: Gauges customer perception of the platform's security.
- SATISFIED_SCORE: Captures general satisfaction levels.
- SPEED_SCORE: Reflects the perceived speed and responsiveness of digital services.
- EASY_SCORE: Indicates the ease of use and navigability of the digital banking platform.

Considering all of the attributes together, the 23 attributes form a robust, multi-faceted dataset which allows for thorough analysis, and development of accurate churn prediction models specific to the digital banking context.

3.3. Data Pre-Processing and Analysis

The data pre-processing and analysis phase was an important step in the preparation of the data we collected to predict customer churn. This step focused heavily on cleaning and transforming the input data set so it would conform to the specific requirements of machine learning. The first step was to clean the data. The first step involved dropping columns that had all missing values, which would help reduce noise and redundancy. Then missing values were addressed step by step. Categorical variables such as REGION and BRANCH_NAME had missing values filled with the most common category. Numerical variables such as BALANCE and AGE had missing values filled with the median to keep the distribution of the data the same but lower the impact from potential outliers. After the data was cleaned, the categorical features were turned into a numeric format using Label Encoding, so that the features could be used with machine learning algorithms. The numerical variables to follow such as BALANCE, AGE, and TENURE also underwent normalization via StandardScaler to help provide a mean of zero and a standard deviation of one for normalization as one of the most important things to look at in a modeling situation is how similar all features were at the start in order to improve convergence and maximize model performance.

The data was split into two sets 80 training and 20 testing sets to evaluate the models' effectiveness using previously unseen data. Considering that customers can possibly have more than one account, data were aggregated at the CUSTOMER_ID level: attributes like TENURE were converted to average values to obtain a more representative value, financial measures like BALANCE were summed to reflect total account holdings, while categorical features were aggregated using the model.

3.4. Sentiment Analysis

Sentiment analysis in this study involved processing and analyzing responses from a comprehensive survey designed to capture customers' emotional and behavioral feedback regarding their experience with digital banking services. The survey included a range of question types: categorical questions and sentiment questions.

To process these diverse question types, a multi-step approach was employed. First, the sentiment questions (e.g., "Do you feel our app or platform provides all the features you need? If not, what is missing?), were analyzed using a BERT-based sentiment analysis model, which classified the textual responses into three categories: positive, neutral, and negative. The model not only classified the sentiment but also provided confidence scores that were used to adjust neutral sentiment dynamically, ensuring more accurate sentiment interpretation. This approach allowed the detection of emotional tone (positive, neutral, or negative) in a variety of contexts, such as customer feelings towards usability, features, or overall satisfaction.

In addition to the sentiment analysis of textual responses, categorical encoding was applied to the categorical responses of the survey, such as satisfaction levels (e.g., very easy, easy), using ordinal encoding. This enabled the conversion of qualitative data into numerical values that were ready for integration with the rest of the dataset. For instance, higher levels of satisfaction were assigned higher numerical values, helping the machine learning model capture the relative intensity of customer opinions.

Finally, all the sentiment-based features such as average sentiment score and sentiment variance were calculated per customer to quantify emotional trends over time. These sentiment features, and the responses to the categorical questions, merged with the main dataset using the CUSTOMER_ID. A combined dataset with transaction, demographic, and sentiment properties allowed for the creation of a richer feature set for the churn prediction model.

By embedding the sentiment and emotional analysis in the overall prediction model, we were able to use not only transactional behavior but also emotional signals from which we could gauge why customers disengaged - churned, from digital banking or lending services. By adding categorical and sentiment responses, the accuracy and relevance of the churn prediction model improved and provided a more complete understanding of the customer experience and feeling.

3.5. Feature Selection

Feature selection represents an important step in the preparation of the data for machine learning, as it allows models to focus on features that have been identified as the most relevant and influential, thereby also improving performance and interpretable results. In this study, feature selection was conducted using Recursive Feature Elimination with Cross-Validation (RFECV), which is a powerful method to look at the most relevant features as it recursively eliminates the least relevant features from the model one at a time and directly evaluates performance each time during validation of the model. The feature selection process started with establishing a RandomForestClassifier model as the base model to rank the different features; the random forest was selected because it is an important ensemble-based classifier that can model non-linear relationships in complex data, and allow interpretability of the model through examining which features were the most important for predicting outcomes due to its in-built feature ranking.

RFECV was performed on the data set to iteratively rank and remove features inconsistent with model accuracy due to cross-validation, and to identify the highest ranked(s) integer ranking of features with the most predictive accuracy in outcomes. Features that were identified as less relevant or redundant will be removed, and the modeling performance will be updated to give the optimal number of significantly relevant features. This procedure meant that only the most important features were kept for further model training, thus reducing the likelihood of overfitting and simplifying the model with only the required complexity in features. Once the features were selected, the process was applied to the training (Train) and the test (Test) datasets, which then transformed the datasets to one that only had the relevant features. This process only permitted the model to learn from the relevant attributes, and hence it will generalize better on unseen data.

In addition to feature selection, class balancing was done to deal with actual/class imbalance in both datasets, which can adversely affect the model's performance, as was the case with churn prediction, where customers that churn may be fewer than those customers that do not. To rectify this issue, the Synthetic Minority Oversampling Technique (SMOTE) was employed on the Train and the Test datasets. SMOTE synthesizes samples of the minority class (churned customers) by interpolating between previous samples of the minority class. SMOTE creates a balanced class distribution. This method aids in reducing bias towards the majority class which can lead to incorrectly identifying customers as having churned when they have not. It also reduced inequality for accuracy measures. By combining RFECV for feature selection, along with SMOTE for class balancing, the preparation of the data was undertaken as mindfully as possible to offer the best chance of the model correctly predicting customer churn, not only considering the features that were relevant but also the equally relevant contribution of data set to the model. The features selected by using the RFECV method included REGION, LAST_LOGIN, BALANCE, NO_DB_TRN, TOTALCHARGE, and NO_BB_TRN, created from transactional, behavioral, and demographic data, as well as CONTINUE_SCORE, SECURE_SCORE, SATISFIED_SCORE, SPEED_SCORE, and EASY_SCORE, which were from the sentiment features.

3.6. Model Training

For developing a successful churn prediction system, the critical step is the process of developing and training each model. This will define how well the model can generalize to data that has not yet been collected while minimizing overfitting. There are several supervised machine learning classification models that are part of the model development process used in this study. Each model was trained and evaluated using a balanced dataset to improve the performance of individual class churn predictions and minimize biases that could happen due to class imbalance.

