Addis Ababa University

College of Natural and Computational Science

School of Information Science

Customer Relationship Management using Machine Learning: The case of Bunna Bank S.C.

By:

Kemal Ali Yimam GSE/7434/12

Advisor

Michael Melese (Ph.D.)

September, 2025

Addis Ababa

Ethiopia

Addis Ababa University

College of Natural and Computational Science

School of Information Science

Customer Relationship Management using Machine Learning: The case of Bunna Bank S.C.

By:
Kemal Ali Yimam GSE/7434/12
Advisor
Michael Melese (Ph.D.)

Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Science

Name and Signature of Members of the Examining Board

Name	Title	Signature	Date
Michael Melesse (PhD)	**Advisor**	_____	_____
Melkamu Beyene (Phd)	**Examiner**	_____	_____
Dereje Teferi (Phd)	**Examiner**	_____	_____

September, 2025

Declaration

I confirm that this thesis has not been submitted previously for the award of any degree, nor is it being presented for a degree at any other university at the same time.

I also state that the study titled “Customer Relationship Management using Machine Learning: The case of Bunna Bank S.C.” is the outcome of my own work, except where specific references are made. The study was carried out independently under my advisor's direction and oversight. All external sources consulted have been properly acknowledged through citations, and a full list of references is included at the end of the thesis.

Signature: _____

Kemal Ali Yimam

The thesis has been submitted for examination with my approval as a university advisor.

Signature: _____

Michael Melese (Ph.D.)

Abstract

Customer Relationship Management (CRM) is critical for enhancing customer satisfaction, loyalty, and profitability in the banking sector. Traditional CRM systems, however, often lack the capability to effectively analyze big volumes of customer data and predict customer behavior. The goal of this research is to create a predictive model for customer classification using machine learning, into high, medium, and low-value segments using Bunna Bank S.C.'s customer dataset to improve CRM practices. In an attempt to develop a customer level prediction model, a total of 64,638 datasets by eleven attributes have been used. The overall accuracy of the model has been used as the evaluation metric for this study to identify the best classifier. Accordingly, supervised machine learning algorithms including Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Deep Neural Networks were applied to evaluate their effectiveness in predicting customer levels. Data preprocessing methods such as cleaning, feature selection, and balancing were employed to enhance model performance. The models were evaluated using standard evaluation methods, including accuracy, precision, recall, F-measure, and confusion matrix.

The best performing from the selected classifier is Deep Neural Network (DNN) with an accuracy of almost 79.7%, a precision of 70.9%, and recalls also 79.8%; then Random Forest (RF) with an accuracy of almost 76.4%, a precision of 70.2%, and recall also 76.4%; ; Logistic Regression (LR), with an accuracy of almost 62.1%, a precision of 51.1%, and recall 60.3%; Support Vector Machine (SVM) with an accuracy of almost 60.5 %, a precision of 51.2%, and recall also 60.8%; and K-Nearest Neighbour (KNN) with an accuracy of almost 51.0%, a precision of 35.2%, and recall 38.3% respectively. The study concludes that machine learning-based CRM significantly improves Customer Relationship Management, enabling Bunna Bank to deliver personalized services, optimize resource allocation, and strengthen customer retention strategies. Furthermore, the adoption of advanced CRM systems can enhance profitability, support financial inclusion, and modernize Ethiopia's banking sector.

Keywords: *Customer Relationship Management, Machine Learning, Banking, Deep Learning.*

Acknowledgment

Above all, I thank and praise Almighty God (ALLAH) for giving me the health, strength, and wisdom that allowed me to successfully complete this thesis. This success would not have been achieved without His direction and blessing.

My sincere appreciation goes out to my advisor, Dr. Michael Melese, for his invaluable guidance, encouragement, and constructive feedback throughout this research. His patience, mentorship, and constant support played a pivotal role in helping me finish this work.

My heartfelt thanks also go to Bunna Bank, particularly the management and staff members who provided me with the necessary data, resources, and domain knowledge that were essential for this study. Their cooperation and willingness to assist me are sincerely appreciated.

I am forever indebted to my beloved family my parents, wife, brothers, and sisters for their endless love, encouragement, and prayers. Their unwavering support has been the foundation of my journey and success.

Lastly, I want to express my gratitude to my fellow students., colleagues, and friends, who shared with me not only academic discussions but also unforgettable memories during this period. Your support and friendship will always be a part of me.

Contents

List of Figure

List of Tables

Acronyms

Abbreviation	Definition
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
ATM	Automated Teller Machine
AUC	Area under the curve
AUROC	Area Under the Receiver Operating Characteristic curve
BERT	Bidirectional Encoder Representations from Transformers
CCPA	California Consumer Privacy Act
CDP	Customer Data Platforms
CLV	Customer Lifetime Value
CNN	Convolutional Neural Network
CNTK	Cognitive Toolkit
CRM	Customer Relationship Management
CSAT	Customer Satisfaction
DNN	Deep Neural Network
DSRM	Design Science Research Method
EDA	Exploratory Data Analysis
FN	False Negative
FP	False Positive
GB	Gigabyte
GBM	Gradient Boosting Machine
GDPR	General Data Protection Regulation
GHz	Gigahertz
GPU	Graphics Processing Unit
IOT	**Internet of Things**
KNN	K-Nearest Neighbor
KPI	key performance indicator (KPI)
LGBM	Light Gradient Boosting Machine
LR	Logistic Regression

LSTM	Long Short-Term Memory
LTV	Lifetime Value
MAE	Mean Absolute Error
ML	Machine Learning
NLP	Natural Language Processing
NPS	Net Promoter Score
POS	Point-of-Sale
RAM	Random Access Memory
RF	Random Forest
RFM	Recency, Frequency, Monetary
RMSE	Root mean square error
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic Curve
ROI	Return on Investment
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VGG	Visual Geometry Group
VIP	Very Important Person
WEKA	Waikato Environment for Knowledge Analysis
XGB	Extreme Gradient Boosting (XGBoost)

CHAPTER ONE

INTRODUCTION

1.1 Background

The banking industry is undergoing a significant shift in the digital era. as a result of evolving consumer expectations and the emergence of modern technologies, the focus on Customer Relationship Management (CRM) has become paramount [1]. This introduction delves into the world of CRM in banking, emphasizing its pivotal role, challenges, and opportunities in creating enduring and mutually beneficial relationships between banks and their customers.

Over the past few years, the banking landscape has witnessed a seismic shift. Consumer expectations, interests, and habits are changing rapidly, driven by factors such as technological advancements, regulatory changes, and globalization [2] [3]. In this context, CRM emerges as a strategic imperative for banks, allowing them to adapt to these changes and maintain their competitiveness. CRM in banking refers to the comprehensive strategy and set of practices that banks employ to manage, analyze, and leverage customer data and interactions. Its core objective is to provide specialised experiences and solutions in order to increase customer happiness, loyalty, and profitability.

A customer-centric approach lies at the heart of CRM in banking [4][5]. This approach entails putting the customer's needs, preferences, and experiences at the forefront of all banking activities. This helps banks build trust, cultivate long-term relationships, and stand out in a competitive market.

The use of data is essential to CRM, within the banking sector. Banks accumulate vast volumes of data encompassing customer profiles, transaction histories, digital footprints, and more [6]. Effective CRM hinges on the intelligent analysis and utilization of this data to understand customer behavior and preferences thoroughly. While data holds immense potential, its management presents significant challenges. Banks must grapple with issues related to data privacy, security, accuracy, and integration across multiple systems. Furthermore, strong compliance methods are required by data regulations like the CCPA and GDPR [7]. Banks can customize their products and services leveraging CRM. Banks can tailor products, marketing initiatives, and communication platforms to each customer's needs by using data-driven insights. In addition to increasing customer satisfaction, personalization enhances revenue.

Today's customers interact with banks through multiple touch point's branches, websites, mobile apps, social media, and more. CRM enables banks to deliver a consistent and seamless experience across these channels, ensuring that customers can access services where and when they prefer [8]. Beyond customer-centricity, CRM assists banks in risk mitigation. Banks can identify odd trends that can point to fraud or credit risk by continually monitoring consumer accounts and transactions. This proactive approach safeguards both the bank and its customers. Assessing the effectiveness of CRM initiatives is crucial. Banks employ various metrics such as Customer Lifetime Value (CLV), Net Promoter Score (NPS), and Customer Satisfaction (CSAT) to gauge their success in managing and nurturing customer relationships. The dynamism of the banking industry requires constant innovation and adaptation in CRM. Banks need to be up to date on new technologies like block-chain, big data analytics, and artificial intelligence (AI) for the purpose of stay competitive and offer state-of-the-art CRM solutions [9].

CRM in banking is not merely a buzzword but a strategic necessity in today's financial landscape [10]. It empowers banks to forge meaningful connections with their customers, respond to their evolving needs, and navigate the complexities of the digital age. CRM will become more and more crucial in determining how banking and customer interaction develop in the future as the banking sector develops.

In today's banking environment, when customer expectations are ever-changing and competition is intense, efficient customer relationship management, or CRM, is crucial [11]. The ability to understand, engage, and cater to customers' unique needs is not only a strategic imperative but also a key differentiator for banks. In this era, where data is often hailed as the new oil, banks are using advanced machine learning algorithms to unlock the latent potential of their customer relationships. This introduction establishes the framework for a thorough review of CRM, which is categorized into high, medium, and low-level customer segments. It also highlights the critical role that machine learning algorithms play in forming these connections.

The banking industry has witnessed a significant change in the last few years, primarily driven by digitization and data proliferation [12]. Banks, including those in Ethiopia like Bunna Bank, are at the forefront of this transformation, gathering an extensive wealth of customer data spanning transactions, interactions, and preferences. This wealth of data forms the bedrock for CRM initiatives that aim to deliver personalized and tailored experiences to customers. Not all customers are created equal. Recognizing this, banks have turned to customer segmentation as a fundamental strategy in CRM. Segmentation divides consumers into high, medium, and low level

customer groups according to a number of criteria, including account balances, spending habits, and marketing campaign receptivity.

High-value customers are the customers with substantial account balances, frequent transactions, and a long history of loyalty [13]. They are the ones that banks covet and, through effective CRM, aim to retain and expand their engagement. Medium-value customers are exhibit moderate spending and transaction patterns, and their potential for growth is substantial. Effective CRM strategies aim to nurture and elevate these customers to higher value tiers. Low-value customers, while contributing to a bank's numbers, do not engage at the same level as their high or medium-value counterparts. They may have limited account activity and lower balances. However, they present opportunities for growth through targeted CRM efforts.

Machine learning and artificial intelligence, are the engines driving modern CRM initiatives [14]. These algorithms analyze vast datasets, identifying patterns, predicting behavior, and recommending actions. In the context of high, medium, and low-value customer segments, machine learning algorithms are instrumental in customizing marketing strategies, predicting level of customer, and optimizing resource allocation. While machine learning algorithms offer immense potential, they also pose challenges related to data privacy, security, and the ethical use of customer data. Banks must proceed with caution, making sure that their CRM programs respect client privacy and comply with regulatory requirements.

However, effective customer relationship management is essential for businesses, particularly banks that want to expand without becoming overly dependent on the highly volatile financial markets. To improve customer relationship management, the marketing, sales, and customer retention departments must strive to ensure that customers are happy, give them incentives, and strategically promote offers. In the highly competitive world of today, effective customer relationship management is a crucial part of banking strategy. Banks' traditional revenue streams are being hampered by their inability to remain competitive globally. Effective customer relationship management gauges a business's ability to satisfy current customers as well as its ability to attract new ones. Additionally, it raises ROI, fosters customer loyalty, and attracts prospective customers. Machine learning algorithms are therefore required to evaluate datasets, derive conclusions from patterns found, or generate predictions.

1.2. Motivation of the Study

During my time working at Bunna Bank, I witnessed firsthand the challenges our teams face in managing customer relationships effectively. Although the bank strives to build strong and lasting

connections with its customers, most of the existing CRM practices still rely on traditional, manual processes. These systems make it difficult to fully utilize the large amount of customer data collected every day. As a result, identifying high-value customers, anticipating their needs, and developing proactive retention strategies often becomes a complex and time-consuming task. Seeing these challenges up close motivated me to explore how technology especially machine learning could help improve the way Bunna Bank manages and understands its customers. Machine learning offers a way to analyze large datasets quickly, detect hidden patterns, and predict customer behavior in ways that traditional CRM methods cannot. By applying such techniques, the bank can make more informed decisions, personalize its services, and strengthen customer loyalty.

While many banks around the world are already using machine learning to transform their CRM systems, this practice is still very limited in Ethiopia. This gap presents a great opportunity to introduce data-driven approaches that can make CRM more efficient and effective in the local context. Inspired by both the global potential and the local challenges I observed at Bunna Bank, I was driven to develop and evaluate a machine learning–based customer predictive model that can help predict customer value levels, support customer retention, and contribute to more technology-driven banking practices in Ethiopia.

1.3. Statement of Problem

Customer relationship management (CRM) is a basic instrument in the contemporary banking sector for promoting customer satisfaction, loyalty, and profitability. The adoption and effective utilization of these data-driven CRM paradigms are markedly uneven across the globe. However, in developing countries like Ethiopia, the adoption of advanced CRM practices remains limited [15].

The absence of a data-driven segmentation framework is one of the main weaknesses in the bank's existing system. Currently, extremely basic metrics like account balance are used to classify customers, and behavioural characteristics like transaction frequency, service usage diversification, or preferred channels of communication are not taken into account. According to research, depending on this kind of one-dimensional segmentation results in a misallocation of resources: chances to cultivate medium- and low-value consumers are underutilised, while high-value customers frequently do not receive the individualised service they need [16].

In many developing economies, including Ethiopia, a considerable technological gap exists between local banking practices and those of advanced financial systems. While banks in

developed nations are increasingly using machine learning (ML) techniques for tasks like implementing hyper-segmentation strategies, predicting customer churn, and evaluating lifetime value, Ethiopian institutions such as Bunna Bank continue to use traditional, primarily manual approaches to customer relationship management [15]. The amount of customer data that is being generated is always growing, and these traditional methods cannot completely process and comprehend it. Consequently, Bunna Bank faces challenges in converting customer data into insights that can be used to enhance competitiveness. This restriction has led to operational inefficiencies, subpar consumer experiences, and financial limitations.

A related concern is the problem of customer attrition; banks retention efforts remain reactive in the absence of predictive tools that can identify early signs of disengagement, such as declining transaction activity, decreased use of digital services, or recurring complaints. The chance to step in is gone once an account is closed, which reduces the bank's consumer over time and compels it to make expensive acquisition efforts [17].

A major challenge for Bunna Bank is the absence of a robust data-driven CRM system. While global banks are increasingly adopting machine learning (ML) technologies to analyze customer behaviors and predict future needs, Bunna Bank's CRM practices remain reactive and fragmented [18]. Without utilizing predictive analytics, the bank cannot fully harness the potential of its customer data to inform decision-making, design tailored products, or personalize interactions. This limits its ability to meet the expectations of modern, digitally savvy customers and stay competitive in the fast-evolving financial industry.

Customer segmentation is another critical area where Bunna Bank faces difficulties [19]. The bank currently lacks a systematic method for classifying its customers into high-value, medium-value, and low-value categories based on behavioral and transactional data. This results in inefficient resource allocation, with high-value customers not receiving the premium attention they deserve and medium- or low-value customers not being effectively nurtured to unlock their potential. Consequently, the bank struggles with high customer churn rates and missed opportunities for cross-selling and upselling.

Finally, there is a significant gap in context-specific studies, most researches on ML-enabled CRM are based on data and conditions from developed economies, which differ greatly in terms of infrastructure, customer expectations, and regulatory environments [20]. Consequently,

Ethiopian banks lack locally validated frameworks to guide adoption, forcing them into a trial-and-error approach that delays progress and increases risk.

Given these realities, it is essential to critically examine Bunna Bank's current CRM practices and design a machine learning framework suited to its operational environment. This study aims to fill that gap by developing a model that supports scientific segmentation, reliable churn prediction, and personalized marketing strategies. Beyond offering practical solutions for Bunna Bank, the findings are expected to provide a context-aware contribution to the limited body of knowledge on technology-driven financial systems in Ethiopia. Consequently, this study focused on creating a predictive model using machine learning techniques, which helps to determine customer levels to fill in the aforementioned gaps. In an attempt to solve the above-stated problems, the research work answers the following research questions:

RQ1. Which are the most determinant features for customer level prediction?

RQ2. Which type of machine learning algorithm is suitable for customers' level prediction model?

RQ3. To what degree does the proposed model accurately predict outcomes at the individual customer level for Bunna Bank?

1.4. Objective of the Study

1.4.1. General Objective

The primary objective of this research is to create a customer-level predictive model for Bunna Bank using machine learning methods, aiming to strengthen customer relationship management and boost profitability.

1.4.2. Specific Objectives

To meet the primary objective of this research, the following specific objectives have been formulated:

- To evaluate and examine relevant documents and earlier works in the area to get an insight into the area and to find out related works and their contributions.
- To collect and prepare the dataset through the use of standard data cleaning and pre-processing techniques, using problem and data understanding.
- To select the best and appropriate tools and machine learning techniques, that is used for building the required models.

- To determine the variables that is crucial for predicting the level of customers.
- To apply the identified machine-learning approaches for model training.
- To evaluate the performance of the machine learning model by evaluation metrics.
- To provide concluding remarks and recommendations for further research works.

1.5. Scope and Limitation of the Study

This study's focus involves developing and examining machine learning models for customer segmentation at Bunna Bank. The main objective is to categorize customers into high, medium, and low-value groups by analyzing demographic and transactional data collected from the bank's East District. The research process includes data collecting, cleaning, feature selection, model development, and performance evaluation of selected supervised algorithms (LR, SVM, KNN, RF, and DNN) using common measures such as accuracy, precision, recall, and F-measure.

This study is subject to certain limitations, the customer data used was drawn only from Bunna Bank's East District, which may not accurately reflect the characteristics of customers in other districts. Although the dataset included 64,638 records, the attributes available were limited, meaning that some important behavioral or transactional features were not captured. This study concentrated on five commonly used supervised learning techniques (LR, SVM, KNN, RF, and DNN) while excluding other modern algorithms such as Gradient Boosting and XGBoost that might have yielded stronger predictive results. Computational resources also posed restrictions, as the experiments were run on Google Colab with modest GPU capacity, limiting extensive parameter tuning and the training of more complex models. Even though oversampling techniques were applied, imbalances among the customer categories remained and affected the accuracy of predictions for less represented groups. Lastly, the data represented a single point in time; hence, the models could not fully account for the way customer preferences and behaviors change over time.

1.6. Significance of the Study

Techniques for customer-level prediction aim to identify the specific behaviors and key variables that indicate how customers are categorized or valued [21], this prediction is used to improve customer relationship management and enhance profitability. If Bunna Bank is willing to deploy the best model of the study, for predicting the levels of customers:

The Bunna Bank will be benefitted to understand customer preferences, spending habits, and engagement patterns more effectively. This categorization allows the bank to implement targeted

CRM strategies, improving customer loyalty, reducing churn, and sustaining long-term profitability. By prioritizing resources on high-value customers while applying cost-effective strategies for others, the bank ensures optimal resource utilization and maximized returns. Additionally, the data-driven insights generated by machine learning enhance decision-making in areas such as loan approvals, credit risk assessments, and customer retention, while also informing the development of new financial products and competitive pricing models. Embracing advanced machine learning techniques positions Bunna Bank as a technology-driven institution, strengthening its competitive advantage and attracting tech-savvy customers while inspiring trust among existing clients.

The study's application of machine learning benefits customers by providing personalized banking services customized to meet each customer's individual requirements and preferences. High-level customers may receive exclusive offerings like investment advice and priority services, while medium- and low-level customers gain access to targeted offers, financial literacy programs, and incentives to deepen engagement. These personalized interactions enhance customer satisfaction by creating a seamless and fulfilling banking experience. Additionally, customers can enjoy increased financial benefits, such as lower interest rates, reduced fees, and predictive tools that help them avoid financial risks. For low-level customers, educational tools and guidance empower them to enable them to make sound financial choices, enhancing their overall financial health and possibly.

The application of machine learning for customer-level prediction offers significant benefits for Ethiopia by advancing financial inclusion through tailored products like microloans, mobile banking solutions, and Sharia-compliant services, bringing underserved populations and small businesses into the formal banking system. This inclusion supports entrepreneurship, job creation, and economic growth by improving access to credit and investments while guiding banks to focus on high-potential sectors. Furthermore, Bunna Bank's adoption of innovative technology fosters competitiveness and inspires other financial institutions to embrace machine learning, leading to continuous service improvements. This study also promotes a data-driven culture across industries, enhancing productivity, efficiency, and technological advancement in Ethiopia's economy.

Besides, the output of this study can be a source of knowledge in the area of customer relationship management in banking and it can be used by future researchers for further studies.

1.7. Organization of the Study

The study's contents are arranged as follows:

Chapter One (Introduction): explains the introduction, motivation, problem statement, goals, research questions, study importance, scope, and limitations.

Chapter Two (Literature Review and Related Works): discusses the literature related to the banking industry and Bunna bank, customer relationship management, and machine learning. This chapter examines and contrasts the earlier research in this field that used machine learning techniques to predict customer levels.

Chapter Three (Design and Methodology): explains in fully the design and approach used to address the research topic. It follows the methodology of Peffer design science, and chapter 3 provides a detailed explanation of each stage.

Chapter Four (Experiment, Result, and Discussion): The study describes the process used to build the system and shows the outcomes. It goes into great detail about the models that were used, their selection, and their application. The validity of the research hypothesis is evaluated by carefully weighing the results against it. The best model is determined by analysing the data. In light of these findings, the main research issue is re-examined, and the hypothesis is scrutinised. Finally, the work reflects on its own strengths and acknowledges its limitations.

Chapter Five (Conclusion and Recommendation): discuss the study problem together with the evaluation and outcome attained. It explains the study’s contribution to the study question and provides a summary of the findings. It also recommends some further research in a related field.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORKS

2.1. Overview

This chapter describes the necessary concepts related to customer-level prediction, customer relationship management (CRM), and machine learning algorithms that aid the banking sector in understanding and managing customer behavior. It also reviews several related studies to explore the research problem further and investigates how machine learning techniques can enhance CRM practices. The chapter ends by identifying limitations in existing research and establishing the objectives for this study.

2.2. Banking Industry

In an economy, banks are essential financial intermediaries. A healthy banking industry promotes general economic stability and makes it easier for the economy to withstand negative shocks. Banks have a major impact on how capital is distributed, how firms expand, and how industries evolve by collecting, allocating, and investing the savings of economic agents. Therefore, the effectiveness and profitability of banks are crucial for the well-being and expansion of the larger economy as well as for individual institutions [22]. This sector of the economy, which includes government regulation of banking operations, as well as services like insurance, mortgages, investment management, and credit card issuance, is dedicated to protecting the financial assets of others and leveraging those assets to create new wealth.

For both individuals and businesses, the banking industry handles cash, credit, and other financial activities. It is a vital part of the global economy because it helps provide the liquidity that businesses and consumers require in order to plan and make investments for the future. The services that banks provide can be used to classify them. Retail banking concentrates on offering loans, deposit accounts, and money management services to individuals and families, whereas commercial banks mostly cater to businesses and private customers [23].

