

Addis Ababa University

College of Natural and Computational Sciences

School of Information Science(Information Science Track)

Predicting Risk Factors for Customer Churn with Explainable Artificial Intelligence

Ermiyas Amare

Sep 2025

Addis Ababa, Ethiopia

Predicting Risk Factors for Customer Churn with Explainable Artificial Intelligence

A thesis submitted to the School of Graduate Studies, College of Natural and Computational Sciences, Addis Ababa University in partial fulfillment for the Degree of Master of Science in Information Science

By

Ermiyas Amare

Advisor: Michael Melese(PhD)

Sep 2025

Addis Ababa, Ethiopia

Addis Ababa University
College of Natural and Computational Sciences
School of Information Science

Graduate Programs Board of Examiners

Thesis Approval Sheet

A thesis by **Ermiyas Amare**, titled **"Predicting Risk Factors for Customer Churn with Explainable Artificial Intelligence"** is approved by the undersigned members of the examining committee.

Name & Role	**Signature**
Michael Melese(PhD) (Advisor)	_____
(Chairperson)	_____
Wondwossen Mulugeta(PhD) (Examiner)	_____
Martha Yifiru(PhD) (Examiner)	_____

Date: _____

Declaration

I declare that this thesis is my original work and has not been presented as a partial degree requirement for a degree in any other university and that all sources of materials used for the thesis have been duly acknowledged.

Ermiyas Amare

Signature: _____

Date: _____

Thesis Advisor

Michael Melese(PhD), Addis Ababa University

Signature: _____

Date: _____

Abstract

Customer retention remains a critical challenge in the banking sector, particularly in emerging economies where customer behavior is rapidly evolving. Despite access to extensive data, Bunna Bank still relies on conventional customer relationship management practices that are often manual, reactive, and lacking analytical depth. This study focuses on predicting the risk factors for customer churn at Bunna Bank using machine learning and explainable artificial intelligence methods, using a real-world dataset of 308,293 customer records with 19 attributes having transactional behavior, account activity, and demographic information. Six machine learning classifiers were evaluated: Random Forest, Gradient Boosting, XGBoost, LightGBM, Extra Trees, and Multi-Layer Perceptron. Before modeling, the dataset was balanced using SMOTE to address class imbalance, and a hybrid feature selection strategy was applied to identify the most relevant attributes. XGBoost achieved the highest accuracy (92.23%), along with strong precision (0.96), recall (0.94), F1-score (0.95), and AUC (0.97), making it the most effective model for identifying high-risk customers. To enhance interpretability, SHAP and LIME were applied to the XGBoost model, providing both global and local explanations. The analysis revealed that features such as Has_Mobile Banking, Is_USSD Active, Transaction Count, Transaction Recency, Age, and ATM Activity were the most influential drivers of churn risk. Active digital banking usage and recent transactions were protective factors, while younger customers with low engagement showed higher churn risk. A prototype was developed using Streamlit to risk factor prediction that integrate predictive modeling with XAI enables accurate and transparent into customer churn.

Keywords: Customer churn, Risk factor prediction, Machine Learning, Model Explainability

Acknowledgment

I would like to express my gratitude to the individuals and organizations who have supported me throughout the process of completing this thesis. Above all, I wish to give my sincere thanks and praise to GOD, for providing me with knowledge, and opportunities throughout my life. I would like to take this opportunity to express my deepest gratitude to my advisor, Michael Melese(PhD.). Your continuous support, valuable guidance, and insightful mentorship have been fundamental to the successful completion of this thesis. I truly appreciate for the knowledge and wisdom you have shared, as well as your constructive feedback, encouraging words, and thoughtful advice throughout this journey. Your dedication and commitment to my growth have been truly inspiring, and I am sincerely thankful for your support. I would like to extend my deepest gratitude to my dear wife, Netsant Lulu, my dear son, Yonael Ermiyas, and my precious daughter, Yoadan Ermiyas. They hold a special place in my heart. My wife, especially, deserves my deep thanks for always supporting me through every situation, and constantly motivating me to persevere. She means the world to me, and am forever thankful for all she has done. I would like to express my sincere thanks to Million Meshesha(PhD.) for his valuable guidance and suggestions during the proposal stage, which had a significant impact on the direction of my thesis. A special thanks to my friends, Ermias Tsegu, Taye Mohammed and Mikiyas Girma, for their continuous help and friendship. I truly appreciate your support. I would also like to thank my classmates and professors in the Information Science Department at Addis Ababa University for making my time at the university such a memorable experience. Lastly, I appreciate the Bank Business Team for providing me with the necessary expert advice.

Table of Contents

Declaration iv

Abstract v

Acknowledgment vi

Table of Contents x

List of Figures xii

List of Tables xii

1 Introduction 1

1.1 Background 1

1.2 Motivation of the Study 3

1.3 Statement of the problem 4

1.4 Research Questions 6

1.5 Objective of the study 6

1.5.1 General Objective 6

1.5.2 Specific objectives 6

1.6 Methodology 6

1.6.1 Research Design 7

1.6.2 Literature Review 7

1.6.3 Data Collection 7

1.6.4 Data Preparation 7

1.6.5 Modeling 8

1.6.6 Evaluation 8

1.6.7 Prototype Implementation 9
1.7 Scope and limitation of the study 9
1.8 Significance of the study 9
1.9 Organization of the study 10

2 Literature Review **11**
2.1 Banking Industry. 11
2.2 **Customer Relationship Management** 12
2.2.1 Overview of Customer Relationship Management 12
2.2.2 Customer Churn 13
2.2.3 Machine Learning Algorithms 14
2.2.4 Machine Learning Applications in Customer Relationship Management 17
2.3 Machine Learning Techniques for Customer Churn Prediction 17
2.3.1 Random Forest 18
2.3.2 Gradient Boosting (GB) 19
2.3.3 Extreme Gradient Boosting (XGBoost) 21
2.3.4 LightGBM 21
2.3.5 Extra Trees 23
2.3.6 Multilayer Perceptron (MLP) 24
2.4 Explainable Artificial Intelligence (XAI) in Predictive Modeling . . . 26
2.4.1 Concept and Importance of XAI 26
2.4.2 The Need for Explainability 27
2.4.3 XAI Methods and Categorization 28
2.4.4 Evaluation Metrics and Tools 28
2.4.5 Types of Explanation 29
2.4.6 Explainable AI Methods 29
2.5 Related Works 30
2.5.1 Research Gap 34

3 Methodology **36**

3.1 Overview . . . 36
3.2 The Proposed Architecture and Method . . . 36
3.3 Data Collection and Understanding . . . 37
3.3.1 Descriptive Analysis . . . 39
3.3.2 Data Quality Verification . . . 41
3.4 Data Preprocessing . . . 41
3.4.1 Feature Selection Based on Domain Expertise . . . 42
3.4.2 Data Integration . . . 42
3.4.3 Attribute Construction . . . 42
3.4.4 Data Cleaning . . . 43
3.4.5 Data Transformation . . . 45
3.4.6 Data Standardization . . . 46
3.4.7 Labeling of Churn Risk . . . 48
3.4.8 Correlation Matrix . . . 50
3.4.9 Feature Selection . . . 51
3.5 Dataset Split . . . 54
3.6 Imbalance Dataset Handling . . . 55
3.7 Select a Predictive Model . . . 55
3.7.1 Random Forest . . . 56
3.7.2 Gradient Boosting . . . 56
3.7.3 Extreme Gradient Boosting . . . 56
3.7.4 LightGBM . . . 57
3.7.5 Extra Trees . . . 57
3.7.6 MLP . . . 57
3.8 Model Evaluation . . . 58
3.8.1 Confusion Matrix . . . 58
3.8.2 Accuracy . . . 59
3.8.3 Precision . . . 60
3.8.4 Recall . . . 60
3.8.5 F1-score . . . 60

3.8.6 Receiver Operating Characteristic 60
3.9 Feature Importance 61

4 Experimental Results and Discussion 62
4.1 Development Environment and Tools 62
4.2 Data Preprocessing Summary 63
4.2.1 Importing Libraries & Dataset Loading 63
4.2.2 Exploratory Data Analysis (EDA) 63
4.3 Results of Feature Selection 64
4.3.1 Model Evaluation with Chi-Square Selected Features 64
4.3.2 Dataset Splitting and Class Balancing 70
4.3.3 Model Development 71
4.4 Model Comparison 79
4.5 Model Explainability 80
4.5.1 XGBoost Explaibability Using SHAP Method: 81
4.5.2 XGBoost Explaibability Using LIME Method: 85
4.6 Low-Risk Customer (Instance 512) 86
4.7 High-Risk Customer (Instance 319) 87

5 Conclusion and Recommendations 89
5.1 Conclusion 89

References 91

List of Figures

2.1 Types of machine learning algorithms. 15

3.1 Proposed System Architecture. 37

3.2 initial extracted data. 38

3.3 Demographic and Behavioral trends. 40

3.4 Final dataset ready for experimentation. 43

3.5 Encoded Categorical Features 47

3.6 Correlation matrix of features. 51

3.7 $\text{confusion}_{m}atrix.png$ 59

4.1 Importing Libraries and Dataset Loading 63

4.2 Feature importance analysis 70

4.3 Performance evaluation for Random Forest 72

4.4 ROC curve for Developed Models 73

4.5 Gradient boosting performance evaluation 74

4.6 XGBoost performance evaluation 76

4.7 XGBoost Confusion Matrix 76

4.8 MLP performance evaluation 77

4.9 LightGBM performance evaluation 78

4.10 Extra Trees performance evaluation 79

4.11 SHAP local waterfall plot 512th instance for low risk class 82

4.12 SHAP local waterfall plot 319th instance for High risk class 83

4.13 global explanation 84

4.14 Feature contributions for Low-Risk customer (Instance 512). 87

4.15 Feature contributions for High-Risk customer (Instance 319). 88

List of Tables

4.1 Performance of classifiers with Chi-Square-selected features 65

4.2 Performance of classifiers with Random Forest-selected features . . . 66

4.3 Performance of classifiers with RFE-selected features for predicting customer churn risk 67

4.4 Performance of classifiers with hybrid-selected features for predicting customer churn risk 68

4.5 Top Features Ranked by XGBoost Importance for Customer Churn risk factor Prediction 69

4.6 Class Distribution Before SMOTE (80/20 Split, Stratified) 71

4.7 Class Distribution After SMOTE on Training Set (80/20 Split, Stratified) 71

4.8 Performance comparison of machine learning models for risk factor prediction 79

4.9 Key feature contributions for Low-Risk instance 512 87

4.10 Key feature contributions for High-Risk instance 319 88

Acronyms

RF Random Forest

AI Artificial Intelligence

MLP Multilayer Perceptron

ML Machine Learning

CatBoost Category Boosting

DNN Deep Neural Network

KNN K-Nearest Neighbors

DBMS Database Management Systemn

USSD Unstructured Supplementary Service Data

F1-score A metric for model evaluation

GMM Gaussian Mixture Models

CRM Customer Relationship Management

DNN Deep Neural Network

SVM Support Vector Machine

EDA Exploratory Data Analysis

LightGBM Light Gradient Boosting

AUC Area under the Curve

GBT Gradient Boosted Trees

XGBoost Extreme Gradient Boosting

SQL Structured Query Language

RFE Recursive Feature Elimination

XAI Explainable Artificial Intelligence

SHAP SHapley Additive exPlanations

LIME Local Interpretable Model-agnostic Explanations

SMOTE Synthetic Minority Oversampling Technique

CBE Commercial Bank of Ethiopia

Chapter 1

Introduction

1.1 Background

According to [1], identifying and predicting customer behavior becomes essential for banking institutions because their customer preferences exhibit constant change and extensive diversity. Among the various analytical techniques available churn prediction stand out as particularly valuable. Customer churn, defined as the gradual yet consistent loss of customers over time, has become a significant challenge for banks seeking to retain their customer base and maintain profitability [2,3]. By accurately identifying customers who are most likely to leave, banks can proactively implement retention measures.

Recent studies have demonstrated that machine learning (Machine Learning (ML)) techniques provide enhanced insights into customer behaviors and preferences, offering improved accuracy and efficiency in churn prediction [2, 4]. Despite these advances, Bunna Bank continues to rely on conventional data analysis methods such as Structured Query Language (SQL) queries and traditional reporting tools which, while useful for initial analysis, become increasingly inadequate as the volume and complexity of data grow. This reliance on legacy systems not only limits the bank's ability to accurately predict churn but also elevates the operational costs associated with extracting actionable insights.

As a result, the financial sector has shifted its focus toward more sophisticated techniques that utilize ML algorithms [5]. These advanced methods minimize manual effort and uncover complex patterns in customer data, allowing banks to remain com-

petitive and better serve their customers' evolving needs. Moreover, as highlighted by [2, 6, 7], artificial intelligence (Artificial Intelligence (AI)) has revolutionized customer relationship management (Customer Relationship Management (CRM)) by using data quality and pattern analysis to address real-world challenges such as customer churn prediction.

Furthermore, ensemble methods have been shown to enhance predictive accuracy and stability by combining the outputs of various models. Such methods, including homogeneous ensembles like Random Forest (RF) and Gradient Boosted Trees (GBT), as well as heterogeneous ensembles of different classifiers (e.g., decision trees and Support Vector Machine (SVM)), have proven effective in the context of churn prediction [8, 9].

Although customer churn prediction has been widely studied, there is still a noticeable gap in the literature. Many existing approaches are mainly concerned with building accurate models, but they often fall short in explaining the reasons behind a customer's decision to leave. For banks, this lack of clarity is a major limitation, since knowing who is likely to churn is not enough on its own. What is equally important is understanding why customers are leaving and which risk factors contribute the most to their decision. Without this interpretability, predictive models remain difficult to apply in practice, as managers cannot easily translate predictions into meaningful retention strategies.

In this study, customer churn risk factor prediction in the banking sector is addressed using advanced machine learning algorithms. Customer data were collected from Bunna Bank, consisting of demographic, behavioral, and transactional attributes. Ensemble-based classifiers were employed to improve predictive performance. To enhance interpretability, post-hoc explainable artificial intelligence (XAI) techniques were integrated to justify the predictions of the machine learning models. The main goal of this study is to develop a churn prediction model that not only forecasts which customers are at risk of leaving but also explains why these predictions are made. Moreover, this explainable forecasting approach is essential for enabling data-driven decision-making in customer retention and supporting more transparent and

trustworthy AI adoption in the Ethiopian banking industry.

1.2 Motivation of the Study

Customer churn poses a significant challenge for the banking sector, as it directly affects revenue and increases the cost of acquiring new clients. While machine learning (ML) algorithms are widely used for predicting churn, most research emphasizes prediction accuracy over understanding the reasons behind the predictions. Without insights into why customers leave, banks face difficulties trusting the results and making informed decisions. This study focuses on explainable artificial intelligence (Explainable Artificial Intelligence (XAI)) methods to shed light on the factors influencing churn, helping banks gain a clearer understanding of customer behavior and supporting more effective, data-driven decision-making.

Moreover, existing studies often overlook the integration of advanced machine learning techniques for churn prediction in the banking sector. Many approaches rely on single feature selection methods, limiting the robustness and overall quality of predictive models. Similarly, the use of ensemble learning remains relatively underexplored, which may constrain the performance and reliability of churn prediction systems.

Another significant gap lies in the interpretability of these models. While many methods can predict which customers are likely to churn, they often fail to provide clear explanations of why these predictions are made. This lack of transparency reduces the practical usefulness of churn prediction, as banks cannot fully understand the driving factors behind customer behavior.

This study is motivated by the need to address these shortcomings by developing a churn prediction framework that leverages hybrid feature selection, ensemble ML, and explainable artificial intelligence (XAI) techniques. The framework aims to enhance the accuracy and interpretability of churn predictions, providing Bunna Bank with actionable insights into customer behavior.

1.3 Statement of the problem

The banking industry has witnessed the adoption of machine learning (ML) and data-driven techniques, which have reshaped customer relationship management (CRM) and operational strategies [10, 11]. Financial institutions worldwide have successfully used advanced algorithms, predictive analytics to enhance customer engagement, improve marketing efficiency, and reduce churn. Despite these global trends, Bunna Bank has not fully utilized the power of data analytics and ML to optimize its customer management and retention efforts. In the current competitive and digitally transforming Ethiopian banking environment, customer churn has become a significant concern, particularly for Bunna Bank. Although exact data on customer attrition for Bunna Bank are not publicly available, increasing competition and evolving customer expectations in the banking sector suggest that retaining customers is a growing challenge. The Bank's Five-Year Strategic Plan (2023/24–2027/28) explicitly emphasizes the need to strengthen customer retention and loyalty through the development of targeted programs supported by an effective CRM system (Bunna Bank, 2024). This underscores the importance of understanding and predicting customer churn to support data-driven strategic decision-making. This strategic initiative demonstrates an institutional recognition of a gap between the bank's current reactive, survey-based retention practices and the planned data-driven, personalized approach. Despite having access to vast amounts of customer data, Bunna Bank still depends largely on traditional tools like customer surveys and manual analysis. These methods do not provide the scalability or accuracy needed to generate clear, actionable insights. Consequently, the bank does not yet have a predictive, explainable churn model to identify and retain customers at high risk of leaving before they actually churn.

These internal capability gaps are further validated at the national level by the National Bank of Ethiopia's *Financial Stability Report* [12], which highlights growing operational and technological risks that threaten customer trust and service quality across the banking sector. The report specifically points out that smaller and mid-sized banks face heightened vulnerabilities due to limited digital infrastructure,

inadequate CRM systems, and underutilized customer data—conditions that closely mirror Bunna Bank's current challenges. Furthermore, the report warns that high deposit concentration among a few clients increases systemic risk and service instability, indirectly contributing to customer dissatisfaction and churn. Therefore, the problems faced by Bunna Bank are not isolated but are aligned with sector-wide weaknesses, further reinforcing the relevance, urgency, and validity of this research. Globally, advanced ML techniques, particularly ensemble-based methods like Gradient Boosting and Extreme Gradient Boosting (XGBoost), have been shown to significantly improve the accuracy of churn prediction models [13, 14]. These techniques have the potential to provide more accurate and reliable churn predictions for Bunna Bank. However, such methods have not been widely applied in the Ethiopian banking sector, and Bunna Bank has yet to explore their potential. Furthermore, the use of explainable artificial intelligence (XAI) to provide transparency and interpretability in predictive models remains largely unexplored, representing a significant gap in the bank's data-driven decision-making capabilities.

Given the challenges of customer churn, Bunna Bank requires a data-driven approach to identify the risk factors contributing to churn. Leveraging advanced ensemble-based ML techniques combined with XAI allows not only the prediction of which customers are at risk of leaving but also provides transparent insights into why these churn predictions occur. Understanding these risk factors enables the bank to make informed, evidence-based decisions about customer management, improving the ability to anticipate churn and respond proactively.

Therefore, this research aims to address gaps in Bunna Bank's customer churn prediction by employing ensemble-based ML techniques, such as Gradient Boosting and XGBoost, to improve prediction accuracy. In addition, XAI methods will be integrated to enhance model transparency and identify the key risk factors contributing to customer churn. By revealing these factors, the study provides a clear understanding of why certain customers are likely to leave, supporting evidence-based decision-making in the bank.

1.4 Research Questions

- What are the most determinant attributes for customer churn?
- Which of the algorithm best suited for predicting customer churn risk?

1.5 Objective of the study

1.5.1 General Objective

The general objective of this research is to design and develop machine learning model for predicting risk factor that contribute for customer churn at Bunna Bank, utilizing Explainable AI techniques to improve model interpretability.

1.5.2 Specific objectives

- Reviewing literature and related work to organize the required data for the study.
- Collect, clean & prepare customer dataset for training
- Determine the key attributes of the dataset that are used for risk factor prediction of customer churn.
- Determine the best prediction model of customer churn used different model evaluation metrics.
- Develop a robust machine learning model and Evaluation.

1.6 Methodology

As stated by Goundar [15], research methodology refers to the systematic processes and techniques employed to collect, analyze, and interpret data on a given topic. It provides a structured approach to problem-solving, ensuring the accuracy and consistency of research outcomes while addressing specific aims, objectives, and questions. In this context, the study utilized a range of methods to develop an effective segmentation and churn prediction model for Bunna customers, aimed at enhancing customer relationship management strategies.

1.6.1 Research Design

According to Kamiri and Mariga [16], in their article *Research Methods in Machine Learning: A Content Analysis*, research design is a structured plan of methods and techniques used to systematically address a research problem. This study will employ an experimental research design, which involves planning and carrying out an experiment to generate results. Typically, the process followed in such experiments includes collecting data, preprocessing it, training the model, testing the model, and finally evaluating its performance. Experimental research controls some variables while testing the effects of others on the subjects involved.

1.6.2 Literature Review

We will intensively review different literature to explore related works on algorithms for predictive analysis, processes, and methods applicable to predictive models.

1.6.3 Data Collection

For this study, historical customer data from Bunna Bank will be collected, including information such as transaction history, age, account details, and customer interactions. During the data understanding phase, this data will be carefully examined to assess its quality and features, which is crucial because the performance of the machine learning model depends heavily on the quality of the data used. Key activities in this phase are reviewing the types and distribution of the data, identifying any missing values or outliers, ensuring data consistency, and exploring the relationships between different variables. This thorough analysis will help to prepare the dataset for building an effective predictive model.

1.6.4 Data Preparation

Data preparation involves cleaning and transforming the data to ensure it is suitable for use with machine learning algorithms, as the quality of this data directly affects the accuracy of the results. Poor preparation can lead to incorrect outcomes and often requires multiple iterations, making it a time-consuming process. Key tasks

during this phase include checking for completeness, correcting errors and outliers, and filling in missing values, creating new features, normalizing and aggregating data, and removing irrelevant or duplicate attributes. Additionally, reducing the dataset size through sampling and selecting relevant features helps improve efficiency. It is also essential to ensure the data is compatible with the tools used. In this case, MS Excel 2013 will be employed to carry out these preparation steps, including data transformation, dimensionality reduction, and feature selection.

1.6.5 Modeling

The modeling phase focuses on predicting the risk factors for customer churn at Bunna Bank. Supervised machine learning techniques will be employed to develop a churn prediction model capable of identifying customers at risk of leaving. To enhance the interpretability of the model's predictions, Explainable AI (XAI) methods will be applied, providing transparent insights into the key factors influencing churn decisions. This approach allows the bank to understand why customers are likely to leave, supporting more informed decision-making.

We will use Python as the primary tool for modeling due to its readability and wide adoption in data science. It offers numerous libraries essential for segmentation and classification tasks, such as Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, and Scipy. Additionally, tools like Streamlit will be used to build interactive prototypes, allowing easy visualization and testing of the models during development.

1.6.6 Evaluation

The algorithms produce models that must be evaluated to identify the most suitable one. This evaluation includes analyzing the results and interpreting outcomes. The performance of the classification algorithms will be assessed through the Confusion Matrix. Metrics such as accuracy, precision, recall, F1 score, and the ROC curve, all derived from the confusion matrix, are used to evaluate the effectiveness of the classification models.