The approach was defined and trained in five different machine learning models, each with tuned hyperparameters, to minimize any potential overfitting. The models were selected to include different types of models, such as simple, complex, and non-linear. The models that were used were:

- Logistic Regression. This is a simple and common model for using binary classification tasks such as churn prediction.
- Decision Tree. This model organizes the data non-linearly, as this model separates the subject data into smaller and smaller groups according to the value of the feature. This allows for an easily interpretable model and is often highly successful for data where there really is a non-linear relationship between the independent and dependent variables.
- Random Forest: An ensemble learning method that combines multiple decision trees to improve accuracy and reduce the risk of overfitting by averaging the results of individual trees.
- Gradient Boosting: An iterative method that builds an ensemble of trees in a sequential manner, where each tree corrects the errors of its predecessor. It is known for its high predictive power and ability to model complex relationships.
- Neural Network (simulated as a Neural Network using MLPClassifier): A deep learning model that simulates the learning process of the human brain. In this case, the Multi-Layer Perceptron (MLP) classifier was used to model non-linear relationships in the data.

In order to minimize overfitting, and improve performance, specific hyperparameters were set for each machine learning model based on the characteristics of the dataset. The strategy encompasses the use of regularization, limitations on model complexity and early stopping decisions employed to obtain valid and reliable outcomes.

The hyperparameters used for the Logistic Regression model were max_iter=1000, random_state=42, C=0.01, penalty='l2', and solver='lbfgs'. The increase in the maximum iteration of 1000 needs to ensure the model has enough time to converge, which is important for use with larger datasets. The reduced regularization parameter of C=0.01 indicates that strong L2 regularization is being applied, penalizing for a large coefficient that will possibly keep overfitting the model. The solver 'lbfgs' was used for maximizing the logistical loss function. The random_state=42 parameter helps us to ensure that each run of the logistic regression is consistent and reproducible, quantified measures of overfit, and reliable future predictions.

The Decision Tree Classifier was run with max_depth=6, min_samples_split=40, min_samples_leaf=20, max_features='sqrt', ccp_alpha=0.005, and random_state=42. Setting the max tree depth to 6 reduced the model's complexity allowing for better generalization when attempting to predicting unseen data. The parameters min_samples_split=40 and min_samples_leaf=20 make sure enough samples are included in each internal node and leaf node resulting in reduced variance as well, having enough samples at each node would produce smoother and less predictable boundaries. The model contains some randomness authentic with using max_features='sqrt', allowing for feature randomness at each split and increasing the diversity and stability of the model. The ccp_alpha=0.005 allowed for cost-complexity pruning, which simply removes insignificant splits in the decision tree that reduce overall complexity. Ultimately, these custom tree settings aimed to keep the decision tree simple, and interpretable, reduce overfitting, and generalize well.

The hyperparameters used by the Random Forest Classifier include: n_estimators=100, max_depth=5, min_samples_split=30, min_samples_leaf=20, max_features='sqrt', max_samples=0.75, and random_state=42. Using 100 trees provides enough ensemble diversity while still keeping the computational cost reasonable. Max depth at 5 and minimum split and leaf sizes of 30 and 20, respectively, help to control each tree's complexity and prevent overfitting. Using max_features='sqrt' gives randomness at each splitting point which able to more robustness in the ensemble. The use of max_samples=0.75 allows for 75% of the training data to be used per tree, providing more diversity and reducing variance.

For the Gradient Boosting Classifier, the values of the hyperparameters were adjusted by having a greater n_estimators value (150) with a small learning rate (0.01) enabling the model to slowly build on itself without adding large mistakes. Shallow trees with a max depth of 2 help mitigate overfitting by keeping individual models simple. We were accessing only 75% of the training data for contributions to each tree (subsample=0.75), which aids in the diversity of the models and lessens overfitting. To help facilitate early stopping, we reserved 15% of the training data for validation (validation_fraction=0.15), where early stopping occurs when there is no improvement for 5 consecutive iterations (n_iter_no_change=5). Finally, in early stopping, the model is iteratively converging through the use of a very small tolerance (tol=1e-4) on the change, in order to get the best possible precision and performance during the convergence process. Overall, the hyperparameters selected allowed the Gradient Boosting model to work toward balancing performance with generalization, which is extremely important, especially with respect to overfitting due to the sensitivity of the boosting algorithms on that behavior.

Lastly, the Neural Network (MLP Classifier) was run with hidden_layer_sizes=40, max_iter=500, alpha=0.01, early_stopping=True, validation_fraction=0.15, tol=1e-4, learning_rate_init=0.001, and random_state=42. A single hidden layer with 40 neurons gives sufficient capacity to learn complex structures. In the case of Alpha, it is a parameter for the regularization term and penalizes the model if weights become large, this enforces generalization. In the case of early_stopping=True, the training will stop if the validation performance does not improve while using 15% of the data for validation (validation_fraction=0.15) is a reasonable way to enable reliable early stopping. The learning rate (learning_rate_init=0.001) for the MLP Classifier forces the model to converge, while a small "tolerance" (tol=1e-4) ensures that the weight optimization continues until only negligible improvements are being made. All of the Neural Network hyperparameters allow the model to learn useful structures in the data while introducing multiple mechanisms for preventing overfitting, whilst also ensuring stable and generalized performance.

In this study, the models did not simply predict if the target variable was a churn or not, but rather predicted churn probabilities, which generated more informative and rich results. It provided more nuanced insights into customer churn because it allowed us to identify customers across a wide range of churn probabilities instead of identifying someone as either churned or non-churned. Therefore, the trained Random Forest model will give us a churn probability for the customer, presented as a percentage of the likelihood that the customer will churn, based on that customer's aggregated historical behavior.

3.7. Model Evaluation

The model evaluation drew upon a few performance metrics appropriate for classification, including Accuracy which indicated the overall prediction accuracy for both churn and no-churn cases. Precision, Recall, and F1-Score are also applied to measure the model’s accuracy in predicting churn cases. The ROC-AUC Score gauges the model’s ability to distinguish between churn cases and no churn cases. Using hyperparameters optimization performance metrics were balanced.

Model performance was compared using a variety of metrics, including Precision, Recall, F1-Score, log loss, and ROC-AUC metrics measuring the model's strengths and weaknesses. A main component of the evaluation was assessing for overfitting by examining the model performance between training and testing sets. In the very last comparisons, model results were compared across all the models to identify the model which model was the one that provided well-balanced performance with generalization. In addition to tuning the hyperparameters and assessing the evaluation metrics, all models benefitted from a comprehensive evaluation process to develop a precise and effective churn prediction model.

3.8. Ethical Considerations

All customer data used in the analysis were anonymized to reduce the impact on individual privacy. Data collection and use were in accordance with appropriate local data protection laws, and any personally identifiable information (PII) was expressly not included in the dataset. Sentiment analysis on volunteer survey responses was used, with informed consent collected prior to collection.

Machine learning models were developed with ethical criteria in mind and SHAP was used to provide interpretability and lower the risk of decision making. The study avoided judgment based on biases or poorly conceived assumptions by using one too many features covering behavioral indicators, demographics, and elements that indicated emotional indicators. Careful consideration was taken in interpreting the results to ensure the proper representation of requesting customers' behaviors is reflected as there are context-specific elements of Ethiopia's socioeconomic and digital context.