2.2.1. Bank Industry in Ethiopia

Modern banking was first introduced in Ethiopia in 1931 G.C. thanks to an agreement established in 1905 G.C. between Emperor Menelik II and Mr. Ma Gillivray, a representative of the British-owned National Bank of Egypt [24]. On February 16, 1906 G.C., the Emperor created the Bank of Abyssinia as the main bank after an official agreement. The Egyptian National Bank was in charge of its operations. The Bank of Abyssinia mostly concentrated on a few limited activities throughout its comparatively short existence, such as handling government accounts, offering

modest export finance, and carrying out certain duties on behalf of the government. However, the bank's inefficiencies and preference for profit over national interests drew harsh criticism. In order to lessen foreign influence over Ethiopia's banking industry and establish a financial organisation that could meet the nation's credit needs, a deal was eventually made to dissolve the bank. Therefore, in G.C. 1931, shortly after Emperor Haile Selassie ascended to power, the Bank of Abyssinia was officially replaced by the Bank of Ethiopia.

The Bank of Ethiopia, established on August 29, 1931, was entirely Ethiopian and marked the first indigenous bank in Africa, starting with a capital of £750,000 [10]. To ensure a seamless transition, the Bank of Egypt gave up its concession rights in return for £40,000. The new organisation kept the Bank of Abyssinia's staff and facilities, including Mr. Collier, the manager. 60% of the bank's shares were owned by the Ethiopian government, and the minister of finance had to supervise all business dealings [24]. Now, the Bunna Bank is one of the known private banks in Ethiopia.

Bunna International Bank S.C. is among the well-established private commercial banks in Ethiopia. It began operations in October 2009 after securing a license from the National Bank of Ethiopia. Over the years the bank has steadily expanded its reach, operating more than 465 branches across the country and employing over 4,100 permanent staff. Financial statements show that Bunna Bank’s total assets have grown to about 54.53 billion Ethiopian birr, with customer deposits rising to around 43.87 billion birr, demonstrating strong growth even during periods of economic uncertainty [25]. The bank continues to broaden its services by introducing additional cash-payment machines and expanding digital banking channels, aiming to provide convenient access for customers in both urban centers and regional towns.

2.3. Customer Relationship Management

In the modern business environment, CRM (customer relationship management), has become a vital tactic for businesses all over the world [26]. Regardless of the business type, managing and fostering strong relationships with customers is vital for organizational success. Poor CRM practices can result in reduced profit margins, diminished customer loyalty, and lost opportunities for referrals from satisfied customers. Effective CRM focuses on understanding customer needs, preferences, and behaviors to nurture loyalty, foster engagement, and ensure sustained satisfaction [27]. By building robust customer relationships, companies can not only retain existing customers but also cultivate loyal customers that supports sustained income expansion and market standing.

2.3.1. The Role and Importance of Customer Relationship Management

CRM is a tactical procedure which aids businesses in preserving and improving customer satisfaction. It plays a pivotal role in managing customer interactions throughout their lifecycle from initial engagement to post-purchase support [28]. Strong CRM practices help companies avoid potential pitfalls like poor communication, inconsistent service, or failure to meet customer expectations. When implemented effectively, CRM enables organizations to anticipate customer needs, resolve issues promptly, and provide a seamless experience, which is crucial for customer retention and repeat business.

Poorly managed customer relationships can lead to financial setbacks due to declining sales and increased costs associated with acquiring new customers. Acquiring new customers typically costs five to ten times more than retaining existing ones. CRM is therefore given top priority by businesses in order to reduce customer attrition and increase customer lifetime value (CLV) [29].

2.3.2. Customer Relationship Management Strategies

CRM encompasses both strategic initiatives and technology solutions [30]. Strategies include:

1. **Personalized Engagement**: Understanding individual customer preferences and tailoring products, services, and communications to meet those needs.
2. **Data-Driven Insights**: Utilizing analytics to discover more about consumer behavior, predict trends, and develop targeted campaigns.
3. **Proactive Communication**: Regularly interacting with customers through multiple channels (e.g., emails, social media, direct calls) to build trust and keep them informed.
4. **Loyalty Programs**: Introducing incentive programs to promote recurring business and improve relationships with customers.

2.3.3. Customer Relationship Management Approaches: Reactive vs. Proactive

Customer Relationship Management (CRM) strategies can be categorized into two primary approaches: **Reactive CRM** and **Proactive CRM** [31]. These approaches differ in their methods of managing customer relationships and addressing customer needs. Both are essential components of a comprehensive CRM strategy and are often used together to ensure customer satisfaction and loyalty.

2.3.3.1. Reactive Customer Relationship Management

Reactive CRM involves responding to customer inquiries, complaints, or issues as they arise [32]. It is focused on solving problems that customers bring to the organization and maintaining service quality through efficient resolution.

This approach's primary characteristic is that the customer always initiates the conversation. The company takes action only after someone reaches out for support, whether it's to solve a challenge or get clarification. The goal is to satisfy the customer and fix the problem. Additionally, reactive CRM depends largely on customer feedback because inquiries and complaints frequently highlight areas in which the organization might make improvements.

There are some advantages to operating in this manner. First, it allows for quick problem solving customers generally appreciate a fast response when something goes wrong. Second, it can be more cost-efficient in some situations because the business doesn't need to put resources into proactive outreach; it simply responds when needed. Additionally, consumer inquiries can highlight trends and reoccurring problems, providing the company with important information about what may require more extensive fixes.

However, there are drawbacks to reactive CRM. It restricts the company's ability to interact with customers in ways other than problem solving because it only addresses problems after they arise. There's also a risk of things escalating if responses are too slow or inadequate, which can damage trust and lead to frustration. And since the focus is on short-term fixes, there's always the chance that the root causes of dissatisfaction go unaddressed. Typical instances of reactive CRM include addressing product flaws via support channels, providing refunds and replacements when items arrive damaged, and correcting billing problems following a customer complaint.

2.3.3.2. Proactive Customer Relationship Management

Proactive CRM predicts what customers may require and addresses possible problems before they occur [32]. It focuses on delivering a seamless customer experience by anticipating issues and responding with tailored interactions.

One of the key strengths of proactive CRM is its anticipatory approach. By using data analytics, predictive modeling, and customer insights, businesses can forecast customer behavior and identify future needs. This makes it possible to engage customers in meaningful ways before they encounter difficulties. At its core, proactive CRM is also about customer-centric engagement,

building long-term relationships by consistently offering value that goes beyond a single transaction. In many cases, companies rely on automation and AI to detect patterns, automate routine communications, and deliver well-timed interventions that strengthen the customer relationship.

The advantages of proactive CRM are clear. It enhances customer loyalty by showing attentiveness and care, even before problems surface. It also helps improve retention rates, since customers who feel understood and valued are less likely to leave. On the operational side, anticipating customer needs allows companies to allocate resources more efficiently, reducing the costs often associated with reactive problem-solving.

However, proactive CRM does come with challenges. It can be resource intensive, requiring investments in advanced technology, skilled staff, and strong data analytics capabilities. There's also the risk of overstepping, too much communication or irrelevant outreach can frustrate customers rather than build trust. Finally, implementing a proactive system is often complex, demanding a solid infrastructure and continuous refinement to stay effective. Practical examples of proactive CRM include sending reminders for service renewals or scheduled maintenance, offering loyalty rewards or discounts based on past purchases, and reaching out to customers affected by product issues before they notice or report them. Businesses may also use this approach to suggest complementary products or services, creating value while deepening the customer relationship.

Table 2.1. Comparison: Reactive vs. Proactive CRM [32]

Feature	Reactive CRM	Proactive CRM
Focus	Problem-solving after issues arise	Anticipating and preventing issues
Interaction Type	Customer-initiated	Business-initiated
Technology Dependence	Minimal	High (requires analytics and automation)
Customer Experience	Resolves dissatisfaction	Builds trust and loyalty
Cost	Typically lower upfront costs	Higher investment in tools and processes
Long-term Impact	Limited, solve immediate concerns	Significant, fosters customer retention

2.4. CRM and Long-Term Business Success

Customer Relationship Management (CRM) is more than just a system for keeping track of interactions with current and potential customers. but a strategic approach that directly influences a company's long-term success [33]. Businesses can improve customer satisfaction and retention,

strengthen relationships with customers, and promote sustainable growth by utilizing CRM effectively. Here's a detailed explanation of how CRM contributes to long-term business success.

2.4.1. Building Strong Customer Relationships

At its core, CRM focuses on understanding and meeting customer needs. Establishing strong relationships with customers ensures loyalty and trust, both of which are vital for long-term success [33]. One of the most effective ways CRM achieves this is through personalized interactions. By collecting and analyzing customer data, CRM systems allow businesses to tailor their communication, offers, and services. Customers are much more likely to remain loyal when they feel appreciated and engagged because to this individualised approach.

Equally important is consistent engagement. CRM helps organizations maintain regular, meaningful communication with their customers. This ongoing interaction strengthens relationships over time and ensures that competitors have fewer opportunities to draw customers away.

2.4.2. Enhancing Customer Satisfaction

Satisfied customers are more likely to come back, buy again, and advocate for a brand; CRM systems ensure satisfaction [27]. One important way CRM systems enhance customer satisfaction is by enabling timely responses. By streamlining customer support processes, businesses can reduce response times and address concerns more quickly and efficiently. This not only resolves issues faster but also reassures customers that their needs are a top priority.

Another valuable feature is proactive problem-solving, with the help of predictive analytics; CRM tools can predict likely problems before they occur. This allows businesses to take early action, prevent disruptions, and deliver a seamless customer experience.

Finally, effective feedback management plays a main role in building satisfaction and loyality. By collecting, analyzing, and acting on customer feedback, businesses show that they are listening and committed to continuous improvement. Over time, this strengthens loyalty and reinforces customer confidence in the brand.

2.4.3. Improving Customer Retention

Long-term profitability depends on maintaining current customers because it often more expensive to acquire new ones [34].

CRM supports customer retention by fostering loyalty, anticipating needs, and reducing churn. Through loyalty programs, businesses can design and manage rewards that encourage repeat purchases and strengthen long-term commitment. By analyzing customer data, CRM systems also anticipate future requirements and deliver timely solutions, making customers feel valued and understood. In addition, CRM helps reduce churn by identifying early signs of dissatisfaction and enabling targeted interventions to re-engage at-risk customers. Together, these features build trust, enhance satisfaction, and ensure lasting relationships with customers.

2.4.4. Driving Sales Growth

CRM systems enhance sales processes by providing insights and tools that empower sales teams to perform more effectively [35]. One major advantage is the ability to run focused advertising campaigns. By segmenting customers based on their interests and actions, CRM ensures that marketing efforts reach the right audience, improving both efficiency and results.

CRM also supports upselling and cross-selling by using data-driven recommendations. Sales teams can identify opportunities to offer complementary products or premium services, which not only boosts revenue but also delivers greater value to customers.

In addition, CRM streamlines sales processes through automation. By reducing manual tasks such as data entry and follow-ups, sales teams can spend more time building meaningful relationships and closing deals. This combination of efficiency and personalization makes CRM a powerful tool for driving sustainable sales growth.

2.4.5. Facilitating Data-Driven Decision Making

CRM tools allow companies to store and manage all customer data in a unified database, empowering them to make well-informed decisions that support long-term goals [36]. One key benefit is access to valuable customer insights. Detailed reports and analytics reveal trends, preferences, and pain points, which guide strategic planning and help businesses design products and services that truly meet customer needs.

CRM data also offers a window into market trends. By analyzing patterns and behaviors across a broad customer base, businesses can anticipate shifts in demand, adapt quickly, and stay ahead of competitors in an ever-changing market.

Another important advantage is performance tracking. With CRM, organizations can monitor key performance indicators such as customer acquisition, retention, and satisfaction. These metrics provide a clear picture of what's working and where improvements are needed, ensuring strategies remain effective and results-driven.

2.4.6. Fostering Collaboration across Departments

Effective CRM implementation promotes cross-departmental collaboration, ensuring that every team works towards common goals [37].

One of the best advantages of CRM is the creation of a unified customer view. By integrating data from marketing, sales, and customer support, it provides a comprehensive picture of each customer, making it easier to personalize interactions and address needs effectively. CRM also enhances coordination between teams, allowing businesses to maintain consistent messaging and service delivery across all touch points, which builds trust and strengthens brand reliability. In addition, it promotes knowledge sharing, as insights from one department such as customer feedback or sales trends can be shared with others, leading to smarter decisions and improved overall organizational performance.

2.4.7. Enhancing Scalability and Adaptability

As businesses grow, managing customer relationships becomes more complex. CRM systems enable this expansion by offering flexibility and scalability [38]. One important advantage is the use of automation and artificial intelligence (AI). Automated workflows and AI-powered tools help businesses manage larger volumes of customer interactions efficiently, ensuring speed and consistency without sacrificing quality. CRM platforms are also designed to be customizable. They can be tailored to fit the unique requirements of each business and adapted as those needs evolve, making them a long-term solution that grows alongside the organization.

In addition, CRM systems support a global reach by enabling multi-channel communication. This allows businesses to engage with customers seamlessly across regions and time zones, ensuring consistent service and stronger relationships no matter where their customers are located.

2.4.8. Strengthening Competitive Advantage

In a competitive marketplace, exceptional customer experience often differentiates leading businesses from the rest. CRM helps businesses gain this edge [39]. One way CRM contributes is by identifying opportunities. Through advanced analytics, businesses can uncover untapped market segments or unmet customer needs that competitors may overlook, giving them a first-mover advantage. CRM also strengthens brand loyalty. By fostering strong, personalized relationships, businesses make it harder for competitors to draw customers away, ensuring long-term retention and trust.

Another important benefit is continuous improvement. With regular feedback collection and performance tracking, CRM allows businesses to stay agile and responsive to changing customer preferences, ensuring they consistently deliver experiences that meet or exceed expectations.

2.4.9. Ensuring Long-Term Financial Stability

The cumulative benefits of CRM enhanced customer satisfaction, improved retention, and increased sales contribute to sustained financial success [34].

One major outcome of CRM is higher customer lifetime value, as satisfied and loyal customers make repeat purchases and are more likely to refer others, steadily increasing revenue over time. It also helps reduce acquisition costs by focusing on nurturing strong relationships with existing customers, allowing businesses to rely less on costly marketing campaigns to attract new ones. In addition, CRM provides predictable revenue streams through accurate forecasting tools, enabling better financial planning, smarter resource allocation, and greater stability even in changing market conditions.

2.5. Modern CRM: A Data-Driven Approach

Modern Customer Relationship Management (CRM) has evolved far beyond the traditional methods of handling customer interactions. In the current digital environment, data plays a central role in CRM systems, transforming how businesses comprehend, interact with, and serve their customers. A data-driven CRM strategy enables businesses to improve business performance, make educated decisions, and customise experiences for every customer [40]. Here is an in-depth review of modern CRM and its data-driven approach:

2.5.1. Data-Driven Customer Relationship Management

Data-driven CRM uses customer data to establish and maintain stronger relationships by identifying patterns, predicting behaviors, and delivering personalized experiences. It integrates data from multiple sources, departments like marketing, sales, and customer support, in order to develop a complete understanding of every customer [40].

2.5.2. Key Components of Data-Driven Customer Relationship Management

In the extremely competitive marketplace of today landscape, leveraging data to build and maintain customer relationships is crucial for business success. Data-driven CRM incorporates technology, analytics, and automation to provide personalized and efficient customer experiences [41]. The following sections outline the core components that make up a data-driven CRM system.

The first component is customer data collection. Modern CRM systems gather information from multiple touch points, including websites, social media, email campaigns, and customer support interactions. The data collected can range from demographics and transaction history to personal preferences and behavioral patterns, such as browsing habits or purchase frequency. This rich dataset forms the foundation for more targeted engagement. The next step is data integration. By consolidating information from different sources into a unified platform, CRM systems eliminate departmental silos and create a complete 360-degree view of the customer. This concludes that all teams have allow to consistent and accurate information, which is crucial for delivering seamless customer experiences.

Advanced analytics is another key element of data-driven CRM. Analytical tools process vast amounts of customer data to uncover trends and actionable insights. Predictive analytics, in particular, can forecast future customer behavior such as the likelihood of churn or potential future purchases allowing businesses to act proactively rather than reactively.

Finally, automation and AI significantly enhance CRM capabilities. Automation simplifies In the current environment of intense competition repetitious duties, such as handling data entry, setting up appointments, or sending follow-up emails, allowing teams to concentrate on more strategic initiatives . At the same time, AI tools provide intelligent recommendations, enable chatbots for instant customer support, and even perform sentiment analysis to better understand customer emotions and needs.

2.5.3. Benefits of a Data-Driven Customer Relationship Management Approach

Adopting a data-driven CRM approach empowers businesses to foster stronger customer relationships and drive growth. By using data and analytics, organizations can deliver more targeted and personalized experiences while improving operational efficiency [42]. The primary advantages of this approach are outlined in the below sections.

Improved customer insights are one of the main advantages. Businesses can better understand consumer preferences, wants, and pain points by analysing customer data. These insights make it possible to create meaningful customer segments and design tailored strategies for each group, ensuring that no opportunity is overlooked. Data-driven CRM also powers personalized customer experiences. Marketing campaigns and product recommendations can be customized based on individual data, which increases engagement and satisfaction. With the ability to track customer journeys in real time, businesses can respond quickly to specific needs and strengthen relationships through timely, relevant interactions.

Improved customer retention is another critical advantage. Predictive analytics helps identify at-risk customers, allowing businesses to intervene early with targeted offers, discounts, or proactive support. Enhanced customer service, guided by accurate data, ensures that issues are resolved effectively, increasing loyalty and trust over time. Streamlined operations further add value; automation reduces repetitive manual tasks, freeing teams to focus on more strategic initiatives. At the same time, data-driven CRM aligns marketing, sales, and customer service, creating smoother collaboration across departments and delivering a consistent customer experience.

Finally, adopting this approach leads to increased revenue. By targeting the right customers with the right offers, conversion rates naturally improve. Data insights also uncover opportunities for cross-selling and upselling, helping businesses maximize value from their existing customer base while continuing to grow.

2.5.4. Key Technologies Enabling Data-Driven Customer Relationship Management

The rapid advancement of technology has revolutionized how businesses manage and leverage customer relationships [43]. Data-driven CRM systems are powered by cutting-edge technologies that enhance data collection, integration, and analysis. The following sections explore the core technologies enabling this transformation.

Big Data and cloud computing play a foundational role. Big data technologies allow CRM systems to process and analyze massive datasets, uncovering valuable patterns and trends that

guide smarter decision-making. At the same time, cloud-based CRM platforms provide scalable and easily accessible solutions, making advanced capabilities available to businesses of all sizes.

CRM capability is further enhanced by machine learning and artificial intelligence (AI). AI-powered systems deliver predictive insights, intelligent recommendations, and automated customer support, while through learning from new data sets, machine learning models consistently increase their accuracy over time. This makes customer interactions smarter, faster, and more relevant. The Internet of Things (IoT) also contributes by generating real-time customer data through connected devices. Information such as product performance, usage behavior, and service needs can be captured and integrated into CRM systems, enabling businesses to enhance offerings and proactively address potential issues before they escalate.

Customer Data Platforms (CDPs) verify data consistency and accuracy. By unifying and standardizing customer data from multiple sources, CDPs provide a reliable foundation that feeds into CRM systems. This unified view strengthens overall functionality and ensures that businesses operate with a complete and consistent understanding of their customers.

2.5.5. Applications of Data-Driven CRM in Modern Business

In the current environment of intense competition, businesses are increasingly leveraging data-driven systems for managing customer relationships (CRM) are used to enhance the overall satisfaction of customers and streamline business processes [44]. These systems allow companies to utilize valuable insights from customer data, driving strategic decisions and fostering stronger relationships. By analyzing patterns and behaviors, organizations can tailor their approaches to meet evolving customer needs effectively.

One of main applications of data-driven CRM in modern business is personalized marketing. By leveraging customer segmentation and behavioral insights, companies are able to create precisely focused campaigns that effectively engage particular customer groups. Dynamic content further adapts in real time tailored according to each person's choices and previous experiences, ensuring that every communication feels relevant. Proactive customer support is another critical advantage. With predictive analytics, businesses can anticipate potential issues before customers even report them. Tools like Chabot’s and virtual assistants supply 24/7 support, improving responsiveness and significantly boosting customer satisfaction.

Data-driven CRM also excels in customer journey mapping. By tracking every touchpoint in the customer journey, organizations can identify areas that need improvement. Optimizing these interactions ensures a seamless, consistent experience, which helps build long-term loyalty. In

sales optimization, CRM systems guide teams by prioritizing leads based on their likelihood to convert. Real-time insights shape sales tactics, enabling teams to focus their efforts on high-potential prospects and ultimately increase close rates.

Feedback and continuous improvement are also central to data-driven CRM. By analyzing customer feedback, businesses can enhance products, services, and overall experiences. Tracking feedback trends also ensures that companies remain agile and responsive to shifting customer expectations.

2.5.6. Challenges in Implementing Data-Driven Customer Relationship Management

The process of putting in place a data-driven Customer Relationship Management (CRM) system comes with its own set of challenges that businesses must properly manage [45]. In order to properly use data for improved customer interactions, the process entails both organizational changes and technology adjustments. Companies need to address technical complexities while fostering a culture that supports data-driven decision-making. Below are some of the critical challenges encountered.

One of the most pressing challenges is data privacy and security. Businesses are required to comply with regulations such as GDPR and CCPA to ensure that customer information is protected. Beyond compliance, companies must also implement strong security measures to safeguard against data breaches and maintain customer trust. Another challenge lies in data quality and integration; ensuring that data is accurate and consistent across multiple sources can be difficult, especially for organizations that deal with large volumes of information. Effective integration often requires advanced tools and specialized expertise to create a unified and reliable dataset.

Change management is equally important. Shifting to a data-driven CRM approach often calls for cultural and operational transformation within the organization. This may involve redesigning workflows, providing comprehensive employee training, and securing stakeholder buy-in to ensure that the implementation is both smooth and sustainable.

2.6. Customer Relationship Management (CRM) in the Banking Industry

These days, Customer Relationship Management (CRM) is a key component of the banking industry's success [46]. As banks face heightened competition, technological disruption, and evolving customer expectations, implementing effective CRM strategies is essential to maintain

customer loyalty, streamline operations, and enhance profitability. The following sections explore a detailed exploration of CRM in the banking sector.

2.6.1. Role of Customer Relationship Management in Banking

In the banking sector, CRM aims to strengthen customer connections by offering tailored services, maintaining effective communication, and making decisions based on data insights [46]. Through CRM, banks gain a deeper understanding of customer preferences and behaviors, enabling them to design and offer tailored financial products and services. This personalized approach not only meets individual needs but also enhances overall satisfaction. CRM also helps banks anticipate customer demands, allowing them to proactively address expectations and create more seamless experiences. In turn, this fosters stronger retention and loyalty, as customers are more likely to remain with a bank that understands and supports their financial goals.