1.6.7 Prototype Implementation

A prototype system has been created to demonstrate the practical application of the model. This prototype accepts new customer data and provides churn predictions along with explanations using XAI methods such as SHAP.

1.7 Scope and limitation of the study

This study focuses on developing a model for predicting risk factor for customer churn in the context of Bunna Bank using historical customer data, which includes demographic details, account activity, transaction history, and digital preferences. The scope covers the application of machine learning algorithms such as Random Forest, XGBoost, LightGBM, Extra trees, Gradient Boosting and MLP (Multilayer Perceptron) for churn prediction. To enhance model transparency and decision-making, Explainable AI (XAI) techniques, specifically SHAP (SHapley Additive exPlanations), will be used to interpret churn prediction results. Limitations of the study include the availability and quality of historical data, possible data imbalance, and the exclusion of real-time data or external factors such as social media behaviors. Additionally, while the model will be evaluated on standard performance metrics, it will not be deployed in a live banking environment, and customer feedback on recommendations is not included in this phase.

1.8 Significance of the study

This study is significant because it provides Bunna Bank with insights into the key factors that influence customer churn. While many existing models focus solely on predicting which customers might leave, they often lack transparency and interpretability. By employing Explainable AI (XAI) techniques alongside advanced machine learning models, this research highlights not only the likelihood of customer churn but also the reasons behind these predictions. These insights enable the bank to better understand customer behavior, support informed decision-making, and enhance the effectiveness of customer management practices. Additionally, this study contributes to the literature by applying modern predictive and interpretability tech-

niques to the Ethiopian banking sector, providing a reference for future research and practice.

1.9 Organization of the study

This research paper is organized as follows. Section 2 presents a review of relevant literature on customer segmentation, churn prediction, basic concepts about machine learning, and explainable AI. Section 3 describes the proposed architecture and methodology of the study that shows the overall process used to carry out the study, which involves several key steps. These include preparing the data, dividing the dataset for training and testing, building predictive models, and evaluating the results. The process also covers identifying churn attributes, developing a working prototype. Additionally, the study makes use of explainable artificial intelligence techniques and emphasizes thorough data preprocessing, such as cleaning, transforming, standardization, and selecting important features. Section 4 provides a detailed summary of the experiments conducted and their outcomes. It covered the processes of selecting important features, building clustering and predictive models, analyzing key drivers, designing the prototype, applying explainable artificial intelligence methods, and discussing the findings. Finally, Section 5 concludes the study by summarizing the key findings and also provides suggestions for areas that future studies can explore.

Chapter 2

Literature Review

This chapter begins with an introduction to the core concepts of churn prediction within the banking sector, emphasizing their strategic importance for financial institutions particularly in the context of Bunna Bank and the broader Ethiopian banking industry. It then explores the application of machine learning techniques in risk factor for customers churn and outlining their potential to enhance decision-making and customer relationship management. Furthermore, the chapter examines the role of Explainable Artificial Intelligence in generating transparent and interpretable insights, which support the development of targeted and effective retention strategies. Lastly, a review of relevant prior studies is presented to identify existing gaps in the literature, thereby justifying the motivation for this research.

2.1 Banking Industry.

The banking industry plays a critical role in the economic development of nations by serving as a financial intermediary between depositors and borrowers. It facilitates capital allocation, supports investment, and promotes financial stability by offering essential services such as savings, credit, payment systems, and wealth management [17]. In emerging economies like Ethiopia, the sector contributes significantly to financial inclusion, enabling both individuals and businesses to access formal financial services.

Commercial banks, in particular, serve as the backbone of modern financial systems. They not only mobilize resources and allocate capital but also act as instruments for implementing monetary policy [18]. In recent years, the Ethiopian banking sector

has experienced rapid growth marked by increased competition, digitization, and regulatory transformation. According to the National Bank of Ethiopia [12], the total number of commercial banks and the volume of banked customers have increased substantially, creating the need for more data-driven decision-making and customer-centric strategies. According to [19], customer behavior in banking is evolving due to increasing expectations for personalized and seamless digital experiences. This shift compels banks to go beyond traditional demographic-based marketing and adopt advanced analytics for customer segmentation, retention, and service delivery. In this context, technologies such as machine learning and explainable artificial intelligence (XAI) are becoming crucial tools for analyzing customer data, predicting attrition, and designing targeted interventions.

Bunna Bank, as one of the private banks in Ethiopia, having more than 5 million customers and operates in this dynamic environment which established to serve a wide spectrum of customers, it has expanded rapidly and adopted various digital banking platforms. With the growing volume of transactional and customer data, there is an opportunity to use machine learning and XAI techniques to improve customer relationship management, reduce churn, and enhance service efficiency.

2.2 Customer Relationship Management

2.2.1 Overview of Customer Relationship Management

Customer Relationship Management uses AI and data-driven insights to enhance customer interactions, streamline operations, and enable personalized experiences, particularly in the financial services sector [20]. Different authors provided several definitions of CRM defines CRM as CRM is a strategy used by businesses to understand customer needs, improve satisfaction, and retain customers through efficient data management and interaction tracking [21, 22].

On the other hand, [23] define CRM as “a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and the supply chain functions of the organization to achieve

greater efficiencies and effectiveness in delivering customer value". According to [24], CRM is strategies that companies adopt to maintain healthy relationship with their customers, to attract them, to retain them, etc. By utilizing customer data and technology tools, companies can better understand their customers' needs and preference [25].

CRM provides businesses with several advantages, including enhanced customer engagement, streamlined business operations, data-driven decision-making, and increased profitability [26]. These benefits emphasize the significance of CRM in helping businesses better understand their customers, personalize interactions, and optimize marketing strategies, ultimately leading to improved business performance and customer satisfaction [27].

2.2.2 Customer Churn

Churn Customer refers to the number of existing customers who may leave the service provider over a given period. These customers can be called as churners. The main aim of churn is to predict the churn at the earliest, to identify the reason for churning. The primary goal of churn analysis is to identify and anticipate churn consumers as soon as possible. This will help us to rectify the issues of the customer. This will be helpful to satisfy the customer needs and will continue to use that service. This will help to meet the needs of the customers, and they will continue to utilize the service [28].

Churn prediction refers to the process of identifying customers who are likely to discontinue their relationship with a business. In banking institutions, customer churn poses a significant challenge as it directly affects revenue, profitability, and long-term business sustainability. Retaining existing customers is often more cost-effective than acquiring new ones, making churn prediction a vital aspect of customer relationship management [29, 30].

By implementing churn prediction models, banks can proactively identify customers at risk of leaving and implement targeted retention strategies. This predictive approach helps in minimizing revenue loss, maintaining customer trust, and enhancing

customer satisfaction [30]. Furthermore, accurate churn prediction allows financial institutions to allocate resources efficiently, ensuring that high-risk customers receive timely interventions such as personalized offers, improved services, or better engagement strategies [29]. In a competitive banking landscape, where customers have multiple choices, the ability to anticipate and mitigate churn can provide a strategic advantage [31].

2.2.3 Machine Learning Algorithms

Machine Learning ML has emerged as a powerful tool for extracting insights from data and is increasingly being applied in various fields such as finance, healthcare, and communication. As stated by [30], ML is a subset of AI that allows computers to learn from data and improve their performance without being explicitly programmed. ML is categorized into supervised, unsupervised, semi-supervised, and reinforcement learning, with each type serving different purposes depending on the data and the problem at hand. Supervised learning, for instance, uses labeled data to train models, whereas unsupervised learning works with unlabeled data to uncover hidden patterns [32].

In addition, machine learning algorithms play a crucial role in automating complex tasks such as pattern recognition and predictive analytics. These algorithms can identify patterns and relationships in data that are often too complex for traditional methods. According to [33], ML is transforming various technological domains, including antenna design, where it is used to optimize design processes and improve performance compared to conventional methods. By mimicking the human learning process, ML algorithms adapt and improve as more data becomes available, leading to more accurate predictions and better decision making. The rapid advancement of machine learning techniques has been driven by the increasing availability of large datasets and powerful computing resources. This has enabled ML algorithms to tackle problems in real-time applications, from speech recognition to autonomous systems, driving the adoption of AI across multiple industries [32, 33].

Types of Machine Learning Algorithms

Machine Learning Algorithms are divided into categories based on their purpose. According to [34, 35], the main categories are: supervised, unsupervised, semi-supervised, and reinforcement learning. The four types of machine learning algorithms are depicted in Figure 1.

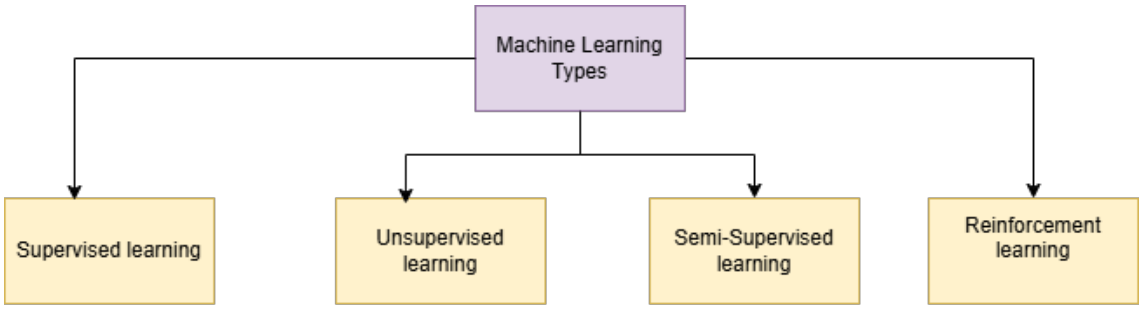


Figure 2.1: Types of machine learning algorithms.

Supervised Learning Supervised learning is a commonly applied method in machine learning. According to [32], supervised algorithms rely on labeled data, which is typically split into two parts: a training set and a testing set. The training set includes known output values that the model aims to predict. The goal is for the algorithm to recognize patterns in the training data and use this understanding to forecast outcomes for the testing data. Popular examples of supervised learning algorithms include Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbors (KNN) [32].

Unsupervised Learning As highlighted by [36], unsupervised learning does not depend on labeled data. Instead, algorithms are developed using input data and are subsequently evaluated with a dataset. In this process, the training dataset is utilized to build and train the model, while the testing dataset helps in making accurate predictions. The model learns from past data and previously introduced features, enabling it to make real-valued predictions. Common examples of unsupervised learning algorithms include the K-means algorithm, the Apriori algorithm, and the SOM algorithm.

Semi-supervised Learning According to [36], semi-supervised learning is highly applicable in real-world situations because it is more efficient than many other machine learning algorithms. It utilizes a combination of labeled and unlabeled data to create an effective classifier. The core idea is to integrate data with known outcomes (labeled) with data lacking predefined results (unlabeled) to uncover relationships between attributes and develop a classification model for prediction. This approach is useful for both classification and regression tasks [37].

Reinforcement Learning Reinforcement learning (RL) is an advanced machine learning algorithm that allows software agents to make decisions and take actions autonomously within a given environment [37]. It focuses on optimizing the agent's behavior by rewarding desired actions and penalizing undesired ones. According to [35], the goal of RL is to learn from environmental interactions and take actions that maximize cumulative rewards or minimize penalties, ensuring better efficiency over time.

As [32] further explains, RL is particularly beneficial in complex and dynamic environments where the agent needs to adapt and learn from its actions. Unlike simpler machine learning techniques, RL thrives in contexts like robotics, autonomous driving, and complex systems optimization, where explicit programming is not feasible. The feedback from the environment enables the agent to fine-tune its decisions and continually improve its strategies, thereby contributing to greater automation and operational efficiency.

In practice, reinforcement learning is employed in a variety of fields where real time decision making is critical, such as supply chain management, robotics, and autonomous systems. However, as emphasized by both [32,35], RL is best suited for complex problems that involve long-term learning and decision-making rather than simple or static tasks.

2.2.4 Machine Learning Applications in Customer Relationship Management

AI driven CRM systems use predictive analytics to forecast customer behaviors and preferences. By analyzing historical customer interactions, AI can predict future actions, such as purchasing habits or likelihood of churn, allowing businesses to proactively engage with customers in a personalized manner [38]. ML algorithms further refine this by automatically segmenting customers based on behavior, interests, and demographics, improving the precision of marketing efforts and customer outreach [20].

This integration of AI and ML into CRM not only enhances the accuracy but also automates routine tasks such as lead scoring and follow-up activities, saving time and resources for sales and marketing teams [39]. Furthermore, as AI continues to evolve, the segmentation process becomes more dynamic, with systems adapting to changing customer behaviors and preferences in real-time [20].

In the financial services sector, for example, AI-enhanced CRM systems allow organizations to predict customer needs, segment audiences, and create tailored solutions, fostering stronger relationships with clients [20]. As AI technology becomes more advanced, the integration of AI-driven segmentation will be an essential tool for companies looking to stay competitive by providing personalized experiences and improving customer loyalty [38]. By using these technologies, businesses can not only improve customer satisfaction but also drive sustainable growth and enhance their overall CRM strategies [38].

2.3 Machine Learning Techniques for Customer Churn Prediction

Customer Churn Prediction is a critical task that involves identifying customers who are likely to discontinue using a service. It is essential for businesses, particularly in competitive sectors such as telecommunications, banking, and e-commerce, to anticipate churn in order to take proactive measures to retain customers [8]. Machine learning techniques have become an invaluable tool in predicting churn, as they can analyze large volumes of historical customer data to identify patterns associated with

both customer retention and attrition [40,41]. These models help businesses optimize their retention strategies by providing data-driven insights into the likelihood of churn.

The type of classification algorithm selection depends on the characteristics of the dataset. As shown in [**?**], the classification algorithms include Logistic Regression, Decision Tree, Naive Bayes, K-Nearest Neighbors, and Support Vector Machines. Several ensemble-based models can boost performance by using various base models in combination. Random Forest (RF), Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), LightGBM, and Extra trees are among the techniques employed in this strategy [42, 43]. Such combination methods collect multiple model predictions to gain enhanced accuracy while minimizing overfitting through excellent performance with complicated classification problems [42, 43].

2.3.1 Random Forest

Random Forest is a supervised machine learning algorithm based on the decision tree algorithm. It uses ensemble learning, which combines multiple classifiers to solve complex problems [44]. Predictions are made by aggregating individual decision trees [45], and the forest is typically trained using the bagging technique.

Algorithm 1: Random Forest Classifier Pseudocode [46]

Input:

- A training set S with OOO samples (each sample has FFF features)
- The number NNN of trees in the ensemble

Output:

- A Random Forest model $RF(x)$ composed of NNN decision trees

1. For $i = 1$ to NNN do
2. Draw a bootstrap sample S_i from the training set S (sampling with replacement)

3. Function `BuildTree`(S_i, FFF):
4. At each node, randomly select a subset of features $F' \subseteq F$
5. Using only F', determine the best split criterion
6. Split the node into child nodes
7. Repeat recursively until stopping conditions are met
8. Return the constructed decision tree t_i
9. End For
10. Define the ensemble as $RF = \{t_1, t_2, \ldots, t_{NNN}\}$
11. For prediction with $RF(x)$:
12. If classification: output the majority vote of all $t_i(x)$
13. If regression: output the average of all $t_i(x)$
14. Return $RF(x)$

2.3.2 Gradient Boosting (GB)

Gradient Boosting functions as a machine learning technique through a sequential learning approach that adds decision trees step by step to refine the existing predictive model [47, 48]. Construction of this model happens through minimizing a loss function like mean squared error by performing gradient descent in function areas instead of parameter areas. A predictive model built by Gradient Boosting proceeds through an iterative process where weak learners train successively to improve ensemble performance based on previous model errors and residuals. The procedure is as follows:

Algorithm 2: Gradient Boosting Pseudocode [49]

Input:

- A training set with N samples and F features

- The number M of trees
- Learning rate ν
- Loss function L

Output:

- Final model $F(x)$

1. Initialize model $F_0(x)$ with a constant value that minimizes the loss function L.
2. For $m = 1$ to M do
3. For each sample $i = 1$ to N do
4. Compute residual $r_i \leftarrow -\nabla L(y_i, F_{m-1}(x_i))$
5. End For
6. Fit a regression tree $h_m(x)$ to the residuals $\{r_i\}_{i=1}^{N}$
7. For each leaf j of $h_m(x)$ do
8. Compute $\gamma_j \leftarrow \arg\min_\gamma \sum_{x_i \in R_j} L(y_i, F_{m-1}(x_i) + \gamma)$
9. End For
10. Update the model:

$$F_m(x) \leftarrow F_{m-1}(x) + \nu \cdot \sum_j \gamma_j \mathbf{1}_{\{x \in R_j\}}$$

11. End For
12. Return $F_M(x)$

Recent enhancements to Gradient Boosting include regularization techniques like data subsampling and feature randomization, which improve the generalization of the model and help prevent overfitting. Gradient Boosting is particularly effective

when the model needs to handle complex relationships between the features and the target, and it generally provides high accuracy in predictive modeling tasks. However, it is computationally intensive and may require careful tuning of hyper parameters such as the learning rate, the number of trees, and the tree depth to avoid overfitting [39].

2.3.3 Extreme Gradient Boosting (XGBoost)

XGBoost represents an enhanced version of Gradient Boosting Machines that operates with supervised learning task requirements. Through parallel tree boosting the developed XGBoost mechanism enables simultaneous decision tree construction which leads to faster and more efficient processing for numerous applications. XGBoost belongs to the ensemble tree family to provide a flexible machine learning approach which resolves both regression problems and classification problems including stroke classification. The system processes structured data with success to make it an ideal multiple learning framework selection. The incorporation of XGBoost into ensemble provides the advantage of processing structured data alongside improved predictive accuracy [44].

$$\text{Obj}(\Theta) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$$

where l is the loss function, Ω is the regularization term, and f_k are the individual trees [50].

2.3.4 LightGBM

LightGBM is a gradient boosting framework developed by Microsoft. It is an open-source, distributed machine learning library designed to be fast, scalable, and efficient. LightGBM improves the efficiency and scalability of the algorithm without sacrificing its inherited effective performance [51]. One of its significant advantages is the use of a histogram-based algorithm for constructing decision trees, allowing it to handle large amounts of data efficiently. Additionally, LightGBM features Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB)

to improve computational efficiency and reduce dimensionality, especially when dealing with high-dimensional and sparse data[67]. It is particularly effective for large-scale applications and has been successfully implemented in fields such as finance, healthcare, marketing, and anomaly detection. Its ability to manage imbalanced datasets by adjusting class weights makes it suitable for tasks like fraud detection and medical diagnosis, where data imbalance is a common issue [52].

Algorithm 3: LightGBM Pseudocode [53]

Input:

- Training features X
- Target values y
- Number of boosting iterations M
- Learning rate α
- Maximum tree depth

Output:

- Final LightGBM model $F(x)$

1. Initialize the model with a constant prediction:

$$f_0(x) = \text{mean}(y)$$

2. For $m = 1$ to M do

3. Compute residuals based on current predictions

4. Fit a regression tree to residuals (with maximum depth = max_depth)

5. Update the model:

$$f_m(x) = f_{m-1}(x) + \alpha \cdot \text{tree}(x)$$

6. End For

7. Define final model:
$$F(x) = \sum_{m=1}^{M} f_m(x)$$

8. Return $F(x)$

2.3.5 Extra Trees

The Extra Trees method, as outlined by Hussen et al. [54], differs from Random Forest by using the entire feature set for each tree and selecting split points randomly rather than optimizing them. This approach, while increasing model variance, lowers bias, making it more robust to noisy data and less prone to overfitting. Key parameters K (attribute selection strength), nmin (noise averaging intensity), and M (variance reduction through aggregation) are crucial for tuning the model. Despite their higher variance, Extra Trees are computationally efficient and capable of capturing complex data patterns, often requiring fewer trees to achieve accurate results.

Algorithm 4: Extra-Trees Classifier Pseudocode [55]

Input:

- A training set S with continuous data and n samples (each sample has f attributes/features)
- The number k of features to consider in each node
- The number m of trees in the ensemble
- The minimum number $n_{\min}$ of values to split on

Output:

- An ensemble of trees $E = \{t_1, t_2, \ldots, t_m\}$

1. Function `BuildExtraTree(`S`, `k`, `m`, `$n_{\min}$`)`
2. If $|S| \leq n_{\min}$ or all classes/attribute values in S are constant, then
3. return class frequencies

4. Else

5. Randomly select k (non-constant valued) attributes $\{a_1, \ldots, a_k\}$ without replacement, among all candidate attributes in S

6. Generate k splits $\{s_1, \ldots, s_k\}$, where each split s_i is a value in the range of attribute k_i in S

7. Select s_0 such that $\text{Score}(s_0, S) = \max_{i \in \{1,\ldots,k\}} \text{Score}(s_i, S)$

8. From s_0, split S into two sets, S_ℓ and S_r

9. Let $t_\ell = \texttt{BuildExtraTree}(S_\ell, k, m, n_{\min})$ and $t_r = \texttt{BuildExtraTree}(S_r, k, m, n_{\min})$

10. Create a node N with the split s_0, attach t_ℓ and t_r as left and right subtrees

11. End

12. Return N

2.3.6 Multilayer Perceptron (MLP)

The multilayer perceptron is a widely used neural network model composed of an input layer, one or more hidden layers, and an output layer, where each layer contains neurons that process inputs through weighted connections. MLPs use activation functions to introduce non-linearity, enabling the network to model complex patterns. Typically, each neuron applies a bias and an activation function to the weighted sum of its inputs. While increasing the number of hidden neurons and layers can enhance model capacity, it also leads to longer computational times. MLPs have been applied in various cloud-based applications, including forecasting daily food prices and improving cancer diagnosis accuracy through transfer learning [56].

Algorithm 5: Multilayer Perceptron (MLP) Classifier Using Backpropagation [57]

Input:

- Training dataset with input vectors $x(n)$ and desired outputs $d(n)$

- Learning rate η and momentum α
- Number of epochs or stopping criterion

Output:

- Trained MLP model with updated synaptic weights

1. Initialize:
    - Randomly set synaptic weights and thresholds from a uniform distribution
    - Choose sigmoid as the activation function
2. Repeat for each epoch until stopping criterion is met:
    (a) For each training example $(x(n), d(n))$:
        i. Forward Computation:
            - For each layer l from input to output:
                – Compute local field for each neuron j in layer l
                – Calculate output signal using sigmoid activation function
                – If $l = 1$ (first hidden layer), use input vector $x(n)$
                – If $l = L$ (output layer), compute error signal: $error_j = d_j - output_j$
        ii. Backward Computation:
            - Calculate local gradients for each neuron using the derivative of the sigmoid function
            - Update synaptic weights w according to:

$$w_{\text{new}} = w_{\text{old}} + \eta \cdot (\text{local_gradient}) \cdot (\text{input_to_neuron}) + \alpha \cdot (\text{previous_weight_change})$$

3. End Repeat

2.4 Explainable Artificial Intelligence (XAI) in Predictive Modeling

2.4.1 Concept and Importance of XAI

Artificial Intelligence (AI) has become an integral part of modern society, assisting various industries such as healthcare, finance, law enforcement, and autonomous systems. However, many AI models make it difficult to interpret their decision-making processes [58]. This lack of transparency raises concerns regarding trust, accountability, and ethical use, particularly in critical applications [58]. As a response to these concerns, Explainable Artificial Intelligence (XAI) has emerged as a field dedicated to developing methods that provide transparency and interpretability in AI models [59].