.

4. Results and Discussion

4.1. Experimentation Setup

The experiments were conducted on a Lenovo ThinkPad laptop (Model: 21JQS2DU00) powered by a 13th Gen Intel® Core™ i7-13700H processor with 14 cores and 20 logical threads running at 2.4 GHz. The system was equipped with 16 GB of RAM and approximately 954 GB of storage, offering adequate computational resources for machine learning tasks. The development environment was based on Python 3.9 and utilized several essential libraries. The entire workflow was developed and executed within Jupyter Notebook.

4.2. Model Performance

To evaluate the predictive accuracy of several machine learning models in predicting churn in digital banking service customers, five classifiers were trained and tested: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Neural Network (MLP). Each model was assessed through stratified cross-validation, and tested on eight performance metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and log loss. Table 2 outlines the results for each of the models on the six evaluation metrics; further subsequent sections present a complete evaluation of the performance of each of the models.

Table 2: Result of Evaluation of Models

Model	Accuracy	ROC-AUC	Log loss	Precision (W)	Recall (W)	F1-Score (W)
Logistic Regression	0.83	0.90	0.39	0.83	0.83	0.83
Decision Tree	0.92	0.95	0.21	0.93	0.92	0.92

Random Forest	0.95	0.98	0.20	0.95	0.95	0.95
Gradient Boosting	0.92	0.98	0.37	0.93	0.92	0.92
Neural Network (MLP)	0.93	0.97	0.20	0.93	0.93	0.93

4.2.1. Logistic Regression

The Logistic Regression model provided steady cross-validation ROC-AUC scores in the 0.89 to 0.91 range, averaging around 0.90. The Logistic Regression model achieved a test accuracy of 83% and also demonstrated precision and recall scores that showed consistent performance across both churned and retained classes. The model reported an ROC-AUC score of 0.90 and a log loss of 0.39 which indicates reasonably strong separability between churned and non-churned customers.

4.2.2. Decision Tree

The Decision Tree Classifier improved results compared to the Logistic Regression model with cross-validated ROC-AUC scores of 0.91 to 0.95. The Decision Tree model reported a test accuracy of 92% and a powerful ROC-AUC score of 0.95, and it also reported precision and recall values above 0.87 in both churned and retained classes. Overall, this model appears to be performing extremely capable, but slightly more susceptible to overfitting.

4.2.3. Random Forest

Random Forest models outperformed the previous models in all of the evaluation metrics. It had a much higher cross-validation ROC-AUC score (from 0.98 to 0.99), and a test accuracy of 95%. With a ROC-AUC of 0.98 and a log loss of just 0.20, the model had very good discriminative power and generalizability. The high recall and precision for both classes of models suggest it was able to accurately identify someone who is a churned customer and someone who is not a churned customer while making small misclassifications.

4.2.4. Gradient Boosting

Gradient Boosting also demonstrated a high predictive capability, with cross-validation ROC-AUC scores above 0.95. The test accuracy of Gradient Boosting was 92%, the same accuracy as the Decision Tree. Although it had a slightly higher log loss of 0.37 compared to Random Forest, its ROC-AUC scores of 0.98 and equal classification metrics make it another strong candidate for capturing nuances in patterns related to customer churn behavior.

4.2.5. Neural Network (MLP)

The Multi-layer Perceptron (MLP) neural network model exhibited strong results, achieving a test accuracy of 93% and an ROC-AUC score of 0.97, with a log loss of 0.20 the least among all models. This means the model was well-tuned and capable of generating reliable probabilistic predictions. The MLP maintained high precision and recall across both classes showing that the model was well-tuned and able to generalize well to unseen data.

4.2.6. Model Comparison Summary

Performance comparisons across all the models show that Random Forest had the highest overall results (accuracy and ROC-AUC scores) so this model was determined to be the most effective model for predicting digital banking customer churn for this study. Neural Networks (MLP) were second overall with comparable performance levels and lower log loss. Logistic Regression provided reasonable results and is effective in reigning as a simple model that could be used as a baseline. Gradient Boosting and spent well-performing models like Decision Tree, also showed their relative risk of overfitting comparatively higher on the training dataset with the Decision Tree model.

Overall, the ensemble method Random Forest outperformed the singular model given its capability to model intricate, nonlinear data patterns. These results ensure the use of ensemble learning as a valid approach to churn prediction in the context of Ethiopian digital banking services.

4.3. Performance under Different Feature Sets

In this section, we describe the performance of the Random Forest model trained with different selections of features. Feature selection was performed to enhance model efficiency by reducing overfitting through the removal of irrelevant features. Several feature sets were tried to observe how the nature and number of features affect the model performance.

The feature selection process that preserved 11 major features produced the best results across the trial. The Random Forest model trained with these 11 features did better overall than the models trained with selected features of varying quantities and qualities.

The table 3 below summarizes the key evaluation metrics—ROC-AUC, Accuracy, Precision, Recall, F1-Score (Weighted Average)—for Random Forest models trained with different selections of features:

Table 3: Model Performance under Different Feature Selection

Feature Set	**ROC-AUC (Test)**	**Accuracy**	**Precision (W)**	**Recall (W)**	**F1-Score (W)**	**Remark**
11	0.9843	0.95	0.95	0.95	0.95	Best Overall Stability & Accuracy
23	0.9632	0.94	0.95	0.94	0.94	High ROC-AUC, Lower Accuracy, recall and F1 Score
16	0.9849	0.94	0.95	0.94	0.94	Balanced, Slightly Lower Stability

8	0.9845	0.94	0.95	0.94	0.94	Similar AUC, Lower Accuracy
19	0.9852	0.94	0.95	0.94	0.94	Higher AUC, But Lower Overall Accuracy

As shown in the table, while some models achieved slightly higher ROC-AUC values, the model trained with 11 selected features achieved the best balance of high ROC-AUC (0.9843) along with superior Accuracy (0.95), Precision, Recall, and F1-Score, all consistently at 0.95.

This confirms that the 11-feature selection provided a better trade-off between model complexity and prediction accuracy, making it the optimal configuration for reliable churn prediction in this study.

4.4. Impact of Sentiment Analysis on Model Performance

The inclusion of sentiment analysis into the churn prediction model resulted in significant performance enhancements across all machine learning algorithms. The tables below provide a summary of five evaluation metrics: Precision, Recall, F1-Score, ROC-AUC, and Log Loss collected before the inclusion of sentiment features and after the inclusion of sentiment features. Each metric is intended to reflect a different aspect of model performance, and the additions across the full board are collectively reflective of the value of including sentiment data in understanding market behaviors.