Finally, CRM contributes to operational efficiency and profitability. By streamlining processes and improving coordination, banks can reduce costs while maximizing value, ensuring both stronger customer relationships and sustainable business growth.

2.6.2. Key Components of Customer Relationship Management in Banking

In the banking industry, customer relationship management (CRM) , has grown essential because it allows organizations to build closer bonds with their customers and provide more specialized services [47]. By leveraging advanced CRM tools, banks can better understand customer needs, improve operational efficiency, and enhance customer satisfaction. These systems play a critical role in providing a competitive edge by aligning banking services with customer expectations.

One of the key components of CRM in banking is customer segmentation. By grouping customers based on demographics, financial behaviors, and preferences, banks can offer more relevant products and services. Typical segmentation factors include credit ratings, savings habits, transaction frequency, and income levels, all of which help in tailoring solutions to specific customer groups. Personalized banking services further enhance customer experiences; CRM makes it possible to provide tailored loan offers, savings plans, or investment advice by analyzing customer data such as transaction history, spending patterns, and even life events. This level of personalization helps banks demonstrate that they understand and value each customer’s unique financial journey.

Another important element is the 360-degree customer view. By integrating data from multiple channels, CRM platforms create a unified profile for each customer. This holistic perspective ensures that interactions remain consistent and seamless across different departments, including retail banking, corporate banking, and customer support. Data-driven decision-making is also central to modern CRM. Advanced analytics tools help banks identify trends, predict customer needs, and guide strategic planning. Predictive analytics, in particular, supports key functions such as assessing credit risk, identifying cross-selling opportunities, and reducing churn by proactively addressing customer concerns.

Finally, CRM strengthens customer support and communication. With instant access to customer histories and preferences, support agents can provide faster, more accurate assistance. Multichannel communication including mobile apps, email, chatbots, and social media ensures customers can engage with their bank anytime, anywhere, resulting in more seamless and satisfying interactions.

2.6.3. Benefits of Customer Relationship Management in Banking

The adoption of Customer Relationship Management (CRM) in banking offers significant advantages, helping institutions to improve their competitiveness and provide better services [48]. By leveraging CRM systems, banks can effectively manage customer interactions, optimize operations, and align their strategies with customer needs. The insights gained from these tools enable banks to build lasting relationships and achieve sustainable growth. Key benefits of CRM in banking include: One of the most important benefits is enhanced customer retention. By understanding customer demands and taking proactive steps to address potential concerns, banks can cultivate loyalty and reduce churn. CRM tools make It becomes simpler to recognize customers who may be at risk and apply focused strategies to retain them, ensuring strong ongoing relationships.

CRM also drives increased revenue by uncovering opportunities for cross-selling and upselling financial products such as credit cards, loans, and insurance. Banks can fine-tune pricing strategies and develop new offerings tailored to specific customer segments, boosting both customer satisfaction and profitability. Another key advantage is improved customer experience. Personalized services, combined with timely and consistent communication, help create positive interactions at every stage. CRM platforms ensure seamless engagement, reducing friction during transactions or issue resolution and enhancing overall satisfaction.

Operational efficiency is also significantly improved through CRM adoption. By automating repetitive processes such as application processing or sending reminders banks can reduce costs and free up staff for higher-value tasks. Integrated systems further promote collaboration across departments; ensuring customers receive consistent and reliable service.

Finally, CRM contributes to regulatory compliance and risk management. By tracking customer interactions and transactions, these systems provide clear audit trails to meet compliance requirements. Advanced analytics also help detect fraudulent activities and support more accurate risk assessments, protecting both the institution and its customers.

2.6.4. Applications of Customer Relationship Management in Banking

Customer Relationship Management (CRM) plays a transformative role in the banking industry, offering tailored solutions for various customer segments and operational needs [46]. Banks can improve their capacity to service both corporate and retail customers by incorporating CRM systems, improve digital banking services, and streamline customer support. These applications enable banks to deliver a seamless experience while fostering stronger relationships and ensuring customer satisfaction.

In retail banking, CRM tools are widely used to manage individual customer relationships by offering personalized products such as savings accounts, loans, and credit cards. Banks also use CRM to track important life events such as marriage, the birth of a child, or retirement allowing them to recommend timely and relevant financial products that match evolving customer needs. For corporate banking, CRM systems play a crucial role in managing relationships with business clients. They enable banks to monitor interactions, contracts, and credit arrangements while providing the insights needed to design customized solutions, including business loans, treasury services, and investment opportunities. This ensures that corporate clients receive tailored support that strengthens long-term partnerships.

Digital banking is another area where CRM adds significant value. With mobile and online platforms becoming central to customer engagement, CRM helps ensure consistency and personalization across digital channels. By analyzing customer behavior during online interactions, banks can refine their digital services, enhance user interfaces, and provide more intuitive experiences.

Finally, CRM enhances customer support by integrating AI-driven chatbots and virtual assistants that deliver instant responses to common queries. At the same time, support agents can access

comprehensive customer histories through CRM systems, allowing them to resolve issues more effectively and provide a higher level of personalized service.

2.7. Customer-Level Prediction: (High, Medium, and Low-Level Customer Segmentation)

In the context of customer relationship management (CRM), predicting customer behavior at a granular level aids companies in developing focused marketing plans, enhance customer service, and maximize profitability [49]. Segmenting customers into high, medium, and low-level categories is an effective way to prioritize resources and optimize engagement strategies. Here is a comprehensive overview of how this prediction can be done and its implications for businesses:

2.7.1. Understanding Customer-Level Segmentation

Customer-Level Segmentation involves categorizing customers based on their behavior, purchasing patterns, engagement level, and overall value to the business [50]. By dividing customers into high, medium, and low levels, businesses can tailor their approaches to maximize customer satisfaction and business outcomes. Here's what each level typically entails:

High-Level Customers: These account holders generate the largest share of a bank’s income and maintain strong, long-term relationships with the institution. They often keep high balances, move significant sums through checking or investment accounts, and make frequent use of services such as mortgages, wealth-management products, or premium credit cards. Their loyalty shows in their consistent activity and willingness to adopt new offerings. To retain them, a bank can provide dedicated relationship managers, priority service channels, attractive interest rates, and early access to exclusive financial products.

Medium-Level Customers: play an important but less dramatic role in overall revenue. They typically maintain steady deposits, use everyday services like debit cards and online banking, and occasionally respond to promotions or seasonal offers. While their engagement is not as deep as that of the top tier, they can be nurtured into higher-value relationships. Personalized savings plans, targeted loan offers, and reward programs for regular account use can encourage these customers to expand their banking activity.

Low-Level Customers**:** interacts with the bank only sporadically and usually holds minimal balances. They may maintain a basic checking account for payroll deposits or occasional bill payments, but they show little interest in additional products and are often sensitive to fees. Retention efforts should be efficient and low-cost: simple mobile-banking tools, periodic fee

waivers, and automated reminders about entry-level savings or credit options can help keep these customers connected without significant expense.

2.7.2. Applications and Benefits of Customer-Level Prediction

Customer-level prediction is a powerful tool that enables businesses to analyze and forecast individual customer behaviors, preferences, and potential value [51]. By leveraging predictive analytics, organizations can make data-driven decisions that enhance customer satisfaction and drive business growth. These predictions empower companies to tailor their strategies, optimize resources, and create highly personalized experiences for their customers.

Customer-level prediction provides businesses with multiple strategic advantages. It enables targeted marketing by allowing companies to tailor their campaigns according to customer segments, offering premium deals to high-value customers while directing cost-effective promotions to lower-level customers. It also supports customer retention, as businesses can identify clients at risk of leaving and implement proactive measures such as personalized communication or loyalty incentives to maintain engagement. Resource optimization is another benefit, with high-value customers receiving priority attention and services, ensuring that efforts are focused where they matter most. Additionally, it drives revenue growth by recognizing potential high-value clients and encouraging medium-level customers to increase their activity. Finally, understanding customer behavior allows for product and service customization, helping businesses refine or create offerings that better align with the preferences and needs of different customer segments.

2.7.3. Customer-Level Prediction in the Banking Industry

Customer-level prediction is a powerful approach for financial institutions looking to tailor their marketing strategies, enhance customer service, and optimize their revenue streams [52]. In the banking business, Understanding the various customer tiers high, medium, and low-level is vital for personalized interaction and efficient resource allocation. Here is an in-depth exploration of how customer-level prediction works in the banking industry and its impact:

Customer-level prediction involves using data analytics and machine learning models to classify customers into different categories based on their behaviors, transaction history, and financial profiles.

In the banking industry, predicting customer levels allows banks to group customers according to their financial significance and engagement. High-level customers are those who contribute the most through substantial deposits, investments, loans, and frequent or large transactions. Medium-level customers bring in moderate revenue and have the potential to increase their activity if supported with personalized offers or incentives. Low-level customers show limited interaction with banking services and generate minimal revenue, making them candidates for low-cost engagement strategies aimed at improving retention or encouraging greater participation.

To predict customer levels effectively, banks must gather and analyze a variety of data points. This includes a detailed record of deposits, withdrawals, and payment patterns; the range and frequency of products used such as savings accounts, credit cards, loans, mortgages, and investment services; personal details like age, income, and occupation that reveal financial capacity; the frequency of interactions with digital channels or customer service representatives; behavioral signals such as responses to marketing offers, interest in specific banking products, and reliability in repaying loans; and finally the individual's credit score and overall risk profile, which indicate lending risk and financial stability. By evaluating these factors together, the bank can form a clear picture of each client's habits and potential, allowing it to classify customers as high, medium, or low value with greater confidence. By classifying customers in this way, banks can better tailor their strategies to maximize customer lifetime value and improve customer retention.

2.7.4. Techniques for Customer-Level Prediction

Accurately predicting customer levels is crucial for banks to optimize their strategies for engagement, retention, and profitability. By using advanced data-driven techniques, financial institutions can better understand customer behavior, preferences, and needs. These methodologies enable banks to segment their customers into high, medium, or low categories, allowing them to offer personalized products and services [53]. For instance, high-value customers can be targeted by means of loyalty schemes and premium services, while medium and low-value customers may benefit from cost-effective promotional offers or financial literacy programs. Machine learning algorithms provide powerful tools to identify patterns and trends in customer data, while scorecard systems provide a straightforward and interpretable way to assign customer scores. Furthermore, proper data preprocessing ensures the accuracy and reliability of predictions, reducing biases and errors in decision-making. By integrating several approaches,

banks can create a comprehensive approach to customer segmentation that drives sustainable growth and competitive advantage in the industry. To categorize customers into high, medium, and low levels, banks use different machine learning techniques.

2.7.5. Challenges in Customer-Level Prediction

Customer-level prediction, while highly beneficial, comes with its set of challenges that businesses must navigate to achieve accurate and actionable insights [54]. Reliable models and high-quality data are necessary for precise and useful customer-level predictions and continuous adaptation to evolving customer behavior. Additionally, privacy concerns, the complexity of model development, and the ability to keep up with changes in the financial landscape pose significant challenges for banks.

Ensuring that predictive models provide accurate and useful insights requires addressing several key challenges. Protecting customer data is a top priority, and banks must comply with regulations like GDPR and CCPA while implementing strong security measures to safeguard financial information. Maintaining high-quality data is also essential, as incomplete or inaccurate records can lead to incorrect predictions; keeping customer information current and comprehensive supports effective segmentation. Managing model complexity is another concern, since sophisticated machine learning models often need substantial computing resources and expert knowledge, and there is a risk of overfitting, where a model performs well on historical data but fails with new data. Additionally, customer behavior can change over time due to shifts in economic conditions, personal circumstances, or financial priorities, making it necessary for models to be regularly updated to reflect these changes and remain reliable.

2.8. Machine Learning

Machine learning is commonly divided into four main types: supervised, unsupervised, semi-supervised, and reinforcement learning [55]. In supervised learning, the model is trained using data that includes both inputs and their corresponding outputs, allowing it to predict results for new, unseen data. Unsupervised learning deals with data that has no predefined labels, helping to uncover patterns, relationships, or clusters within the dataset. Semi-supervised learning uses a mix of labeled and unlabeled data, which can improve accuracy when labeling all data is costly or difficult. Reinforcement learning is a process where an agent gains experience by interacting with its surroundings, using the feedback of rewards or punishments to progressively refine its decision-making approach.

Machine learning, specifically in the form of classification algorithms, plays a pivotal role in customer segmentation for enhancing customer relationship management (CRM). The bank can utilize machine learning models, such as decision trees, random forests, or neural networks, to predict and categorize customers into high, medium, or low-value segments based on their transaction history, demographics, and behavioral data [56]. These predictive models can provide valuable insights into customer behaviors, helping Bunna Bank to tailor its offerings and marketing strategies to different customer groups. By implementing machine learning, Bunna Bank can achieve more effective CRM by personalizing services for high-value customers, offering targeted promotions or financial products to medium-level customers, and designing engagement programs for low-level customers. Machine learning enables the bank to optimize resource allocation by focusing marketing and customer service efforts where they are most likely to yield the highest return, thus improving customer satisfaction and retention.

Moreover, machine learning algorithms can identify patterns and trends in customer behavior that might not be evident right away, enabling Bunna Bank to anticipate and respond to customer demands and anticipate changes in their financial activities. This predictive capability not only improves operational efficiency but also enhances decision-making related to credit risk evaluations, loan approvals, and retention of customers' strategies.

2.8.1. Machine Learning Techniques

In machine learning, the "No Free Lunch" theorem highlights that there is no one algorithm that performs best for all types of problems [57]. This concept is particularly important in supervised learning, where models are trained on labeled data to make predictions. In this paper, supervised machine learning methods were chosen because the dataset contains labeled information, allowing the use of algorithms designed for predictive tasks. It is not accurate to claim that one algorithm, such as neural networks, will always outperform others like decision trees, as performance depends on multiple factors, including the size, quality, and structure of the dataset. Various machine learning techniques have been applied in prior studies addressing similar problems, such as predicting customer levels, demonstrating that different approaches can be effective under different circumstances.

A variety of machine learning methods have been applied in past studies for predicting outcomes at the individual customer level. Based on the previous literature in this area five supervised machine learning techniques will be compared when aiming to predict customer level, the five

techniques are Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Nighbor (KNN), and Deep Neural Network (DNN).

2.8.1.1. Logistic Regression (LR)

Logistic regression is a type of supervised machine learning method used primarily for classification problems, when the dependent variable is categorical. It models examines how one or more independent factors are connected to a categorical outcome by calculating the likelihood of each outcome through a logistic function [58]. This method is widely employed in various fields, including healthcare, finance, and marketing, to predict outcomes such as disease presence, loan approval, or customer level. For scenarios involving, when the outcome variable has more than two categories, multinomial logistic regression extends binary logistic regression to handle situations with multiple classes. This approach generalizes the logistic model to handle multiclass problems, predicting the probabilities of different possible outcomes concentrate on a set of independent variables. It is particularly useful when the dependent variable has more than two unordered categories, such as customer levels ((Low, Medium, High) or product preferences.

For example, one variable in the dataset could be the customer's "Balance," while another might be the "Tenure" of the customer. The logistic regression model will then estimate how these factors balance and tenure affect the likelihood of a customer falling into a particular level.

Logistic Regression Equation and Sigmoid Function

Provided below is the Logistic Regression function:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\beta 4X4+\ldots+\beta nXn)}}$$

Where:

β0 to βn are various coefficients

X1 to Xn are the independent variables impacting the dependent variable

P(Y=1|X) is the probability of a positive outcome.

Notice to the exponent in the function this is precisely the aspect where linear regression becomes relevant:

β0+β1X1+β2X2+β3X3+β4X4+...+βnXnβ0+β1X1+β2X2+β3X3+β4X4+...+βnXn

The logistic regression model uses a sigmoid function, which produces outputs ranging from 0 to 1. The graph of this function shows a gradual, smooth change between the two levels, making it well-suited for predicting outcomes that have only two possible categories. Based on the value

of the variables, the output can be interpreted as a probability, which can then be threshold to predict a class label of 1 or 0. This corresponds to the probability of an event occurring, such as a customer leaving the company or continuing with it.

To predict outcomes with more than two categories, multinomial logistic regression is an extension of binary logistic regression [58]. It models the probability of each possible class based on input variables, with one category typically chosen as a reference, and estimates the odds of the other categories relative to it. The model outputs probabilities for each class, which can then be used to assign a predicted category by selecting the highest probability. This method is widely applied in customer segmentation; allowing businesses to classify customers into groups such as Low, Medium, or High level customer based on factors like deposit history, demographics, and engagement, enabling targeted marketing strategies and personalized services.

2.8.1.2. Random Forest (RF)

Random Forest (RF) is a versatile and user-friendly supervised machine learning technique that comprises a group of decision trees. It can be applied to both classification and regression tasks. The algorithm operates by creating a "forest," which is an ensemble of multiple decision trees, typically constructed using the bagging approach to improve accuracy and stability [59]. The bagging method is a technique that combines many learning models to improve overall outcomes. The class target that is the highest selection result of the target output by each decision tree is the output of this model, which is made up of many decision trees.

Combinations of tree predictors known as Random Forests (RF) distribute equal values of vector samples to every tree in the forest, with each tree relying on the values of randomly chosen vector samples. The generalised inaccuracy of a forest and its trees depends on the strength of each individual tree in the forest and the association between them [59].

Numerous decision trees are constructed by the Random Forest algorithm; the number of trees to be constructed is a parameter that can be chosen during the learning stage. A random subset of features from a selected training set with replacement is used to learn each decision tree. The votes cast by each individual tree determine the final result. A tree algorithm is used to classify the input data samples in order to construct each decision tree. The testing data will then be categorised using each tree. Labeling any testing data is a decision made by each tree. We refer to this designation as a vote. Following the collection of votes, the forest ultimately determines

the classification outcome of the testing data, returning the most popular class [59]. This scenario is illustrated in Figure 2.1 diagrammatically.

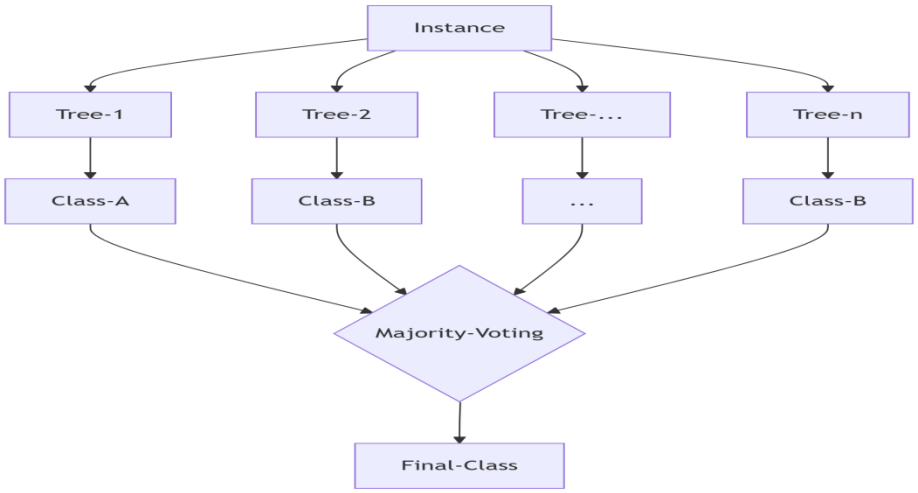

Figure 2.1 Ensemble Classifiers for Random Forest [59]

A Random Forest of the observations in the training dataset may be chosen by the model builder while creating a single decision tree. Additionally, only a small portion of the available factors are taken into consideration at each node in the decision tree construction process. The computational requirement is greatly decreased as a result. It can handle a large dataset with increased dimensionality and is also appropriate when there are few observations and a lot of input variables. RF algorithms identify outliers and create trends [59]. The generalisation error of the forest and its trees depends on the strength of each individual tree and the correlation between them. To complete its categorisation, the RF Algorithm follows this series of actions.

Step 1: Decide how many trees (T) to grow.
Step 2: Select the m variables that will be used to divide each node. The input variables are m ~ M, M.
Step 3: Grow T trees in step three. When growing each tree do

- Construct a bootstrap sample of size n sampled from Sn with the replacement and grow a tree from this bootstrap sample.
- At each node select m random variables and use them to find the best split.

Step 4: The tree should grow as much as possible without being pruned.

Step 5: Gather votes from each tree in the forest to classify point X. Then, utilise majority voting to determine the class label.

2.8.1.3. K-Nearest Neighbor (KNN)

One of the most basic machine learning techniques is the K-Nearest Neighbour (KNN) classifier. The KNN algorithm's goal is to predict the classification of a new sample point by using a database whose data points are divided into many distinct classes. Classifying an object involves examining its closest examples. A majority vote and any distance metric can be used to make the measurement. K specifies the number of neighbours that will be counted for voting. An object will be assigned the same class as its closest example, according to the simple statement K=1. As K grows, we must categorise a specific instant according to how similar all of the K instances [60]. But as a rule of thumb, take $K \leq \sqrt{n}$, where n is total number of dataset.

Generally speaking, the study begins with a collection of data, each of which belongs to a recognised class. Based on the established classifications of the dataset's observations, it will then be able to predict the class of a new data point [60]. The process of choosing the class of the new observation is known as the classification problem.

We must find a method of object comparison before we can determine whether two observations are comparable. The issue with this is that there are numerous ways to measure similarity between different sorts of data, such as a number, colour, location, or a true/false (boolean) response to a question. To guarantee that we can compare observations, the first step is to pre-process the data in the database. After that, observations are transformed into points in space, and we can use a suitable metric to interpret the distance between them as their similarity.

How to determine which database observations are sufficiently similar to new observations for us to consider their classification when classifying the new observation is the other issue at hand [61]. One of the popular used metrics is the Euclidean distance. The Euclidian distance between two instances (X1, X2, X3,…Xn) and (U1, U2, U3,….Un) is given by the following formula:

$$\sqrt{(\mathrm{X1}-\mathrm{U1})+\ (\mathrm{X2}-\mathrm{U2})+\cdots+(\mathrm{Xn}-\mathrm{Un})}$$

Where

X1, X2, X3, and Xn are predictors of instance #1

U1, U2, U3, and Un are a predictor of instance #2.

2.8.1.4. Support Vector Machine (SVM)

One supervised learning model that can be applied to classification, prediction, and clustering issues is Support Vector Machines (SVM). An SVM creates a model that can categorise fresh observations into one or more classes based on a collection of input observations and related binary outputs.

The SVM model divides the observation sets linearly by mapping the training observations as points in space [62]. It is determined that the margin surrounding this linear separation should be as wide as feasible. A nonlinear partition in the output space is equivalent to a linear hyper-plane partitioning in a higher dimensional space. Any mathematical surface can define this higher dimensional segmentation, which is referred to as the SVM kernel. Linear, quadratic, polynomial, and Gaussian radial basis functions are a few of the most popular kernels. Using a hyper-plane, an SVM attempts to linearly segregate a set of observations. A straightforward line can be created to divide the data in two dimensions. Figure 2.2 shows that the data is linearly separated by the bold line. The distance from the middle line to the sidelines shows the margins.