According to Ridley [60], Explainable Artificial Intelligence (XAI) is a subfield of artificial intelligence (AI) that provides explanations for the predictions, recommendations, and decisions of intelligent systems. XAI is a set of techniques, processes, and strategies that libraries should adopt and advocate for to ensure that machine learning appropriately serves librarianship, the academy, and the public interest. XAI is needed to:

- Generate trust, transparency, and understanding;
- Ensure compliance with regulations and legislation;
- Validate models and validate explanations generated by XAI.

XAI consists of testable and unambiguous proofs, various verification and validation methods that assess influence and veracity, and authorizations that define requirements or mandate auditing within a public policy framework. Ridley also notes that taxonomies of XAI types are classified based on their scope and mechanism, such as local explanations, global explanations, black-box or model-agnostic explanations, and white-box or model-specific explanations.

Oladele [61] defines XAI in the context of financial market prediction, stating that XAI techniques aim to bridge the gap between model performance and interpretability. The paper reviews various XAI methods, such as SHAP (SHapley Additive

exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and attention mechanisms, assessing their effectiveness in elucidating the decision-making processes of complex models. These techniques provide insights into feature importance, model behavior, and the relationships between input variables and predictions. Incorporating explainability into market prediction models enhances stakeholder confidence and facilitates regulatory compliance.

The ability to understand and explain how a machine learning model arrives at its predictions or decisions is crucial for building trust and ensuring that stakeholders can make informed choices based on model outputs [59]. As machine learning techniques become increasingly complex, especially with the rise of deep learning architectures, there is a growing need for transparent and explainable models to facilitate decision-making processes and comply with ethical standards [62]. Without clear interpretability, models risk being perceived as "black boxes," which undermines user confidence and can hinder their adoption in critical applications.

According to Salih et al. [59], achieving model interpretability typically involves employing explainable artificial intelligence techniques that convert complex model outputs into human-understandable insights. Among the most widely used XAI methods are SHAP and LIME (Local Interpretable Model-Agnostic Explanations), both of which offer ways to explain predictive models by breaking down their decision-making processes. While SHAP uses game theory to assign contributions to each feature, LIME approximates models locally to explain individual predictions.

2.4.2 The Need for Explainability

The complexity of AI models often increases their predictive power but decreases their interpretability [63]. Traditional machine learning models such as decision trees and logistic regression are relatively easy to understand due to their transparent structure. However, deep learning models consist of numerous layers and parameters, making it challenging to trace how specific inputs lead to particular outputs. The trade-off between model performance and interpretability is a central issue in AI research.

2.4.3 XAI Methods and Categorization

To address the interpretability challenge, XAI methods can be categorized into four main axes [58]:

1. **Data Explainability** – Focuses on understanding data characteristics and feature importance before training a model.

2. **Model Explainability** – Aims to design inherently interpretable models, such as decision trees and Bayesian networks.

3. **Post-hoc Explainability** – Involves explaining complex models after training using techniques like SHAP, LIME, and Grad-CAM.

4. **Assessment of Explanations** – Evaluates the effectiveness and reliability of explanations through metrics such as fidelity, stability, and consistency.

2.4.4 Evaluation Metrics and Tools

Various metrics have been proposed to assess the quality of explanations, including faithfulness, comprehensibility, and robustness [63]. Additionally, several open-source XAI frameworks, such as SHAP and LIME, provide tools for implementing explainability methods in machine learning applications.

Fidelity According to [64], fidelity measures how well an explanation reflects the actual behavior of the machine learning model it is trying to explain. High fidelity means the explanation closely follows the model's logic, while low fidelity suggests the explanation might be misleading or inaccurate.

Stability Stability in Explainable AI means how consistent the explanations are when the input is slightly changed [65]. If two inputs are very similar and the model gives similar results, then the explanations should also be similar, indicating that the explanation method is reliable. To check stability, small changes are added to the input, and we see if the explanation stays almost the same [66]. A common way to measure this is using a formula like $S = \lambda \times (1 - \text{similarity})$, where higher

similarity means better stability. Stability helps to ensure that the explanations are not random but make sense even when the input changes slightly.

2.4.5 Types of Explanation

Understanding the different types of explanations in Explainable AI (XAI) helps us see how AI models make decisions [67]. According to [67], the two main types are global explanations and local explanations.

Global Explanations Global explanations aim to provide a broad understanding of how an AI model behaves across all possible inputs [68]. They focus on identifying general patterns and determining which features most significantly influence the model's predictions overall. These explanations are especially useful in domains like finance and healthcare, where understanding systemic model behavior and detecting potential biases is critical.

Local Explanations Local explanations focus on clarifying the reasoning behind a specific prediction made by the model [68]. They help users interpret individual decisions by showing how input features contribute to a particular outcome. Tools such as LIME and SHAP are widely used for this purpose, offering context-specific insights that can enhance trust and accountability in sensitive areas like medical diagnosis or criminal justice.

2.4.6 Explainable AI Methods

There are two popular methods used in Explainable Artificial Intelligence (XAI) to explain model predictions: SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) [67]. Both methods help us understand why a model made a certain decision, especially for complex models.
LIME is model-agnostic, which means it can work with any type of model without needing to change the model itself [69]. It works by changing the input data slightly and seeing how these changes affect the prediction, then builds a simple linear model around the specific example to explain the decision.

SHAP uses game theory to fairly assign the contribution of each feature in the prediction [67, 69]. It calculates Shapley values, which measure how much each feature affects the result by looking at all possible feature combinations. SHAP can explain model decisions both for single cases (local) and for the whole model (global). Although SHAP gives more stable and trustworthy explanations, it requires more computing power, especially when the model has many features. Both SHAP and LIME are widely used in important areas like healthcare and finance, helping users trust AI by making predictions clearer.

2.5 Related Works

This section presents a review of relevant studies carried out by researchers that investigate the use of machine learning techniques for customer segmentation and churn prediction. It also includes an analysis of the existing research gap, which forms the basis for the current study's motivation.

Muhamed Hassen Seid and Michael Melese [2] conducted a study aimed at predicting customer churn at the Commercial Bank of Ethiopia (CBE) using various machine learning algorithms, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Deep Neural Networks (DNN). The study utilized a dataset consisting of 204,161 instances with 11 attributes, incorporating features such as ATM usage and mobile banking utilization. A significant contribution of the study was the focus on data preparation, including handling class imbalance using the SMOTE technique. Among the methods tested, the DNN model exhibited the best performance, achieving an accuracy of 79.32%, a precision of 85.08%, and a recall of 78.19%. While the study addressed class imbalance and performed feature selection using feature importance and correlation matrix, it primarily focused on the predictive accuracy of individual algorithms without delving into model interpretability or combining segmentation with churn prediction. Furthermore, although the study contributed to the customer churn prediction domain, it lacked the integration of advanced techniques such as ensemble methods and Explainable Artificial Intelligence, which is crucial for practical

business applications. Additionally, the study lacked an ensemble learning approach that could leverage the strengths of multiple algorithms to enhance predictive performance. The authors also acknowledged the need to incorporate additional attributes and explore diverse machine learning techniques in future work.

Edemealem and Tulu [70] developed a customer churn prediction model for Lion Insurance, leveraging machine learning techniques to maintain competitiveness by identifying potential churners. The study utilized a dataset of 12,007 instances with nine features, clustered using the K-means++ algorithm to label data. Due to the imbalance in clusters, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Among the five machine learning algorithms tested, including Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Deep Neural Networks (DNN), the DNN model demonstrated superior performance with an accuracy of 98.81%, utilizing a structure of (9-55-55-55-55-55-1). The study highlighted the need for further exploration of other branches of Lion Insurance and additional insurance companies to generalize the model. Moreover, the researchers recommended developing a web-based application to predict churn using customer data, as well as exploring other deep learning algorithms and clustering techniques for enhanced accuracy. The study acknowledged limitations related to data collection from a single branch and suggested incorporating data from multiple branches and insurance companies in future work.

Sharmila et al. [28] designed a customer churn prediction model for the telecom sector using machine learning techniques to reduce customer loss and enhance retention strategies. The study employed classification algorithms, including Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree Classifier, using large-scale telecom data. The main finding is that the Random Forest classifier achieved the highest accuracy of 99.09%, with a precision and recall of 99%, indicating high predictive performance. However, the study does not address class imbalance, data scaling, or preprocessing, which may lead to overfitting and limit the model's generalizability. The absence of techniques like SMOTE and cross-validation suggests that the reported accuracy might not reflect real-world performance.

Liu et al. [71] proposed a hybrid customer churn prediction model for the telecom sector that integrates clustering algorithms (k-means, k-medoids) with classification methods (Gradient Boosted Trees, Decision Trees, Random Forest, Deep Learning, Naïve Bayes). They used ensemble learning techniques, such as stacking and bagging, to improve prediction accuracy, achieving up to 96% accuracy on the Orange dataset and 93.6% on Cell2Cell. Their results demonstrate that combining clustering and classification outperforms single methods. However, the study focuses primarily on predictive performance without incorporating explainability or actionable retention strategies, which limits its direct business applicability. This contrasts with my work, which combines customer segmentation and churn prediction with explainable AI to not only predict churn but also provide understandable insights and data-driven retention recommendations in the banking context.

Seyoum [72] conducted a study on customer churn prediction for the Commercial Bank of Ethiopia (CBE) using machine learning algorithms. The study applied K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes, and Logistic Regression on a dataset of 48,051 customer records. After thorough data preprocessing and model training using R Studio, the KNN classifier achieved the highest accuracy of 99.91%, followed by Logistic Regression at 93.8%, SVM at 92.4%, and Naïve Bayes at 83.8%. Key predictors identified included total number of transactions, POS usage, ATM usage, internet banking, mobile banking, and industry type. The study highlights the potential of integrating predictive models with Customer Relationship Management (CRM) systems to proactively manage churn. The authors also suggest that expanding datasets across multiple banks with more diverse attributes could further enhance model performance.

Another research studied customer churn in the telecom sector using machine learning to support customer retention [73]. The study pointed out that churn is a serious problem caused by factors such as better offers from competitors and poor service quality, which can reduce a company's income and customer lifetime value. To address this, the researchers developed a system that uses machine learning models specifically Logistic Regression, Support Vector Machine (SVM), Random Forest,

and XGBoost to identify customers who are likely to leave. The process involved collecting telecom data, cleaning and preparing it, selecting the most relevant features, and applying the models to predict churn. Among the algorithms tested, XGBoost performed the best due to its high accuracy and speed. The system not only predicted potential churners but also provided reasons for their possible departure, offering actionable insights to help telecom companies improve service delivery and customer satisfaction. The study concluded that machine learning can play a significant role in minimizing customer churn, and suggested that future work could extend the approach to other industries or involve advanced techniques such as deep learning to further enhance customer retention strategies.

Sagala and Permai [74] investigated the effectiveness of boosted tree algorithms for customer churn prediction in the banking sector. The study evaluated three boosting-based models: Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Category Boosting (CatBoost) on a single bank dataset. Following exploratory data analysis and feature engineering including quantile-based discretization, one-hot encoding, and Robust Scaling the dataset was balanced using a hybrid SMOTE-Tomek method. Models were trained on 80% of the data and optimized using Grid Search with 10-fold cross-validation. LightGBM outperformed other algorithms, achieving an accuracy of 91.4%, precision of 94.8%, and recall of 87.7%. The authors suggest their approach provides valuable insights for banks to manage customer churn effectively but note the limitation of using a single dataset and recommend future comparisons with other predictive methods.

Adiputra and colleagues [75] developed a customer churn prediction framework using a weighted average ensemble of XGBoost and Random Forest models. The study applied multiple preprocessing techniques and evaluated performance on public datasets from two sectors: telecom and insurance. The model achieved an F1-score of 0.850 on the telecom dataset and 0.947 on the insurance dataset, with a processing time of 28 seconds. Compared to recent studies, their ensemble model showed improved F1-score and efficiency on the telecom dataset but slower execution on the insurance dataset. The authors note limitations related to time consumption

on large datasets and absence of customer segmentation for strategic use. They recommend future work to improve model efficiency and effectiveness on large data and suggest integrating unstructured data, such as customer feedback from social media, to enhance predictive power and better reflect customer satisfaction.

Peng and Peng [50] proposed a churn prediction model for telecom customers using a genetic algorithm to optimize XGBoost, combined with SHAP for explaining the model's decisions. They addressed the common problems of imbalanced data by using the ADASYN method to create more samples of the minority class. Their model performed well, with an AUC score of 0.9561, better than several other machine learning methods. SHAP helped to show which features influenced customer churn and how, making the results easier for managers to understand. However, the study used a small dataset from Kaggle, which may limit how well the results apply to other cases. The paper also did not explain much about how they chose or prepared the features, and it only used SHAP for explanation without exploring other methods. Still, their combination of model tuning and explanation is useful for improving churn prediction and customer retention efforts.

2.5.1 Research Gap

Several studies have applied machine learning techniques to predict customer churn across various sectors, including banking, telecom, and insurance. These studies primarily focus on developing predictive models that achieve high accuracy, often by leveraging algorithms such as Logistic Regression, Random Forest, Gradient Boosting, and Deep Neural Networks. While these models demonstrate strong predictive performance, a critical limitation is their lack of interpretability; most models operate as "black boxes," making it difficult for business stakeholders to understand the underlying factors that drive customer churn. Consequently, organizations may be able to predict churn, but they struggle to derive actionable insights to prevent it effectively.

Some recent studies have attempted to address this issue by analyzing feature importance or applying Explainable AI techniques, such as SHAP values, to provide local

or global explanations of model outputs. However, these efforts are often limited in scope and are not integrated into a systematic framework that combines predictive accuracy with interpretability. Many approaches either focus on a single dataset or sector, neglect ensemble learning strategies, or provide explanations at a level that is too technical for practical decision-making. Moreover, while domains such as telecom and insurance have received considerable attention, the banking sector remains underexplored, particularly in terms of understanding the key risk factors influencing customer churn through interpretable, data-driven models.

Addressing these limitations, this study aims to develop a comprehensive framework that integrates ensemble-based machine learning techniques with Explainable AI (XAI). The proposed framework is designed not only to achieve high predictive accuracy in identifying potential churners but also to provide transparent and actionable insights into the primary risk factors driving churn. By combining advanced machine learning algorithms with explainability methods, this study seeks to bridge the gap between predictive modeling and practical decision-making, enabling banking professionals to make informed, data-driven strategies to enhance customer retention.

Chapter 3

Methodology

3.1 Overview

This chapter outlines the research methodology adopted for designing and developing a machine learning model to predict risk factors for customer churn using explainable artificial intelligence (XAI). It describes the step-by-step approach used in the study, including data collection, preprocessing, feature selection, and the development of churn prediction models. The chapter also explains the application of XAI methods, such as SHAP and LIME, to interpret model predictions and provide insights into the key factors driving customer churn. Furthermore, it details the experimental setup, including procedures for training, validating, and testing the models, as well as the tools and algorithms employed in each phase.

3.2 The Proposed Architecture and Method

In this study, we develop a prototype for churn prediction as shown in Figure 3.1 below. The architecture includes key components that are integrated such as data preparation, data preprocessing, dataset split, model development, user input, model prediction, use of XAI. By following this architecture, we can effectively prepare the data, develop a reliable predictive model, evaluate its performance, and seamlessly integrate it with the prototype for churn prediction.

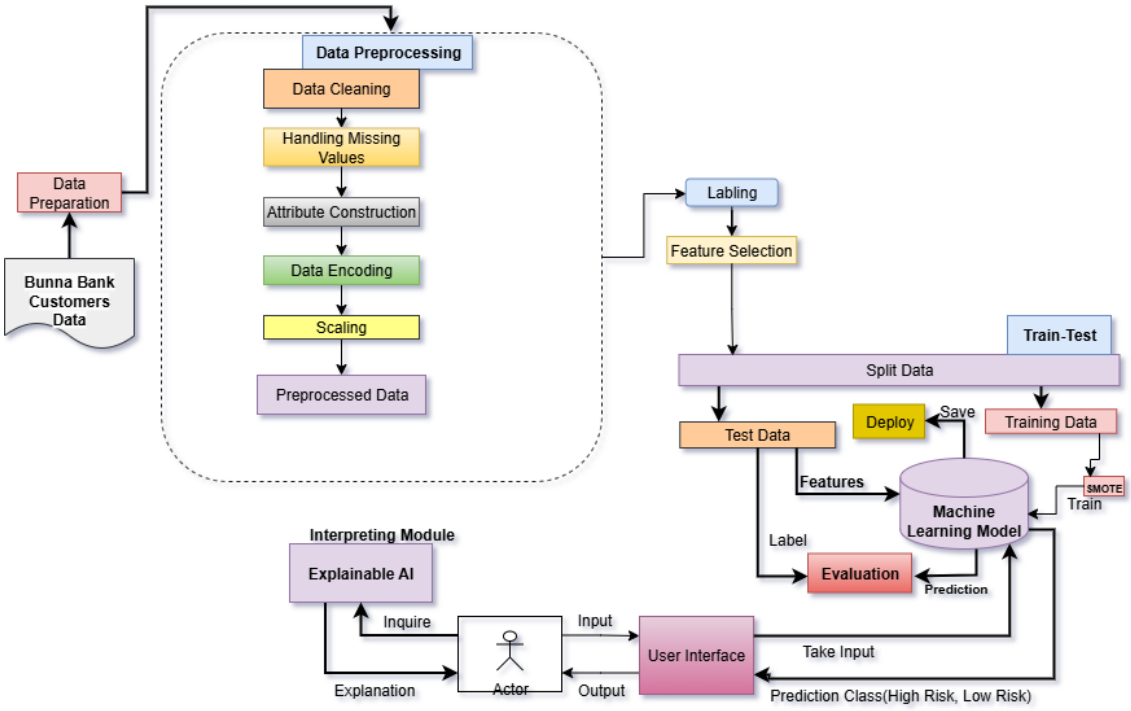


Figure 3.1: Proposed System Architecture.

The above Figure 3.1 shows that the general approach of the risk factor of customer churn prediction model. This model has six major components. The first part involves input the collected data and then preparing the data using different preprocessing techniques, which consists of feature selection, data transformation, and data balancing. The third step involves splitting our dataset into training-validating and testing sets. The fourth section involves model building by training and validating the machine learning models (RF, GBoost, XGBoost, MLP,LightGBM , and Extra tress). Finally, models explain ability is applied to the selected model.

3.3 Data Collection and Understanding

In this study, we used customer historical data extracted from Bunna Bank core banking system database, which serves as the main data source and it was accessed and retrieved specifically for this research. The data includes customer demographic details, account status, transaction history, and service usage patterns. These records were systematically stored and maintained within the bank's database infrastructure. To obtain the dataset, structured SQL queries were written to extract relevant fields from the core banking system. The data reflects actual cus-

tomer interactions with the bank and was used without any manual input from customers during this research. All attributes were extracted from the target tables of the Oracle in CSV format. Accordingly, more than 308,293 rows of behavioral data records and 3 months of transactional data of the corresponding customers were extracted in CSV format. The initial extracted data contains 19 attributes. These attributes include: CIF, Age, Gender, Income, Marital_Status, Occupation, Location, Account_Type, Current_BAL, TXN_Count, Total_Monthly_Deposit, Total_Monthly_Withdrawal, Tenure, ATM_CARD_FLAG, MOBILE_BANKING_FLAG, USSD, Repayment_History, Debit_Transactions_Last_3_Months, Credit_Transactions_Last_3_Month and Last_TXN_DATE. The researcher converted these data into Excel format to perform data pre-processing. Figure below shows some part of the initial data collected.

No	Features	Description	Data Type
1	CIF	Customer ID	Numeric
2	Age	Age of customers	Numeric
3	Gender	Gender of Customers	Nominal
4	Income	Monthly or yearly income.	Numeric
5	Marital_Status	Customer's marital status	Nominal
6	Occupation	Job type	Nominal
7	Location	Customer's account opening branch	Nominal
8	Account_Type	Type of bank account	Nominal
9	Current_BAL	Current account balance	Numeric
10	TXN_Count	Total Number of transactions.	Numeric
12	Total_Monthly_Deposit	Total money deposited monthly	Numeric
13	Total_Monthly_Withdrawal	Total money withdrawn monthly.	Numeric
14	Tenure	How long the account has been active	Date Time
15	ATM_CARD_FLAG	Whether customer has an ATM card (Yes/No).	Nominal
16	MOBILE_BANKING_FLAG	If customer uses mobile banking (Yes/No).	Nominal
17	USSD	If customer uses USSD banking (Yes/No).	Nominal
18	Repayment_History	Customer's loan repayment status.	Ordinal
19	Debit_Transactions_Last_3_Months	Total number of transactions in last 3 months.	Numeric
20	Credit_Transactions_Last_3_Months	Total number of credit transaction in last 3 months	Numeric

Figure 3.2: initial extracted data.

3.3.1 Descriptive Analysis

The dataset contains a total of 308,293 customer records and 19 attributes. These features include demographic information, banking behavior, and service utilization metrics. The customer data from the Bunna Bank core banking database system reveals notable demographic and behavioral trends. Figure 3.3 Distribution of key categorical features (a) Gender, (b) Marital Status, (c) Occupation, (d) Location, (e) Account Type, (f) Loan Ownership, (g) ATM Usage, (h) Mobile Banking, (i) USSD Activation, (j) Repayment History.



Figure 3.3: Demographic and Behavioral trends.

Males make up the majority of the customer base, accounting for approximately 65.8%, while females represent 34.2%. Marital status is dominated by married individuals (47.9%), followed by singles (30.4%) and widowed customers (21.7%). Unemployment is the most common occupation (29.9%), with significant representation from agricultural/manual laborers (25.3%) and professionals (16.6%). Geographically, the highest concentration of customers comes from Amhara (29%) and Addis Ababa (27.8%), indicating a strong urban and central regional presence. Regard-

ing account types, Youth & Diaspora Accounts are the most prevalent, slightly ahead of General Savings and Special Purpose Accounts, showing a balanced approach to customer segmentation. The majority of customers (62.9%) do not have active loans, suggesting either limited credit access or risk aversion. ATM usage is widespread, with over 77% actively using the service. Mobile and USSD banking channels are adopted by 61.2% and 64.6% of customers respectively, though a significant proportion remains offline or under digitized. Lastly, repayment history indicates that 62.9% of customers have no loan record, while only a small fraction display good repayment behavior, and about 17.8% are categorized under default or poor, highlighting areas for credit support and financial literacy improvement.