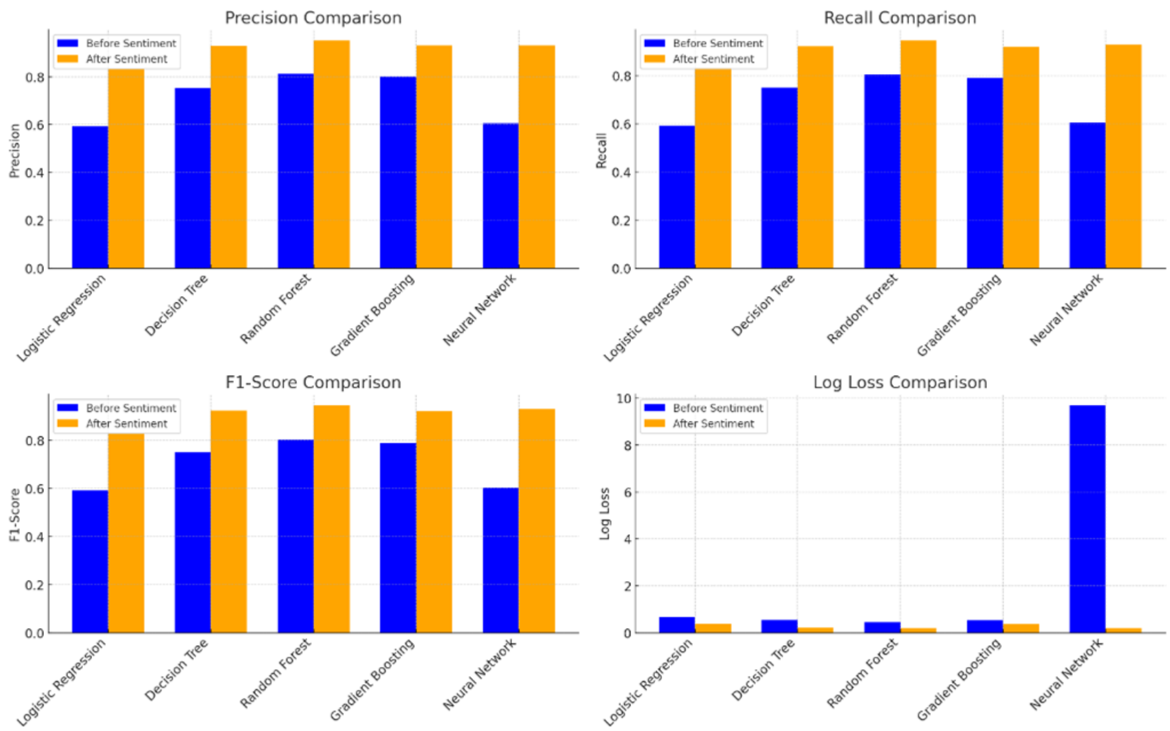

Figure 5: Model Performance Comparison Before vs. After Sentiment Integration

In Figure 5, the bar chart shows five machine learning models (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Neural Network [MLP]) based on four performance metrics: Precision, Recall, F1-Score, and Log Loss, before and after incorporating sentiment analysis into the models. It shows each result of models performance using evaluation metrics. For precision, which identifies the proportion of true churners out of the predicted churners, all machine learning models showed increased precision with the addition of sentiment features. The Neural Network and Logistic Regression models had the largest increases, with the Neural Network showing an increase from about 0.60 to more than 0.93 precision. The increased precision can be illustrate as a decreased number of false positives when those were previously present in the model, meaning the model was likely more successful in predicting customers likely to churn.

Recall is the measure of how well the model identifies all actual churners. After adding sentiment analysis, all the models were better at identifying customers at risk of leaving, while the Neural Network (from ~0.60 to ~0.93) and Logistic Regression (from ~0.59 to ~0.83) had the largest increases in recall. In this case, it indicates that all models were improved at detecting potential churners and reducing the chance of missing high risers along the way.

There were notable improvements in the F1-Score (harmonic mean of precision and recall) across all models. The models that previously underperformed, such as Neural Network and Logistic Regression, had big improvements, showing that sentiment data helped these models balance out false alarms and churner detection. Even though Random Forest and Gradient Boosting are stronger classifiers, sentiment data continued to help these models.

Log Loss measures the cost of false classifications that are very certain. Every model had lower log loss after sentiment data integration. The Neural Network had the largest log loss drop from an unstable 9.69 to 0.20, a very large improvement in the confidence of the predictions. Similarly, Decision Tree and Logistic Regression showed significant log loss improvements suggesting that the models had become better and more certain in their predictions.

Figure 6, compares ROC curve model performance before and after sentiment analysis. All models showed improved class separation, with curves shifting closer to the top-left corner. Notably, the Neural Network model's AUC increased from 0.63 to 0.97, and Logistic Regression improved from 0.63 to 0.90, showing substantial gains.

Even strong baseline models like Random Forest and Gradient Boosting improved, reaching 0.98. This highlights how sentiment features enhance the models' ability to distinguish churners from non-churners effectively.