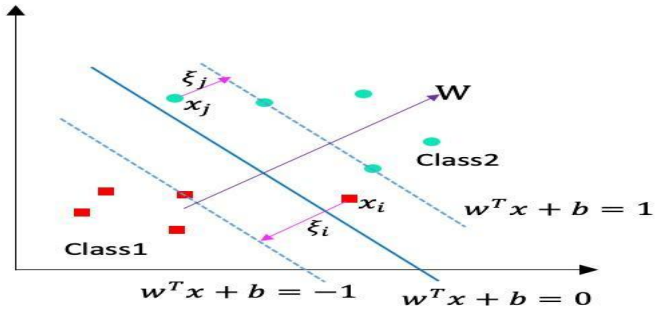

Figure 2.2 Support Vector Machine classification [62].

As illustrated in Figure 2.2 above, any hyperplane may be expressed as the collection of locations X that fulfil wTX + b = 0, where b is the hyperplane's offset, wTX + b = 0 from the original point along the direction of w, and w is a normal vector perpendicular to the hyperplane. We display the labels as yi € {1, −1} given the labels of data points X for two classes (class 1 and class 2). In the meantime, we use the sign of the function f (X) = sin (wTX + b) to categorise data X into

either class 1 or class 2 given a pair of (wT, b). Therefore, the following equation 2.1 can be used to define the linear separability of the data X in these two classes.

$x_i \cdot w + bx + 1$, for $y_i = +1$ ---------- (2.1)

$x_i \cdot w + bw - 1$, for $y_i = -1$ ---------- (2.2)

The above two equations can be combined and form the following equation 2.3.

Yi $(x_i \cdot w + b)x1$ _---------- (2.3)

Furthermore, $r = (wT\ x + b)/||w||$ can be used to calculate the distance between the data point and the separator hyperplane, $wT\ x + b = 0$. The data points that are closest to the hyperplane are referred to as support vectors. The margin of the separator, which is simply $2/||w||$, is the distance between support vectors, as shown in Figure 3.2. The quadratic optimisation problem is expressed in the following equation to solve linear SVM [62].

$$\text{Minimize } w,b \left(\frac{1}{2}||w||^2\right)$$

$$\text{Subject to } y_i(w^T x_i + b) \geq 1 \text{ ---------- (2.4)}$$

In the case of linearly separable data, the solution is represented as a linear combination of only the data points that are on the margin of the optimal separating hyperplane; other data points are discarded. Because the SVM learning process typically selects a small number of support vectors, the number of features found in the training data has no bearing on the model complexity of SVM. Because of this, SVMs are ideally suited to learning problems in which there are a lot of characteristics compared to training cases.

2.8.1.5. Deep Neural Network (DNN)

Neural network research and applications are particularly active in the classification domain. An ANN is a computer model that draws inspiration from the information processing capabilities of biological neural networks found in the human brain [63]. ANN, also known as processing elements is nodes that are joined to build a network of nodes to form artificial neural networks. Physically cellular systems, neural networks are able to learn, store, and use experimental knowledge.

Now, the amount of data increases, and because of this, it becomes increasingly difficult for data analysts to analyse such vast amounts of data. Consequently, machine learning systems for

customer level prediction were crucial. By learning from historical data or situations, a variety of research methods ensure that machine learning is effective in making predictions [63]. Deep learning models address another learning worldview in artificial intelligence (AI) and machine learning (ML).

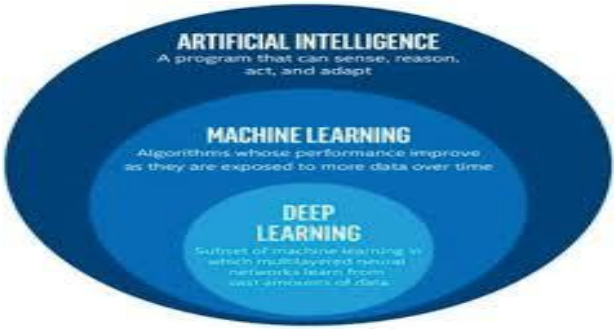

Figure 2.3 Artificial Intelligence, Machine Learning, and Deep Learning [63].

A deep neural network helps to deal with the big volume of bank data more realistically and accurately because of layering architecture. Layering architecture of customer level prediction the model prompts hierarchical learning by shaping several hidden layers.

Deep Learning (DL), a subset of Machine Learning, employs layered algorithms to process data, simulate human reasoning, and create abstract representations. These models use multiple layers to interpret information, enabling capabilities such as speech comprehension and visual recognition. Data flows sequentially through these layers, with each layer's output serving as the input for the next. The first layer is called the input layer, the final layer is the output layer, and all layers in between are termed hidden layers. Deep Neural Networks (DNNs) are particularly effective in analyzing complex datasets, including customer behaviors and transaction histories, to classify customers into categories such as high, medium, or low value. By leveraging multiple hidden layers and non-linear activation functions, DNNs improve the accuracy of identifying high-value and at-risk clients, supporting targeted marketing, customized services, and optimal resource allocation. Furthermore, these networks can promote financial inclusion by revealing underserved market segments and enabling tailored solutions such as microloans. Integrating DNNs into Bunna Bank's operations can therefore strengthen CRM practices, enhance competitive advantage, and drive economic development in Ethiopia through advanced technological adoption.

Table 2.2 Machine learning classification algorithms and their pros and cons

Algorithm	Strengths	Limitations
Logistic Regression (LR)	Straightforward to implement and interpret; provides clear probability estimates; efficient for linearly separable data.	Performs poorly when the true decision boundary is highly non-linear; results can be distorted by multicollinearity or extreme outliers.
Random Forest (RF)	Combines many decision trees to improve accuracy; handles non-linear relationships and high-dimensional features well; resistant to overfitting and noisy data; offers measures of variable importance.	Less transparent than a single tree; training and prediction can be memory-intensive for very large datasets.
k-Nearest Neighbours (KNN)	Conceptually simple with no explicit training stage; naturally accommodates multi-class problems; adapts to complex boundaries given enough data.	Prediction slows as dataset size grows; sensitive to the choice of kk and distance metric; performance drops in high-dimensional spaces.
Support Vector Machine (SVM)	Effective with high-dimensional inputs; kernels allow flexible modelling of non-linear boundaries; generally robust to overfitting when well-regularised.	Parameter tuning and kernel selection require care; training time can be long on large datasets; probability outputs need extra calibration.
Deep Neural Network (DNN)	Capable of modelling very complex and non-linear patterns; well suited to large volumes of data and unstructured inputs such as images or text; highly adaptable architectures.	Demands substantial labelled data and computational resources; training can be lengthy; internal operations are difficult to interpret.

2.9. Related Works

The prediction of customer level for customer relationship management (CRM) has been the subject of numerous studies worldwide; most of these included using machine learning algorithms to customer data in order to anticipate customer levels and customer attrition. This section discusses some of the studies.

The study H. Amnur [64] focuses on the integration of Customer Relationship Management (CRM) with Machine Learning (ML) technologies to enhance customer interaction and retention strategies. CRM is presented as an essential tool for businesses to foster strong customer relationships, improve profitability, and implement targeted marketing efforts. By leveraging data-driven approaches, CRM systems enable companies to identify, engage, and retain their most valuable customers. This study specifically examines how ML algorithms can be utilized to support these goals, providing deeper insights into customer behavior and optimizing resource allocation for improved business outcomes. The study is based on a dataset from Bunna Bank, consisting of five years of customer data related to special savings products. This dataset serves as the foundation for analysis, focusing on identifying patterns and trends in customer behavior. By employing Support Vector Machines (SVM), a robust ML algorithm known for its ability to handle nonlinear data relationships, the study aims to improve the capacity for prediction of CRM systems. SVM is particularly effective in classifying and segmenting customers based on their value to the organization.

To measure customer value, the study applies two key metrics: Lifetime Value (LTV) and RFM (Regency, Frequency, and Monetary) analysis. These methods provide a comprehensive understanding of customer contributions over time, enabling the bank to categorize its clients into three distinct segments: high-value, medium-value, and low-value customers. The primary objectives are to identify high-value customers who generate the most profit, predict customer responses to retention strategies, and allocate resources effectively to maximize profitability. The findings highlight the importance of focusing on high-value customers, as they provide the greatest long-term benefits to the organization. The study emphasizes the use of incentives, such as gifts and bonuses, to strengthen relationships with these customers and ensure their loyalty [64]. By integrating predictive analytics into CRM strategies, the bank can anticipate customer needs, enhance decision-making, and implement retention techniques more effectively. The study underscores the potential of ML-powered CRM systems to transform customer management practices. By using algorithms like SVM, organizations can identify valuable customers, optimize

profits, and maintain competitive advantages in the market. The integration of CRM and ML not only increases the efficiency of marketing and retention efforts but promotes long-term growth strengthening customer relationships.

The limitation of the study H. Amnur [64] exclusively employs Support Vector Machines (SVM) without comparing its performance against other machine learning models like Random Forest, Logistic Regression, or Neural Networks. This restricts the analysis from determining whether alternative algorithms could yield better results for CRM tasks. Furthermore, the paper does not report performance metrics such as accuracy, precision, recall, or F1 scores for the SVM model, making it difficult to assess the effectiveness of the proposed solution objectively.

The study by N. Singh et al. [65] explores the application of machine learning (ML) in Customer Relationship Management (CRM) to improve customer identification, attraction, retention, churning prediction, and customer development. It examines 35 studies from 2016 to 2020 in detail, ocusing on the types of ML techniques used in CRM and their impact on customer behavior analysis. The authors identify supervised learning (e.g., Decision Trees, Support Vector Machines) as the most utilized approach (48.48%), followed by deep learning techniques (27.27%) and unsupervised learning (15.15%). The study emphasizes the transition from traditional CRM methods to advanced ML-driven approaches, emphasizing deep learning for better predictive analytics and customer insights. Key findings include the use of RFM (Recency, Frequency, and Monetary) models for customer segmentation and the adoption of algorithms such as K-Means, Random Forest, and Gradient Boosting for specific CRM tasks. The study also underscores the significance of deep learning in improving customer lifetime value and managing churn prediction.

However, despite its comprehensive scope, the review by Singh et al. [65] may lack technical depth. The study provides a broad overview of ML algorithms in CRM but does not delve into the specifics of model architectures, their practical implementation challenges, or a critical assessment of their strengths and limitations within real-world systems. Additionally, the absence of concrete case studies or real-world examples limits its applicability, making it challenging for professionals to comprehend how these theoretical concepts translate into actionable strategies. Furthermore, the paper could be criticized for overemphasizing generalities about the potential of machine learning in CRM without offering novel insights or perspectives, which may diminish its value to readers seeking a deeper, more practical understanding of the topic.

The study by B. Bhattacharjee [66] evaluates the application of machine learning (ML) models to classify customer sentiments in the banking sector. It highlights the importance of sentiment analysis in improving customer service and decision-making by examining feedback from various sources like surveys, social media, and app reviews. The study compares conventional ML algorithms (Logistic Regression, Naive Bayes, Support Vector Machines, Random Forest) with advanced deep learning models (Long Short-Term Memory Networks, BERT).

The dataset consists of approximately 100,000 customer feedback entries categorized by services such as loans, online banking, and customer support. The performance of each algorithm is measured using metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. The results highlight that deep learning models surpass conventional algorithms in handling complex and nuanced sentiments. BERT achieves the highest accuracy (88%), followed by LSTM (85%), Random Forest (83%), SVM (82%), Logistic Regression (80%), and Naive Bayes (79%). BERT also excels in recall, precision, and F1 score due to its bidirectional contextual understanding.

Despite its valuable comparative analysis, the study by Bhattacharjee [66]may suffer from several weaknesses. A primary limitation is its lack of a clearly stated objective, which could confuse readers and undermine the impact of its conclusions If the machine learning algorithms or approaches are not explained in sufficient detail, readers may struggle to understand the methodology. Additionally, the study might overlook critical aspects of machine learning ethics and fairness by not addressing potential biases in the data or models. Furthermore, if the interpretability of the models is not considered, especially in customer-related applications where transparency is vital, it could be seen as a significant oversight.

The study by V. Anandi & M. Ramesh [67] focuses on analyzing customer relationship management (CRM) using machine learning techniques to improve customer retention and satisfaction. The research compares various machine learning models, including Random Forest, AdaBoost, Decision Trees, Support Vector Machines, and others, to predict customer churn and improve CRM systems. Random Forest and AdaBoost showed the highest performance with a 97% accuracy rate. The study emphasizes that understanding customer behavior through website traffic, feedback analysis, and addressing errors is critical to improving customer satisfaction. These studies underline the importance of leveraging advanced analytical methods to optimize customer segmentation and retention strategies in CRM.

Despite its valuable exploration of machine learning in CRM, the study by Anandi and Ramesh [67] has several significant limitations. A primary drawback is its reliance on a dataset from a single website's traffic logs, which restricts the findings' applicability to other industries or broader contexts and limits the generalizability of the results. Furthermore, the analysis shows only weak correlations between key variables, such as good traffic and customer satisfaction (CSAT) or exception traffic and CSAT, raising concerns about the robustness of the conclusions. The study also focuses heavily on quantitative data from website logs, overlooking qualitative aspects such as customer emotions, preferences, and cultural nuances that are integral to understanding satisfaction. Furthermore, while Random Forest and AdaBoost are identified as top-performing models, the study lacks a detailed explanation of why these models outperform others, leaving questions about their context-specific utility. The hypotheses presented are broad and somewhat overlapping, which could dilute the research's focus, and the static dataset fails to capture dynamic customer behaviors or seasonal trends, both of which are critical in real-world CRM scenarios. Addressing these issues would enhance the study's practical relevance and reliability.

The study by P. Kagwa [56] examines how machine learning (ML) algorithms transform CRM systems by improving customer engagement, retention, and operational efficiency. Techniques like Decision Trees, Support Vector Machines (SVM), K-Means Clustering, and Neural Networks are highlighted for their ability to enhance customer segmentation, predict churn, and automate routine tasks. For instance, SVM achieves over 85% accuracy in churn prediction, while Long Short-Term Memory (LSTM) models provide approximately 90% accuracy for sequential customer behavior prediction. K-Means Clustering effectively groups customers for tailored marketing, improving loyalty by 15%. The study emphasizes ML's transformative role in both developed and developing economies, citing increased customer retention in Kenya (28%) and improved loyalty in South Africa (26%). By using these algorithms, companies may examine enormous datasets, find hidden trends, and make data-driven decisions that will ultimately increase customer satisfaction and foster long-term relationships. The integration of ML into CRM systems also supports resource optimization through automation, further driving efficiency and profitability.

Despite these valuable insights, the study by P. Kagwa [56] is subject to several limitations. First, it relies on a desk study methodology using secondary data, which limits real-world validation of findings and practical applicability. Second, the focus on Rwanda and select global examples restricts the generalizability of the conclusions to other regions with diverse cultural, economic,

and technological contexts. Additionally, the study does not adequately address the challenges of implementing ML in CRM, such as data quality issues, algorithmic complexity, and computational resource demands. Moreover, the limitations of specific ML techniques, such as the scalability challenges of K-Means Clustering for high-dimensional data or the potential overfitting of Neural Networks, are underexplored. The lack of a comparative analysis between ML algorithms limits understanding of their suitability for different CRM contexts. Addressing these gaps would enhance the study's relevance and applicability across broader settings.

The study by A. Khaliq et al. [68] examines the role of machine learning (ML) in predicting customer churn in the banking sector. It evaluates the performance of multiple classifiers, including Random Forest (RF), Support Vector Classifier (SVC), XGBoost (XGB), LightGBM, and Logistic Regression (LR). Through feature selection and oversampling, the study tackles this problem using a churn modeling dataset that has an imbalance between retained and churned customers. The results show that Random Forest achieves the highest accuracy at 95%, followed by Light GBM with 80% and XGB with 79%. The feature selection process identified "Age," "Number of Products and "Is Active Member" as the most important predictors. The study emphasizes the potential of ML to improve early detection of customer churn, enabling proactive retention strategies. The findings demonstrate the importance of balancing datasets, using effective feature selection, and optimizing hyper parameters to enhance model performance. This work contributes to the growing body of research on ML-based churn prediction, offering insights into the practical application of these methods in the banking industry.

Despite its valuable insights, the study by Khaliq et al. [68] is subject to certain limitations. A primary concern is its reliance on a relatively small dataset of 10,000 samples, which may lack the complexity and scale of real-world banking data, potentially affecting the model's generalizability and robustness. Additionally, the dataset's imbalanced nature (with churned customers accounting for only 20%) necessitates oversampling, which, while effective, can sometimes lead to overfitting and reduced generalizability. The study also lacks a detailed analysis of why certain models, like Random Forest, outperform others in this context. Furthermore, while feature selection is discussed, the exclusion of potentially relevant features, such as customer feedback or transactional trends, might limit the model's predictive power. Finally, the research does not address the challenges of deploying these ML models in real-time banking environments, such as computational costs, data security concerns, and model interpretability. Addressing these gaps would enhance the study's practical applicability and robustness.

The study by O. Id et al. [69] explores customer retention in commercial banking as a classification task using machine learning (ML) methods. The researchers used a number of different algorithms, including logistic regression, decision trees, random forests, support vector machines (SVM), k-nearest neighbors (k-NN), and the naive Bayes algorithm, to predict customer churn. The models were assessed on a dataset of 10,000 bank customers, with key features such as age, credit rating, and service usage, using metrics like accuracy, precision, recall, and AUROC. Among the models, random forests demonstrated the highest predictive performance with an AUROC of 0.8609, achieving superior precision, recall, and balanced accuracy compared to other approaches. The study highlights that predictive modeling can significantly reduce customer churn by identifying at-risk customers, allowing banks to intervene with retention techniques such as discounts and special offers. Furthermore, it emphasizes the potential for ML-based techniques to enhance cost efficiency and profitability in the banking sector.

However, the findings of O. Id et al. [69] are tempered by several methodological shortcomings. A significant limitation is the dataset's pronounced imbalance, with churn cases representing only 20.37% of the total. Despite the use of the ROSE technique to address this, the risk of model bias toward the majority class remains a concern that could affect the reliability of the results. The feature set relied heavily on financial and demographic data, neglecting behavioral factors that are crucial for churn prediction. Additionally, models like k-NN and naive Bayes performed poorly, indicating potential issues with feature selection or model tuning. While random forests showed high accuracy (AUC-ROC 0.8609), the study lacked validation on real-world data, raising concerns about overfitting.

The study by Z. Revesai et al. [70] investigates customer churn in Zimbabwean banks using classical machine learning algorithms (e.g., Logistic Regression, Decision Trees, Random Forest, XGBoost) and deep neural networks. The study emphasizes the importance of customer retention for minimizing revenue losses, reducing acquisition costs, and protecting reputation. Using demographic and transactional data, the authors developed predictive models and evaluated their effectiveness based on accuracy, precision, recall, and ROC-AUC. Random Forest attained the highest accuracy (70.7%) and recall (99.4%), indicating strong predictive performance, while Logistic Regression and XGBoost also performed well. Deep learning showed moderate performance with a recall of 81.7% and accuracy of 62.6%, making it less competitive in this context. The article emphasizes the practical utility of predictive models for identifying at-risk customers and implementing retention strategies.

Despite these contributions, the study by Revesai et al. [70], has several limitations. First, the dataset was limited in size (the dataset used for customer churn analysis included 1,000+ customer records from Zimbabwean bank) and scope, potentially affecting the generalizability of the models. Additionally, the feature set focused primarily on demographic and banking-related data, overlooking other influential factors such as customer sentiment and behavioral patterns. While Random Forest demonstrated the highest accuracy (70.7%) and recall (99.4%), the study lacked detailed exploration of hyper-parameter optimization and model interpretability, which could enhance predictive insights. The deep learning model underperformed compared to classical algorithms, likely as a result of limited training data or feature engineering. Furthermore, the economic implications of churn interventions were not extensively analyzed, limiting the practical application of the results. Addressing these gaps, including expanding data sources and refining model architectures, could improve the study's relevance and reliability for broader contexts.

The study by A. Todupunuri [71], applies machine learning techniques to predict Customer Lifetime Value (CLV) using a dataset of over 100,000 banking customers. The dataset includes customer demographics, account activity, transaction history, and product usage. The article uses algorithms such as Linear Regression, Decision Trees, Random Forests, Gradient Boosting Machines (GBM), Neural Networks, and K-means Clustering. The Random Forest and GBM algorithms attained the best predictive accuracy, with R-squared (R^2) values exceeding 0.85, and Mean Absolute Error (MAE) improving by over 20% compared to traditional methods. These models were evaluated using AUC and R^2, demonstrating strong performance in customer segmentation and CLV prediction.

However, a significant limitation of Todupunori's work [71] is the lack of a detailed explanation for the model selection process. Furthermore, the justification and interpretation of key evaluation metrics, such as AUC and R^2, are insufficient, which is essential for comprehensively validating the effectiveness and reliability of the presented machine learning models. Without these proper evaluation criteria, the reliability and robustness of the model could be questioned. Additionally, if the article does not address model interpretability, it could be a significant weakness, especially considering that deep learning models in particular can be complex and difficult to interpret. Banking institutions require transparent models for decision-making and regulatory compliance, and failing to discuss this aspect could limit the practical applicability of the proposed models.

The study by Y. W. Wardana & I. R. Pratama [72] examines how various factors Internet banking, product advantages, and customer relationship management (CRM) impact customer loyalty within the context of Bank Syariah Indonesia in Semarang City. The study collects data through surveys and questionnaires distributed to customers of Bank Syariah Indonesia. While the exact sample size is not explicitly mentioned in the summary provided, the analysis relies on customer responses, making it crucial to know the number of participants to gauge the representativeness and reliability of the findings. Statistical techniques, such as regression or correlation analysis, are used to evaluate the strength and significance of relationships between the variables (Internet banking, product advantages, and CRM) and customer loyalty. However, the article does not give explicit measures of accuracy, such as R^2 values, p-values, or confidence intervals, which are critical for assessing the statistical robustness and predictive power of the model.

A primary weakness of this study [72]is the lack of clarity regarding the sample size and dataset characteristics, which makes it difficult to assess the statistical validity and reliability of the results. Furthermore, the article omits essential statistical measures such as R^2 values, p-values, or confidence intervals. The absence of these metrics undermines the ability to evaluate the model's explanatory power, the significance of the findings, and the overall robustness of the conclusions. Moreover, the absence of detailed accuracy metrics, such as R^2 values, p-values, or confidence intervals, reduces the ability to evaluate the robustness of the statistical analysis. These weaknesses collectively impact the study’s broader relevance and the strength of its conclusions.

The study by M. Rahman & V. Kumar [73] focuses on leveraging machine learning (ML) techniques to predict customer churn in the banking sector. The study aims to identify factors contributing to customer attrition and implement predictive models to enhance customer retention strategies in banking. The research utilizes a publicly available dataset containing customer demographic, transactional, and behavioral information relevant to the banking sector. The dataset comprises 10,000 customer records with multiple features, such as account balance, transaction frequency, and credit scores, to enable the development of predictive models. The study implements various machine learning algorithms, including Decision Trees, Random Forest, Logistic Regression, and Support Vector Machines (SVM). Among these, the Random Forest model achieves the highest accuracy, with a reported accuracy of 89.3% in predicting customer churn. The model's performance is also evaluated using additional measures like precision, recall, and F1-score.