3.3.2 Data Quality Verification

The dataset used in this study originates from digital sources (mainly from core banking databases), ensuring a relatively high level of reliability and record completeness. Prior to analysis and prediction, the data was thoroughly examined for the completeness and correctness of required attributes, as well as overall integrity. Demographic and financial attributes considered essential for understanding customer behavior are included; however, the researcher acknowledges potential limitations in the data. Despite its generally good quality, some entries may be missing, incomplete, or irrelevant. These issues will be systematically addressed during the data preprocessing phase to ensure the dataset is suitable for modeling and analysis.

3.4 Data Preprocessing

This stage is fundamental in the machine learning process, aiming to gain a comprehensive view of the data sources, structure, quality, and distribution. It begins with data collection and includes various activities to identify quality issues, extract insights, and prepare the data for analysis. Data preprocessing involves cleaning, transforming, balancing, and selecting relevant features to ensure the dataset is well-structured and reliable for modeling. Effective preprocessing enhances the performance of machine learning algorithms by removing noise, redundancy, and

irrelevant or unreliable information [76]. Without proper preprocessing, achieving accurate and meaningful model outcomes during training becomes significantly more difficult [76].

3.4.1 Feature Selection Based on Domain Expertise

In this study, we focused was on understanding the actual customers who left the bank and identifying the factors that contributed to their churn. Instead of automatically reducing features, we worked closely with domain experts to select the most relevant attributes for analysis.To ensure meaningful insights and maintain interpretability, we carefully examined the data of customers who had actually left the bank, analyzing their behavioral, demographic, and transactional characteristics. Based on this detailed understanding, we identified the features most informative for predicting churn and assessing the associated risk factors.

This expert-driven feature selection approach allowed us to maintain data quality, improve model efficiency, and ensure that the resulting predictions are both accurate and interpretable. By focusing on the attributes most relevant to customers who churned, the model provides actionable insights into the key factors influencing customer churn.

3.4.2 Data Integration

Data integration is merging data from multiple data stores might be required [77]. For this study, demographic and behavioral data were stored in separate tables hence the data was integrated using a common identifier, the "CIF ID." This process merges customer demographics with their corresponding transaction data.

3.4.3 Attribute Construction

Attribute construction is a task that can improve the result of machine learning model accuracy, which is creating new attributes from the existing attributes [78]. In this phase of the study, new attributes were created from existing ones to enhance the data's analytical value. For example, the "Account_Age" attribute was derived by subtracting the "Tenure" (account opening date) from the reference date which

is the last date of taking data from database (March 22, 2025), to represent the age of the account in months. To further support the analytical goals of this research, an additional attribute was constructed:

- **Transaction_Recency:** This attribute represents the number of days since the customer's last transaction, calculated by subtracting the "Last_TXN_DATE" from the reference date.

These engineered features were designed to better capture customer behavior over time. The original date fields were subsequently removed after transformation. The major objective of these transformations was to prepare a more informative dataset for model training and insightful visualizations.

Age	Gender	Income	Marita	Occupa	Locatio	Accour	Current_B	TXN_Cc	Is_Loa	Total_N	Total_N	Tenure	Is_ATM	Has_M	Is_USS	Repayn	Debit_1	Credit_	Last_TX	DATE
36	Male	1000000	Single	Unemploy	Addis Aba	Special Pu	3634.3	3	Yes	51400		########	No	Yes	Yes	Good	3	0	########	
40	Male	87000	Married	Agricultur	Addis Aba	Youth & D	866871.51	67	Yes	1741362	2393513	########	Yes	Yes	No	Good	49	18	3/5/2025	
40	Male	19532	Married	Agricultur	Addis Aba	Youth & D	16168.26	14	No	51511.9	66465.5	########	Yes	Yes	No	No Record	12	2	3/7/2025	
40	Male	10000	Married	Agricultur	Addis Aba	Special Pu	83574.03	83	Yes	354800	384891	########	No	No	No	Good	60	23	3/7/2025	
40	Male	1	Married	Agricultur	Addis Aba	Special Pu	514490.14	65	Yes	768131	817613	########	Yes	No	No	Default	49	16	3/7/2025	
40	Male	369985	Single	Agricultur	Addis Aba	General Sa	50132.74	98	Yes	305501	318540	########	No	No	Yes	Default	48	50	########	
40	Female	30000	Married	Private Se	Addis Aba	Special Pu	46928.33	1	No	50		########	Yes	No	Yes	No Record	1	0	3/3/2025	
45	Male	10000	Married	Unemploy	Amhara	General Sa	5000	2	Yes	32280.6		########	Yes	Yes	No	Good	2	0	########	
40	Female	1500	Married	Agricultur	Addis Aba	Youth & D	15847.29	107	Yes	196753	199968	########	Yes	Yes	No	Good	87	20	3/6/2025	
40	Male	10000	Married	Agricultur	Addis Aba	Special Pu	1621.56	12	No	174370	175142	########	Yes	Yes	No	No Record	8	4	2/7/2025	
40	Male	0	Married	Unemploy	Addis Aba	General Sa	8746.41	13	Yes	1.2E+08	1.2E+08	########	Yes	No	Yes	Good	9	4	1/6/2025	
51	Male	3000	Married	Private Se	Addis Aba	Youth & D	96579.26	33	Yes	1.5E+07	1.5E+07	########	Yes	Yes	Yes	Good	16	17	########	
40	Female	23400	Married	Agricultur	Addis Aba	General Sa	52305.22	6	No	2203.84	1220.14	########	Yes	No	Yes	No Record	3	3	########	
40	Male	52241	Widowed	Agricultur	Addis Aba	Special Pu	446.5	3	No	10500	5160.71	########	Yes	No	Yes	No Record	2	1	########	
40	Female	5200	Married	Agricultur	Addis Aba	Youth & D	21480.08	1	No	5		########	Yes	Yes	Yes	No Record	1	0	2/7/2025	
45	Male	20000	Married	Unemploy	Amhara	Special Pu	17577.02	8	Yes	2494872	500000	########	Yes	Yes	No	Good	6	2	########	
43	Male	45000	Single	Private Se	Addis Aba	General Sa	80130.23	108	Yes	265533	220898	########	Yes	No	No	Default	96	12	3/7/2025	
34	Male	4000	Widowed	Professior	Addis Aba	Youth & D	3814.6	24	Yes	4600000	4600600	########	No	Yes	Yes	Good	12	12	########	
44	Male	10000	Married	Unemploy	Addis Aba	Youth & D	878.96	5	Yes	517267	517913	########	Yes	No	Yes	Good	3	2	3/7/2025	
40	Male	8500	Married	Agricultur	Addis Aba	Special Pu	62792.67	7	No	18000	49043.6	########	Yes	No	Yes	No Record	5	2	########	
38	Female	25000	Widowed	Unemploy	Addis Aba	Youth & D	268229.28	55	Yes	2639433	2463019	########	Yes	No	No	Default	43	12	3/3/2025	
40	Male	22000	Married	Agricultur	Addis Aba	Special Pu	442414.85	1	No		409885	########	Yes	Yes	No	No Record	0	1	1/6/2025	
40	Male	1500	Married	Agricultur	Addis Aba	Special Pu	799.95	1	No		18.91	########	Yes	No	Yes	No Record	0	1	########	

Figure 3.4: Final dataset ready for experimentation.

3.4.4 Data Cleaning

Data cleaning serves as a foundational step in building high-quality machine learning models, as the performance of these models heavily depends on the quality of the data used [79]. It is a vital process that transforms raw data into a reliable and accurate form for analysis [79]. This step addresses multiple data quality challenges by improving the dataset through actions such as handling missing values, removing outliers, reducing noise, and resolving inconsistencies. To deal with missing data, different strategies were applied, including imputing the missing values.

Missing Values

Missing values are a frequent issue in data analysis and can be addressed through various strategies. In this study, appropriate techniques were applied based on the extent and nature of the missing data. Upon assessing the dataset, several variables were found to contain missing values, necessitating targeted imputation strategies.

The *Marital_Status* and *Location* attributes exhibited relatively low levels of missing values. Given the minimal extent of the gaps and the categorical nature of these features, the mode (most frequent value) was used to impute missing entries, ensuring data integrity with minimal distortion.

The *Income* variable contained only three missing entries. To address this, K-Nearest Neighbors (KNN) imputation was implemented using an odd-numbered k range from 3 to 1999 (`k_values = range(3, 2000, 2)`). This technique estimates missing values based on similarities among neighboring data points. Throughout the range of k values tested, the imputation consistently filled all missing entries, particularly at higher k values. As a result, the *Income* field was rendered complete and ready for further analysis, with minimal risk of distortion.

In the case of *Total_Monthly_Deposit*, the distribution showed a highly skewed pattern with a significant proportion of repeated low values. As the missingness was not substantial, imputation was carried out using the mode, which was identified as 0.01. To facilitate deeper financial behavior analysis, the variable was subsequently categorized into meaningful bins (e.g., 0–1K, 1K–5K, etc.). After validation confirmed zero remaining missing entries, the temporary categorical column was dropped.

For *Total_Monthly_Withdrawal*, missing values were also imputed using the mode of the column. To enhance the accuracy of this process, a temporary categorization was applied to segment withdrawal amounts, allowing for conditional mode imputation within each bin. Once all missing values were addressed and no further imputation was required, the auxiliary category column was removed.

The *Repayment_History* variable posed a more complex challenge due to a considerable volume of missing entries, particularly among customers flagged as loan holders (*Is_Loan* = Yes). Rather than defaulting to statistical imputation, a rule-

based method informed by domain knowledge was employed. Criteria such as income level, account balance, transaction frequency, deposit volume, and mobile banking engagement were used to infer repayment behavior. This allowed for tailored imputations classified into "Good," "Poor," or "Default" categories. The logic was implemented via a custom function applied to each record, successfully imputing all missing values in the column.

After addressing the missing values, the dataset was rendered complete, with no null entries remaining across any variables. Initially, it contained 308,293 records, with notable proportions of missing values in features like *Total_Monthly_Withdrawal* (13.45%), *Repayment_History* (6.44%), and *Total_Monthly_Deposit* (3.10%). Through the application of appropriate imputation techniques discussed above, these missing values were successfully handled. The dataset is now fully cleaned and prepared for subsequent preprocessing steps and predictive model development.

Inconsistent Data Values

Inconsistent data values were identified in the *Occupation* and *Location* columns. For example, the occupation "SERVICE" appeared truncated as "SERVI," and banking roles were inconsistently spelled as "BANKE." In the *Location* field, the capital city "ADDIS ABABA" was abbreviated or misspelled as "ADDIA." These values were systematically corrected to maintain consistency and support reliable categorization in downstream analysis.

3.4.5 Data Transformation

Data transformation refers to the process of converting raw data into a structured format that is appropriate for analytical tasks and model development [79]. Machine learning algorithms tend to perform more efficiently when numeric data is provided as input. Therefore, categorical data must be converted into numeric values, where each category is assigned a corresponding number [79]. For this study, we used hot label and hot encoding methods to encode categorical features.

Algorithm 6: Encoding Categorical Variables

Input:

- Dataset with categorical features
- List `OHE_columns` of columns to apply One-Hot Encoding
- List `LE_columns` of columns to apply Label Encoding

Output:

- Dataset with encoded categorical features

1. **One-Hot Encoding:**
   (a) For each column `col` in `OHE_columns`:
      i. Identify all unique categories in `col`.
      ii. For each unique category except the first (to avoid multicollinearity):
         - Create a new binary column named `col_category`.
         - For each row, set the new column to 1 if the value in `col` equals the category, else 0.
      iii. Remove the original column `col` from the dataset.
2. **Label Encoding:**
   (a) For each column `col` in `LE_columns`:
      i. Identify all unique categories in `col`.
      ii. Assign an integer label (0, 1, 2, ...) to each unique category.
      iii. Replace each value in `col` with its assigned integer label.
3. **Return** the encoded dataset.

3.4.6 Data Standardization

As explained in the review by Fan et al. [79], data scaling is an important step in preparing data, ensures input features are on comparable scales and the validity for predictive models. It mainly transforms numerical values to standardized formats,

Attributes	Old values	New values
Gender	[male, female]	Male → 0 (base), Female → 1
Marital Status	[Single, Married, Widowed]	Single → [0,0] (base), Married, Widowed (OHE)
Occupation	Unemployed, Agricultural/Manual Labor, Private Sector, Professional/Academic, Public Sector	Unemployed → [0,0,0,0], others (OHE)
Location	Addis Ababa, Amhara, Tigray, Oromia, SNNPR, Gambela, Harari, Benishangul-Gumuz, Somali, Dire Dawa, Afar	10 columns (excluding Addis Ababa), e.g., Location_Amhara...
Account Type	Special Purpose & Targeted Accounts, Youth & Diaspora Accounts, General Savings Account	Special Purpose → [0,0], others (OHE)
Is Loan	no, yes	0, 1
Is ATM Active	no, yes	0, 1
Has Mobile Banking	no, yes	0, 1
Is USSD Active	no, yes	0, 1
Repayment History	Good, No Record, Default, Poor	0, 1, 2, 3

Figure 3.5: Encoded Categorical Features

improving the stability and performance of algorithms which helps by changing different variables to the same range, so no one variable dominates because of its size. Two common methods are min-max scaling and z-score standardization. Min-max scaling changes data to a range between 0 and 1 by using the minimum and maximum values of each variable. However, this method can be affected by outliers, which are extreme values. On the other hand, z-score standardization changes data to have a mean of zero and standard deviation of one, making it less affected by outliers and useful when data is normally distributed.

Two common methods are min-max scaling and z-score standardization.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3.1}$$

$$z = \frac{x - \mu}{\sigma} \tag{3.2}$$

This study used min-max scaling to ensure all input variables are on the same scale.

3.4.7 Labeling of Churn Risk

To train the model, customers were labeled according to their likelihood of leaving the bank, using a High Risk and Low Risk classification. This labeling was based on a detailed examination of the actual customers who had left Bunna Bank.
By analyzing behavioral, demographic, and transactional characteristics of churned customers, domain experts helped define meaningful thresholds and rules to assign risk labels. This expert-driven approach ensures that the labels accurately reflect real-world patterns of customer attrition, providing a reliable target for the predictive models [80].Accordingly, the following algorithm assigns churn risk labels to customers based on their profiles and behavioral characteristics

Algorithm 7: Risk Assignment Based on Customer Profile

Input:

- Customer dataset with features including: `Age`, `Marital_Status`, `Income`, `Occupation`, `Account_Type`, `Current_BAL`, `TXN_Count`, `Total_Monthly_Deposit`, `Total_Monthly_Withdrawal`, `Debit_Transactions_Last_3_Months`, `Credit_Transactions_`, `Is_ATM_Active`, `Has_Mobile_Banking`, `Is_USSD_Active`, `Repayment_History`, `Account_Age`, `Transaction_Recency`

Output:

- Risk label for each customer: `High_Risk` or `Low_Risk`

1. Initialize `rules_satisfied` = 0 for each customer.
2. **Demographics:**
   - If `Age < 35` or `Age > 50`, increment `rules_satisfied`.
   - If `Marital_Status` is `Single` or `Widowed`, increment `rules_satisfied`.
3. **Income and Occupation:**
   - If `Income <= 0`, increment `rules_satisfied`.

- If Occupation is in ['Agriculture', 'Manual Labor'], increment rules_satisfied.

4. **Account Usage:**

   - If Account_Type is Special Purpose or Youth, increment rules_satisfied.
   - If Current_BAL < 100, increment rules_satisfied.

5. **Transactions and Deposits:**

   - If TXN_Count < 20, increment rules_satisfied.
   - If Total_Monthly_Deposit <= 0, increment rules_satisfied.
   - If Total_Monthly_Withdrawal > Total_Monthly_Deposit, increment rules_satisfie
   - If Debit_Transactions_Last_3_Months = 0, increment rules_satisfied.
   - If Credit_Transactions_Last_3_Months = 0, increment rules_satisfied.

6. **Channel Usage:**

   - If Is_ATM_Active = 'No', increment rules_satisfied.
   - If Has_Mobile_Banking = 'No', increment rules_satisfied.
   - If Is_USSD_Active = 'No', increment rules_satisfied.

7. **Credit Behavior:**

   - If Repayment_History is Poor or Default, increment rules_satisfied.

8. **Tenure and Activity:**

   - If Account_Age < 6 months, increment rules_satisfied.
   - If Transaction_Recency > 365 days, increment rules_satisfied.

9. **Final Decision:**

   - If rules_satisfied >= 3, assign High_Risk.
   - Else, assign Low_Risk.

3.4.8 Correlation Matrix

Correlation matrix is a table displaying the correlation coefficients between all pairs of variables in a dataset, revealing the strength and direction of their relationships[96]. Values range from -1 to +1, where +1 indicates a strong positive correlation, -1 a strong negative correlation, and 0 no correlation [81, 82]. This matrix helps identify how variables are related, uncover patterns, and guide feature selection by highlighting which variables are most relevant especially those strongly correlated with the target variable thus aiding in better understanding the data and improving predictive models. In the figure offers insights into how numerical features relate to each other, helping us identify trends and patterns that may influence customer behavior. This matrix is valuable for feature selection, as it highlights which variables move together. Accordingly, Total Monthly Deposit and Total Monthly Withdrawal show a strong positive relationship, indicating customers who deposit more also tend to withdraw more. Similarly, transaction count is positively related to both debit and credit transactions over the last three months. Age and Income have little to no correlation with other variables, which might reduce their importance in prediction tasks. Diagonal values represent each feature's correlation with itself, which is always 1. By studying these relationships, we can select the most relevant features for churn prediction while considering potential redundancies. Features with stronger correlations may play a more influential role in modeling, while those with weaker links might be less useful.

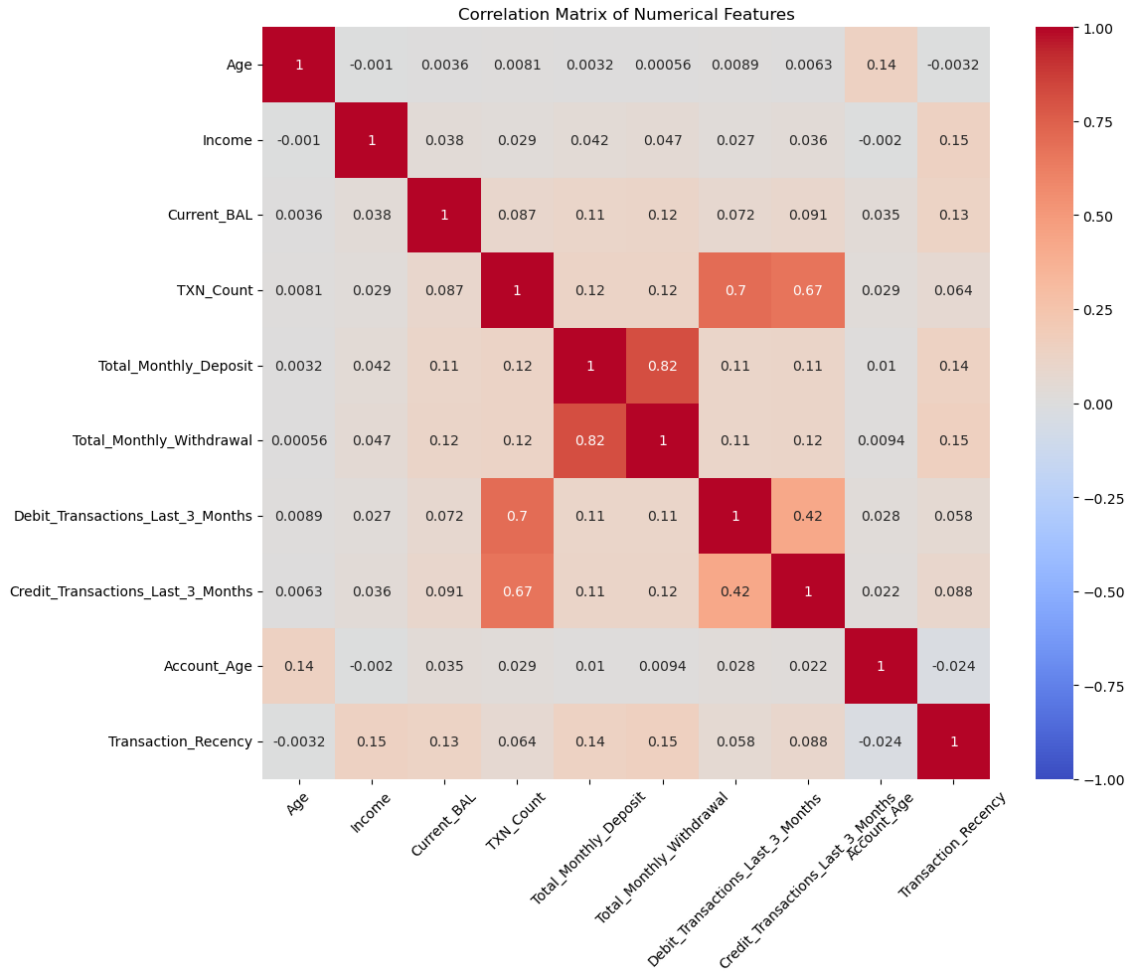


Figure 3.6: Correlation matrix of features.

3.4.9 Feature Selection

Selecting appropriate features is a vital aspect machine learning. The main objective is to pinpoint the variables that have the greatest influence on the model construction and effectiveness. [83, 84].When done properly, feature selection not only enhances predictive accuracy but also minimizes the risk of overfitting and makes the model easier to interpret.A range of strategies can be applied to carry out this task, including filter approaches, wrapper techniques, embedded methods, and hybrid methods.

Filter Method

The filter method is an approach in data preprocessing for selecting the most influential features. It operates independently of any predictive model, relying instead on statistical metrics to evaluate feature relevance [85]. Techniques such as corre-

lation analysis, mutual information, and chi-square tests are typically employed to assess the strength of the relationship between each feature and the target variable. This method enhances model efficiency, improves interpretability, and reduces the risk of overfitting by removing irrelevant or redundant data inputs. Because filter methods are fast and scalable, they are particularly useful in high-dimensional datasets [85].The Chi-square test is a statistical method used for feature selection, particularly with categorical data [85]. It evaluates the independence between each feature and the target variable. Features with high Chi-square values are strongly associated with the target and are more useful for model training, while features with low values are less relevant.

The Chi-square statistic is calculated as:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \tag{3.1}$$

Where: χ^2 = Chi-Square statistic. O = Observed frequency. E = Expected frequency.

Higher Chi-square scores indicate a stronger relationship with the target, making the feature more important for classification tasks. This technique helps select the most relevant features by comparing observed and expected frequencies, improving model accuracy and interpretability [86].