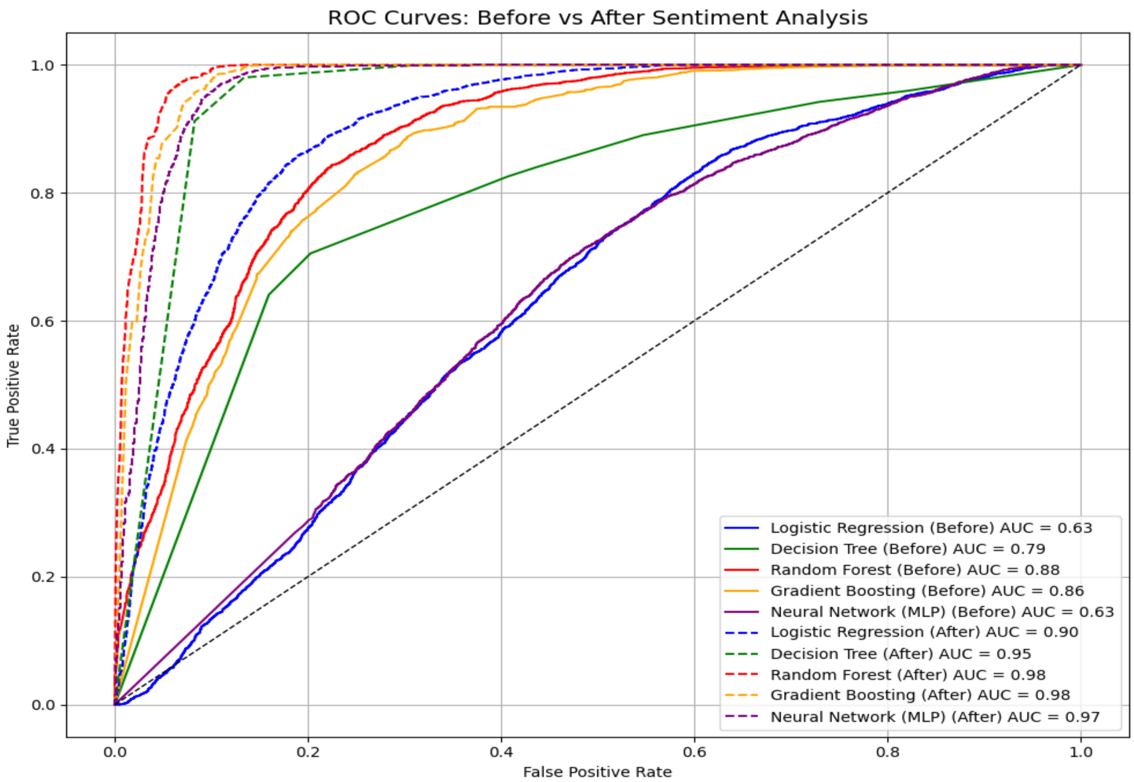

Figure 6: ROC-AUC Curve Before vs After Sentiment Analysis

This bar chart and ROC-AUC curve effectively demonstrate that adding sentiment features significantly enhanced model performance across all key metrics. Models that initially struggled, benefited the most, and even strong classifiers like Random Forest and Gradient Boosting achieved more stable and accurate predictions. This emphasizes the critical role of customer emotion and sentiment analysis in strengthening churn prediction in digital banking environments, particularly in emerging markets like Ethiopia.

4.5. Churn Probability Results

After training and evaluating the churn prediction models, the next step involved estimating individual customer churn probabilities. This was achieved by aggregating each customer's behavioral and transactional data into a single summarized input, which was then passed to the trained machine learning model specifically, the Random Forest classifier, which demonstrated the highest overall performance across evaluation metrics.

- Extracted all data rows associated with a specific CUSTOMER_ID from the full dataset.
- Aggregated the numerical features using the mean to produce a representative customer profile.
- Matched the structure of the input to the feature set expected by the model (i.e., 11 features).
- Used the predict probability method to output the likelihood of churn as a percentage.

This method enables customer-level churn risk analysis, facilitating targeted intervention strategies.

4.5.1. Sample Prediction Results

Some example customers were selected to demonstrate the output of churn probability predictions.

Table 4 results were generated using the aggregated feature inputs, and the model successfully handled both high- and low-risk profiles.

- Customer 9837947 showed a churn probability of 81.84%, indicating a strong signal that this customer is at risk of leaving. This suggests the customer may have exhibited signs such as low engagement, poor sentiment feedback, reduced digital activity, or a high branch-to-digital banking ratio.

- Customer 8983987 registered a 0.0% churn probability, highlighting a highly satisfied and engaged customer. Such customers may consistently use digital platforms, report high satisfaction, and show positive behavioral trends.

Table 4: Churn Probability Result

Customer ID	Churn Probability	Interpretation
9837947	81.84%	High likelihood of churn; requires immediate attention.
8983987	0.00%	Very low risk; likely to remain engaged.

The approach provides personalized churn risk scores, enabling the bank to:

- Proactively retain high-risk customers through targeted campaigns.
- Focus loyalty programs on low-risk customers to maintain engagement.
- Prioritize resources by risk segment.

This individualized churn probability prediction method is a valuable tool for customer-centric decision-making. By translating historical data and sentiment insights into actionable probabilities, the bank can deploy data-driven retention strategies with higher precision. As shown in the examples, the model effectively distinguishes between high and low churn risk profiles, which enhances its practical value in a real-world banking context.

4.6. Feature Importance

Understanding the relative importance of different features in predicting customer churn is critical for guiding business interventions and enhancing customer retention strategies. In this study, feature importance was assessed using the Random Forest model, which consistently outperformed others across key evaluation metrics. Additionally, SHAP (Shapley Additive explanations) values were employed to provide a model-agnostic, interpretable explanation of each feature's contribution to the model's predictions.

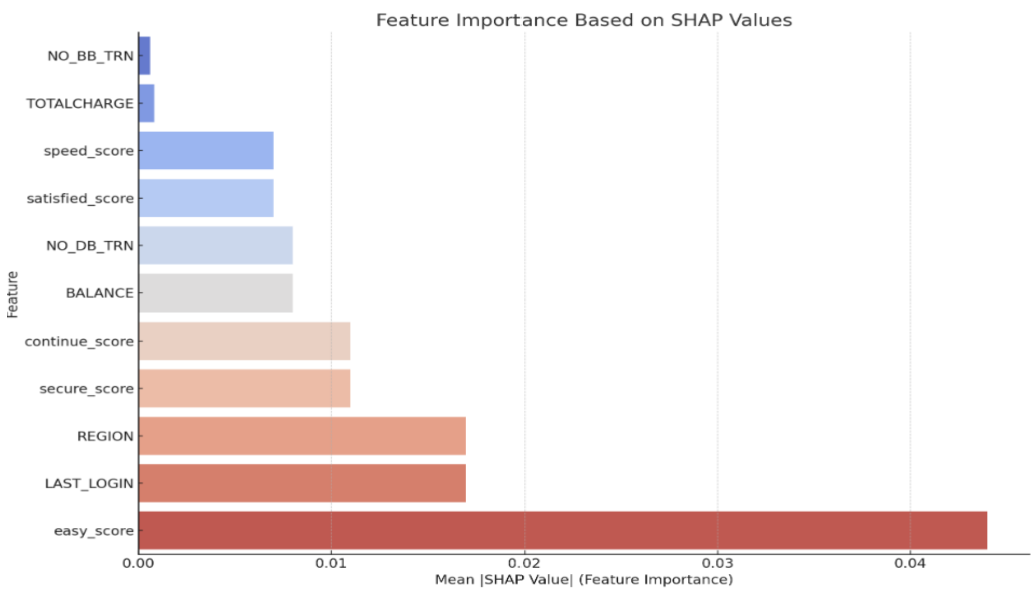

Figure 7: Feature Importance Result on SHAP Values

Figure 7 shows the feature importance result, to identify the most impactful predictors, we focused on features with SHAP values greater than 0.01, a commonly accepted threshold indicating significant influence on the model's output. The following features met or exceeded this threshold:

- EASY_SCORE (0.0438): This was the most influential feature. Derived from sentiment analysis, it reflects the customer's perceived ease of using digital banking platforms. Its prominence underscores the importance of user-friendly interfaces in fostering continued usage.

- REGION (0.0170) and LAST_LOGIN (0.0170): These features highlight the importance of customer demographics and engagement behavior. Regional disparities may point to differences in service accessibility or satisfaction, while LAST_LOGIN serves as a direct proxy for recent digital activity, a known churn indicator.
- CONTINUE_SCORE (0.0106) and SECURE_SCORE (0.0106): These sentiment-based scores represent the customer's perceived intention to continue using the service and their confidence in the platform's security. Their inclusion affirms that trust and long-term confidence are vital in retaining digital customers.