One of the main weaknesses of the study by M. Rahman & V. Kumar [73] is the reliance on small number of dataset, which may not encompass the full range of customer behaviors and patterns, leading to overfitting or reduced accuracy when applied to larger or more diverse customer populations. Additionally, a limited dataset can hinder the ability to validate the model effectively across various scenarios, reducing its applicability in actual banking environments. Additionally, the research does not provide a detailed discussion on feature engineering or preprocessing steps, which are crucial for understanding how the raw data was transformed for analysis. While the models achieve high accuracy, the article does not address the interpretability of these models, which is vital for practical implementation in the banking sector. Furthermore, the lack of a comparison with deep learning techniques leaves room for improvement in exploring advanced methods for churn prediction.

The study by P. Verma [74], explores the application of machine learning techniques to predict customer churn in the savings account segment of the banking industry. The research aims to assist banks in identifying at-risk customers and developing strategies to retain them. The study uses a dataset consisting of 100,000 customer records from the State Bank of India. The dataset includes customer demographic data, account activity, transaction history, and other behavioral attributes. These features form the basis for building and training predictive machine learning models. Several machine learning algorithms are applied, including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting. Among these, the Gradient Boosting model performs the highest accuracy, with a reported accuracy of 91.4% in predicting churn. The model's performance is also evaluated using additional measures like precision, recall, and F1-score.

The primary weakness of the study by P. Verma [74] is that the article does not explore advanced deep learning models, which could potentially improve predictive performance even further. While the accuracy metrics are promising, the study does not address the interpretability of the models, which is crucial for decision-making by bank managers. Furthermore, the article lacks a detailed discussion of preprocessing steps and feature engineering, making it difficult to replicate the study or assess its robustness.

The study by I. A. Adeniran et al. [75], explores the application of machine learning (ML) methods to forecast customer churn in the telecommunications industry. It emphasizes the importance of retaining customers in a highly competitive market by leveraging predictive analytics to determine at-risk customers and implementing effective retention strategies. The research utilizes a dataset consisting of 5,000 customer records from a telecommunications

company. The dataset includes customer demographics, call history, service usage patterns, and payment behavior, which serve as input features for model training and evaluation. The study uses several machine learning algorithms, including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting. The Gradient Boosting model achieves the highest predictive performance, with a reported accuracy of 87.5%. The research also evaluates the models using other metrics, such as precision, recall, and F1-score, to ensure a comprehensive assessment.

One of the main limitations of the study by I. A. Adeniran et al. [75] is the relatively small dataset size, which may decrease the generalizability of the findings to larger and more diverse customer populations. Additionally, the article focuses exclusively on conventional ML techniques and does not explore the potential of deep learning models, which could provide improved performance for complex, high-dimensional datasets. The study also lacks a detailed discussion on feature selection and preprocessing steps, which are critical for ensuring robust and reproducible results. Furthermore, while accuracy metrics are reported, the article does not provide insights into the interpretability of the models, which is essential for practical implementation in a business context.

Table 2.3. Summary Selected Related Work

Citation	Author(s)	Algorithm(s)	Dataset	Industry	Accuracy / Result
[64]	H. Amnur	Support Vector Machines (SVM)	5 years of customer data	Banking	Not reported
[65]	N. Singh et al.	Decision Trees, SVM, Deep Learning	35 studies reviewed (various)	Various	No specific accuracy reported (review paper)
[66]	B. Bhattacharjee	BERT, LSTM, Random Forest, SVM, Logistic Regression, Naive Bayes	100,000 customer feedback entries	Banking	BERT 88%, LSTM 85%, Random Forest 83%, SVM 82%
[67]	V. Anandi & M. Ramesh	Random Forest, AdaBoost	Website traffic logs	General CRM	Random Forest & AdaBoost: 97% accuracy
[56]	P. Kagwa	Decision Trees, SVM, K-Means, LSTM	Secondary data (various)	Various	LSTM ~90%, SVM > 85%, K-Means: loyalty improved by 15%
[68]	A. Khaliq et al.	Random Forest, SVC, XGBoost, LightGBM, Logistic Regression	10,000 customer churn cases (imbalanced)	Banking	Random Forest 95%, LightGBM 80%, XGBoost 79%
[69]	O. Id et al.	Logistic Regression, Decision Trees, Random Forest, SVM, k-NN, Naive Bayes	10,000 bank customers	Banking	Random Forest: AUROC 0.8609
[70]	Z. Revesai et al.	Logistic Regression, Decision Trees, Random Forest, XGBoost, Deep Neural Networks	1,000+ Zimbabwean bank customers	Banking	Random Forest 70.7% accuracy, 99.4% recall
[71]	A. Todupunuri	Random Forest, Gradient Boosting Machines (GBM), Neural Networks	100,000+ banking customers	Banking	$R^2 > 0.85$, MAE improved by 20%
[72]	Y. W. Wardana & I. R. Pratama	Regression, Correlation Analysis	Survey responses (size not stated)	Banking (Indonesia)	Not reported
[73]	M. Rahman & V. Kumar	Decision Trees, Random Forest, Logistic Regression, SVM	10,000 public customer records	Banking	Random Forest 89.3% accuracy
[74]	P. Verma	Logistic Regression, Decision Trees, Random Forest, Gradient Boosting	100,000 customer records (State Bank of India)	Banking	Gradient Boosting 91.4% accuracy
[75]	I. A. Adeniran et al.	Logistic Regression, Decision Trees, Random Forest, Gradient Boosting	5,000 customer records	Telecommunications	Gradient Boosting 87.5% accuracy

2.10. Summary of the Related Work

Customer Relationship Management increasingly relies on machine learning as a key tool for improving customer interactions, insights, enhancing customer segmentation, retention, and profitability. Research has shown the effectiveness of various ML techniques, such as Support Vector Machines (SVM), Decision Trees, and deep learning models, in determining high-value customers, predicting churn, and analyzing customer behavior. For example, one study [64] employs SVM with metrics like Lifetime Value (LTV) and RFM analysis to allocate resources effectively, though it is limited by its exclusive focus on SVM without comparing alternative models. Another review [65], highlights supervised learning techniques, such as Decision Trees and SVMs, alongside the potential of deep learning for predictive analytics, but lacks real-world applications and technical depth. Similarly, a study on sentiment analysis in banking [66] shows that deep learning models like BERT and LSTM outperform traditional methods in analyzing customer feedback but fails to address data biases and explainability. Other works, such as those on churn prediction [67] [68], emphasize the high accuracy of models like Random Forest and AdaBoost but face challenges related to limited dataset diversity and oversampling techniques, raising questions about robustness and scalability. Collectively, these studies underline ML's transformative role in CRM while exposing critical gaps, including the need for diverse datasets, real-world validation, and interpretable models to ensure ethical and effective applications.

However, this study aims to develop a robust predictive model for Customer Relationship Management (CRM) tailored to Bunna Bank in Ethiopia, by using a dataset of 64,638 records with 11 attributes, including a class attribute for customer level (high, medium, low). To address the imbalance in customer-level data, the Synthetic Minority Oversampling Technique (SMOTE) is applied, ensuring balanced class distributions for improved model performance. Leveraging machine learning approaches such as Logistic Regression, Random Forest, Support Vector Machines (SVM), and Deep Neural Networks, the study classifies customers effectively according to attributes like transaction history, account balances, product usage, and engagement patterns. By applying metrics to evaluate models such as accuracy, precision, recall, and F1-score, the research identifies the most effective algorithm for enhancing customer segmentation, retention, and profitability. This study fills gaps in Ethiopian CRM research while following international best practices, providing a scalable, data-driven solution for resource optimization and decision-making in the banking sector.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Overview

This chapter presents the overall research methodology used in the study, including the research design, system architecture, data preprocessing techniques, data description, correlation analysis, and evaluation metrics. The study follows a design science research approach, which emphasizes developing and evaluating practical solutions for real-world problems. In this case, the approach is applied to design and build a predictive model that supports customer relationship management in the banking sector. Each stage of the design process, problem identification & motivation, objective a solution, design & development, demonstration, evaluation, and communication are described to show how the study was systematically carried out.

All experiments were conducted using the Python programming language due to its strong data analysis and machine learning capabilities. The collected data was carefully preprocessed to handle missing values, reduce noise, and address class imbalance, ensuring accurate and reliable inputs for model training. Correlation analysis was performed to explore relationships among customer attributes and guide feature selection. The proposed system's architecture illustrates how data flows through various components to produce customer-level predictions. Finally, model performance was assessed using metrics such as accuracy, precision, recall, F1-score providing a clear evaluation of the model's effectiveness and reliability.

3.2. Research Design

A paradigm for problem-solving, Design Science Research (DSR) aims to strengthen human knowledge through the production of new artifacts. To put it simply, DSR aims to strengthen scientific and technological knowledge bases by producing innovative artifacts that address issues and enhance the environment in which they are implemented [76]. The overall methodology of the study follows a design science approach, where the main procedures involve collecting and organizing customer transaction data. This dataset includes customer-related information such as gender, age, occupation, number of debit transactions over six consecutive months, number of credit transactions over the same period, and current account balance. However, no personally identifiable information such as customer names or addresses was used, which has been organized and made available for report and administration purposes and has been utilized as input data for the model.

The final outcome of DSR encompass both the newly designed artifacts and design knowledge that has a fuller understanding of designed theories of why the artifacts enhance (or, disrupt) the appropriate scenarios for application. The foundations of the Design Science Research (DSR) paradigm can be found in both engineering and science [77]. It is essentially a paradigm for solving problems. Through the production of innovative artifacts and the generation of design knowledge through creative solutions to pressing issues, DSR aims to strengthen human knowledge [76]. As a result, the research paradigm has gained a lot of attention in the last 20 years, especially due to its potential to support organizations' capacity for innovation and the much-needed sustainability transformation of society.

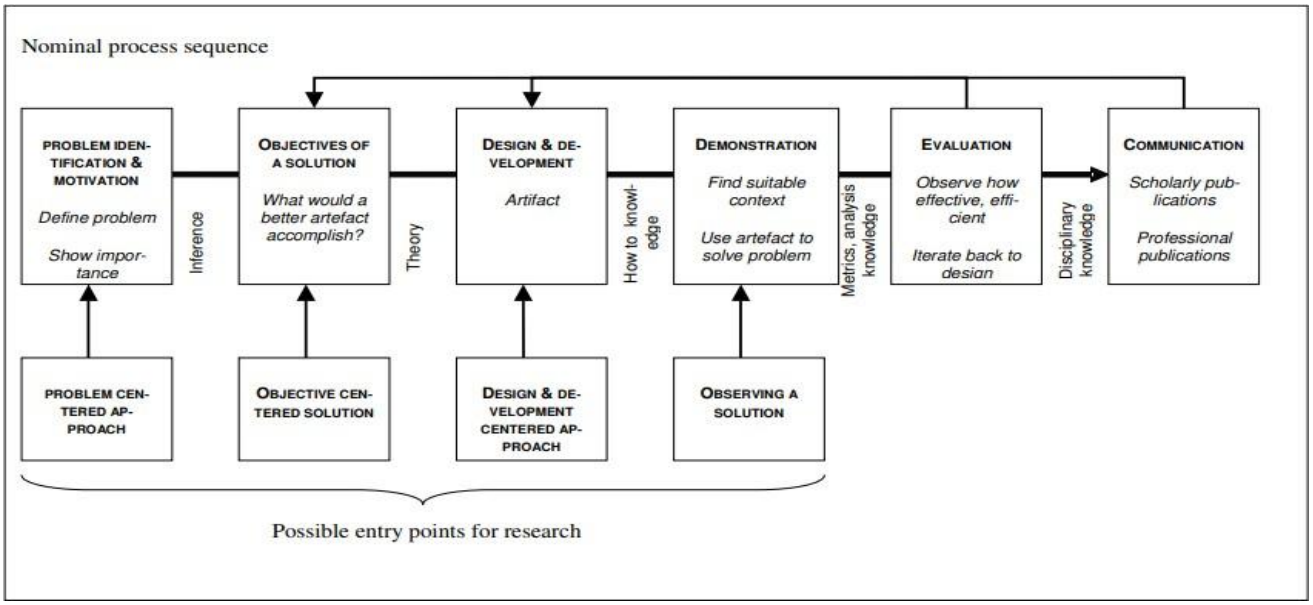

Figure 3.1 Design science research process (DSRP) model [78]

Although it is generally acknowledged that design science research generates design artifacts, the processes by which these artifacts are produced vary from one study to another. For instance, Peffers [78], developed a six-step approach to design science research. This study comprises the following steps: problem identification and motivation, objectives of the solution, design and development, demonstration, evaluation, and communication.

3.2.1. Problem Identification and Motivation

Define a problem and emphasize the significance of a solution in this activity. Chapter one of the thesis reports outlined the issue and provided evidence for the need for an answer under this step. Consequently, chapter one can discuss the specifics of this phase.

3.2.2. Objectives of a Solution

The second finding the goals of the suggested solution is the second step in the DSR process model, whereas identifying the problem in general is the first step. These goals can be qualitative,

like outlining how a new artifact is anticipated, or quantitative, such stating that a desired solution would be superior to the ones that are now in place to support solutions to issues that have not yet been addressed [78].

According to [78] the objectives should be inferred rationally from the problem specification. Based on this, the objective of the customer level prediction model is for CRM to assist the Bunna Bank in properly manipulating datasets, is to designing and developing a predictive model for customer levels in Bunna Bank through machine learning approaches. Then this research model will help to understand the behavior and wants of customer levels so that management may take corrective action, it helps keep those anticipated customers, reducing the need to find new ones, and it also helps the bank earn its greatest profit and succeed overall. Accordingly building dataset such as data collection, data pre-processing, data splitting and data sampling are incorporated in this phase.

3.2.2.1 Data Collection

The aim of this research is to build the best model for customer level prediction to for effective customer relationship management. So, it needs a dataset from Bunna Bank to build the appropriate model.

Data is essential for machine learning tasks [79], therefore, data collection is necessary for this study. This study focused on the Bunna Bank; because Bunna Bank has many customers, it is one of the privet banks in Ethiopia, and increasing its revenue of the bank. Therefore, the data collection phase of the research methodology in the initial dataset has been collected from the possible sources of data at Bunna Bank by a simple random sampling method from East district. In the bank, there is a computerized system that accommodates all the customers' account data.

To carry out machine learning research, careful preparation and a thorough comprehension of the data were necessary. Data collecting, data exploration, data quality assessment, and drawing conclusions from the data to develop a hypothesis are the main objectives of the second phase of the DSR process. The original customer information was gathered from Ethiopia's Bunna Bank. Python-based statistical and visual analytic tools were used to analyse all of the data.

The Bunna Bank customers' dataset before pre-processing (i.e. before feature selection and resampling) contains 11 variables; which contained information of 64,638 customers of Bunna Bank. The target variable was 'CUSTOMER_LEVEL' which was three values that denotes whether a customer was Low, Medium or High level customer. A few of the variables are discussed in detail:

Table 3.1 Sample of the Dataset from Header

INTERNET	MOBILE_B	BOOK_TYI	AGE	SEX	INDUSTRY	NO_CH	NO_DEI	TENURE	CUSTOME	CUST_OW	BALANCE	CUSTOMER_LEV
USER	USER	Wadiah Ordinary Sa		MALE	Employed	1	38	7	Private	Individual	73693	M
USER	USER	Women Saving Acco		FEMALE	Employed	1	7	9	Private	Individual	274	M
USER	USER	Women Saving Acco		FEMALE	unemploy	1	7	9	Private	Individual	28	L
NOT_USEI	USER	Women Saving Acco		FEMALE	Employed	1	7	7	Private	Individual	30602	M
USER	USER	Saving Account		MALE	Other indi	1	2	3	Private	Individual	20407	M
USER	USER	Women Saving Acco		FEMALE	Other indi	1	23	5	Private	Individual	52	M
USER	USER	Saving Account Staff		MALE	Other indi	1	2	1	Private	Individual		M
USER	USER	Saving Account		MALE	Other indi	1	30	0	Private	Individual	1	M
NOT_USEI	NOT_USEI	Labbaik Sa	24	MALE	Other indi	1	21	0	Private	Individual	716	M
USER	NOT_USEI	Saving Account		FEMALE	Other indi	1	6	0	Private	Individual	9340	M
NOT_USEI	USER	Saving Account		MALE	Other indi	1	30	1	Private	Individual	282786	H
NOT_USEI	USER	Saving Account		MALE	Other indi	1	30	0	Private	Individual	1	M

a) Mobile_Banking_Status (User, No)

'Mobile_Banking_Status' is a significant variable that describes whether the customer uses mobile banking or not? From the given dataset 62,226 customers use mobile banking; the remaining 2,412 customers do not use mobile banking.

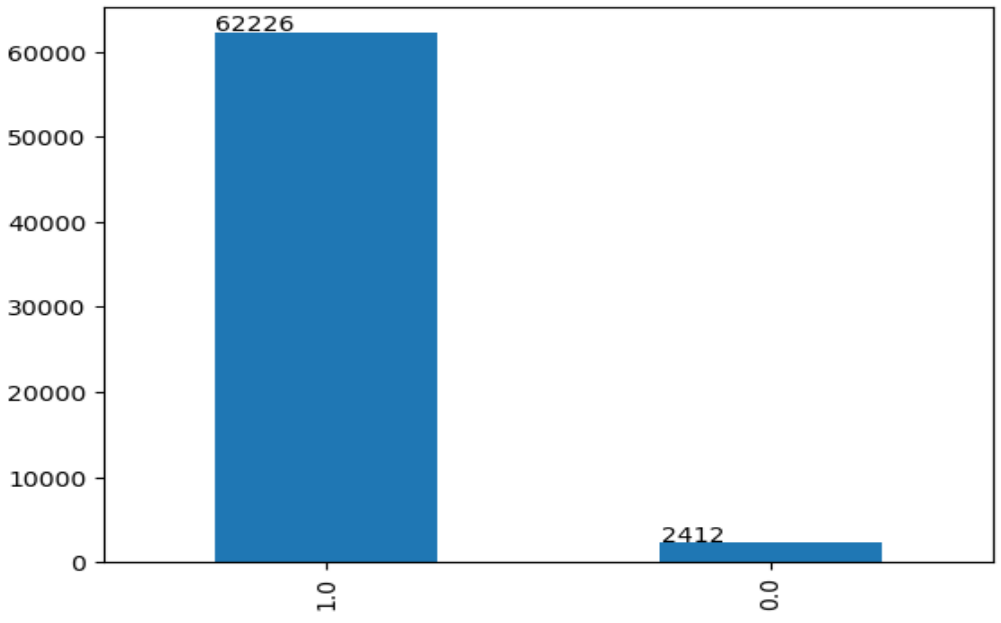

Figure 3.2 Mobile_Banking_Status

b. Book_Type

"Book_Type" is a multinomial category categorical variable. The distribution of "Book_Type" in the customer level prediction for CRM was shown in the graph below.

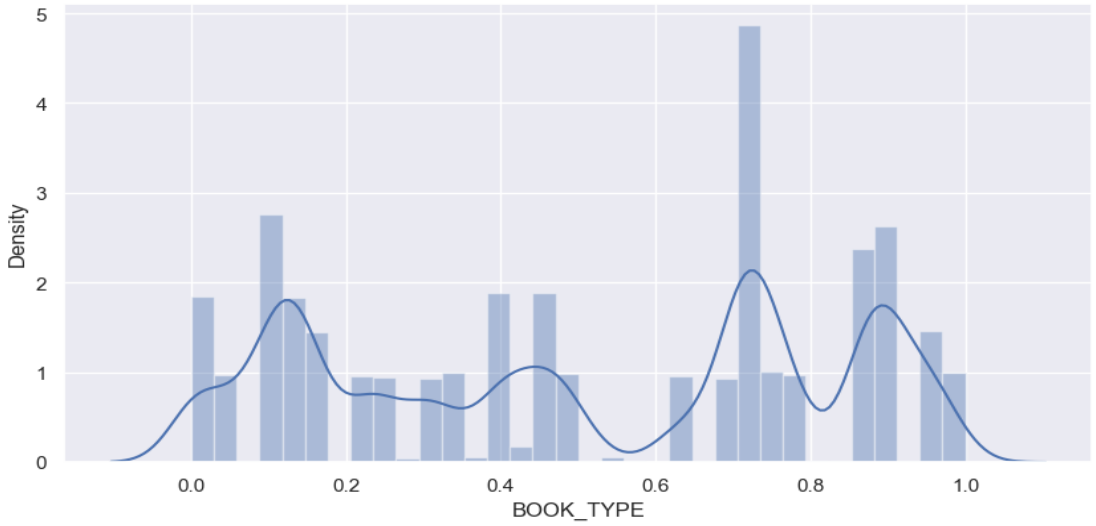

Figure 3.3 Book_Type frequency distributions

c. Age

The age of Bunna Bank's customers was represented by the numerical variable "Age." Its range, which is displayed below, was 19 to 72:

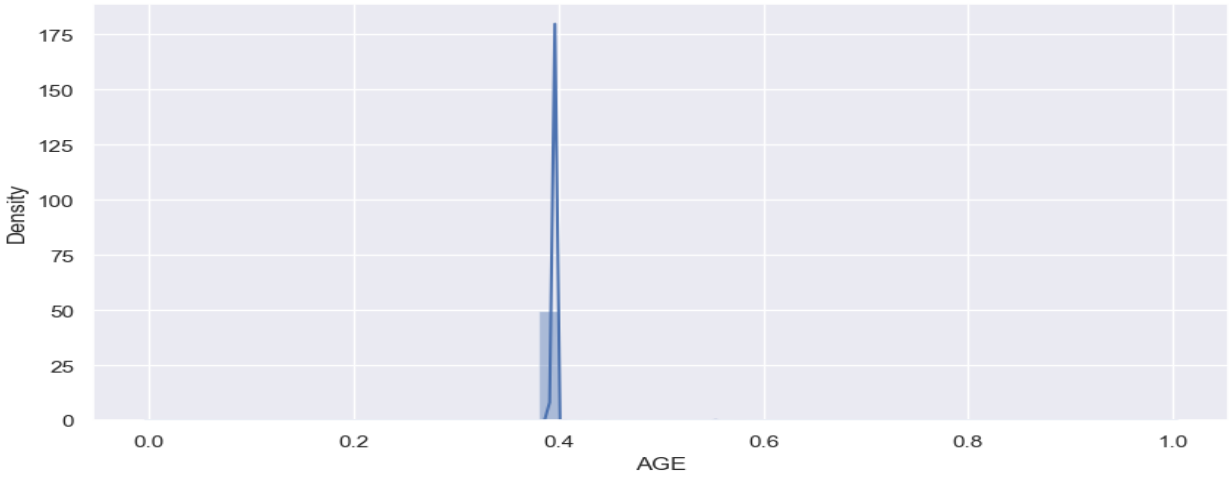

Figure 3.4 Age variable Histogram

d. SEX

'SEX' is also one of the significant variables which describe the customer gender. From the given dataset 2,104 customers' are females (0.0), and the remaining 62,534 customers are males (1.0).