Algorithm 8: Chi-Square Feature Selection

Input:

- X: Feature matrix
- y: Target variable

Output:

- `selected_features`: List of features selected based on Chi-Square test

1. Initialize an empty list: `selected_features = []`.
2. **Apply Chi-Square Test:**

(a) Initialize `chi2_selector = SelectKBest(chi2, k='all')`.

(b) Fit the selector on the data: `chi2_selector.fit(X, y)`.

(c) Obtain p-values for all features: `chi2_p_values = chi2_selector.pvalues_`.

3. **Select Features:**

(a) For each feature i in `range(len(chi2_p_values))`:

- If `chi2_p_values[i] < 0.05`, add `X.columns[i]` to `selected_features`.

4. Print "Selected Features based on Chi-Square Test".

5. **Return** `selected_features`.

Wrapper Methods

Wrapper methods in machine learning involve repeatedly training and evaluating a model to identify the most relevant features for the task [85]. Techniques like forward selection, backward elimination, and Recursive Feature Elimination (RFE) are common examples. These methods generally enhance model accuracy by fine-tuning the selection of features but can be computationally expensive, making them less suitable for large datasets [85]. Since wrapper methods assess subsets of features by evaluating the model's performance, they can be resource-intensive and may require substantial computational time, especially with a high number of features. Recursive Feature Elimination (RFE) is a popular wrapper method where features are removed iteratively based on their contribution to model performance [83, 85]. It starts with all features and eliminates the least important ones, often using metrics like feature coefficients or importance scores. RFE is effective when used with algorithms such as linear regression, support vector machines, or tree-based models. Below is the pseudocode [87]:

Embedded Methods

Embedded methods perform feature selection as an integral part of model training, allowing them to be both efficient and predictive. Unlike filter and wrapper

techniques that operate before or around the modeling process, embedded methods identify and select important features during the learning phase itself [85]. A key advantage of this approach is that feature selection is directly aligned with the model's objective, ensuring that only the most relevant variables are retained. One prominent example of an embedded method is Random Forest, a tree-based ensemble algorithm that evaluates feature importance based on how effectively a feature contributes to decision splits [88]. Features that consistently reduce impurity in the trees are ranked higher, making Random Forest not only a powerful predictive tool but also an efficient method for automatic feature selection.

Below is the pseudocode for Random Forest Feature Selection [88]:

Hybrid Feature Selection

Hybrid feature selection methods combine multiple techniques such as filter, wrapper, and embedded approaches to identify the most relevant features for a given task [86, 89]. This strategy is designed to harness the advantages of each method while compensating for their individual limitations, improving the overall effectiveness and reliability of feature selection. Practitioners often begin with a fast, computationally efficient filter method to reduce the initial pool of features, followed by a more refined selection using a wrapper method. This layered approach balances computational efficiency with model performance, making hybrid methods especially suitable for complex or large-scale datasets.

Below is the pseudocode [86, 89]:

3.5 Dataset Split

Data splitting is a crucial step in the machine learning model validation, primarily used to assess a model's performance [90, 91]. This process involves dividing the dataset into separate, non-overlapping subsets, typically designated as training and testing sets. The training set is used to fit the model, while the testing set is held out to evaluate how well the model performs on unseen data [91]. A widely used approach follows an 80:20 ratio, allocating 80% of the data for training and 20%

for testing. The choice largely depending on the dataset size, problem complexity, and model type. Although the 80:20 split is often inspired by the Pareto principle, it remains a heuristic guideline rather than a rigid standard. Selecting an appropriate split ratio is essential to ensure balanced training and reliable performance evaluation [91].

3.6 Imbalance Dataset Handling

Class imbalance is a common problem in machine learning, where one class (usually the minority class) has far fewer instances than the other classes. This can cause models to be biased toward the majority class, resulting in poor performance on the minority class.

To address this issue, the SMOTE (Synthetic Minority Oversampling Technique) method is used. Instead of simply duplicating existing minority instances, SMOTE generates new synthetic examples by interpolating between existing samples [92]. This helps expand the decision space of the minority class and reduces the risk of overfitting.

3.7 Select a Predictive Model

Following the preprocessing of the dataset, Segmenting, Attrition status labeling and splitting, a combination of ensemble and deep learning algorithms including Random Forest, Gradient Boosting (GBoost), Extreme Gradient Boosting (XGBoost), LightGBM, Extra Trees, and Multilayer Perceptron (MLP) were employed to develop the predictive model. These algorithms were selected due to their robustness and flexibility in managing diverse data structures while delivering high predictive accuracy [93]. Their widespread use stems from their strong performance and ability to model complex patterns, making them especially suitable for tabular datasets. Here is how each algorithm contributes effectively to modeling tasks involving tabular datasets.

3.7.1 Random Forest

Random Forest trees are independently constructed using a random subset of features, making them capable of handling both categorical and continuous data. This characteristic makes them highly effective for tabular datasets. Furthermore, Random Forest models are robust against overfitting and demonstrate strong generalization to unseen data [52, 94, 95].

3.7.2 Gradient Boosting

Gradient Boosting: As explained by Khan et al. [52], Gradient Boosting is a powerful ensemble learning method that builds a strong predictive model by sequentially combining weak learners trained on the pseudo-residuals of a chosen loss function. Applicable to both classification and regression tasks, it iteratively minimizes error by directing each learner to correct the mistakes of its predecessors. While effective in capturing complex patterns in data and achieving high accuracy, especially in high-dimensional settings, it requires careful tuning of hyper parameters such as learning rate and number of stages to prevent overfitting. Gradient Boosting has demonstrated broad applicability across domains, although its training can be computationally intensive and less straightforward to parallelize.

3.7.3 Extreme Gradient Boosting

As stated by Khan et al. [52], XGBoost is an advanced variant of Gradient Boosting that incorporates sophisticated regularization techniques, including L1 (lasso) and L2 (ridge) regularization, to improve model generalization. Compared to traditional Gradient Boosting, XGBoost offers superior computational efficiency due to its ability to parallelize training across clusters, making it highly suitable for large-scale datasets and high-dimensional problems. Despite its computational intensity, XGBoost remains highly popular for tasks like classification and regression, particularly in competitive environments due to its high accuracy and scalability. However, careful hyper parameter tuning is essential to mitigate overfitting and achieve optimal performance [52].

3.7.4 LightGBM

According to Zhang and colleagues [51, 52], Light Gradient Boosting Machine is a powerful gradient boosting framework that excels in handling large datasets while maintaining high computational efficiency. It utilizes a leaf-wise tree growth strategy rather than the traditional level-wise approach, allowing it to optimize splits more effectively and reduce computational time. One of its key innovations is the gradient-based one-side sampling (GOSS) technique, which selectively samples data during tree construction, thereby preventing overfitting and improving efficiency. LightGBM is particularly effective for tasks involving classification, regression, and ranking, and its ability to manage imbalanced datasets by assigning different weights to classes makes it an appealing choice for problems like fraud detection, anomaly detection, and medical diagnosis.

3.7.5 Extra Trees

As stated by Hussen et al. [54, 96], extremely randomized tree stands out as a highly effective choice for classification tasks due to its unique approach to split selection. Unlike traditional decision trees, which meticulously search for the optimal split, extra trees randomly selects splits, thereby introducing increased randomness that enhances model diversity and robustness. This characteristic not only mitigates overfitting but also improves the model's ability to generalize, especially when dealing with complex or noisy data. Additionally, the computational efficiency of extra tree makes it suitable for handling large datasets without compromising accuracy. By leveraging ensemble averaging and random feature selection, effectively balances bias and variance, achieving reliable classification performance with fewer trees compared to more conventional methods.

3.7.6 MLP

The Multilayer Perceptron (MLP), as described by [96], is a widely utilized neural network architecture known for its structured composition of an input layer, one or more hidden layers, and an output layer, where each layer consists of interconnected

neurons. These neurons process inputs through weighted connections and biases to generate outputs. Due to its capability to learn complex patterns, MLP is frequently employed in for classification tasks, such as predictive modeling and medical image analysis. And therefore, when appropriately trained and optimized, the MLP model can accurately categorize data into separate classes, making it a highly effective solution for diverse classification applications.

3.8 Model Evaluation

After the models are built, the next critical step in the machine learning process is to evaluate the performance of the model. In this study, accuracy, precision, recall, ROC curve, confusion matrix, and F1-score are employed to evaluate the performance of classification models.

3.8.1 Confusion Matrix

As pointed out by [97], a confusion matrix is a tool for evaluating the performance of a model by summarizing predicted and actual values which helps to know the weakness of the model for model enhancement. For binary classification problems, the confusion matrix defines the base for performance measures. Most of the performance metrics are derived from the confusion matrix, for example, accuracy, precision, F1 score, and recall [98]. The confusion matrix allows to identify accurately classified, incorrectly classified, and their relationship:

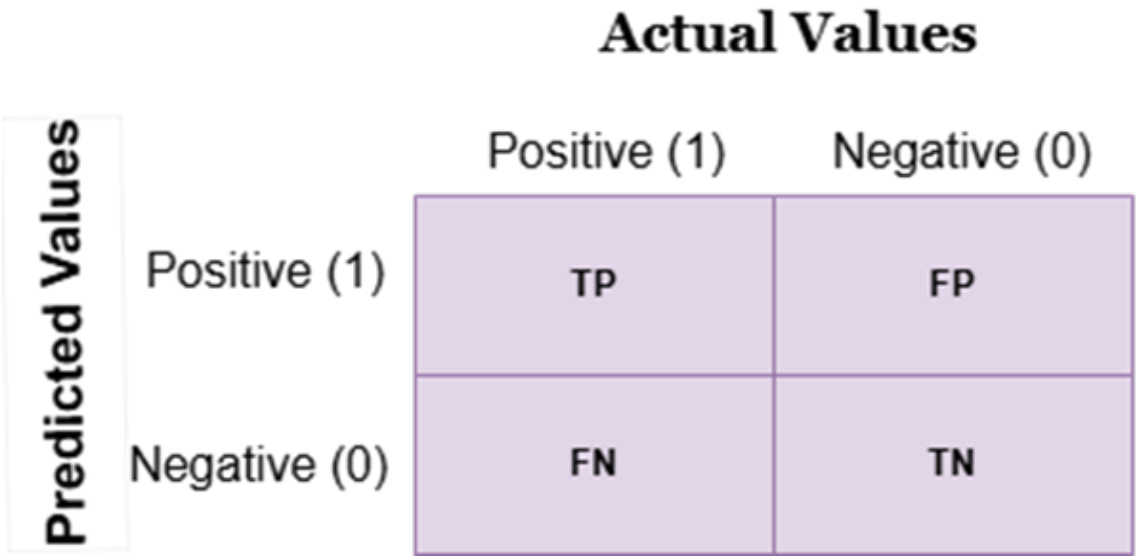


Figure 3.7: confusion$_m atrix.png$

- True Positive (TP): An instance for which both predicted and actual values are positive.
- False Positive (FP): An instance for which predicted value is positive but actual value is negative.
- True Negative (TN): An instance for which both predicted and actual values are negative.
- False Negative (FN): An instance for which predicted value is negative but actual value is positive.

3.8.2 Accuracy

Accuracy quantifies the proportion of correct predictions made by the model out of the entire test dataset. As a fundamental and essential metric, it helps evaluate the model's performance. The formula for calculating accuracy is as follows

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.3}$$

3.8.3 Precision

Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It helps assess the reliability of the model's positive predictions. Precision is calculated as follows

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.4}$$

3.8.4 Recall

Recall measures the proportion of actual positive cases that the model correctly identified. It indicates how well the model captures true positives. Recall is calculated as follows

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.5}$$

3.8.5 F1-score

The F1 score is a weighted average that combines both precision and recall (true positive rate). It evaluates the performance of a classification model by summarizing precision and recall using their harmonic mean, based on values from the confusion matrix

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.6}$$

3.8.6 Receiver Operating Characteristic

As stated by [99, 100], a Receiver Operating Characteristic (ROC) curve is a graphical tool that plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. It helps evaluate and compare classifiers by showing their ability to distinguish between classes as the decision threshold changes. On the plot, the x-axis represents the false positive rate, while the y-axis shows the true positive rate. The diagonal line running from the bottom-left to the top-right corner indicates random guessing; a good classifier's curve should lie above this line. The Area under the Curve (AUC) is a widely used metric to summarize the classifier's overall effectiveness, with higher AUC values indicating better per-

formance. By analyzing the ROC curve, it's possible to determine the most suitable threshold for predictions based on specific goals and limitations.When evaluating classifier performance, researchers consider metrics such as accuracy, precision, F1-score, recall, and the ROC curve to choose the best model.

3.9 Feature Importance

Feature importance analysis is a crucial aspect of developing accurate and reliable predictive models. In the context of customer churn prediction, analyzing feature importance helps identify the key variables that significantly impact the likelihood of customer churn [101]. By evaluating the contribution of each feature within the model, we can gain insights into the factors that most strongly influence whether a customer stays or leaves. In this study, we utilized feature importance techniques within the XGBoost algorithm to measure the contribution of each feature to the prediction of customer churn. XGBoost was selected as the top performing model, as it consistently outperformed other algorithms in various evaluation metrics. The feature with the highest importance value was considered the most influential in predicting churn, while the feature with the lowest value was considered the least influential.

Chapter 4

Experimental Results and Discussion

This chapter discusses the dataset preparation and analysis, tools, and experimental setup used to develop the prototype to implement the model. We will discuss the results acquired from the experiment of the model and give an interpretation of the experimental findings.

4.1 Development Environment and Tools

This study aims to develop customer segmentation and churn prediction model with XAI. As hardware, we used Intel(R) Core(TM) i5-5300U, CPU @ 1 GHz 1.190 GHz, 8 GB memory, 1TB hard disk. As software tools, the thesis was prepared entirely using LaTeX on the Overleaf platform to ensure consistent formatting and adherence to academic standards, Python programming language for the experiment since Python libraries are readily available for all our ML workflow. We use Jupyter Notebook through Anaconda distribution as the development environment for writing and running Python code and testing machine learning models. Microsoft Office Presentation 2023 was used for thesis presentation, and Microsoft Office Excel 2023 for understanding the datasets manually. This research is implemented using an Anaconda navigator. Once the dataset is prepared, preprocessed, and features selected, the next step is to build the optimal predictive model that can segment and identify customer churn status.

After the dataset is prepared, preprocessed, segmented, and relevant features are selected, the next step is to develop the optimal predictive model for identifying churn status. In this study, we conducted six experiments with Random Forest,

Gradient Boosting, Extreme Gradient Boosting, LightGBM, Extra Trees, and MLP models. The next section provides a descriptive overview of the dataset.

4.2 Data Preprocessing Summary

Following the descriptive analysis phase, the next critical step involved preparing the experimental environment and conducting exploratory data analysis (EDA) as a foundation for machine learning modeling.

4.2.1 Importing Libraries & Dataset Loading

Initially, all essential Python libraries and specific modules were carefully imported to support various aspects of the analysis. A snapshot of this process is presented in Figure 4.1.

```
import pandas as pd
import numpy as np
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler, LabelEncoder

data_path = r"C:\Users\BIB-Ermi\OneDrive\Desktop\Msc\Prototype\expalainbable\thesis\lab\14 4 2025\15 04 2025\BunnaHistoricaldata_
df = pd.read_excel(data_path)
```

Figure 4.1: Importing Libraries and Dataset Loading

In this phase, the `pandas` library was employed for data manipulation and analysis, while `numpy` provided capabilities for numerical operations and linear algebra computations. The dataset utilized in this research comprises extensive customer records containing both behavioral and transactional attributes. The dataset was loaded from a locally stored `.xlsx` file. Figure 4.1 illustrates the code segment used to read the dataset into a DataFrame.

4.2.2 Exploratory Data Analysis (EDA)

EDA served as a crucial step in understanding the structure, quality, and underlying patterns within the data. To begin, the `head()` function was executed to display the initial records of the dataset. Subsequently, the `describe()` method was used to generate summary statistics for all numerical features, including mean, standard

deviation, minimum, and maximum values. In parallel, the `info()` method was applied to display structural metadata about the DataFrame.

4.3 Results of Feature Selection

This section outlines the methods used for feature selection and presents the resulting outcomes. Various techniques were applied to identify the most significant features, ensuring the model's performance was optimized while minimizing complexity. The selected features were then evaluated for their impact on the predictive accuracy of the model.

Experiment 1: Filter Method – Chi-Square Test In this experiment, a filter-based feature selection approach was applied using the Chi-Square test to evaluate the statistical relationship between input features and the churn risk labels (High Risk, Low Risk). A total of 20 features were initially considered, and the Chi-Square test helped identify the most influential attributes. Based on the results, 16 features were retained for further modeling, ensuring that only the most relevant variables were included in the predictive framework.

```
['Age', 'Gender', 'Marital_Status', 'Occupation', 'Location',
 'Account_Type', 'Account_Age', 'TXN_Count', 'Total_Monthly_Withdrawal',
 'Debit_Transactions_Last_3_Months', 'Credit_Transactions_Last_3_Months',
 'Transaction_Recency', 'Is_ATM_Active', 'Has_Mobile_Banking',
 'Is_USSD_Active', 'Repayment_History']
```

These features were selected iteratively based on statistical significance and relevance to domain knowledge, particularly in the context of regional and demographic diversity in Ethiopian banking customers.

4.3.1 Model Evaluation with Chi-Square Selected Features

To evaluate the predictive power of the Chi-Square-selected features, several classifiers were trained and compared, including Random Forest, Gradient Boosting, XGBoost, Multi-Layer Perceptron (MLP), LightGBM, and Extra Trees. Each model

was trained on the reduced feature set, and performance was measured using accuracy, precision, recall, F1-score, and confusion matrices.

Among the models, Gradient Boosting and XGBoost achieved the highest accuracy (approximately 74%), followed closely by LightGBM and MLP, while Random Forest and Extra Trees also delivered competitive results. This confirms that the selected features provide a solid foundation for predicting customer churn in the Ethiopian banking context. Importantly, ensemble models demonstrated superior performance, highlighting their robustness in handling complex feature interactions.

Table 4.1: Performance of classifiers with Chi-Square-selected features

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Random Forest	0.734	0.74	0.73	0.73
Gradient Boosting	0.736	0.75	0.74	0.74
XGBoost	0.741	0.75	0.74	0.74
MLP	0.733	0.74	0.73	0.73
LightGBM	0.738	0.76	0.74	0.74
Extra Trees	0.732	0.74	0.73	0.73

These findings validate the utility of filter-based feature selection, specifically the Chi-Square test, as an effective preprocessing step in churn prediction tasks. They also highlight the potential of ensemble learning models to deliver strong and interpretable predictions in banking applications.

Experiment 2: Embedded Method – Random Forest Feature Selection

In this experiment, we employed a wrapper-based feature selection approach using the Random Forest algorithm to identify the most informative features for churn prediction. Unlike filter methods, wrapper techniques consider feature interactions and model-specific performance metrics, providing a more comprehensive evaluation of feature relevance in the predictive context.

A total of 12 features were available for modeling:

```
['Age', 'Income', 'Current_BAL', 'TXN_Count', 'Total_Monthly_Deposit',
'Total_Monthly_Withdrawal', 'Is_ATM_Active', 'Has_Mobile_Banking',
'Is_USSD_Active', 'Debit_Transactions_Last_3_Months',
'Transaction_Recency', 'Marital_Status_Single']
```

Random Forest feature importance analysis highlighted Total_Monthly_Deposit, TXN_Count, and Income as the most predictive features, while channel engagement features like Is_ATM_Active and Is_USSD_Active were also important in other models such as XGBoost.

Table 4.2: Performance of classifiers with Random Forest-selected features

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.924	0.93	0.93	0.93
Gradient Boosting	0.892	0.93	0.89	0.90
XGBoost	0.928	0.94	0.94	0.96
MLP	0.866	0.91	0.86	0.88
LightGBM	0.900	0.93	0.90	0.91
Extra Trees	0.910	0.92	0.91	0.91

The models trained on these features demonstrated strong performance. XGBoost achieved the highest test accuracy of 0.9278, followed by Random Forest (0.9236), Extra Trees (0.9099), and LightGBM (0.8999). Gradient Boosting and MLP performed slightly lower with accuracies of 0.8922 and 0.8662, respectively. Classification reports and confusion matrices indicated high precision, recall, and F1-scores for the churn class (label 1), confirming the sufficiency of these features for accurately identifying high-risk customers.

Overall, the Random Forest-based embedded feature selection method effectively identified a compact and informative subset of features, capturing financial behavior, transactional activity, and engagement channels. This enabled efficient and accurate churn prediction while preserving interpretability, which is essential for practical banking applications.

Experiment 3: Wrapper Method – Recursive Feature Elimination (RFE)

In this experiment, we applied a wrapper-based feature selection method using Recursive Feature Elimination (RFE) to identify the most relevant features for predicting customer churn risk. RFE recursively removes the least important features based on model performance, resulting in a compact subset of highly informative predictors. Random Forest was used as the estimator due to its ability to capture complex interactions and provide feature importance insights, which are critical for explainable AI (XAI) in risk prediction.

RFE selected the following ten features as most significant:

```
['TXN_Count', 'Is_ATM_Active', 'Has_Mobile_Banking', 'Is_USSD_Active',
'Debit_Transactions_Last_3_Months', 'Credit_Transactions_Last_3_Months',
'Account_Age', 'Transaction_Recency', 'Marital_Status_Single',
'Marital_Status_Widowed']
```

These features reflect a combination of transactional activity, channel engagement, account tenure, and demographic characteristics, all of which contribute to assessing the risk of customer churn.

Table 4.3: Performance of classifiers with RFE-selected features for predicting customer churn risk

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.8905	0.90	0.92	0.89
Gradient Boosting	0.8426	0.90	0.84	0.86
XGBoost	0.9013	0.91	0.94	0.90
MLP	0.8397	0.91	0.83	0.86
LightGBM	0.8631	0.91	0.87	0.88
Extra Trees	0.8860	0.90	0.92	0.89

The models trained on these features demonstrated strong predictive performance. XGBoost achieved the highest accuracy of 0.9013, followed by Random Forest (0.8905), Extra Trees (0.8860), LightGBM (0.8631), Gradient Boosting (0.8426), and MLP (0.8397). Classification reports indicated high precision, recall, and F1-scores for the churn class (label 1), confirming that the selected features effectively identify high-risk customers.

Feature importance rankings provided actionable insights for risk prediction. For example, Credit_Transactions_Last_3_Months, TXN_Count, Transaction_Recency, and Debit_Transactions_Last_3_Months were consistently top predictors, highlighting that recent financial activity strongly influences churn risk. Engagement features such as Is_ATM_Active, Has_Mobile_Banking, and Is_USSD_Active were also important, indicating that customer interaction with banking channels is a key risk factor.