In addition to these top contributors, several features had moderate SHAP values slightly below the 0.01 threshold, yet were retained due to their contextual relevance in the Ethiopian digital banking landscape:

- NO_DB_TRN (0.0080) and BALANCE (0.0080): While not as impactful as sentiment or engagement features, transaction frequency, and account balances still offer valuable insights into a customer's financial behavior and engagement level.
- SPEED_SCORE (0.0067) and SATISFIED_SCORE (0.0067): These sentiment-based features provide feedback on customer satisfaction and perceived platform performance, supporting their inclusion as supplementary behavioral signals.
- TOTALCHARGE (0.0005) and NO_BB_TRN (0.0005): These features contributed minimally to the model’s predictions. Their low SHAP values suggest that in this context, digital churn is more closely tied to customer experience and digital engagement than to transactional fees or branch-based activity.

This study ascertains that, by integrating SHAP values into the feature selection process, a more interpretable and transparent model can be formed that utilizes significant, explainable features. Anytime the SHAP values are above 0.01 (e.g., EASY_SCORE, REGION, LAST_LOGIN, CONTINUE_SCORE, SECURE_SCORE), these features serve to improve predictive quality and enable the formation of more informed decisions and strategies. In addition, features with a moderate SHAP value, (e.g., NO_DB_TRN, BALANCE), were also included due to their relevance to the customer banking domain and contributive predictive power as an added feature on top of behavioral and emotional indicators. This included moderate features due to their relation to the customer banking domain and to observably enhance predictive power as an additional feature related to behavioral indicators and emotional categorization.

The results propose that banks should focus on providing a good user experience, consistent digital engagement, and security to reduce customer churn. Moreover, incorporating real-time sentiment analysis and regional personalized services can potentially lead to elevated customer loyalty in digital banking services.

4.7. How Results Answer Research Questions

This research was designed with the important questions that were fundamental to understanding customer churn in the Ethiopian digital banking landscape. First, we wanted to know which machine learning model could best predict customer churn. The results indicate that the Random Forest method almost unconditionally performed better than all models, in terms of different metrics, not only proving to be the best model in this case, but it is reasonably indicative of being robust to variation in data, and can accurately predict churners and non-churners. The second question aimed to determine the extent to which sentiment analysis improves the ability to predict churn. The results revealed that the addition of sentiment markedly enhanced the accuracy of every model. For example, precision increased by about 20% - 30%, recall increased by 20% - 35%, F1 increased by 20% - 35%, ROC-AUC increased by 20% - 35%, and log loss decreased by roughly 30% - 50% after sentiment was added. This significant improvement demonstrates the strong impact of customer sentiment signals to better provision the models to categorize possible churners with fewer false alarm rates.

Finally, this study identified the key predictors of churn as part of the Ethiopian digital banking context, which are: EASY_SCORE, REGION, LATEST_LOGIN, CONTINUE_SCORE, SECURE_SCORE, NO_DB_TRN and BALANCE. The significance of these variables demonstrates the pivotal role they may play in distinguishing customers at risk and the wider implications for financial institutions to create customer retention strategies.

4.8. Discussion

The chapter presented and analyzed the results of machine learning models that were built for the prediction of customer churn in the context of Ethiopian digital banking. Understanding the importance of the data with respect to the study's objective and the wider context of digital banking was the central objective of this discussion. The comparative analysis of the various supervised learning models enabled the selection of an appropriate approach for churn prediction. Given the performance differences across models, the study also generated insights into the importance of the various features of the behavioral, demographic, and sentiment-based factors associated with churn through its feature importance and prediction output analysis. The study offered a great deal of evidence of the behavioral trends of customer engagement patterns, patterns of service, and patterns of satisfaction among Ethiopian digital banking users.

Perhaps even more interesting than the many observations made based on these model results was the inclusion of individual churn probability scores, which lead to a customer-focused perspective that could be used to apply positive termination strategies leading to individualized retention strategies - resulting in risk-based decisions and opening the door for personalized churn strategies rather than a binary base decision-making model.

The result also emphasized the benefits of incorporating local features and sentiment analysis to increase predictive accuracy and contextual sensitivity. These findings highlight the study's goal of designing a locally validated and practical churn prediction model.

Overall, the findings provide a launching pad for more informed, data-driven customer retention strategies as the digital banking context continues to evolve in Ethiopia.

5. Conclusion and Recommendations

5.1. Summary of Findings

This dissertation was undertaken to develop and assess machine learning models for predicting customer churn in the context of digital banking in Ethiopia. Integrating and collecting various data including transactional behavior, demographics, system usage, and sentiment analysis, the study produced several findings:

The inclusion of Ethiopian-oriented variables including regional digital usage patterns, local sentiment, and behavioral scoring were all keys to achieving insights and usable, meaningful findings. The study trained and evaluated the performance of five machine learning models, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Neural Network (MLP), and identified the most effective modeling approach to predicting churn. Out of these models, the Random Forest model produced the best performance. The comparative analysis indicated ensemble methods, particularly Random Forests, were well-suited to modeling churn behavior in a digitally disruptive environment like Ethiopia. It also demonstrates incorporating sentiment analysis significantly enhanced the predictive performance of all models. The paper clearly stated that customer perceptions and emotional feedback are crucial indicators of churn. This shows the value of unstructured textual feedback in refining predictive models.

Utilizing the selected model, churn probabilities were calculated at the customer level, and we calibrated this with high confidence. The results indicated clear differences in churn probabilities as expressed in terms of Local Sensitivity—thus highlighting the potential for models used in targeted customer retention. Through SHAP analysis, the key predictors of churn were identified in the context of Ethiopian digital banking. These were EASY_SCORE, REGION, LATEST_LOGIN, CONTINUE_SCORE, SECURE_SCORE, NO_DB_TRN, and BALANCE. This shows the relevance of usability, perceptions of trust, digital engagement, and geographic factors associated with churn.

Therefore, the development of this work included different elements of machine learning, sentiment analysis, churn probability, identifying key predictors of churn, and localized behavioral features offering a strong framework to model and understand customer churn.

Overall, these findings provide the empirical foundation for actionable strategies to improve digital customer retention for Ethiopian banks.

5.2. Contributions

This research contributes to the field of churn prediction by emphasizing the power of merging traditional behavioral characteristics with sentiment analysis designed in a developing economic context, like Ethiopia. This research incorporates customer feedback and emotion-based predictors from contextual surveys. Further, it offers real-world feature selection based on SHAP values, which further supports both the explainability and performance of the model. It also validates churn prediction models with several machine-learning algorithms. The study provides a clear picture of the uplift sentiment provided and demonstrates a comprehensive framework that could be useful for churn prediction in Ethiopia as well as other developing markets with the same digital adoption barriers. Additionally, the study's aggregative method of churn probability per customer provides a practical way to assess the risk of churn for individual customers, providing a realistic operational risk measurement.