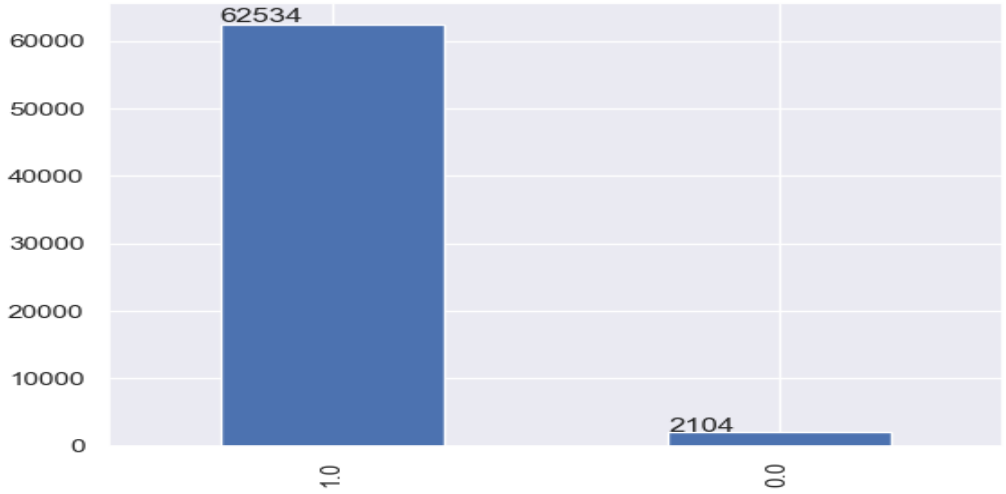

Figure 3.5 SEX attribute

e. CUSTOMER_JOB

'CUSTOMER_JOB' is a categorical variable with multinomial categories. The distribution of "CUSTOMER_JOB" between customer levels was shown in the graph below.

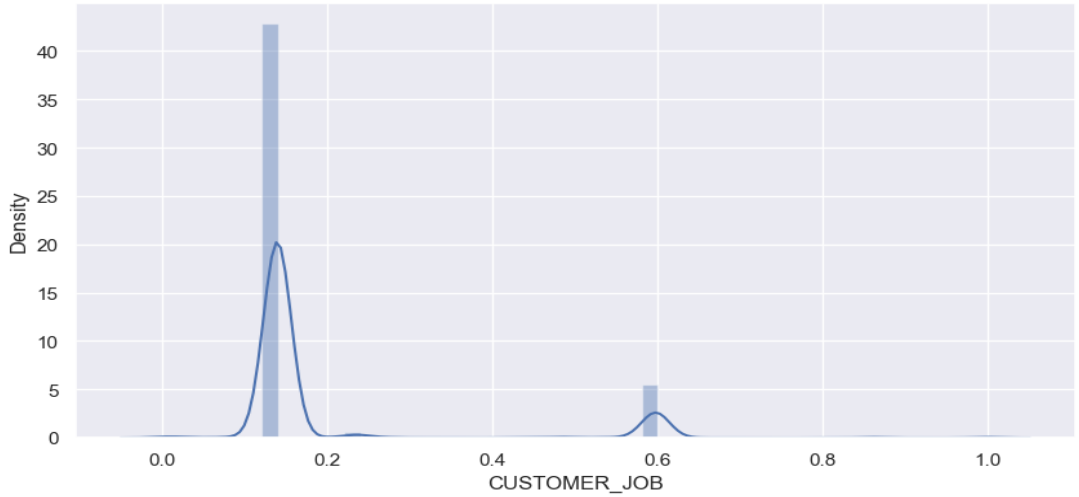

Figure 3.6 Sector_Name frequency distributions

f. NO_CREDIT_TXN

The study has used the number of credit transactions for consecutive six months considered as one of the attributes of the dataset. Since bank customers can identify levels of customers within six months [80]. The distribution was shown in the graph below of 'NO_CREDIT_TXN' in customer levels.

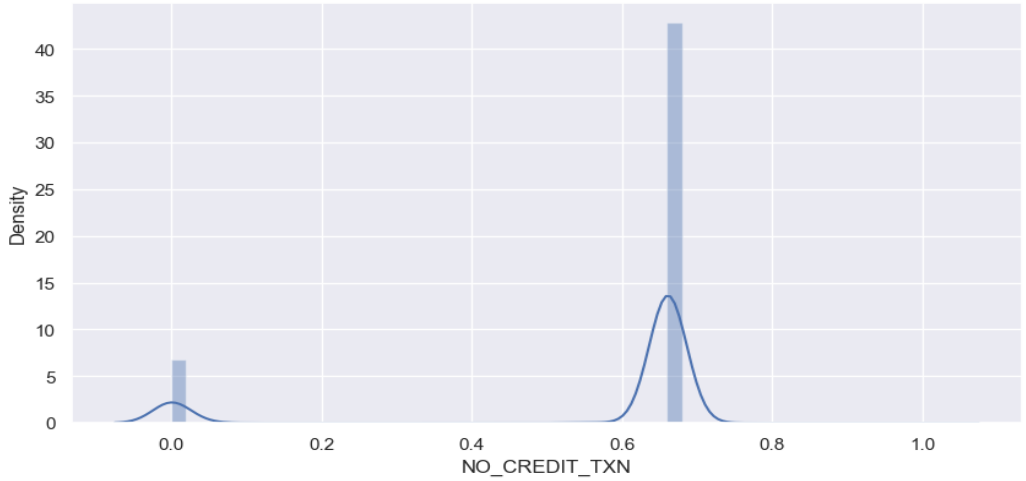

Figure 3.7 No_of_Credit_Txn frequency distributions

g. No_of_Debit_Txn

The study has used the number of debit transactions for consecutive six months considered as one of the features of the dataset. Since bank customers can identify levels of customers within six months. The below graph depicted the distribution of 'No_of_Debit_Txn' in customer levels.

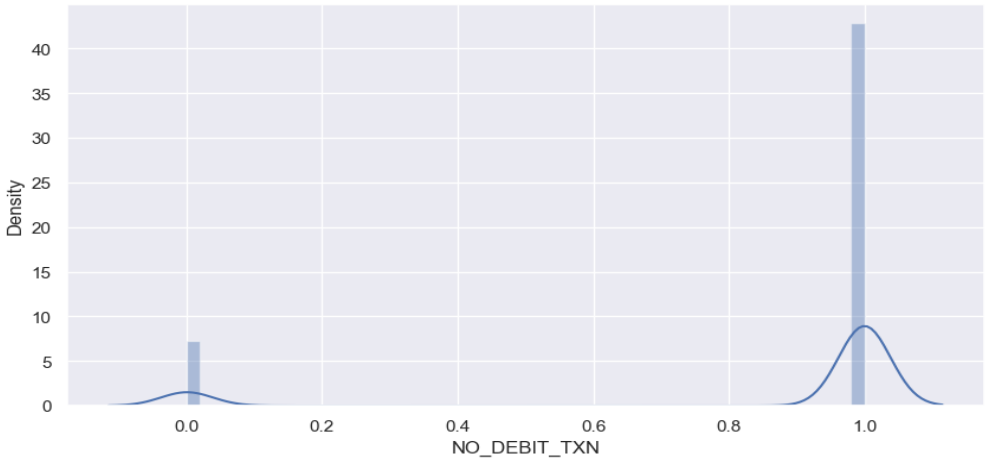

Figure 3.8 No_of_Debit_Txn frequency distributions

h. Tenure

The study has used 'Tenure' considered as one of the features of the dataset. The below graph depicted the distribution of 'Tenure' for customer levels.

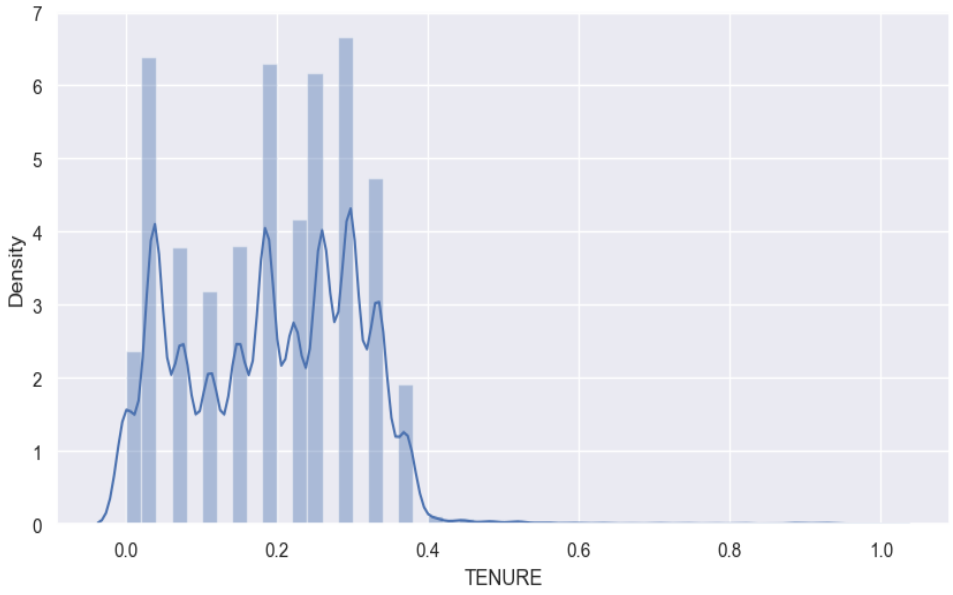

Figure 3.9 Tenure attribute

i. Internet_Banking

'Internet_Banking_Status' is a significant variable that describes whether the customer uses internet banking or not? From the given dataset 53,078 customers use internet banking; the remaining 11,560 customers do not use mobile banking.

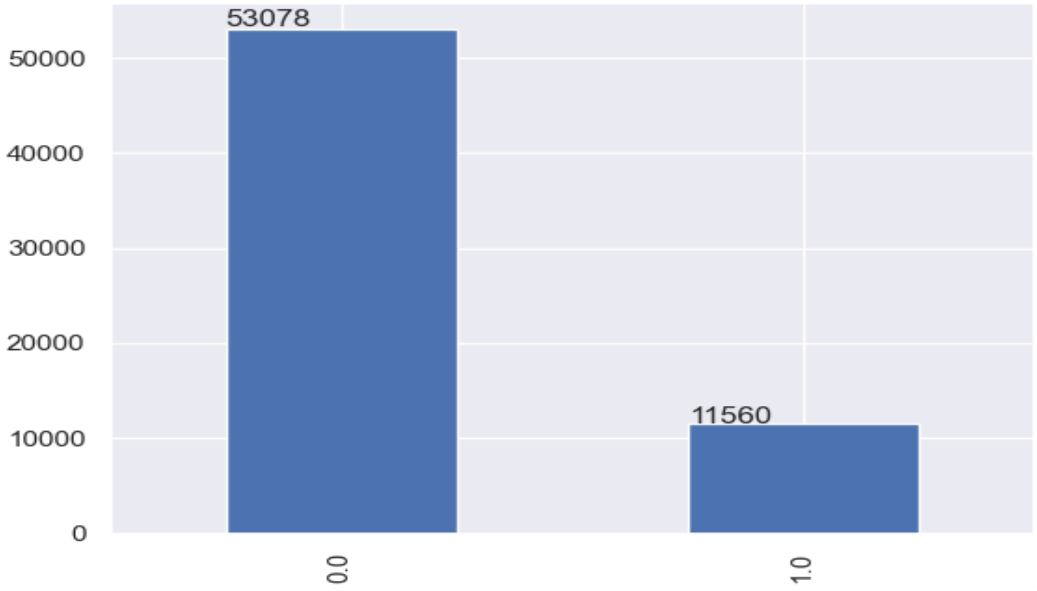

Figure 3.10 Internet banking

j. Balance

The study has used 'Balance' considered as one of the features of the dataset which the bank book has balance in birr. The distribution was shown in the graph below of 'Balance' in customer levels.

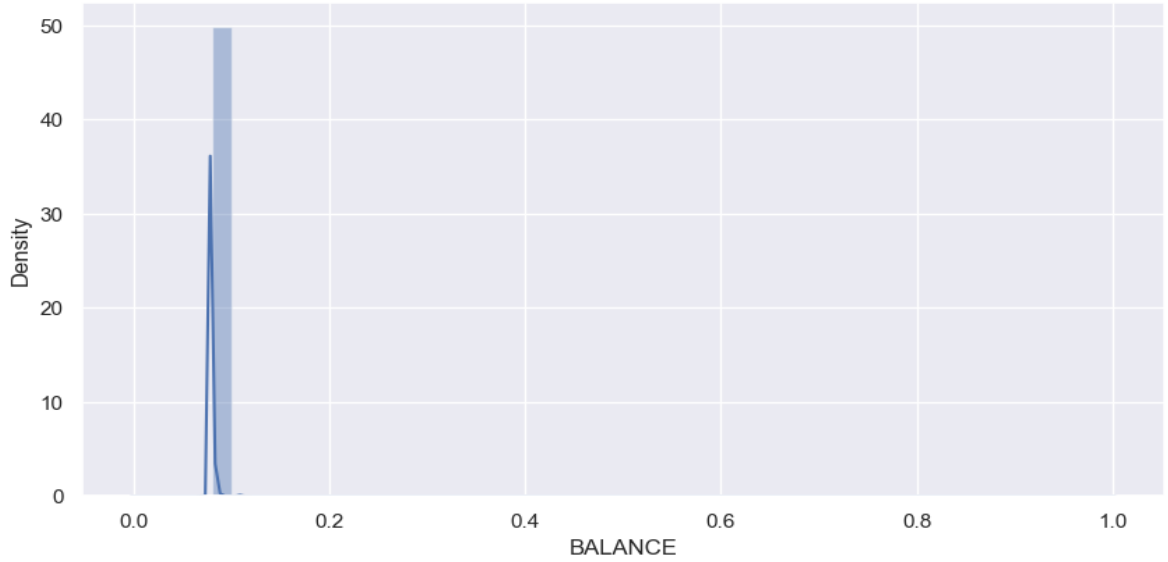

Figure 3.11 Balance frequency distributions

k. Customer_Level

The 'customer_level' is the target variable from the total dataset with values as '0' for those who have Low level cusomers (4,644 customers), '1' for those who are Middle level customers (44,022 customers) and , '2' for those who are High level customers (15,972 customers) of Bunna Bank, the data distribution is as shown below:

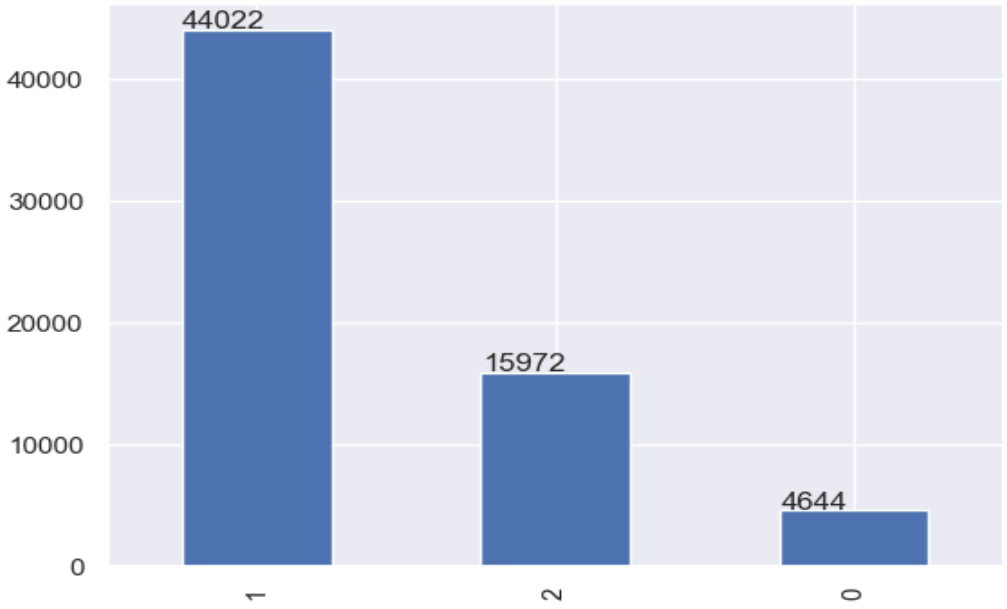

Figure 3.12 Customer level attribute

3.2.2.2. Data Preparation

Data pre-processing is one of the most important stages in machine learning [81][82]. At this phase, the source of the data required for the experiment has been collected, described, and examined for a deeper comprehension. However, modelling cannot be done directly with the collected data. This is due to:

- Data may have errors, outliers, and other quality issues that must be fixed beforehand.
- The achievement of the desired goal may not be correlated with the significance of every attribute and data domain. Thus, data selection may be required.
- From the available data, attributes and even tables may need to be derived.
- It would become necessary to integrate data from two or more tables, and

This indicates that all tasks necessary to transform the raw data into the final dataset were completed during the data preparation phase, which may be used to construct models and feed into modelling algorithms. This phase involved a number of tasks, including data transformation, data cleansing, outlier removal, missing value imputing, and the development of new characteristics.

Based on the knowledge of the data from the previous phase, a number of data pre-processing processes were performed in order to prepare the final dataset for the experiment. Handling missing values was part of the data pre-processing phase. Other steps included encoding, normalizing the data, extracting features, standardizing the data, splitting the data, and, lastly, applying the SMOTE technique in this study.

3.2.2.2.1. **Handling Missing Values**

Since many machine learning algorithms do not support data with missing values, managing missing values is crucial. The dataset that has been provided contains numerous missing values. A number of different causes could be the cause of this. The improper collection of the data could be one of the causes. It is possible to eliminate variables (columns) from the final dataset if more than 60% of their values are missing [83].

Mean values were used to impute values for continuous variables, which might have any value between their minimum and maximum values. Missing values ranged from 2% to 30%. For randomly chosen observations from a normal distribution, the mean is a fair approximation. A number of causes can lead to missing values. Missing data caused a number of issues; it decreased statistical power, which could result in a wrong evaluation of the hypothesis. Additionally, it can make the sample less representative. It might make data analysis more difficult. Few algorithms do not work with missing data. The most common methods for imputing missing values in

categorical variables with labels were the maximum likelihood approach and the last observation carried forward methodology [84]. The values that occurred most frequently were used to impute the missing values in order to maximize likelihood. The prior observation was imputed to the missing value in the last observation carried forward approach.

There were missing values in the customer dataset, according to the descriptive analysis of the dataset. The mode technique was used to impute the missing values in categorical variables [85]. Python is used to impute missing values in categorical variables by using mode values to determine the dataset's missing values. Mean values were used to impute values for continuous variables that had missing values ranging from 2% to 30% [86].

3.2.2.2.2. **Encoding**

The way we process and feed various kinds of variables into the model affects how well the majority of machine learning and deep learning models perform, in addition to the model and hyper-parameters. Pre-processing the categorical data becomes essential because deep learning models only accept numerical variables. In order for the model to understand and extract relevant information, we have to transform these category variables into numerical values. The LabelEncoder function in Sklearn was used to convert the categorical data into numerical values. The process of converting category data into numerical data is called encoding. The most common data type in the dataset was a categorical one. Typically, these variables were maintained as text values. Because machine learning algorithms are based on mathematical equations, they can only operate on numerical data. Therefore, the categorical variables had to be converted into numeric form because they could not be kept in their current form.

The categorical variables in this study were encoded using the Label Encoding technique. The learning process is not slowed down by employing this strategy, which converts each category to a number without adding more columns. A label encoder was used in this study to encode data that is categorical. The variables Book_Type, Sex, Customer_job, Internet_Banking_Status, Mobile_Banking_Status, and Customer_Level attributes were encoded.

3.2.2.2.3. **Normalizing Data**

Data normalization is a method used in data pre-processing for machine learning model construction. Finding an appropriate scale for the values is the goal of normalization. Each numerical column was evaluated for skew and kurtosis; if the values were outside of the range of +/-2, the variable was considered skewed data. Using the sklearn MinMaxScaler() algorithm, these attributes were normalised.

The continuous variables' skew and kurtosis were determined. The variable was not normally distributed if the skew and kurtosis values were outside of the [0, 1] range. It is possible to determine whether or not the variables were normally distributed even from the histogram plot. By using the sklearn MixMaxScaler() method on the specific variables in the dataset, the Bunna Bank dataset has been normalised.

3.2.2.2.4. **Feature Selection**

Feature selection is used to the dataset as there are a large number of dependent variables in the given dataset. Relevant features for the model creation can be identified through feature selection. To determine the correlation between the independent and dependent variables, a correlation matrix was employed. This allowed for identification of the multicollinearity problem and the most relevant attributes for output prediction.

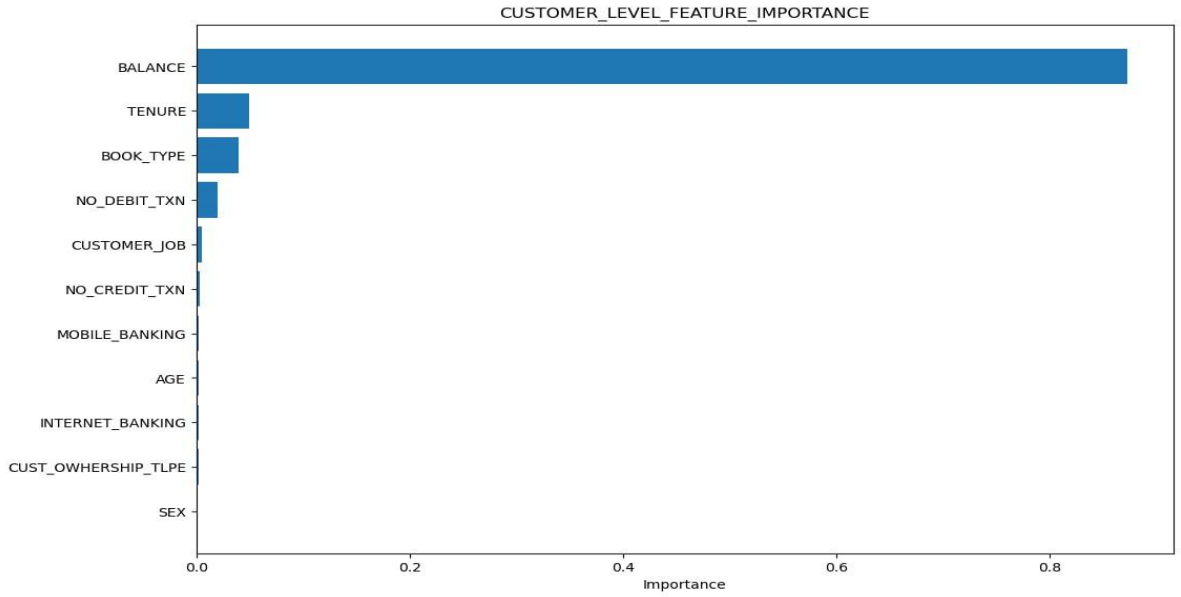

Figure 3.13. Feature Selection

As seen from the above graph also the most important feature in predicting the target was ‘Balance’ followed by ‘Tenure’, ‘Book_Type’, and so on.