Overall, the RFE approach yielded a compact, interpretable, and high-performing feature set, enabling both accurate churn prediction and transparent identification

of risk factors. This aligns with the goals of XAI, providing banks with actionable insights to prioritize retention strategies and mitigate customer attrition effectively.

Experiment 4: Hybrid Feature Selection Strategy In this experiment, we applied a hybrid feature selection approach combining the strengths of Chi-Square, Random Forest, and Recursive Feature Elimination (RFE) to identify the most informative and interpretable features for predicting customer churn risk. The final selected features included:

```
['Transaction_Recency', 'Age', 'TXN_Count', 'Debit_Transactions_Last_3_Months',
'Total_Monthly_Withdrawal', 'Is_USSD_Active', 'Has_Mobile_Banking',
'Credit_Transactions_Last_3_Months', 'Account_Age', 'Is_ATM_Active',
'Marital_Status_Single', 'Marital_Status_Widowed', 'Repayment_History',
'Location', 'Account_Type_Youth & Diaspora Accounts']
```

These features reflect transactional behavior, account engagement, demographic attributes, and repayment history, all of which are key risk factors in customer churn prediction.

Table 4.4: Performance of classifiers with hybrid-selected features for predicting customer churn risk

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.9181	0.92	0.94	0.95
Gradient Boosting	0.8952	0.93	0.89	0.94
XGBoost	0.9223	0.93	0.94	0.95
MLP	0.9001	0.92	0.90	0.94
LightGBM	0.8996	0.93	0.90	0.94
Extra Trees	0.9102	0.92	0.93	0.95

Models trained on this hybrid feature set demonstrated strong performance. XGBoost achieved the highest accuracy of 0.9223, followed by Random Forest (0.9181), Extra Trees (0.9102), LightGBM (0.8996), Gradient Boosting (0.8952), and MLP (0.9001). Classification reports indicated high precision, recall, and F1-scores for the churn class (label 1), confirming that the selected features effectively identify high-risk customers. For example, Random Forest achieved a recall of 0.94 and F1-score of 0.95 for churners, while XGBoost achieved similar high metrics.

Table 4.5: Top Features Ranked by XGBoost Importance for Customer Churn risk factor Prediction

Rank	Feature	Importance
1	Marital_Status_Single	0.1397
2	Marital_Status_Widowed	0.1370
3	Is_ATM_Active	0.1295
4	Is_USSD_Active	0.1168
5	Has_Mobile_Banking	0.1021
6	Credit_Transactions_Last_3_Months	0.0844
7	Debit_Transactions_Last_3_Months	0.0635
8	Repayment_History	0.0624
9	Age	0.0548
10	TXN_Count	0.0545
11	Transaction_Recency	0.0364
12	Total_Monthly_Withdrawal	0.0109
13	Account_Age	0.0055
14	Location	0.0013
15	Account_Type_Youth & Diaspora Accounts	0.0013

According to feature importance analysis shown below using XGBoost revealed that customer demographic and engagement features, such as Marital Status (Single/Widowed) and channel activity (Is ATM Active, Is USSD Active, Has Mobile Banking), were the most influential in predicting churn. Transaction-related features, including Credit and Debit Transactions in the Last 3 Months, Repayment History, Age, TXN Count, and Transaction Recency, also contributed to model predictions, though to a lesser extent.

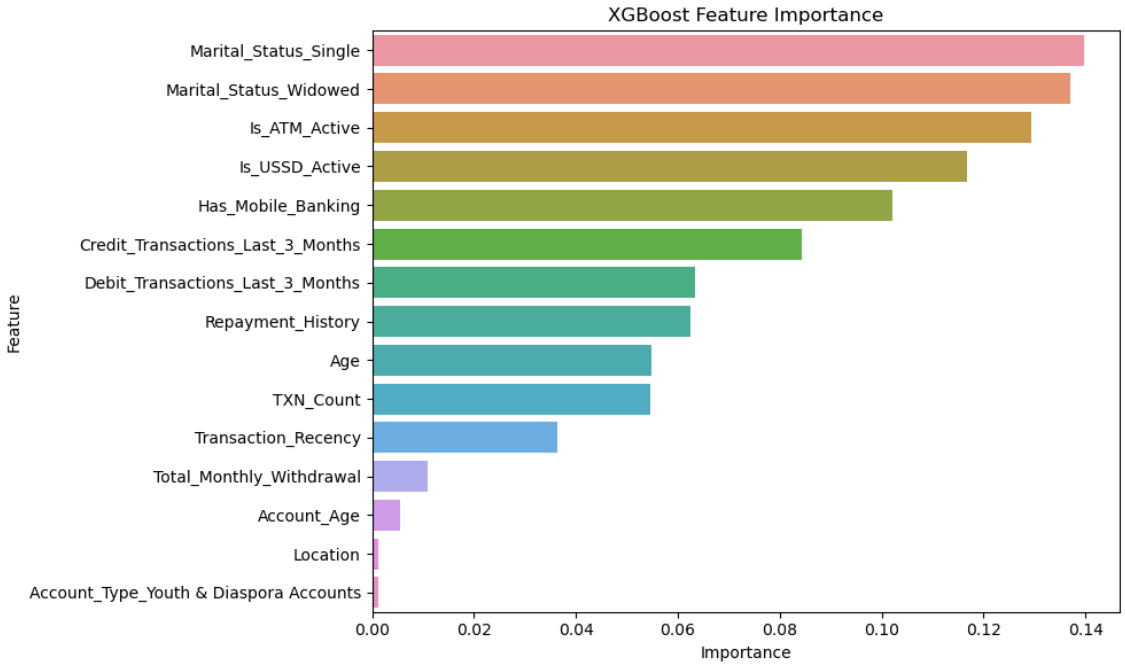


Figure 4.2: Feature importance analysis

These results highlight that customer demographics, engagement channels, and recent transaction behavior are key risk factors for churn, aligning with XAI principles by providing actionable insights for banking decision-makers.

The hybrid feature selection strategy combined with XGBoost achieved high predictive performance (accuracy: 0.9223, AUC: 0.97) while preserving interpretability, making the model suitable for real-world churn risk prediction and intervention planning.

4.3.2 Dataset Splitting and Class Balancing

To evaluate the predictive models' ability to generalize to unseen data, the dataset was partitioned into separate training and testing sets using an 80/20 split, where 80% of the data was used for training and 20% for testing. The target variable is **Risk_Label**, indicating whether a customer is **high-risk** (1) or **low-risk** (0). Features for training were selected through a rigorous hybrid feature selection process, combining filter, wrapper, and embedded methods to capture the most informative predictors. Table 4.6 shows the class distribution in the original training and test sets before applying any class-balancing technique. The training set is highly imbalanced, with more low-risk customers than high-risk customers.

Table 4.6: Class Distribution Before SMOTE (80/20 Split, Stratified)

Dataset	High-Risk (1)	Low-Risk (0)	Total
Training Set	212,228	34,406	246,634
Test Set	53,058	8,601	61,659

The initial training data exhibited a significant imbalance, with the majority of customers being low-risk. To address this, the **Synthetic Minority Oversampling Technique (SMOTE)** was applied to the training set only, generating a balanced class distribution. The test set was kept untouched to provide an unbiased evaluation of model performance.

Table 4.7: Class Distribution After SMOTE on Training Set (80/20 Split, Stratified)

Dataset	High-Risk (1)	Low-Risk (0)	Total
Training Set	212,228	212,228	424,456
Test Set	53,058	8,601	61,659

Balancing the training set allows models to learn equally from high-risk and low-risk customers, enhancing their ability to accurately identify potential churners. The test set reflects the original distribution to assess real-world performance. This methodology supports **explainable AI** principles, enabling interpretable predictions of risk factors and actionable insights for churn intervention strategies.

4.3.3 Model Development

We selected 15 essential features through a hybrid feature selection approach to predict customer churn. A series of experiments were performed to evaluate the predictive performance using these features. The experiments included models such as Random Forest, Gradient Boosting, XGBoost, LightGBM, Extra Trees, and MLP. The data set used for training contained 308,293 customer instances, categorized into two groups: High_Risk and Low_Risk. To build a predictive model, three main stages are involved: training and testing. During the training phase, the model is exposed to the data multiple times to learn the patterns and predict the desired outcomes. Finally, in the testing phase, the trained model is applied to new unseen data to assess its performance. The complete data set is divided into subsets for these phases, with 80% allocated for training, 20% for testing.

Experiment 1: Random Forest As shown in Figure 4.2, Experiment 1 was conducted using the Random Forest classifier, an ensemble method that combines multiple decision trees to achieve strong classification performance. In this experiment, Random Forest achieved an accuracy of 91.81%, with a precision of 0.92, recall of 0.92, and F1-score of 0.92 on average. These metrics indicate that the model effectively captures complex, non-linear relationships between the input features and the target variable, a task where linear models such as logistic regression typically fall short.

```
Random Forest Accuracy: 0.9180654892229845
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.68      0.79      0.73      8601
           1       0.96      0.94      0.95     53058

    accuracy                           0.92     61659
   macro avg       0.82      0.86      0.84     61659
weighted avg       0.92      0.92      0.92     61659

Random Forest Confusion Matrix:
[[ 6793  1808]
 [ 3244 49814]]
```

Figure 4.3: Performance evaluation for Random Forest

The confusion matrix reveals that 49,814 out of 53,058 high-risk customers were correctly classified, while 3,244 were misclassified as low-risk. For low-risk customers, 6,793 out of 8,601 were correctly identified, with 1,808 misclassified as high-risk. The high recall for high-risk customers demonstrates the model's robustness in detecting potential churn, making Random Forest highly suitable for predicting customer risk and informing targeted retention strategies.

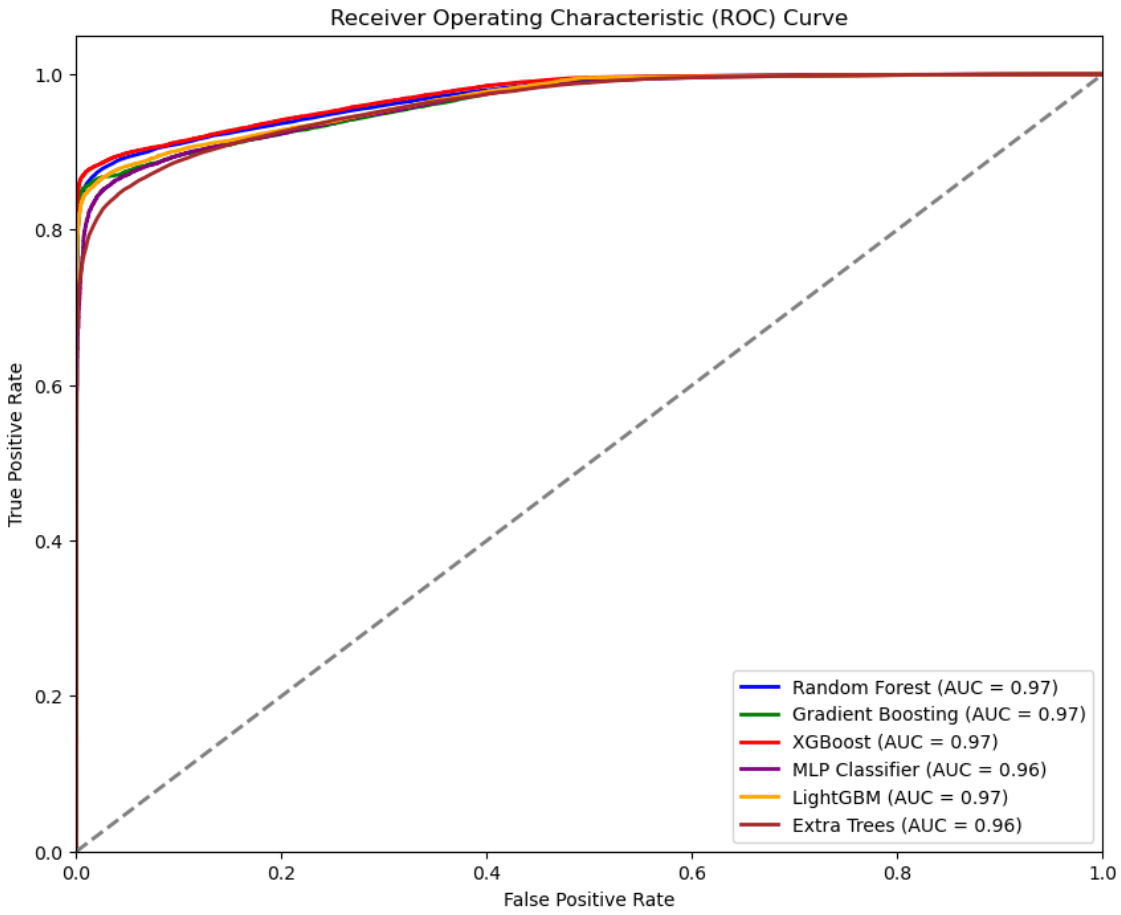


Figure 4.4: ROC curve for Developed Models

According to the Random Forest ROC curve (Figure 4.3), the model achieved an AUC of 0.97, demonstrating excellent capability in distinguishing between high-risk and low-risk customers. In the context of Predicting risk factor for customer churn, this result confirms Random Forest as a reliable model for identifying at-risk customers and informing retention strategies. Its strong precision (0.92), recall (0.92), and F1-score (0.92) indicate that the model not only captures the majority of high-risk customers but also minimizes false alarms for low-risk customers.

From an explainable AI (XAI) perspective, the Random Forest feature importance analysis highlights key drivers of churn risk, such as transaction activity (*TXN_Count*, *Debit Transactions Last 3 Months*, *Credit Transactions Last 3 Months*), engagement channels (*Is_USSD_Active*, *Has_Mobile_Banking*, *Is_ATM_Active*), and demographic factors (*Age*, *Marital Status*). These interpretable insights provide transparency into how customer behavior and account characteristics influence churn, ensuring that the model's predictions are both accurate and actionable.

Experiment 2: Gradient Boosting As shown in Figure 4.4 , the Gradient Boosting model achieved an accuracy of 89.52%, with strong performance across other metrics: precision of 93%, recall of 90%, and an F1-score of 90%. These results highlight the model's ability to capture complex, non-linear relationships between input features and the churn risk label.

The confusion matrix further illustrates this performance. Out of 53,058 high-risk customers, the model correctly identified 47,401, while 5,657 were misclassified as low-risk. For the 8,601 low-risk customers, 7,797 were correctly classified, with only 804 misclassified as churned. This balance demonstrates the model's robustness, particularly its high recall for at-risk customers, which is essential for proactive intervention.

Gradient Boosting provides insights into the key predictors driving churn risk. Features related to financial activity (*TXN_Count*, *Credit Transactions Last 3 Months*, *Debit Transactions Last 3 Months*), engagement channels (*Has_Mobile_Banking*, *Is_USSD_Active*), and demographic characteristics (*Age*, *Marital Status*) were among the most influential. The ROC curve (AUC = 0.97) confirms the model's ability to reliably distinguish between high-risk and low-risk customers.

```
Gradient Boosting Accuracy: 0.8952139995783259
Gradient Boosting Classification Report:
              precision    recall  f1-score   support

           0       0.58      0.91      0.71      8601
           1       0.98      0.89      0.94     53058

    accuracy                           0.90     61659
   macro avg       0.78      0.90      0.82     61659
weighted avg       0.93      0.90      0.90     61659

Gradient Boosting Confusion Matrix:
[[ 7797   804]
 [ 5657 47401]]
```

Figure 4.5: Gradient boosting performance evaluation

Experiment 3: Extreme Gradient Boosting (XGBoost) As shown in Figure 4.5, the XGBoost model achieved the highest performance among the tested al-

gorithms, with an accuracy of 92.23%. The model maintained strong balance across metrics, achieving 93% precision, 94% recall, and a 95% F1-score for the high-risk class. These results highlight XGBoost's capacity to handle complex, non-linear feature interactions while preserving predictive accuracy.

The confusion matrix further illustrates this performance. Out of 53,058 high-risk customers, XGBoost correctly identified 50,097, with only 2,961 misclassified as low-risk. For the 8,601 low-risk customers, 6,773 were correctly predicted, while 1,828 were incorrectly flagged as high-risk. This demonstrates that XGBoost is highly sensitive to detecting at-risk customers, a critical requirement for proactive churn prevention strategies.

XGBoost also provides valuable insights into the key predictors of churn risk. Features such as *Marital_Status* (Single and Widowed), *Is_ATM_Active*, *Is_USSD_Active*, and *Has_Mobile_Banking* ranked highest in importance, highlighting the role of both demographic and engagement channel variables in customer retention. Financial behavior variables such as *Credit Transactions Last 3 Months*, *Debit Transactions Last 3 Months*, *TXN_Count*, and *Transaction Recency* were also influential, indicating that churn risk is strongly tied to both activity level and transactional patterns.

Overall, XGBoost not only achieved the strongest predictive performance but also produced interpretable results that align with the objectives of churn risk modeling. Its ability to accurately identify high-risk customers while providing transparency into the driving factors makes it particularly valuable for explainable, actionable banking applications.

```
XGBoost Accuracy: 0.9223308843802203
XGBoost Classification Report:
              precision    recall  f1-score   support

           0       0.70      0.79      0.74      8601
           1       0.96      0.94      0.95     53058

    accuracy                           0.92     61659
   macro avg       0.83      0.87      0.85     61659
weighted avg       0.93      0.92      0.92     61659

XGBoost Confusion Matrix:
[[ 6773  1828]
 [ 2961 50097]]
```

Figure 4.6: XGBoost performance evaluation

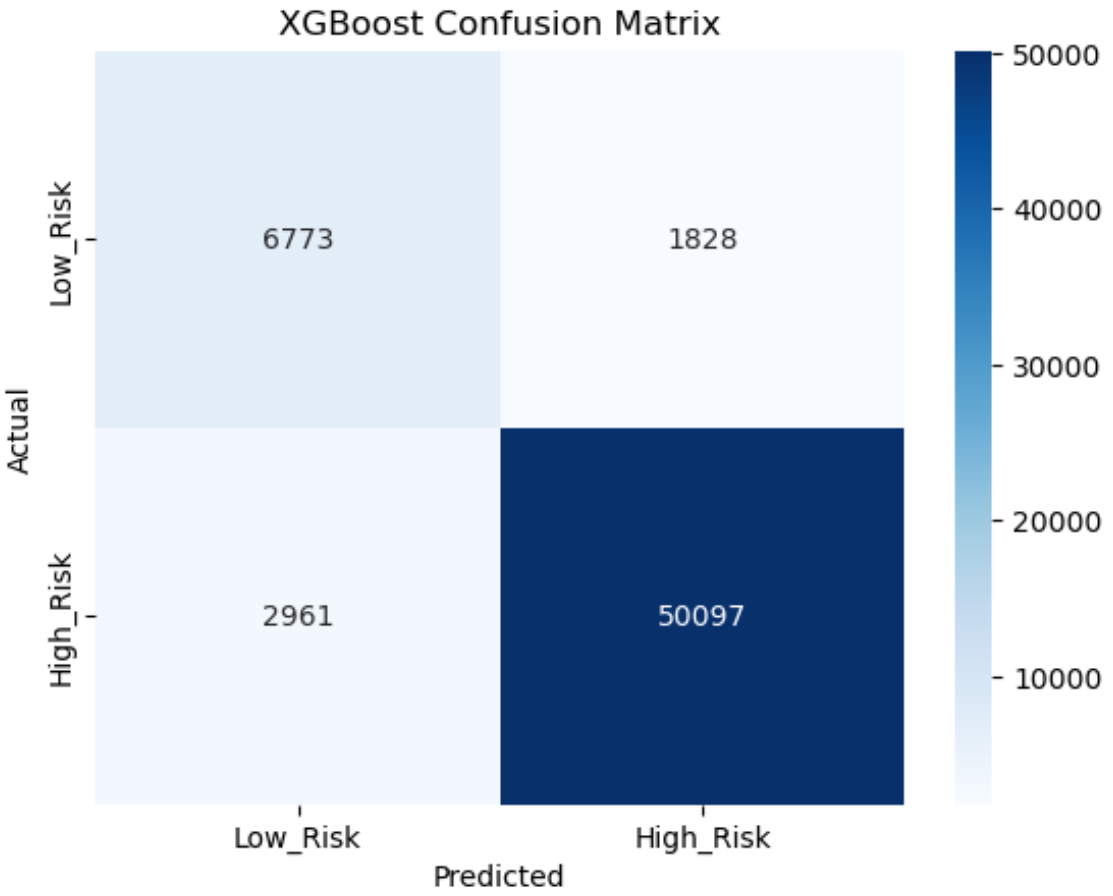


Figure 4.7: XGBoost Confusion Matrix

Experiment 4: MLP Classifier As shown in Figure 4.7 , the Multilayer Perceptron (MLP) model achieved an accuracy of 90.01%, which is slightly lower than

XGBoost, Gradient Boosting, and Random Forest. The model demonstrated strong recall for the churn (high-risk) class, correctly identifying 47,991 out of 53,058 at-risk customers, with 5,067 misclassified as low-risk. For the 8,601 low-risk customers, the model correctly classified 7,506 but incorrectly labeled 1,095 as high-risk.

The classification report shows that the High_Risk class attained a precision of 98%, recall of 90%, and F1-score of 94%, confirming the model's ability to reliably capture high-risk customers. However, the precision for the non-churn class was lower at 60%, reflecting a higher false-positive rate compared to tree-based models. This means that while MLP is effective at minimizing false negatives (critical for retaining high-risk customers), it also flags more Low_Risk customers incorrectly, which could lead to additional retention costs.

```
MLP Test Accuracy: 0.9000632511068943

Confusion Matrix (MLP):
[[ 7506  1095]
 [ 5067 47991]]

Classification Report (MLP):
              precision    recall  f1-score   support

           0       0.60      0.87      0.71      8601
           1       0.98      0.90      0.94     53058

    accuracy                           0.90     61659
   macro avg       0.79      0.89      0.82     61659
weighted avg       0.92      0.90      0.91     61659
```

Figure 4.8: MLP performance evaluation

Experiment 5: LightGBM As shown in the figure below, the LightGBM model achieved a test accuracy of 89.96% and a strong AUC of 0.97, highlighting its capability to discriminate between high-risk and low-risk customers. The confusion matrix indicates that 47,611 high-risk (churn) customers were correctly identified, with 5,447 misclassified as low risk. On the other hand, 7,858 low-risk customers were correctly classified, while 743 were incorrectly flagged as high risk.