5.3. Practical Implications

The results of this study lay the groundwork for Ethiopian banks to leverage learnings toward moving their retention efforts to data-driven, customer-centric practices. The results highlight the opportunity to improve digital experiences for users, especially in relationship to usability and security interface features that are significant predictors of churn prediction. Banks should invest in user-friendly mobile apps, offer multilingual interfaces, and enhance platform stability to build digital trust. Secondly, the significant role of regional and sentiment-based factors suggests that customer engagement strategies should be localized, considering both cultural and emotional aspects of customer behavior. Banks can use real-time sentiment monitoring tools to proactively address dissatisfaction before customers disengage. For policymakers, the results emphasize the need to promote responsible data usage, support infrastructure development for digital banking, and facilitate partnerships between financial institutions and tech providers to boost digital financial inclusion. These actions collectively support a more resilient and inclusive banking ecosystem.

5.4. Future Work

Future research can build on this study by further deepening the personalization of churn prediction models. One promising direction is the development of customer-centric frameworks for identifying churn reasons from the sentiment attributes. Additional directions include time-series modeling to capture behavioral trends and integrating real-time feedback loops to enable continuous model learning. These advancements would improve accuracy and make churn prediction a proactive tool for customer engagement and innovation.

References

1. Malyshev, A. *What Is Digital Banking? [2025 Guide & Trends]*. Banking Software 2025 06/1/25 [cited 2025 April 06]; Available from: https://sdk.finance/what-is-digital-banking/.

2. Wewege, L., et al., *Disruptions and digital banking trends.* 2020. **10**(6): p. 15-56.

3. Singh, B. and C. Kaunert, *Futuristic Digital Banking and e-Financial Strategies Stimulating Financial Inclusion: Satellite SDG-1 Alleviating Poverty Compliance Sustainable Development Lensing Global Perspectives*, in *E-Financial Strategies for Advancing Sustainable Development: Fostering Financial Inclusion and Alleviating Poverty*. 2024, Springer. p. 199-214.

4. Balkan, B.J.T.I.o.A.I.o.G., Economics and V.I. Finance, *Impacts of Digitalization on Banks and Banking.* 2021: p. 33-50.

5. Wewege, L., *The digital banking revolution*. 2017: Lulu. com.

6. Barnes, S.J. and B.J.I.j.o.m.c. Corbitt, *Mobile banking: concept and potential.* 2003. **1**(3): p. 273-288.

7. Online Banking Statistics 2024 By Finance, T., Growth. *Global Online Banking Adoption*. 2024 May 8, 2024 [cited 2024 02 November 2024]; Available from: https://scoop.market.us/online-banking-statistics/#:~:text=As%20of%202021%2C%20it%20is,reach%203.6%20billion%20by%202024.

8. Laukkanen, T. and M.J.J.o.F.S.M. Pasanen, *Mobile banking innovators and early adopters: How they differ from other online users?* 2008. **13**: p. 86-94.

9. Entele, B.R.J.E.J.o.S. and S. Development, *Mobile Banking Technology in Ethiopia.* 2019. **6**(2): p. 61-71.

10. Million, M.M.A.i.E.C.H.t.N. *Mobile Money Accounts in Ethiopia Climb High to Nearly 60 Million*. 2023 02 November 2024]; Available from: https://shega.co/news/mobile-money-accounts-in-ethiopia-climb-high-to-nearly-60-million.

11. *ETHIOPIA'S DIGITAL PAYMENTS PERFORMANCE*, in *PAYMENT SYSTEM AND DIGITAL FINANCIAL SERVICES*. 2024, NBE: Ethiopia.

12. Kaur, B., et al., *Digital banking in Northern India: The risks on customer satisfaction.* 2021. **9**(11): p. 209.

13. Zaidi, S.H. and R.J.R. Rupeika-Apoga, *Liquidity synchronization, its determinants and outcomes under economic growth volatility: Evidence from emerging Asian economies.* 2021. **9**(2): p. 43.

14. Tran, H., et al., *CUSTOMER CHURN PREDICTION IN THE BANKING SECTOR USING MACHINE LEARNING-BASED CLASSIFICATION MODELS.* 2023. **18**.

15. Revesai, Z., et al., *Customer Churn Analytics using Classical Machine Learning Algorithms and Deep Neural Networks: A Case of Zimbabwe Banks.*

16. Ashraf, R.J.A.a.S., *Bank Customer Churn Prediction Using Machine Learning Framework.* 2024.

17. Seid, M.H. and M.M. Woldeyohannis. *Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia*. in *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*. 2022. IEEE.

18. Kebede, D.A.J.J.o.I.T., Logistics and Law, *Digital Financial Products and Services in Ethiopian Banking Industry: An Enquiry on its Hindrances of Adoption and Usage.* 2020. **6**(2): p. 1-12.

19. Asfaw, T.J.T.S.T., *Customer churn prediction using machine-learning techniques in the case of commercial bank of Ethiopia.* 2023. **14**(03): p. 618-624.

20. Goodie-Okio Jennifer, A.J.I.A.J.o.M. and Marketing, *Mobile Banking Service Quality and Customer Satisfaction of Deposit Money Banks in Port Harcourt.* 2022.

21. Ranjan, S. and S.J.J.W.E.T. Sood, *Sentiment analysis based telecom churn prediction.* 2020. **7**(1): p. 2455-1880.

22. Lappeman, J., et al., *What social media sentiment tells us about why customers churn.* 2022. **39**(5): p. 385-403.

23. Kilimci, Z.H., H. Yörük, and S. Akyokus. *Sentiment analysis based churn prediction in mobile games using word embedding models and deep learning algorithms*. in *2020 international conference on innovations in intelligent systems and applications (INISTA)*. 2020. IEEE.

24. Wu, S.J.H.i.S., Engineering and Technology, *Customer Churn Prediction in Telecom Based on Machine Learning.* 2024. **94**: p. 113-118.

25. Rabih, R., W. Sun, and M. Ayoubi, *Enhancing Customer Churn Prediction in Telecom: A Hybrid Strategy with Quartile Calculation and Random Forest.* 2024.

26. Sina Mirabdolbaghi, S.M., B.J.D.D.i.N. Amiri, and Society, *Model optimization analysis of customer churn prediction using machine learning algorithms with focus on feature reductions.* 2022. **2022**(1): p. 5134356.

27. Peng, K., Y. Peng, and W.J.P.o. Li, *Research on customer churn prediction and model interpretability analysis.* 2023. **18**(12): p. e0289724.

28. Taboada, M.J.A.R.o.L., *Sentiment analysis: An overview from linguistics.* 2016. **2**(1): p. 325-347.

29. Medhat, W., A. Hassan, and H.J.A.S.e.j. Korashy, *Sentiment analysis algorithms and applications: A survey.* 2014. **5**(4): p. 1093-1113.

30. Wankhade, M., A.C.S. Rao, and C.J.A.I.R. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges.* 2022. **55**(7): p. 5731-5780.