3.2.2.3. Data Splitting

The final dataset was separated into train and test datasets following basic preprocessing processes. The table below depicts the final dataset that was obtained following,handling of missing values, data sampling, and normalization. Based on previous studies, the final dataset in this study was divided into 80% (51,710) training datasets and 20% (12,928) test datasets [87]. The data was divided 40 times into train and test datasets using the for-loop and the sklearn train_test_split algorithm. The data was divided at random each time the loop ran, models were

made, and various accuracy levels were recorded in the list. The best model was selected to predict the customer level after the supervised machine learning models were compared according to their accuracy scores.

3.2.2.4. Data Sampling

In several real-world applications, class imbalance is one of the most prevalent data issues. Typically, most instances in this study are labeled as belonging to the medium-level customer class, while fewer instances are categorized as lower-level customers. This issue is referred to as a class imbalance. The issue of class imbalance is seen in many application domains [88]. In this particular case, the instances are labeled by the Bunna Bank MIS work unit and are based on the available balance and transaction activity of customers. Specifically, the labeling criteria are as follows:

- Low-level customers: Customers who have not conducted any transaction within the last six months.
- Medium-level customers: Customers who have transacted within the last six months and have an account balance of less than one million birr.
- High-level customers: Customers who have transacted within the last six months and maintain an account balance of greater than one million birr.

Of the 64,638 Bunna Bank dataset, there were 4,644 (7.2%) low level, 44,022 (68.1%) medium level, and 15,972 (24.7%) high level customers in this study. During the study's data pre-processing phase, the data level strategy for addressing the problem of class imbalance was either random under-sampling or random over-sampling. In this study, class imbalance issues were addressed using the SMOTE approach.

The course of action that can be taken in the data pre-processing phase of the project was either the random under-sampling or random over-samplings which addressed the issue of class imbalance at the data level. The SMOTE approach was employed in this study to address issues of class imbalance.

3.2.2.2.5. **SMOTE Technique**

SMOTE (Synthetic Minority Over-sampling Technique) generates synthetic samples from the minority class, which has been applied only for the training dataset (80% of the dataset). It was used to balance the training dataset synthetically which was then used to train the classifier [89]. Rather than making duplicates of data from the minority class, it produced synthetic samples from the minor class. To balance the data, it selected comparable records from minority classes and

made arbitrary modifications to each record, one column at a time. The records were solely added to the minority class records.

Class imbalance is the most frequent issue with data. The distribution table of the "Customer_Level" variable and the figure below both show the class disparity of the training dataset.

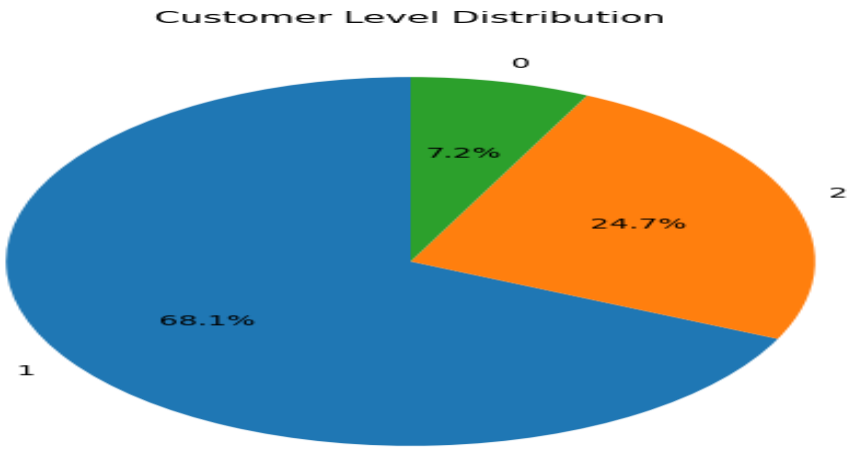

Figure 3.14 Proportion of training dataset before resampling

Table 3.2 Target variable count for training data

Customer Level			
Low (0)	**Medium (1)**	**High (2)**	**Total**
3,738	**35,161**	**12,811**	**51,710**
7.2%	**68.1%**	**24.7%**	**100%**

The Synthetic Minority Over-Sampling Technique (SMOTE) was used in this study to address the class imbalance problem of the training dataset based on the previous research [89]. New minority instances were created between the preexisting minority instances using the SMOTE approach. In order to balance the data, this function was utilized.

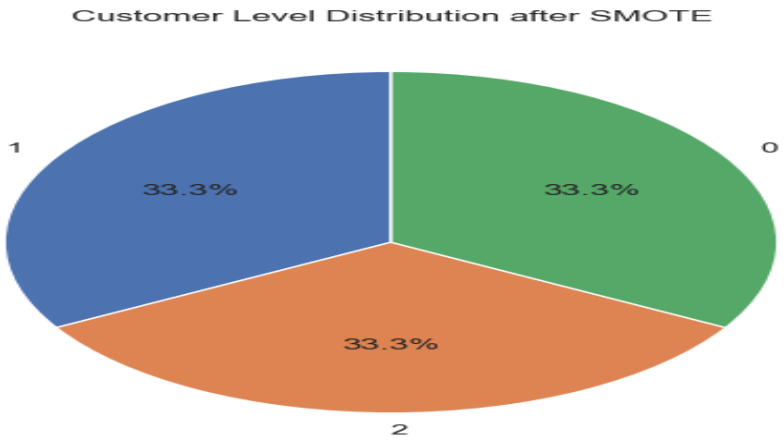

Figure 3.15 Proportion of Low (0), Medium (1), High (2) level customer after resampling

After applying the SMOTE technique, the training target column (y_balance) consisted of 105,483 entries, with each category (low, medium, and high-level customers) containing 35,161 instances. This ensured that the training set was balanced across all categories.

3.2.3. Design and Development

The artifact's development is the definition of this activity [78]. It entails determining the required functionalities of the artifact as well as its architecture, followed by the actual creation of the artifact. According to scholars Hevner & Chatterjee, 2010 [90], these artifacts could be instantiations, models, constructions, or techniques. The customer level predictive model, or artefact, in this study is a model that was created and developed with the previously specified goals in mind.

As described above, this part is carried out by taking a dataset from the Bunna Bank. Then, the information collected is translated and included to the set of patterns and principles being developed. Finally, the suggested customer level predictive model for the Bunna Bank is formed. The designed model illustrates each component and its functionalities to solve the company's customer relationship management problems. Chapter three contains additional information.

3.2.3.1. Architecture

This research explains the architecture and proof-of-concept implementation of models for customer level prediction based on different machine learning algorithms, with Bunna Bank customer levels as the background scenario. The use of machine learning methods offers the promise of detecting new patterns of customer level in the given training datasets.

The machine learning architecture outlines the main processes that are utilised for transforming raw data into training data sets that can be used to inform a system's decision-making. It also describes the different approaches that are engaged in the machine learning cycle. The overall

architectural layout is shown in Figure 3.16 below.

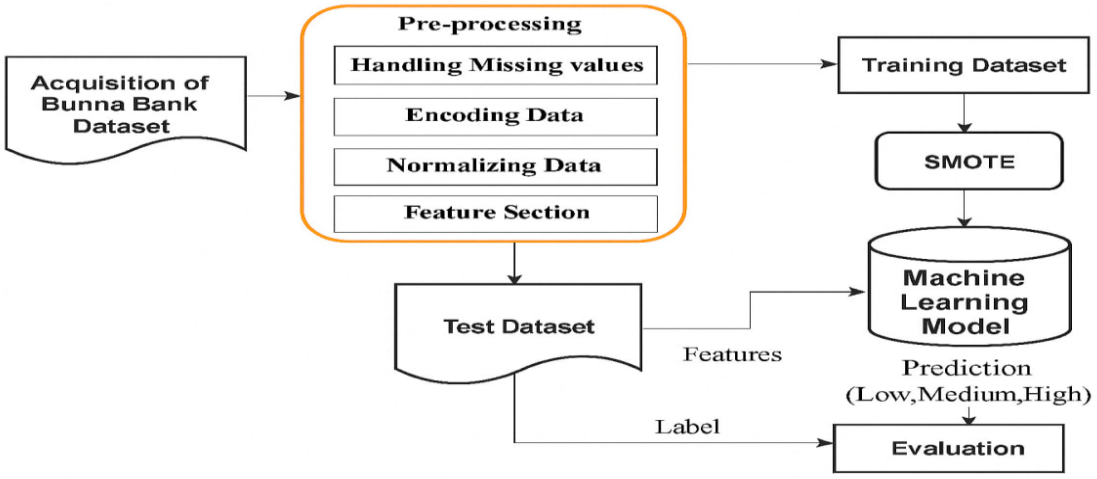

Figure 3.16 Architecture of customer level prediction

As shown in figure 3.16 above, which depicts the architecture of the customer level, data is initially collected from Bunna Bank and then pre-processed, including feature selection, encoding, handling missing values, and normalisation. Thus, following pre-processing, the dataset was divided into training and test datasets, with 80% for training and 20% for testing. Only the training dataset was subjected to the SOMTE technique, and the imbalanced training data used to construct a machine learning model was then balanced using SMOTE. Lastly, use the test dataset to assess and choose the best model.

3.2.4. Demonstration

This phase illustrates how to solve one or more problem cases using the artefact, which is the customer level prediction model. This can entail using it in the chapter four experimentation activities.

3.2.5. Evaluation

The evaluation measures how well the model or artefact helps in a problem-solving strategy. Prior knowledge bases such as scientific theories and methodologies, experience, and expertise are used in this process, as well as design products and processes, to verify innovation (to make sure that the designs created are research contributions rather than routine designs based on the use of known design processes and the appropriation of known design artefacts). As a result, several kinds of metrics have been employed to evaluate the models developed in this study.

subsequently is essential to test the models developed in the previous phase to ensure that they

are applicable and realistic. If at all possible, the bank's representative authorities participate in this model evaluation process. Depending on Bunna Bank's interest and readiness, the models can be utilised to categorise the current data in accordance with the model once they are determined to be realistic, possible, and interesting. Also, various ways to check the performance of machine learning; to evaluate classifiers' Performance in customer level prediction for different schemes with their appropriate parameters, the paper will use the measures of precision, recall, accuracy, and F-measure, determined from the contents of the confusion matrix; Which means several measures have been used to evaluate the models created in this study. Accuracy, precision, recall, and specificity were computed once each model's confusion matrix was constructed. False-positive (FP), false-negative (FN), true-positive (TP), and true-negative (TN) values are obtained [91]. A confusion matrix can be utilised for calculating the following metrics:

Table 3.3 Confusion matrix

		Actual Values	
		Positive (1)	**Negative (0)**
Positive (1)		TP	FP
Negative (0)		FN	TN

Where,

TP (True Positive): is the amount of the specified class detected when it is that class Predicted values correctly predicted as actual positive or result where the model correctly predicts the positive class.

TN (True Negative): This refers to all the correct positive predictions made by the model, except for the true positives of the specific class we're focusing on. It can also refer to true negatives cases where the model correctly identified something as not belonging to the positive class..

FP (False Positive): - In contrast, it can also describe the total number of predictions made for a specific class, excluding the correct ones (true positives). These are the instances where the model predicted something to be part of a class, but in reality, it wasn't meaning it incorrectly labeled negative cases as positive.

FN (False Negative): - This refers to the total number of actual instances of a class, excluding the correctly predicted ones (true positives). In other words, it's the count of real class members that the model mistakenly predicted as belonging to other classes meaning positive cases that

were incorrectly classified as negative.

Table 3.3 presents a confusion matrix for a binary classification problem. The values along the diagonal, running from the bottom-right to the top-left, indicate the correct predictions made by the model, while the values outside this diagonal represent the errors. Using this matrix, we can calculate several commonly used evaluation metrics. The overall accuracy of the classifier is determined by dividing the number of correctly predicted positive and negative cases by the total number of samples. Additional performance metrics such as recall (also known as sensitivity), precision, and the F-measure provide further insight into how well the model performs. These metrics help evaluate the effectiveness of different classification methods.

3.2.6. Communication

The dissemination of design science research findings is one of the most important components of the study [92]. One instance of this kind of communication is found in scholarly research articles; the appropriate stakeholders will be informed about and given an introduction to this research. In other words, this document serves as a thesis research publication in the Addis Ababa University thesis archive and a means of communication with professionals at Bunna Bank for exercise. The results of this study will also be published in scholarly articles.

Summary

There are a lot of challenges that this study faced in getting the best model for customer level prediction by the given datasets. Data collection is challenging, in the case of confidentiality. This study's lack of high-quality data makes the data pre-processing stage one of its major difficulties. For the data pre-processing phase, the entire process was very stressful due to unclean and noisy data. Integration and transformation of the dataset are particularly time-consuming.

This study aimed to minimize inaccurate or faulty predictions by the algorithm, the study tried to enhance the output or the quality of the model by the given data. Therefore, the bank must have a data warehouse that can hold important customer information so that future machine learning research can be carried out without resource constraints.

CHAPTER FOUR

EXPERIMENT, RESULTS, AND DISCUSSION

4.1. Overview

This chapter outlines how the experiment was performed, based on the steps mentioned in chapter three. It includes experimental setup, experimental setting, results, and discussion obtained from the experiments. This research aimed at building and comparing the machine learning techniques using a Bunna Bank customer dataset to predict customer level. The Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), and Deep neural network (DNN) models were built, and the results were compared. Finally, this chapter put the strengths of the research.

4.2. Experimental Setup

For the implementation of the customer level prediction model, many tools and techniques are used. In order to execute the suggested model, all experiments were conducted on a laptop running Windows 10 with an Intel(R) Core i7 Processor running at 1.80GHz and 8GB of RAM. Python as a programming language and algorithms have been implemented using software tools, along with the Scikit-learn,TensorFlow and Keras anaconda environment libraries. These Python-based utilities satisfy all requirements for selection.

Tensorflow: is a free and open-source tool developed by Google, widely recognized as one of the leading deep learning libraries available today [93]. It supports the full process of building machine learning models from handling and preparing data, to designing, training, and evaluating models. At its core, TensorFlow works with something called tensors, which are just multi-dimensional arrays used to manage and pass data through the system.

Anaconda: In this study, we used the Python programming language for data preprocessing, building the model, and evaluating its performance specifically through the Anaconda distribution (version ‘conda 4.11.0’). Python is a top choice for predictive tasks because it's easy to write, beginner-friendly, has a clean syntax, and offers a wide range of built-in libraries and tools. Anaconda made it easier to manage packages and environments, especially for data science and deep learning work. It comes bundled with helpful tools like Jupyter Notebook and Spyder, which are popular for writing and running code. For this project, we mainly used Jupyter Notebook, as it's simple to use and runs directly in a web browser.

Scikit-learn: Scikit-learn was used to build the machine learning model in this study. It's a Python library that's widely used for machine learning tasks and is built on top of the SciPy ecosystem.

Scikit-learn works seamlessly with other key Python libraries such as:

- NumPy – for handling multi-dimensional arrays and numerical operations,
- Matplotlib – for creating a wide range of 2D and 3D plots and visualizations,
- Pandas – for managing and analyzing structured data with ease, etc.

Keras: is a high-level neural network API written in Python that runs on top of backends like TensorFlow, Theano, or Microsoft Cognitive Toolkit (CNTK). It's designed to be user-friendly and flexible, making it easy to build and experiment with deep learning models. One of its key advantages is the ability to quickly create models with minimal code. Keras also comes with a collection of pre-trained models like VGG16 and InceptionV3, which were used during our experiments. It supports various types of networks, including CNNs, RNNs, or combinations of both, making it a powerful and accessible tool for deep learning tasks [94].

At this phase, the cleaned data was used to build machine learning models aimed at predicting customer levels. The goal was to create several supervised models, compare how well they performed, and identify which one gave the best results. The algorithms tested included Logistic Regression, Support Vector Machine, Random Forest, K-Nearest Neighbors, and Deep Neural Networks. We ran the training using Google Colab, taking advantage of its GPU accelerator with 1.63GB of memory to speed things up.

4.3. Experimental Setting

Based on the related works above and previous researches, this study has applied five experiments. After applying different parameters; the following best hyper-parameters were used.

i. Logistic Regression

Logistic Regression is a widely applied classification method that estimates the probability of an observation belonging to a certain class using the logistic function. It is particularly useful when the relationship between the dependent and independent variables is approximately linear. In this study, the solver was set to lbfgs, which provides stable results for multiclass problems and is efficient for larger datasets.

ii. Random Forest

Random Forest is an ensemble learning approach that generates a collection of decision trees and aggregates their predictions to improve accuracy and limit overfitting. Each tree is built using a random subset of the data and features, which increases robustness against noise. For this work, the model was tuned with 100 trees (n_estimators = 100), the Gini index as the splitting criterion, and a maximum tree depth of 5 to control complexity.

iii. K-Nearest Neighbor (KNN)

KNN is an instance-based method that assigns a class label to a data point by examining the majority class among its closest neighbors. Classification is influenced by the choice of distance metric and the number of neighbors considered. The optimal setting in this case was found at n_neighbors = 324, which provided a suitable balance between underfitting and overfitting.

iv. Support Vector Machine (SVM)

SVM operates by identifying the hyper-plane that maximizes the separation between different classes. It can be adapted to both linear and nonlinear problems by applying different kernel functions. For the dataset used in this study, the linear kernel produced the best outcome, ensuring simpler computation while effectively handling the class separation.

v. Deep Neural Network (DNN)

DNNs consist of multiple layers of interconnected neurons designed to capture complex and nonlinear relationships within data. Each neuron processes input values through a weighted transformation followed by an activation function. The network developed for this study was configured with 7 hidden layers, each containing ten neurons, and applied the ReLU activation function. The model was trained with a batch size of 40 over 80 epochs, using the Adam optimizer to speed up convergence.

4.4. Experiment

To identify the model that works best for the dataset, all five supervised machine learning algorithms were implemented. The dataset was separated into two portions: a training set and a testing set. Eighty percent (80%) of the data was allocated for training, enabling the models to learn the underlying patterns and relationships between the input features and the target categories. The remaining twenty percent (20%) was kept for testing, which was used to assess how well the models performed on data that had not been seen during training. This division helped ensure that the evaluation results accurately reflected each model's ability to generalize to new, unseen instances.

4.4.1. Experiment 1: Logistic Regression (LR) Classification

The LogisticRegression algorithm imported using the sklearn.linear_model class in Python was used to construct the Logistic Regression model. Sklearn, often known as Scikit Learn, is an open-source Python machine learning package. It contains numerous supervised and unsupervised learning techniques. The model was trained and tested with only significant variables included and all insignificant variables removed. Every supervised machine learning algorithm used the

same set of features with default parameter settings. The following outcomes were obtained from the logistic regression.

Table 4.1 Logistic Regression confusion matrix

Actual \ Predicted	Low (0)	Medium (1)	High (2)	Row Total
Low (0)	620	190	96	906
Medium (1)	1,300	6,000	1,561	8,861
High (2)	950	800	1,411	3,161
Total Predicted	2,870	6,990	3,068	**12,928**

As indicated in 4.1 the average classification performance metrics for Logistic Regression are 62.1% accuracy, 51.1% precision, and 60.3% recall. The reason for the outcome may be the target label has no linear correlation with the features. In these situations, even with training data, logistic regression (or linear regression for regression problems) is unable to accurately forecast targets. because Logistic Regression is not suitable for complicated data because to its linear decision surface.

4.4.2. Experiment 2: Random Forest (RF) Classification

The Random Forest was built in python using sklearn ensemble class. Ensemble learning is the foundation of the Random Forest algorithm. Ensemble learning uses several machine learning models to enhance predictions on a dataset.

The "gini" criterion, which was the default function for evaluating the quality of split, was selected in this study; n_estimators was set at 100, meaning that 100 random decision trees were ensembled to create the Random Forest. The max_depth was selected as 10 which mean the tree can expand till the maximum depth of 10. Ultimately, "balanced" was selected as the class weight, representing the weight assigned to each class. By default, the weight of Random Forest is inversely proportional to the frequency of occurrence of the class in the data. The following results were observed from the Random Forest.

Table 4.2 Random Forest confusion matrix

Actual \ Pred	Low (0)	Medium (1)	High (2)	Total
Low (0)	**692**	214	0	906
Medium (1)	2,092	**6,769**	0	8,861
High (2)	0	746	**2,415**	3,161
Total Predicted	2,784	7,729	**2,415**	12,928

Based on the table 4.2 the average Random Forest classification performance metrics on average 76.4% accuracy, 70.2% precision, and 76.4% recall score are obtained. The reason for the result may be the dataset is complex, in such a case the Random Forest algorithm can't predict targets with good accuracy.

4.4.3. Experiment 3: Support Vector Machine (SVM) Classification

The SVM model was built using the SVM function imported from sklearn in python. Because of the efficiency of the running time, the SVM model in this study has a "linear" kernel parameter. The kernel function is the most crucial parameter in the Support Vector Machine model. A single line is used to linearly separate the classes in this study because the kernel function was specified to be linear. It helps when the dataset is large. The linear kernel's main advantage was its rapid processing speed. The outcomes of the Support Vector Machines were as follows:

Table 4.3 Support Vector Machines confusion matrix

Actual \ Predicted	Low (0)	Medium (1)	High (2)	Row total
Low (0)	560	250	96	906
Medium (1)	1,450	5,400	2,011	8,861
High (2)	650	650	1,861	3,161
Total Predicted	2,660	6,300	3,968	**12,928**

Based on the table 4.3 the average classification performance parameters for Support Vector Machines are 60.5% accuracy, 51.2% precision, and 60.8% recall. The Support Vector Machine (SVM) doesn't perform well with a large number of the dataset. In such cases, SVM can't predict targets with good accuracy (even on the training data), and it doesn't perform well when the data comes with more noise which means target classes are overlapping.

4.4.4. Experiment 4: K-Nearest Neighbor (KNN) Classification

The KNN model was built using the KNeighborsClassifier function imported from sklearn in python. The most significant parameter in the K-Nearest Neighbours model is the n-neighbors function. Using the rule of tump [95], which is the square root of the training datasets (105,483 datasets used for training), n-neighbors were chosen for the KNN model "(K=324)" in this work. Due to its ease of use, the KNN method was frequently employed. Determining the optimal value of K has the primary benefit of indicating how many nearest neighbours should be included in the majority voting process. The outcomes of the K-Nearest Neighbour were as follows:

Table 4.4 K-Nearest Neighbor confusion matrix

Actual \ Predicted	Low (0)	Medium (1)	High (2)	Total Actual
Low (0)	300	450	156	906
Medium (1)	1,200	4,700	2,961	8,861
High (2)	450	1,800	911	3,161
Total Predicted	1,950	6,950	4,028	**12,928**

As indicated in 4.4 the average K-Nearest Neighbour classification performance measures are 51.0% accuracy, 35.2% precision, and 38.3% recall. The accuracy of the K-Nearest Neighbor may be decreased, when we use high number of instances and for complex data, and also in the

case of defining parameter k. It is unable to accurately predict targets in these situations.