The classification report further confirms this performance, with precision of 0.98

and recall of 0.90 for high-risk customers, and a recall of 0.91 for the low-risk group. These results demonstrate that LightGBM effectively prioritizes recall for high-risk customers, ensuring that most at-risk individuals are identified, even if this comes at the cost of slightly higher false positives.

```
Confusion Matrix (LightGBM):
[[ 7858   743]
 [ 5447 47611]]

Classification Report (LightGBM):
              precision    recall  f1-score   support

           0       0.59      0.91      0.72      8601
           1       0.98      0.90      0.94     53058

    accuracy                           0.90     61659
   macro avg       0.79      0.91      0.83     61659
weighted avg       0.93      0.90      0.91     61659
```

Figure 4.9: LightGBM performance evaluation

Experiment 6: Extra Trees Classifier As shown in the figure below, the Extra Trees model achieved a test accuracy of 91.01% with an AUC of 0.96, demonstrating strong predictive capability for identifying High_Risk. The confusion matrix reveals that 49,377 high-risk customers were correctly identified, with 3,681 misclassified as low risk. Conversely, 6,743 low-risk customers were correctly classified, while 1,858 were incorrectly flagged as high risk.

The classification report supports these findings, with a precision of 0.96 and recall of 0.93 for high-risk customers, and precision of 0.65 with recall of 0.78 for the low-risk group. This balance indicates that the model is highly effective at detecting customers most likely to High_Risk, even if it results in a moderate rate of false positives.

```
Extra Trees Test Accuracy: 0.9101672099774567

Confusion Matrix (Extra Trees):
[[ 6743  1858]
 [ 3681 49377]]

Classification Report (Extra Trees):
              precision    recall  f1-score   support

           0       0.65      0.78      0.71      8601
           1       0.96      0.93      0.95     53058

    accuracy                           0.91     61659
   macro avg       0.81      0.86      0.83     61659
weighted avg       0.92      0.91      0.91     61659
```

Figure 4.10: Extra Trees performance evaluation

4.4 Model Comparison

As stated by [102], comparing models is an important part of the construction and evaluation of machine learning models. It helps researchers determine which algorithm works best for a specific problem. To do this, they train several models using the same data set and measure their performance using metrics such as accuracy, precision, recall, and the F1 score. In the end, this process helps choose the most suitable algorithm, leading to better accuracy and performance. In this study, six experiments were conducted to develop a model for churn prediction and compare different algorithms to find the best one.

Table 4.8: Performance comparison of machine learning models for risk factor prediction

Evaluation Metric	Random Forest	Gradient Boosting	XGBoost	LightGBM	Extra Trees	MLP
Accuracy	91.81%	89.52%	92.23%	89.96%	91.02%	90.01%
Precision	0.96	0.98	0.96	0.98	0.96	0.98
Recall	0.94	0.89	0.94	0.90	0.93	0.90
F1-score	0.95	0.94	0.95	0.94	0.95	0.94
AUC	0.97	0.97	0.97	0.97	0.96	0.96

As shown in Table 4.8, XGBoost achieved the highest accuracy (92.23%) among all the evaluated classifiers, indicating it is the most effective at correctly predicting high-risk and low-risk customers. Random Forest (91.81%) and Extra Trees

(91.02%) follow closely, while Gradient Boosting (89.52%) and LightGBM (89.96%) show slightly lower accuracy.

Precision values were high across all models (0.96–0.98), demonstrating that the models reliably identify high-risk customers without many false positives. Recall, which measures the model's ability to detect actual high-risk customers, was also strong for XGBoost and Random Forest (0.94), indicating their effectiveness in capturing churn cases. F1-scores further confirm balanced performance between precision and recall, with XGBoost and Random Forest achieving the highest scores (0.95).

All models reported similarly high AUC values (0.96–0.97), highlighting their ability to distinguish between high-risk and low-risk customers. Taken together, these results indicate that XGBoost is the top-performing model for predicting customer churn risk factors, making it highly suitable for deployment in practical banking scenarios.

4.5 Model Explainability

In this study, explainable AI (XAI) techniques were employed to enhance the interpretability of the XGBoost model's predictions for customer churn risk. Understanding which features drive the model's decisions is essential for identifying high-risk and low-risk customers and informing targeted retention strategies.

Using SHAP (SHapley Additive exPlanations), we quantified the contribution of each feature to the predicted churn risk. This approach allows stakeholders to see how factors such as transaction volume, account activity, and customer demographics influence the likelihood of a customer being classified as high-risk. By providing transparency into the model's reasoning, SHAP supports trust in the predictions and enables actionable insights for business decision-making.

Integrating XAI with XGBoost ensures that the risk prediction model is not only accurate but also interpretable, helping banks understand why certain customers are at higher risk of churn and how interventions can be prioritized effectively.

4.5.1 XGBoost Explaibability Using SHAP Method:

SHAP (Shapley Additive Explanations) is a model-agnostic XAI method that provides both local and global interpretability by calculating Shapley values, which are based on game theory [62]. SHAP explains how each feature contributes to the model's prediction by measuring the contribution of individual features across all possible feature combinations. This method is particularly useful for understanding the global behavior of the model by ranking features according to their importance in making predictions.

For predicting risk factors of customer churn, SHAP provides a clear understanding of which features most strongly influence the model's predictions. For instance, SHAP dependence plots reveal how individual feature values—such as transaction recency, debit/credit transactions in the last three months, or account activity—affect the likelihood of a customer being classified as high-risk. Customers with low transaction frequency or inactive engagement channels may have higher predicted churn risk, as indicated by the SHAP values.

SHAP also offers local explanations, breaking down predictions for individual customers. For example, the model might indicate that a customer's low mobile banking usage and recent drop in transaction count significantly increase their churn risk. Waterfall plots can visualize this process, showing step by step how each feature's value contributes to the final prediction, allowing banks to identify actionable risk factors for proactive retention strategies. In Figure 4.10 below, the SHAP local waterfall plot illustrates the contribution of each feature to the final prediction of the churn risk class, with features ordered by their impact. Positive contributions (highlighted in red) increase the likelihood of customer churn, whereas negative contributions (highlighted in blue) decrease the likelihood of churn. This provides a clear explanation of which features act as risk enhancers versus risk mitigators for churn in this instance.

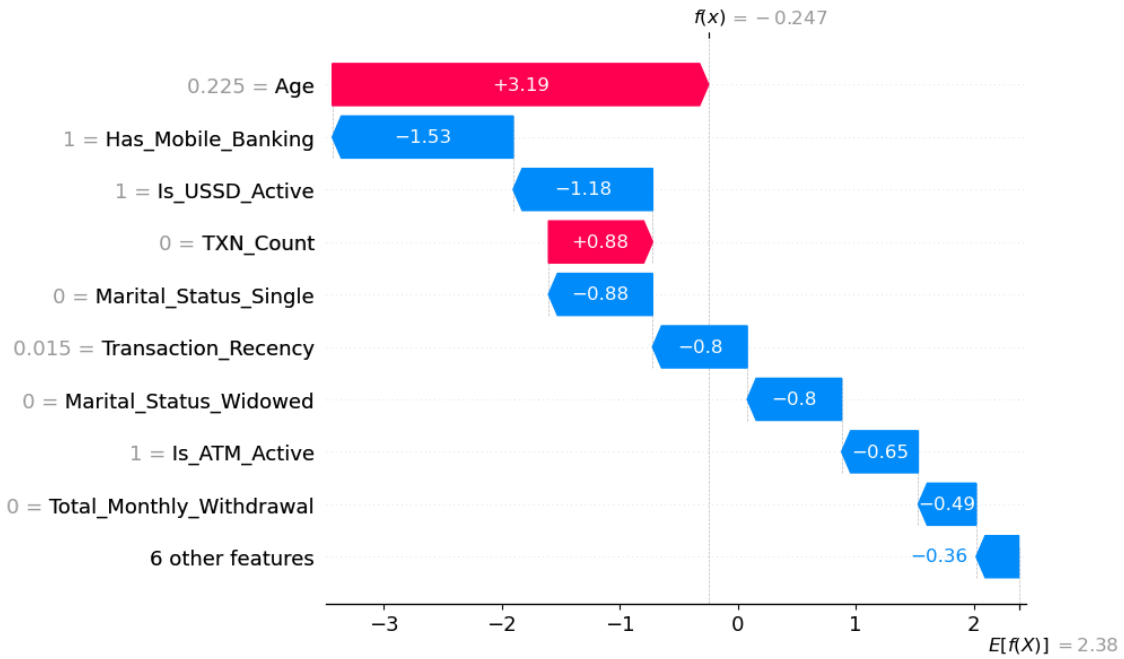


Figure 4.11: SHAP local waterfall plot 512th instance for low risk class

The analysis shows that Age (+3.19) is the most significant positive contributor to churn risk, followed by Transaction Count (+0.88). This suggests that younger customers with higher transaction activity exhibit stronger churn signals. These demographic and activity-related factors can be interpreted as key drivers of churn behavior, making them high-risk indicators in customer retention analysis.

Conversely, several features act as strong protective factors against churn. These include Has Mobile Banking (-1.53), Is USSD Active (-1.18), Marital Status – Single (-0.88), Transaction Recency (-0.80), Marital Status – Widowed (-0.80), and Is ATM Active (-0.65). The negative contributions of these features imply that customers actively using digital channels such as mobile banking, USSD services, and ATMs, or those who have made recent transactions, are less likely to churn. Such engagement-related features therefore play a critical role in customer retention strategies.

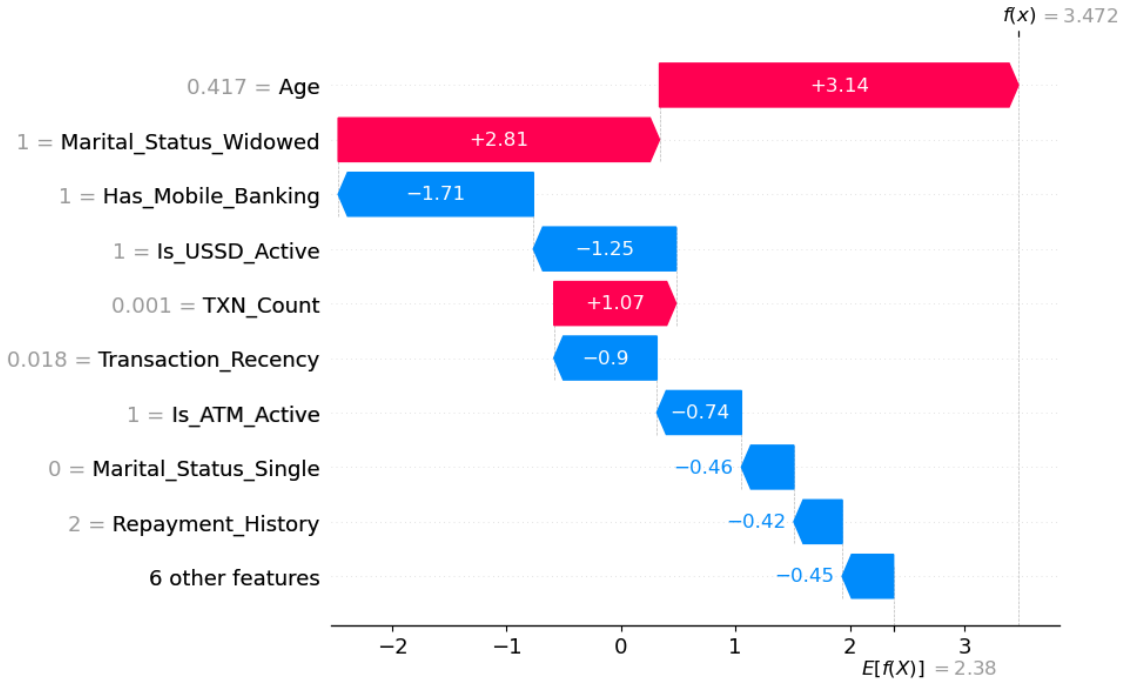


Figure 4.12: SHAP local waterfall plot 319th instance for High risk class

The overall prediction of this instance shifts from the baseline value of E[f(X)] = 2.38 to the final output of f(x) = -0.247, driven by the cumulative effect of these positive and negative feature contributions. Although Age substantially increases churn risk, this is offset by the strong mitigating effects of mobile and USSD banking activity, ultimately lowering the predicted churn probability for this customer.

From a managerial perspective, this analysis emphasizes the importance of digital engagement as a protective factor against churn, while also highlighting the increased vulnerability of certain demographic segments such as younger customers. Targeted retention campaigns could therefore focus on promoting mobile banking adoption and encouraging recent activity, particularly among high-risk groups.

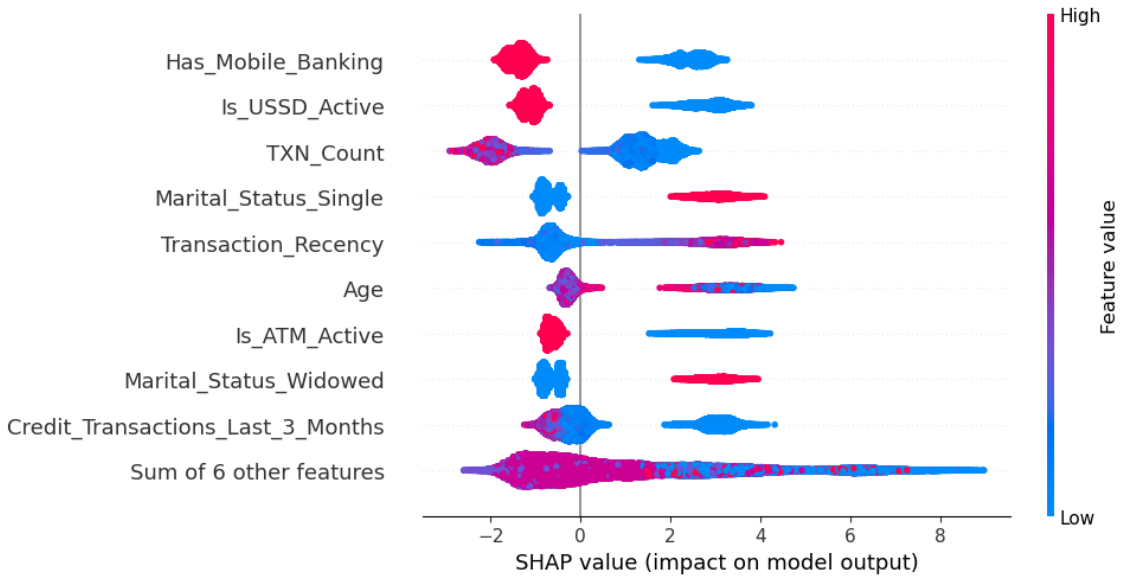


Figure 4.13: global explanation

The Beeswarm plot in Figure 4.13 presents the global explanation of the XGBoost model for predicting risk factors of customer churn. The figure ranks the features in descending order of importance, while also showing how both higher and lower values of each feature influence the likelihood of churn. Each dot represents a single customer, with the position on the x-axis determined by its SHAP value, which quantifies the feature's contribution to churn prediction. The color gradient distinguishes feature values, where red indicates higher feature values and blue indicates lower values. Positive SHAP values correspond to an increased risk of churn, whereas negative SHAP values reduce the churn risk.

From the analysis, Has Mobile Banking, Is USSD Active, and Transaction Count emerged as the most influential risk factors. Customers who lack mobile banking usage (blue values) or do not engage with USSD services (blue values) were strongly associated with increased churn risk, highlighting the protective role of digital banking adoption. Conversely, customers who actively use these channels (red values) showed reduced churn risk. Transaction Count demonstrates a mixed effect: in certain cases, high transaction activity (red values) elevated churn risk, while in others, lower activity (blue values) was associated with reduced churn.

Other important features include Transaction Recency, Age, ATM Activity, and Marital Status (Single and Widowed). Customers with recent transactions (blue

values in Transaction Recency) show a lower likelihood of churn, reinforcing the view that sustained engagement with banking services reduces churn risk. Younger customers (red values in Age), on the other hand, exhibit stronger churn tendencies, suggesting that demographic groups are not equally stable in their banking relationships. Similarly, ATM activity (red values) reduces churn risk, while inactivity increases the risk of churn.

Moderate but relevant contributions are observed from features such as Credit Transactions in the Last 3 Months, Repayment History, Debit Transactions, and Total Monthly Withdrawal, which capture financial behavior and repayment consistency. These variables influence customer loyalty, with low transaction intensity or poor repayment history linked to heightened churn risk. Features such as Account Age, Location, and Account Type (Youth & Diaspora Accounts) appear less influential, as indicated by their smaller SHAP values, but they still exert context-dependent effects on churn risk.

The vertical dispersion of the dots for several features, particularly Transaction Count and Age, indicates heterogeneity in their effects across different customers. This suggests that churn risk is highly individualized, and while some features are strong overall predictors, their impact varies at the customer level.

In conclusion, the global SHAP analysis highlights that digital engagement (mobile banking, USSD, ATM use), transaction recency, and age are the most critical risk factors driving churn predictions in the XGBoost model. This reinforces the importance of designing retention strategies aimed at customers with low digital activity and younger demographic profiles, who represent the highest churn risk segment.

4.5.2 XGBoost Explaibability Using LIME Method:

LIME is another XAI technique that focuses on local explanations. LIME works by generating perturbations of the input features and fitting an interpretable model (e.g., linear regression) to approximate the black-box model locally [62]. Unlike SHAP, which attributes importance to each feature globally, LIME focuses on providing an explanation for a specific instance. For churn prediction, LIME can explain

why a particular customer is predicted to churn by showing how each feature contributed to the final prediction. For example, LIME may show that a customer's high number of customer support calls and low monthly bill are significant factors in predicting their churn risk. The visualization typically includes a breakdown of the feature coefficients in the surrogate linear model, highlighting the most important features for the specific instance.

In this research, we utilized the Local Interpretable Model-agnostic Explanations (LIME) framework to interpret a machine learning model's prediction on risk factor for customer churn. The goal was to understand the factors contributing to the prediction for a specific customer instance, which can offer valuable insights into why the model classifies certain customers as "High_Risk" or "Low_Risk." A LIME explainer was initialized using the training data (X train) and the trained model (xgb model). The explainer is configured to handle classification problems, with the target labels being "Low_Risk" (0) and "High_Risk" (1).

Local Interpretable Model-Agnostic Explanations (LIME) is used to explain the predictions of complex machine learning models such as XGBoost, a powerful gradient boosting framework. While XGBoost achieves high predictive accuracy for customer churn, it is often considered a "black-box" model, meaning that the internal decision-making process is not easily interpretable. LIME provides local, instance-level explanations by approximating the model with an interpretable surrogate around individual predictions.

4.6 Low-Risk Customer (Instance 512)

Figure 4.14 shows the feature contributions for predicting the **Low_Risk** class. The feature with the highest positive contribution is `Age <= 0.27`, indicating that younger customers are more likely to be classified as low risk. Other significant positive contributors include `TXN_Count <= 0.00` and `Debit_Transactions_Last_3_Months <= 0.00`. Negative contributors include Is USSD Active, Has Mobile Banking, Marital Status Single, Marital Status Widowed, and Repayment History.

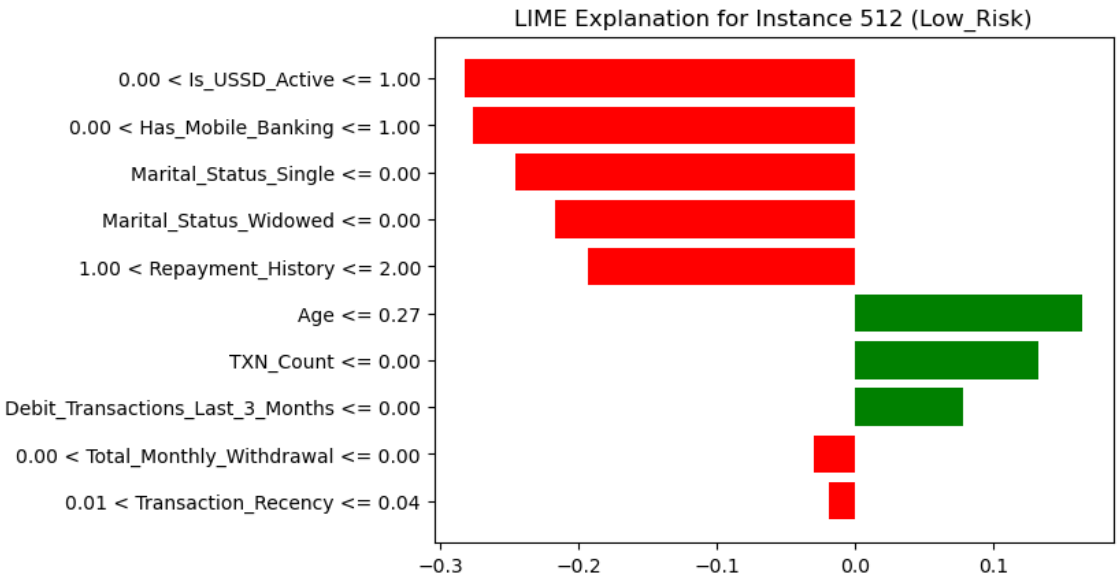


Figure 4.14: Feature contributions for Low-Risk customer (Instance 512).

Table 4.9: Key feature contributions for Low-Risk instance 512

Feature	**Contribution Type**	**Weight**
Age ¡= 0.27	Positive	0.165
TXN_Count ¡= 0.00	Positive	0.133
Debit_Transactions_Last_3_Months ¡= 0.00	Positive	0.079
Is_USSD_Active	Negative	-0.282
Has_Mobile_Banking	Negative	-0.276
Marital_Status_Single	Negative	-0.246
Marital_Status_Widowed	Negative	-0.217
Repayment_History	Negative	-0.193
Total_Monthly_Withdrawal	Negative	-0.030
Transaction_Recency	Negative	-0.019

The predicted probability for **Low_Risk** is 0.56, while for **High_Risk** it is 0.44, indicating moderate confidence in this prediction.

4.7 High-Risk Customer (Instance 319)

Figure 4.15 shows the feature contributions for predicting the **High_Risk** class. The highest positive contributor is `Marital_Status_Widowed > 0.00`, indicating widowed customers are more likely to churn. Other positive contributors include `TXN_Count <= 0.00`, `Age > 0.34`, and `Account_Age between 0.19 and 0.34`. Negative contributors include `Has_Mobile_Banking`, `Is_USSD_Active`, `Marital_Status_Single`, `Repayment_History`, and transaction recency.