Appendices

Survey Questionnaires

1. Write the location of branch you visit? (የጎበኟቹትን አካባቢ ቅርንጫፍ ይጻፉ?)

2. Enter your customer number. (የደንበኛ ቁጥርዎን ያስገቡ።)

3. Select your age range.(የዕድሜ ክልልዎን ይምረጡ።)

A. 18-24

B. 25-34

C. 35-44

D. 45-54

E. 55+

4. How satisfied are you with your overall experience using our digital banking services? (የዲጂታል ባንኪንግ አገልግሎታችንን በመጠቀም ባላችሁ አጠቃላይ ልምድ ምን ያህል ረክተዋል?)

A. Very Unsatisfied (በጣም አልረካሁም)

B. Unsatisfied (አልረካሁም)

C. Neutral (ገለልተኛ)

D. Satisfied (ረክቻለሁ)

E. Very Satisfied (በጣም ረክቻለሁ)

5. What do you like the most about our digital banking services? (በ ዲጂታል ባንኪንግ አገልግሎታችን በጣም የሚወዱት ምንድነው?)

6. What do you dislike the most about our digital banking services? (በዲጂታል ባንኪንግ አገልግሎታችን በጣም የሚጠሉት ምንድነው?)

7. How easy is it to navigate and use our mobile banking application? (የሞባይል ባንኪንግ መተግበሪያችንን ማሰስ እና መጠቀም ምን ያህል ቀላል ነው?)

A. Very Difficult (በጣም ከባድ)

B. Difficult (ከባድ)

C. Neutral (ገለልተኛ)

D. Easy (ቀላል)

E. Very Easy (በጣም ቀላል)

8. Have you faced any technical issues (e.g., app crashes, slow performance) while using our digital banking platform? (የእኛን ዲጂታል የባንክ ፕላትፎርም እየተጠቀሙ ሳለ ማንኛውም ቴክኒካዊ ችግሮች (ለምሳሌ፣ የመተግበሪያ ብልሽቶች፣ ዝግ ያለ አፈጻጸም) አጋጥመውዎታል?)

A. Yes (አዎ)

B. No (አይ)

C. Neutral (ገለልተኛ)

9. If your answer for the above question was "yes", please specify what was the issue that you have faced? (ከላይ ላለው ጥያቄ የሰጡት መልስ "አዎ" ከሆነ እባኮትን ያጋጠመዎት ችግር ምን እንደሆነ ይግለጹ?)

10. How would you rate the speed of transactions (e.g., transfers, Airtime top-up) on our digital banking services? (በዲጂታል ባንኪንግ አገልግሎታችን ላይ የግብይቶችን ፍጥነት (ለምሳሌ፣ ማስተላለፎች፣ የአየር ሰዓት ክፍያ) እንዴት ይመዝኑታል?)

A. Very Slow (በጣም ቀርፋፋ)

B. Slow (ቀርፋፋ)

C. Neutral (ገለልተኛ)

D. Fast (ፈጣን)

E. Very Fast (በጣም ፈጣን)

11. Do you feel our app or platform provides all the features you need? If not, what is missing? (የእኛ መተግበሪያ ወይም መድረክ የሚፈልጉትን ሁሉንም አገልግሎቶች እንደሚሰጥ ይሰማዎታል? ካልሆነ ምን ይጎድላል?)

12. How secure do you feel when using our digital banking services? (የእኛን ዲጂታል የባንክ አገልግሎት ሲጠቀሙ ምን ያህል ደህንነት ይሰማዎታል?)

A. Not Secure (ደህንነቱ የተጠበቀ አይደለም)

B. Somewhat Secure (በተወሰነ ደረጃ ደህንነቱ የተጠበቀ)

C. Neutral (ገለልተኛ)

D. Secure (ደህንነቱ የተጠበቀ)

E. Very Secure (በጣም አስተማማኝ)

13. Do you find our digital banking services accessible regardless of your location (e.g., rural or urban)? (የትም ቦታ (ለምሳሌ ገጠር ወይም ከተማ) የኛን የዲጂታል የባንክ አገልግሎት ተደራሽ ተድርገ አግኝተውታል?)

A. Yes (አዎ)

B. No (አይ)

C. Neutral (ገለልተኛ)

14. If your answer for the above question was "No", please specify where was the location you did not able to access the digital banking service? (ከላይ ላለው ጥያቄ የሰጡት መልስ "አይ" ከሆነ እባክዎን የዲጂታል ባንኪንግ አገልግሎት ማግኘት ያልቻሉበት ቦታ የት እንደነበረ ይግለጹ?)

15. Are there any obstacles (e.g., language, internet connectivity) that limit your use of digital banking? (የዲጂታል ባንኪንግ አጠቃቀምን የሚገድቡ መሰናክሎች (ለምሳሌ ቋንቋ፣ የበይነመረብ ግንኙነት) አሉ?)

16. Which words best describe how you feel about our digital banking services? (You can select multiple selection) (ስለ ዲጂታል የባንክ አገልግሎታችን ያለዎትን ስሜት የሚገልጹት የትኞቹ ቃላት ናቸው? (ብዙ ምርጫዎችን መምረጥ ይችላሉ))።

A. Convenient (ምቹ)

B. Frustrating (የሚያበሳጭ)

C. Secure (ደህንነቱ የተጠበቀ)

D. Unreliable (የማይታመን)

E. Fast (ፈጣን)

F. Complicated (የተወሳሰበ)

17. How likely are you to continue using our digital banking services in the future? (ለወደፊቱ የዲጂታል ባንኪንግ አገልግሎታችንን የመቀጠል ዕድሉ ምን ያህል ነው?)

A. Not Likely (አይቀጥልም)

B. Somewhat Likely (በመጠኑም ቢሆን)

C. Neutral (ገለልተኛ)

D. Likely (ሊሆን ይችላል)

E. Very Likely (በጣም እቀጥላለው)

18. Have you ever considered switching to a different bank or mobile banking service? (ወደ ሌላ የባንክ ወይም የሞባይል ባንክ አገልግሎት ለመቀየር አስበው ያውቃሉ?)

A. Yes (አዎ)

B. No (አይ)

C. Neutral (ገለልተኛ)

19. If yes, what factors influenced your decision to consider switching? (አዎ ከሆነ፣ ለመቀየር ባደረጉት ውሳኔ ላይ ተጽዕኖ ያሳደረባቸው ነገሮች ምንድን ናቸው?)

A. Poor customer service (ደካማ የደንበኞች አገልግሎት)

B. Limited features (ውስን አገልግሎቶች)

C. Technical issues (ቴክኒካዊ ጉዳዮች)

D. Fees and charges (የአገልግሎት ክፍያዎች)

E. Security concerns (የደህንነት ስጋቶች)

F. Other (ሌላ)

20. Any other comments or suggestions for us? (ሌላ ሚሰጡን አስተያየት ወይም ማስተካከያ አለ?)