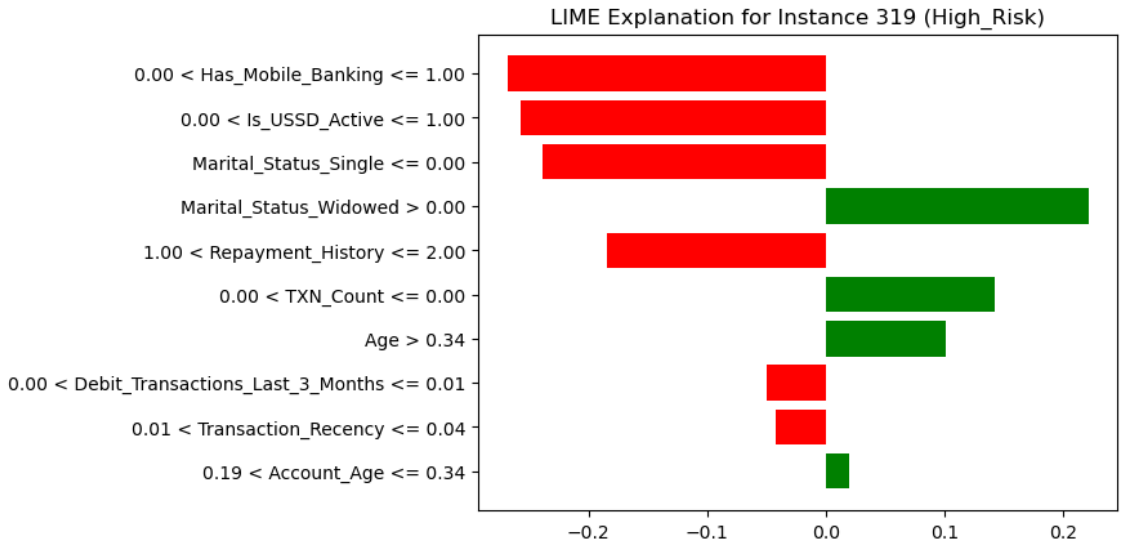


Figure 4.15: Feature contributions for High-Risk customer (Instance 319).

Table 4.10: Key feature contributions for High-Risk instance 319

Feature	**Contribution Type**	**Weight**
Marital_Status_Widowed ¿ 0.00	Positive	0.221
TXN_Count ¡= 0.00	Positive	0.142
Age ¿ 0.34	Positive	0.101
0.19 ¡ Account_Age ¡= 0.34	Positive	0.019
Has_Mobile_Banking	Negative	-0.268
Is_USSD_Active	Negative	-0.256
Marital_Status_Single	Negative	-0.239
Repayment_History	Negative	-0.185
Debit_Transactions_Last_3_Months	Negative	-0.050
Transaction_Recency	Negative	-0.042

The predicted probability for **High_Risk** is 0.97, while for **Low_Risk** it is 0.03, showing very strong confidence in this prediction.

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

Customer churn poses a critical challenge to financial institutions, as retaining existing customers is often more cost-effective than acquiring new ones. Predicting risk factors for customer churn enables organizations to proactively identify customers at risk of leaving and implement targeted retention strategies. This research focused on developing a model for predicting customer churn risk factors using machine learning and explainable artificial intelligence (XAI) methods.

A dataset consisting of 308,293 customer records with 20 attributes was used, integrating transactional behavior, demographic information, and account activity. To improve model performance, data preprocessing techniques were applied, including data cleaning, transformation, feature selection, balancing, and dataset partitioning. Several machine learning algorithms were implemented and evaluated, including Random Forest, Gradient Boosting, XGBoost, LightGBM, Extra Trees, and Multilayer Perceptron (MLP).

This research made significant contributions by addressing two key research questions:

RQ1: What are the most determinant attributes for customer churn? Through a hybrid feature selection approach combined with SHAP (SHapley Additive Explanations) analysis, the study identified the most influential predictors of churn. Features such as Has Mobile Banking, Is USSD Active, Transaction Count, Transaction Recency, Age, and ATM Activity emerged as the strongest risk factors. Customers

lacking mobile or USSD activity, those with infrequent transactions, or younger customers were more likely to churn. Conversely, active use of digital banking channels and recent transactions acted as protective factors against churn. Additional features, such as repayment history, income, and account balance, provided further insights into financial behavior influencing churn risk.

RQ2: Which algorithm is best suited for predicting customer churn risk factors? Among all models tested, the XGBoost classifier achieved the highest accuracy (92.23%), outperforming Random Forest (91.81%), Extra Trees (91.02%), LightGBM (89.96%), Gradient Boosting (89.52%), and MLP (90.01%). XGBoost also demonstrated superior balance across precision (0.96), recall (0.94), F1-score (0.95), and AUC (0.97), making it the most effective model for predicting customer churn risk factors in this context.

To enhance interpretability, SHAP and LIME were employed to explain model predictions at both global and local levels. These explainable AI methods allowed stakeholders to understand how specific features influenced the classification of customers as high-risk or low-risk, thereby building trust in the model and enabling actionable insights for churn management. Overall, the study demonstrated that integrating predictive modeling with explainable AI provides not only high predictive accuracy but also transparency and interpretability, which are essential for deploying machine learning in real-world banking scenarios.

Future Work Future work should focus on expanding the dataset to incorporate additional behavioral and psychographic attributes such as customer satisfaction indices, service usage patterns, and sentiment analysis from customer feedback. these features can provide a more holistic understanding of customer behavior and improve the robustness of churn risk factor prediction. Moreover, exploring advanced learning paradigms such as reinforcement learning could enable the design of adaptive and personalized retention strategies, where the model not only predicts churn risk factors but also learns optimal intervention policies over time. Such approaches would bridge predictive modeling with prescriptive analytics, thereby offering more practical and strategic value for financial institutions.

Bibliography

[1] M. Alizadeh, D. S. Zadeh, B. Moshiri, and A. Montazeri, "Development of a customer churn model for banking industry based on hard and soft data fusion," *IEEE Access*, vol. 11, pp. 29 759–29 768, 2023.

[2] M. H. Seid 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)*, Bahir Dar, Ethiopia, 2022, pp. 1–6.

[3] A. Saxena, A. Singh, and G. M., "Analyzing customer churn in banking: A data mining framework," *Multidiscip. Sci. J.*, vol. 5, p. 2023ss0310, 2023.

[4] D. I. Durojaye and G. Obunadike, "Analysis and visualization of market segmentation in banking sector using kmeans machine learning algorithm," *Fudma J. Sci.*, vol. 6, no. 1, pp. 387–393, Apr 2022.

[5] R. S. Subramanian, J. Aswini, S. N. Bushra, G. Nirmala, and B. Maheswari, "Customer analysis using machine learning with feature selection approaches: A comparative study," in *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Trichy, India, Nov 2022, pp. 196–202.

[6] S. Kumar, S. Datta, V. Singh, S. K. Singh, and R. Sharma, "Opportunities and challenges in data-centric ai," *IEEE Access*, vol. 12, pp. 33 173–33 189, 2024.

[7] R. O. Panchariya, "Ai-enabled customer insights: Transforming marketing and customer relationship management in modern era."

[8] A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A review on machine learning methods for customer churn prediction and recommendations for business practitioners," *IEEE Access*, vol. 12, pp. 70 434–70 463, 2024.

[9] J. L. O. W. Kit, V. S. Asirvadam, and M. F. Hassan, "Enhanced ensemble models for predictive modeling: A conceptual framework," in *2021 IEEE 17th International Colloquium on Signal Processing & Its Applications (CSPA)*, Langkawi, Malaysia, Mar 2021, pp. 24–28.

[10] T. Muhammad, S. Ness, M. Volkivskyi, and Y. Gong, "Exploring ai and machine learning applications in banking: A comprehensive review of literature," Feb 2024.

[11] S. Kalyani and N. Gupta, "Is artificial intelligence and machine learning changing the ways of banking: a systematic literature review and meta analysis," *Discov. Artif. Intell.*, vol. 3, no. 1, p. 41, Dec 2023.

[12] NBE, "Financial-stability-report_nov2024," www.nbe.gov.et, Nov 2024.

[13] A. Habibi, M. R. Delavar, M. S. Sadeghian, B. Nazari, and S. Pirasteh, "A hybrid of ensemble machine learning models with rfe and boruta wrapper-based algorithms for flash flood susceptibility assessment," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 122, p. 103401, 2023.

[14] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated churn prediction and customer segmentation framework for telco business," *IEEE Access*, vol. 9, pp. 62 118–62 136, 2021.

[15] S. Goundar, "Research methodology and research method," https://www.researchgate.net/publication/333015026, 2012.

[16] J. Kamiri and G. Mariga, "Research methods in machine learning: A content analysis," *Int. J. Comput. Inf. Technol.-0764*, vol. 10, no. 2, 2021.

[17] F. S. Mishkin and S. Eakins, *Financial Markets and Institutions: Global Edition*, 9th ed. Pearson, 2018.

[18] F. Allen and E. Carletti, "The roles of banks in financial systems," in *Oxf. Handb. Bank.*, 2012.

[19] N. A. Faisal, J. Nahar, M. Waliullah, and R. S. Borna, "The role of digital banking features in bank selection an analysis of customer preferences for online and mobile banking," *Innov. Eng. J.*, vol. 1, no. 01, pp. 41–58, 2024.

[20] N. S. Egbuhuzor, A. J. Ajayi, E. E. Akhigbe, O. O. Agbede, C. P.-M. Ewim, and D. I. Ajiga, "Ai and data-driven insights: Transforming customer relationship management (crm) in financial services," *Gulf J. Adv. Bus. Res.*, vol. 3, no. 2, pp. 483–511, 2025.

[21] A. U. Khan and E. A. Khan, "2 publications 7 citations see profile," 2021.

[22] T. I. Ijomah, U. S. Nwabekee, E. E. Agu, and O. Y. Abdul-Azeez, "The impact of customer relationship management (crm) tools on sales growth and customer loyalty in emerging markets," *Int. J. Manag. Entrep. Res.*, vol. 6, no. 9, pp. 2964–2988, 2024.

[23] A. Parvatiyar and J. N. Sheth, "Customer relationship management: Emerging practice, process, and discipline."

[24] R. Nandal, "Application of data mining in customer relationship management: A review," vol. 5, no. 1, 2018.

[25] M. A. Camilleri, "The use of data-driven technologies for customer-centric marketing," *Int. J. Big Data Manag.*, vol. 1, no. 1, p. 50, 2020.

[26] W. Reinartz, M. Krafft, and W. D. Hoyer, "The customer relationship management process: Its measurement and impact on performance," *J. Mark. Res.*, vol. 41, no. 3, pp. 293–305, 2004.

[27] S. Nasır, "Customer relationship management strategies in the digital era," in *Advances in Marketing, Customer Relationship Management, and E-Services*. IGI Global, 2015.

[28] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results Control Optim.*, vol. 14, p. 100342, 2024.

[29] M. H. Seid and M. M. Woldeyohannis, "Customer churn prediction using machine learning: Commercial bank of ethiopia," in *EAI Int. Conf. ICT Dev. Afr.*, 2022.

[30] H. D. Tran, N. Le, and V.-H. Nguyen, "Customer churn prediction in the banking sector using machine learning-based classification models," *Interdiscip. J. Inf. Knowl. Manag.*, vol. 18, pp. 087–105, 2023.

[31] Y. Liu, "Bank customer churn prediction using machine learning."

[32] F. Tajdini and S. Bayat, "Machine learning algorithms - a review."

[33] H. M. E. Misilmani and T. Naous, "Machine learning in antenna design: An overview on machine learning concept and algorithms," in *2019 International Conference on High Performance Computing Simulation (HPCS)*, Dublin, Ireland, 2019, pp. 600–607.

[34] N. Dutta, S. Umashankar, V. K. A. Shankar, S. Padmanaban, Z. Leonowicz, and P. Wheeler, "Centrifugal pump cavitation detection using machine learning algorithm technique," in *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe).* Palermo, Italy: IEEE, Jun 2018, pp. 1–6.

[35] H. K. Albahadily, M. Q. Mohammed, and A. K. Mohammed, "A practical guide of machine learning algorithms and applications," *Int. J. Appl. Inf. Syst.*, vol. 12, no. 40, pp. 8–13, 2023.

[36] P. Xu, "Review on studies of machine learning algorithms," *J. Phys. Conf. Ser.*, vol. 1187, no. 5, p. 052103, 2019.

[37] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, 2021.

[38] R. T. Potla and V. K. Pottla, "Artificial intelligence and machine learning in crm - leveraging data for predictive analytics," https://www.researchgate.net/publication/388177970, 2024.

[39] W. Kuryłek, "Artificial neural networks and gradient-boosting decision trees in time series forecasting of earnin."

[40] I. A. Adeniran, C. P. Efunniyi, O. S. Osundare, and A. O. Abhulimen, "Implementing machine learning techniques for customer retention and churn prediction in telecommunications," *Comput. Sci. IT Res. J.*, vol. 5, no. 8, pp. 2011–2025, 2024.

[41] S. Shumaly, P. Neysaryan, and Y. Guo, "Handling class imbalance in customer churn prediction in telecom sector using sampling techniques, bagging and boosting trees," in *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, 2020, pp. 082–087.

[42] M. Shah, K. Gandhi, K. A. Patel, H. Kantawala, R. Patel, and A. Kothari, "Theoretical evaluation of ensemble machine learning techniques," in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2023, pp. 829–837.

[43] A. F. Jadama and M. K. Toray, "Ensemble learning: Methods, techniques, application," 2024.

[44] B. T. Mashi, M. Hamada, J. J. Tanimu, P. Robert, and T. J. Samson, "An ensemble approach for stroke prediction," in *2024 IEEE 17th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, Kuala Lumpur, Malaysia, 2024, pp. 381–388.

[45] F. Ehsani and M. Hosseini, "Customer churn prediction using a novel meta-classifier: an investigation on transaction, telecommunication and customer churn datasets," *J. Comb. Optim.*, vol. 48, no. 1, p. 7, 2024.

[46] T. Asfaw, "Customer churn prediction using machine-learning techniques in the case of commercial bank of ethiopia," *Sci. Temper*, vol. 14, no. 03, pp. 618–624, 2023.

[47] D. Boldini, F. Grisoni, D. Kuhn, L. Friedrich, and S. A. Sieber, "Practical guidelines for the use of gradient boosting for molecular property prediction," *J. Cheminformatics*, vol. 15, no. 1, p. 73, 2023.

[48] S. Fatima, A. Hussain, S. B. Amir, S. H. Ahmed, and S. M. H. Aslam, "Xgboost and random forest algorithms: An in-depth analysis," *Unknown Journal*.

[49] F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis, "A novel approach to estimate the weight of food items based on features extracted from an image using boosting algorithms," *Sci. Rep.*, vol. 13, no. 1, p. 21040, 2023.

[50] K. Peng and Y. Peng, "Research on telecom customer churn prediction based on ga-xgboost and shap," *J. Comput. Commun.*, vol. 10, no. 11, pp. 107–120, 2022.

[51] J. Zhang, D. Mucs, U. Norinder, and F. Svensson, "Lightgbm: An effective and scalable algorithm for prediction of chemical toxicity-application to the tox21 and mutagenicity data sets," *J. Chem. Inf. Model.*, vol. 59, no. 10, pp. 4278–4286, Sep. 2019.

[52] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. Appl.*, vol. 244, p. 122778, Jun. 2024.

[53] M. A. Muslim, Y. Dasril, M. Sam'an, and Y. N. Ifriza, "An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, p. 1002, 2022.

[54] A. Hussen, T. A. Munshi, L. N. Jahan, and M. Hashan, "Advanced machine learning approaches for predicting permeability in reservoir pay zones based on core analyses," *Heliyon*, vol. 10, no. 12, p. e32666, 2024.

[55] D. Melanson, "Extremely randomized trees with multiparty computation," 2024.

[56] K. Y. C. et al., "Deep neural networks in the cloud: Review, applications, challenges and research directions," *Neurocomputing*, vol. 545, p. 126327, 2023.

[57] M. M. A. Mia, S. K. Biswas, M. C. Urmi, and A. Siddique, "An algorithm for training multilayer perceptron (mlp) for image reconstruction using neural network without overfitting," vol. 4, no. 02, 2015.

[58] S. A. et al., "Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence," *Inf. Fusion*, vol. 99, p. 101805, 2023.

[59] A. M. S. et al., "A perspective on explainable artificial intelligence methods: Shap and lime," *Adv. Intell. Syst.*, vol. 7, no. 1, p. 2400304, 2025.

[60] M. Ridley, "Explainable artificial intelligence (xai): Adoption and advocacy," *Inf. Technol. Libr.*, vol. 41, no. 2, 2022.

[61] O. Oladele, "Explainable ai techniques for market prediction models."

[62] M. Panda and S. R. Mahanta, "Explainable artificial intelligence for healthcare applications using random forest classifier with lime and shap," in *Explainable, Interpretable, and Transparent AI Systems*, 1st ed. CRC Press, 2024, pp. 89–105.

[63] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis.

[64] M. Miró-Nicolau, A. J. i Capó, and G. Moyà-Alcover, "A comprehensive study on fidelity metrics for xai," *Information Processing & Management*, vol. 62, no. 1, p. 103900, 2025.

[65] L. Coroama and A. Groza, "Evaluation metrics in explainable artificial intelligence (xai)," in *Advanced Research in Technologies, Information, Innovation and Sustainability*, ser. Communications in Computer and Information Science, T. Guarda, F. Portela, and M. F. Augusto, Eds. Cham: Springer Nature Switzerland, 2022, vol. 1675, pp. 401–413.

[66] K. N. Pellano, I. Strümke, and E. A. F. Ihlen, "From movements to metrics: Evaluating explainable ai methods in skeleton-based human activity recognition," *Sensors*, vol. 24, no. 6, p. 1940, Mar 2024.

[67] K. Kalasampath, K. N. Spoorthi, S. Sajeev, S. S. Kuppa, K. Ajay, and A. Maruthamuthu, "A literature review on applications of explainable artificial intelligence (xai)," *IEEE Access*, vol. 13, pp. 41 111–41 140, 2025.

[68] S. S. Band *et al.*, "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," *Inform. Med. Unlocked*, vol. 40, p. 101286, 2023.

[69] M. Saarela and V. Podgorelec, "Recent applications of explainable ai (xai): A systematic literature review," *Applied Sciences*, vol. 14, no. 19, p. 8884, 2024.

[70] E. D. Kingawa and T. T. Hailu, "Customer churn prediction using machine learning techniques: the case of lion insurance," *Asian J. Basic Sci. Res.*, vol. 4, no. 4, pp. 60–73, 2022.

[71] R. Liu *et al.*, "An intelligent hybrid scheme for customer churn prediction integrating clustering and classification algorithms," *Appl. Sci.*, vol. 12, no. 18, p. 9355, 2022.

[72] B. G. Seyoum, "Bank customer churn prediction model: The case of commercial bank of ethiopia."

[73] Y. Bharambe, P. Deshmukh, P. Karanjawane, D. Chaudhari, and N. M. Ranjan, "Churn prediction in telecommunication industry," in *2023 International Conference for Advancement in Technology (ICONAT)*. IEEE, 2023, pp. 1–5.

[74] N. T. M. Sagala and S. D. Permai, "Enhanced churn prediction model with boosted trees algorithms in the banking sector," in *2021 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, 2021, pp. 240–245.

[75] I. N. M. Adiputra and P. Wanchai, "Customer churn prediction using weight average ensemble machine learning model," in *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2023, pp. 90–94.

[76] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Front. Energy Res.*, vol. 9, p. 652801, 2021.

[77] M. Lenzerini, "Data integration: a theoretical perspective," in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2002, pp. 233–246.

[78] M. Kampouridis and F. E. B. Otero, "Using attribute construction to improve the predictability of a gp financial forecasting algorithm," in *2013 Conference on Technologies and Applications of Artificial Intelligence*. IEEE, 2013, pp. 55–60.

[79] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers in Energy Research*, vol. 9, p. 652801, Mar 2021.

[80] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2354–2364, 2011.

[81] A. Akkache *et al.*, "Machine learning-based prediction of immunomodulatory properties of polymers: Toward a faster and easier development of anti-

inflammatory biomaterials," *Adv. NanoBiomed Res.*, vol. 4, no. 3, p. 2300085, 2024.

[82] N. Meghanathan, "Assortativity analysis of real-world network graphs based on centrality metrics," *Comput. Inf. Sci.*, vol. 9, no. 3, p. 7, 2016.

[83] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, p. 52, 2020.

[84] N. A. Nnamoko, F. N. Arshad, D. England, J. Vora, and J. Norman, "Evaluation of filter and wrapper methods for feature selection in supervised machine learning," *???*

[85] X. Cheng, "A comprehensive study of feature selection techniques in machine learning models," *Insights in Computer, Signals and Systems (Ins. Comput. Signal Syst.)*, vol. 1, no. 1, 2024, review. [Online]. Available: https://soapubs.com/index.php/ICSS

[86] A. Abdo, R. Mostafa, and L. Abdel-Hamid, "An optimized hybrid approach for feature selection based on chi-square and particle swarm optimization algorithms," *Data*, vol. 9, no. 2, p. 20, 2024.

[87] J. Sheng, M. Shao, Q. Zhang, R. Zhou, L. Wang, and Y. Xin, "Alzheimer's disease, mild cognitive impairment, and normal aging distinguished by multimodal parcellation and machine learning," *Sci. Rep.*, vol. 10, no. 1, p. 5475, 2020.

[88] D. M. Reif, A. A. Motsinger, B. A. McKinney, J. E. Crowe, and J. H. Moore, "Feature selection using a random forests classifier for the integrated analysis of multiple data types," in *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*. IEEE, 2006, pp. 1–8.

[89] H. Jeon and S. Oh, "Hybrid-recursive feature elimination for efficient feature selection," *Appl. Sci.*, vol. 10, no. 9, p. 3211, 2020.

[90] H. Bichri, A. Chergui, and M. Hain, "Investigating the impact of train / test split ratio on the performance of pre-trained models with custom datasets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, 2024.

[91] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 15, no. 4, pp. 531–538, 2022.

[92] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, 2018. [Online]. Available: https://www.researchgate.net/publication/320442366

[93] J. Jia, W. Liang, and Y. Liang, "A review of hybrid and ensemble in deep learning for natural language processing," 2024, arXiv:2312.05589.

[94] H. A. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," *Babylon. J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024.

[95] L. Breiman, *Random Forests.* Kluwer Acad. Publ., 2001, vol. 45, no. 1.

[96] M. Yousefi, V. Oskoei, H. R. Esmaeli, and M. Baziar, "An innovative combination of extra trees within adaboost for accurate prediction of agricultural water quality indices," *Results Eng.*, vol. 24, p. 103534, 2024.

[97] Željko Đ. Vujović, "Classification model evaluation metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, 2021.

[98] S. Sathyanarayanan, "Confusion matrix-based performance evaluation metrics," *Afr. J. Biomed. Res.*, pp. 4023–4031, 2024.

[99] L. Gonçalves, A. Subtil, M. R. Oliveira, and P. D. Z. Bermudez, "Roc curve estimation," *REVSTAT-Stat. J.*, pp. 1–20, 2014.

[100] S. Yang and G. Berdine, "The receiver operating characteristic (roc) curve," *Southwest Respir. Crit. Care Chron.*, vol. 5, no. 19, p. 34, 2017.

[101] F. Ehsani and M. Hosseini, "Customer churn analysis using feature optimization methods and tree-based classifiers," *J. Serv. Mark.*, vol. 39, no. 1, 2025.

[102] P. Giudici, A. Gramegna, and E. Raffinetti, "Machine learning classification model comparison," *Socioecon. Plann. Sci.*, vol. 87, p. 101560, 2023.