4.4.5. Experiment 5: Deep Neural Network (DNN) Classification

A Deep Neural Network is a form of artificial neural network with two or more hidden layers. Deep neural networks have the advantage of being able to execute intensive tasks and learn complex features by carrying out more complex operations at once. In contrast to other conventional algorithms using the same dataset, a deep neural network approach is employed in this study to predict Bunna Bank's Low, Medium, and High level customers. As a rule of thumb, the DNN algorithm uses two-thirds ($\frac{2}{3}$) of the total attributes in its hidden layers [96]. Thus, seven hidden layers were used in this study. The outcomes of the Deep Neural Network were as follows:

Table 4.5 Deep Neural Network confusion matrix

Actual \ Predicted	Low	Medium	High	Total
Low (0)	722	150	34	906
Medium (1)	500	7,062	1,299	8,861
High (2)	50	586	2,525	3,161
Total Predicted	1,272	7,798	3,858	**12,928**

As indicated in 4.5 the performance of Deep Neural Networks delivered an average of 79.7% accuracy, 70.9% precision, and 79.8% recall.

The accuracy of a Deep Neural Network decreases when the data is complex, and when the data become a small number of attributes because DNN depends on a lot of training data.

Table 4.6 Results of Supervised Machine Learning Models

Model	Accuracy	Precision	Recall	Remark
Logistic Regression	**62.1**	51.1	60.3	3
Random Forest	**76.4**	70.2	76.4	2
Support Vector Machine	**60.5**	51.2	60.5	4
K-Nearest Neighbour	**51.0**	35.2	38.3	5
Deep Neural Network	**79.7**	70.9	79.8	1

As shown in Table 4.6 results were obtained from each experiment (selected supervised machine learning techniques).

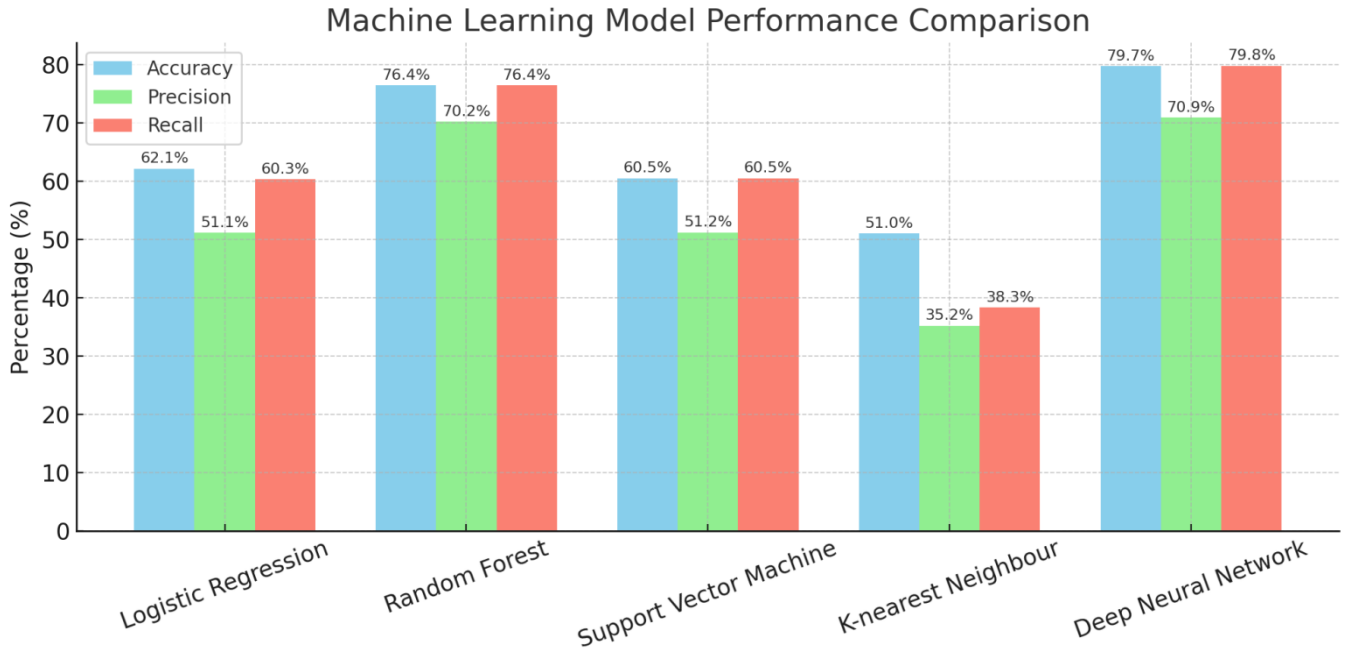

Figure 4.1 Comparisons of Evaluation Metrics

As shown in Figure 4.1 and Table 4.6, evaluation results were obtained for five classifiers which are Deep Neural Network (DNN), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) with corresponding accuracies of 79.7%, 76.4%, 62.1%, 60.5%, and 51.0%, respectively. The experiments aimed to determine the best-performing supervised machine learning model for predicting customer levels using the Bunna Bank customer dataset. All models were trained on a balanced dataset generated using the SMOTE technique.

Among the models, DNN outperformed RF, LR, SVM, and KNN by margins of 3.26%, 16.91%, 17.62%, and 28.65%, respectively. These results show that the Deep Neural Network model achieved the highest accuracy, precision, and recall scores overall. This performance underscores the advantage of using deep learning techniques, especially when dealing with complex and high-dimensional datasets. Hence, DNN proves to be the most effective algorithm for customer level prediction in the Bunna Bank dataset scenario.

4.5. Answers to the Research Questions

RQ1: Which are the most important features for customer level prediction?

Answer: In this study, the Internet_Banking, Mobile_Banking, Book_Type, Age, Sex, Customer_Job, No_of_Credit_Txn, No_of_Debit_Txn, Tenure, Balance, and Customer_Level are the basic variables to build a customer level prediction model.

Using correlation matrix and model-based feature importance, the analysis revealed that 'Balance' is the most significant factor in predicting the target variable. It is followed by 'Tenure', 'Book_Type', 'Number of Debit Transactions', 'Customer Job', 'Number of Credit

Transactions', 'Mobile Banking Status', 'Age', and 'Internet Banking Status'. These features collectively play a crucial role in understanding customer behavior and improving the accuracy of the prediction model.

RQ2: Which type of machine learning model is suitable for customers' level prediction model?
Answer: The Deep Neural Network algorithm gives best results than the other machine learning classifier for predicting customer level for the Bunna Bank customer dataset.

RQ3: To what extent do machine learning techniques predict customer levels in Bunna Bank?

Answer: As seen in the above experiments Deep Neural Network (DNN) model has the highest accuracy of 79.7%, the precision of 70.9%, and recall of 79.8%. So that this model was selected as the best model for predicting customer levels in the Bunna Bank dataset.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

This chapter explains the findings of the experiment conducted to predict customer level and provides a summary of the study that was conducted. It is divided into two main sections: conclusion, future work, and recommendations. It provides an overview of the findings in relation to the initial study research question: 'Which machine learning algorithm: Logistic Regression, Random forest, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), or Deep Neural network (DNN); can best predict the customer level of Bunna Bank with the best accuracy?'.

5.1. Conclusion

This paper demonstrated the significant potential of machine learning techniques in improving Customer Relationship Management (CRM) with the banking sector, using Bunna Bank as a case study. By leveraging predictive models such as Random Forest, Support Vector Machines, and Deep Neural Networks, the study successfully categorized customers into high, medium, and low-value tiers based on behavioral and transactional data. This segmentation enables the bank to implement tailored strategies, fostering customer loyalty, improving retention, and optimizing resource allocation. The findings highlight how data-driven decision-making can enhance profitability by prioritizing high-value customers while nurturing medium-value customers and re-engaging low-value segments. Despite challenges related to data quality, privacy concerns, and computational requirements, the proposed model lays the foundation for smarter, more personalized banking experiences. Moreover, the study contributes to advancing Ethiopia's banking sector by promoting financial inclusion and supporting underserved populations with targeted services. Overall, the research underscores the transformative impact of integrating machine learning into CRM systems, equipping financial institutions to navigate a competitive and evolving digital landscape. The Deep Neural Network model accuracy has 79.7% and also it scored good results in all measurements; the accuracy value of 76.4% for the Random Forest algorithm comes in second place, and so on.

5.2. Future Work and Recommendation

There are several areas for future work that came up during this research. In this study, only one district from the Bunna Bank dataset was analyzed. Future studies could expand the scope by exploring data from other districts. Additionally, more research is needed to incorporate and analyze additional variables that may improve the model's performance and insights.

It's also recommended that the model be trained and tested using a larger and more complex dataset with more complex structures. While this study focused on five machine learning methods applied to the Bunna Bank dataset, future work could explore other techniques to see how they perform. There's still room to try different algorithms and carry out deeper data analysis for better results.

The insights gained from this study can help in building a predictive system to support and strengthen customer relationship management. Based on what was discovered, the following points are recommended for future consideration.

- ✓ This study tested several classification algorithms and found that the Deep Neural Network (DNN) outperformed others like Random Forest, K-Nearest Neighbors, Logistic Regression, and Support Vector Machine. Based on these results, the study recommends using the DNN model for predicting customer levels.
- ✓ Based on the feature ranking algorithms the customer data Internet_Banking, Mobile_Banking, Book_Type, Age, Sex, Customer_Job, No_of_Credit_Txn, No_of_Debit_Txn, Tenure, and Balance are the basic features can be used at a time of decision making as they had shown strong prediction of Customer_Levels power which can help to reduce the customer churn.
- ✓ The findings of this study can serve as valuable input for the company, and can be integrated with existing customer data to enhance service delivery and strengthen customer relationship management.

Since data preparation is often time-consuming, it's recommended that large organizations especially banks invest in or establish a proper data warehouse. This would allow them to store all their data in an organized way, based on key features, making it easier to apply machine learning techniques. Doing so can help streamline daily operations and improve service delivery.

Bibliography

[1] Aung, Tin Htun, Thu Ra Mon, and Amiya Bhaumik. "The Impact of Business Intelligence on Customer Relationship Management in the Banking Sector: A Financial Analysis." Advancement in Management and Technology (AMT) 4, no. 4 (2024): 1-11.

[2] K. Gillon, S. Aral, C. Y. Lin, S. Mithas, and M. Zozulia, "Business analytics: Radical shift or incremental change?," *Commun. Assoc. Inf. Syst.*, vol. 34, no. 1, pp. 287–296, 2014, doi: 10.17705/1cais.03413.

[3] A. Beckett, "From branches to call centres: new strategic realities in retail banking," *Serv. Ind. J.*, vol. 24, no. 3, pp. 43–62, May 2004, doi: 10.1080/0264206042000247759.

[4] A. Perwej, "Effective Management of Customer Relationship Management (CRM) In Banking Industry," *YOJNA Manag. J. KITE Gr.*, no. 3, 2010, [Online]. Available: https://www.researchgate.net/publication/230872506

[5] B. Pokharel, "Customer Relationship Management: Related Theories, Challenges and Application in Banking Sector," *Bank. J.*, vol. 1, no. 1, pp. 19–28, 1970, doi: 10.3126/bj.v1i1.5140.

[6] H. Kong, "Pr ep rin t n ot pe er re v iew Pr ep t n ot pe er v," 1862.

[7] R. Practice, "The consumer-data opportunity and the privacy imperative," no. April, 2020.

[8] M. Peterson, F. Gröne, K. Kammer, and J. Kirscheneder, "Multi-channel customer management: Delighting consumers, driving efficiency," *J. Direct, Data Digit. Mark. Pract.*, vol. 12, no. 1, pp. 10–15, 2010, doi: 10.1057/dddmp.2010.16.

[9] H. U. Khan *et al.*, "Transforming the Capabilities of Artificial Intelligence in GCC Financial Sector: A Systematic Literature Review," *Wirel. Commun. Mob. Comput.*, vol. 2022, p. 8725767, 2022, doi: 10.1155/2022/8725767.

[10] A. Ghazian, M. H. Hossaini, and H. Farsijani, "The Effect of Customer Relationship Management and its Significant Relationship by Customers' Reactions in LG Company," *Procedia Econ. Financ.*, vol. 36, no. 16, pp. 42–50, 2016, doi: 10.1016/s2212-5671(16)30014-4.

[11] J. Cvijovic, M. Kostic-Stankovic, and M. Reljic, "Customer relationship management in banking industry: Modern approach," *Industrija*, vol. 45, no. 3, pp. 151–165, 2017, doi: 10.5937/industrija45-15975.

[12] O. A, "The Digital Transformation in Banking and The Role of FinTechs in the New Financial Intermediation Scenario," *Int. J. Financ. Econ. Trade*, no. 85228, pp. 1–6, 2017, doi: 10.19070/2643-038x-170001.

[13] J. J. Kim, L. Steinhoff, and R. W. Palmatier, "An emerging theory of loyalty program dynamics," *J. Acad. Mark. Sci.*, vol. 49, no. 1, pp. 71–95, 2021, doi: 10.1007/s11747-020-00719-1.

[14] M.-H. Huang and R. T. Rust, "A strategic framework for artificial intelligence in marketing," *J. Acad. Mark. Sci.*, vol. 49, no. 1, pp. 30–50, 2021, doi: 10.1007/s11747-020-00749-9.

[15] "Challenges of information technology audit: A case study on selected private commercial banks in Ethiopia," no. March, 2023.

[16] Olalekan Hamed Olayinka, "Data driven customer segmentation and personalization strategies in modern business intelligence frameworks," *World J. Adv. Res. Rev.*, vol. 12, no. 3, pp. 711–726, 2021, doi: 10.30574/wjarr.2021.12.3.0658.

[17] U. C. L. A. L. Review *et al.*, "U.C.L.A. Law Review," vol. 232, no. 2018, pp. 232–279.

[18] E. Lemma, "CORE BANKING SYSTEM EFFECTIVENESS IN ETHIOPIA: THE CASE OF BUNNA INTERNATIONAL BANK," *Int. J. Manag. Res. Rev.*, vol. 6, pp. 267–277, Feb. 2016.

[19] B. S. C. Casestudy, I. N. Addis, and A. Branches, "ST . MARRY ' S UNIVERSITY SCHOOL OF GRADUATE STUDIES By : DEMEKE TILAHUN ZELALEM," 2019.

[20] S. A. Shuvo, M. Tabassum, N. Tafannum, and S. Chadni, "Machine Learning in Business Intelligence: From Data Mining To Strategic Insights in Mis," *Rev. Appl. Sci. Technol.*, vol. 04, no. 02, pp. 339–369, 2025, doi: 10.63125/drb8py41.

[21] W. J. Reinartz and V. Kumar, "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *J. Mark.*, vol. 67, no. 1, pp. 77–99, Jan. 2003, doi: 10.1509/jmkg.67.1.77.18589.

[22] M. O'Connell, "Bank-specific, industry-specific and macroeconomic determinants of bank profitability: evidence from the UK," *Stud. Econ. Financ.*, vol. 40, no. 1, pp. 155–174, 2023, doi: 10.1108/SEF-10-2021-0413.

[23] D. B. Datta, "Retail Banking Operations at Union Bank PLC .," 2024.

[24] L. Puddu, "A contested financial frontier: Banking and empire building in Eritrea, c.1952-73," *Africa (Lond).*, vol. 91, no. 5, pp. 852–873, 2021, doi: 10.1017/S0001972021000619.

[25] "ANNUAL," 2023.

[26] M. K. Shukla and P. N. Pattnaik, "Managing Customer Relations in a Modern Business Environment: Towards an Ecosystem-Based Sustainable CRM Model," *J. Relatsh. Mark.*, vol. 18, no. 1, pp. 17–33, 2019, doi: 10.1080/15332667.2018.1534057.

[27] N. Rane, S. Choudhary, and J. Rane, "Metaverse for Enhancing Customer Loyalty: Effective Strategies to Improve Customer Relationship, Service, Engagement, Satisfaction, and Experience," *SSRN Electron. J.*, no. 05, pp. 427–452, 2023, doi: 10.2139/ssrn.4624197.

[28] M. Čavlin, V. Bulović, and B. Šmitran, "Strategic Benefits of a Customer-Centric Approach in Customer Relationship Management System," *J. Agron. Technol. Eng. Manag.*, vol. 7, no. 3, pp. 1114–1123, 2024, doi: 10.55817/blvs9823.

[29] M. Majka and N. T. Poland, "Marketing with Customer Lifetime Value," no. October, 2024.

[30] N. K. Dubey, P. Sharma, and P. Sangle, "Implementation and adoption of CRM and co-creation leveraging collaborative technologies: An Indian banking context," *J. Indian Bus. Res.*, vol. 12, no. 1, pp. 113–132, 2020, doi: 10.1108/JIBR-09-2019-0284.

[31] H. A. Al-homery, H. Asharai, and A. Ahmad, "The Core Components and Types of CRM I . Introduction," *Pakistan J. Humanit. Soc. Sci.*, vol. 7, no. 1, pp. 121–145, 2019.

[32] A. Okunola, "Unlocking Proactive Customer Support with Machine Learning : Methods for Anticipating and Resolving Issues," no. November, 2024.

[33] Z. Soltani, B. Zareie, F. S. Milani, and N. J. Navimipour, "The impact of the customer relationship management on the organization performance," *J. High Technol. Manag. Res.*, vol. 29, no. 2, pp. 237–246, 2018, doi: 10.1016/j.hitech.2018.10.001.

[34] A. A. Kankam Boadu, "Customer Relationship Management and Customer Retention," *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3472492.

[35] B. G. P. B. G. L. I. S. S. S. Sandeep Saxena, "Analysing Sales Enablement Technologies and Their Role in Enhancing Sales Teams," *Tuijin Jishu/Journal Propuls. Technol.*, vol. 44, no. 4, pp. 4473–4480, 2023, doi: 10.52783/tjjpt.v44.i4.1693.

[36] P. Agarwal, “Harnessing the Power of Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) Systems for Sustainable Business Practices,” *Int. J. Comput. Trends Technol.*, vol. 72, no. 4, pp. 102–110, 2024, doi: 10.14445/22312803/ijctt-v72i4p113.

[37] S. Payne and P. Frow, "A strategic framework for customer relationship management," Journal of Marketing, vol. 69, no. 4, pp. 167–176, Oct. 2005.

[38] T. Alexander, “Proactive Customer Support: Re-Architecting A Customer Support/Relationship Management Software System Leveraging Predictive Analysis/AI and Machine Learning,” *Eng. Open Access*, vol. 2, no. 1, pp. 39–50, 2024, doi: 10.33140/eoa.02.01.04.

[39] J. Sheth, V. Jain, and A. Ambika, “Repositioning the customer support services: the next frontier of competitive advantage,” *Eur. J. Mark.*, vol. 54, no. 7, pp. 1787–1804, 2020, doi: 10.1108/EJM-02-2020-0086.

[40] T. V. Iyelolu, E. E. Agu, C. Idemudia, and T. I. Ijomah, “Improving Customer Engagement and CRM for SMEs with AI- Driven Solutions and Future Enhancements Improving Customer Engagement and CRM for SMEs with AI- Driven Solutions and Future Enhancements,” no. September, 2024.

[41] Uloma Stella Nwabekee, Oluwatosin Yetunde Abdul-Azeez, Edith Ebele Agu, and Tochukwu Ignatius Ijomah, “Digital transformation in marketing strategies: The role of data analytics and CRM tools,” *Int. J. Front. Res. Sci. Technol.*, vol. 3, no. 2, pp. 055–072, 2024, doi: 10.56355/ijfrst.2024.3.2.0047.

[42] Chinazor Prisca Amajuoyi, Luther Kington Nwobodo, and Ayodeji Enoch Adegbola, “Utilizing predictive analytics to boost customer loyalty and drive business expansion,” *GSC Adv. Res. Rev.*, vol. 19, no. 3, pp. 191–202, 2024, doi: 10.30574/gscarr.2024.19.3.0210.

[43] P. Kumar, “Ultrasoft B-Open Sets and Ultrasoft B-Continuous Functions,” *J. Res. Sci. Eng.*, vol. 6, no. 7, pp. 81–85, 2024, doi: 10.53469/jrse.2024.06(07).16.

[44] A. S. Crm, “JBFE Enhancing Customer Experience and Business Intelligence : The,” vol. 1, no. 2, pp. 1–8, 2024.

[45] M. Naslednikov, “The Impact of Artificial Intelligence on Customer Relationship Management (CRM) Strategies,” *Haaga-Helia Univ. Appl. Sci.*, 2024, [Online]. Available: https://www.theseus.fi/bitstream/handle/10024/858753/Naslednikov_Mikhail.pdf?sequence=2&isAllowed=y

[46] T. H. Aung, T. R. Mo, and A. Bhaumi, “The Impact of Business Intelligence on Customer Relationship Management in the Banking Sector: A Financial Analysis,” *Adv. Manag. Technol.*, vol. 04, no. 04, pp. 01–10, 2024, doi: 10.46977/apjmt.2024.v04i04.001.

[47] L. Anderson, “Doctor of Business Administration,” *Train. Dev.*, vol. 58, no. 6, pp. 3–75, 2004.

[48] B. J. K. Mbatia, “Influence Of Electronic Customer Relationship Management, Competitive Advantage And Performance Of Commercial Banks In Nairobi City County Kenya,” 2019.

[49] N. Rane, S. Choudhary, and J. Rane, “Hyper-personalization for enhancing customer loyalty and satisfaction in Customer Relationship Management (CRM) systems,” *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4641044.

[50] M. R. B, “Adoption Of Consumer Behaviour Metrics Effective Virtual Retail Marketing,” *Educ. Adm. Theory Pract.*, vol. 30, no. 4, pp. 1887–1889, 2024, doi: 10.53555/kuey.v30i4.1780.

[51] X. Zhao and P. Keikhosrokiani, “Sales prediction and product recommendation model through user behavior analytics,” *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 3855–3874, 2022, doi: 10.32604/cmc.2022.019750.

[52] "Bank CRM Optimization Using Predictive Classification."

[53] E. Ok, J. Eniola, C. Islands, and C. Islands, "Enhancing Customer Segmentation Strategies for Improved Personalization in the Banking Industry," no. January 2024, 2024.

[54] N. Bose, A. Chopra, P. Joshi, and A. Reddy, "Leveraging Reinforcement Learning and Predictive Analytics for Enhanced Customer Lifetime Value Optimization Authors :," pp. 1–30.

[55] S. Raschka, V. Mirjalili, and A. Khan, Machine Learning with PyTorch and Scikit-Learn, 2nd ed. Birmingham, U.K.: Packt Publishing, 2022

[56] P. Kagwa, "Role of Machine Learning Algorithms in Enhancing Customer Relationship Management Systems in Rwanda," *Am. J. Data, Inf. Knowl. Manag.*, vol. 5, no. 2, pp. 1–12, 2024, doi: 10.47672/ajdikm.2348.

[57] H. Salman, J. Grover, and T. Shankar, "Hierarchical Reinforcement Learning for Sequencing Behaviors," vol. 2733, no. March, pp. 2709–2733, 2018, doi: 10.1162/NECO.

[58] F. Abramovich, V. Grinshtein, and T. Levy, "Multiclass classification by sparse multinomial logistic regression," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4637–4646, 2021, doi: 10.1109/TIT.2021.3075137.

[59] A. Sharma, R. Kumar, and V. Mansotra, "Proposed Stemming Algorithm for Hindi Information Retrieval," *Int. J. Innov. Res. Comput. Commun. Eng. (An ISO Certif. Organ.*, vol. 3297, no. 6, pp. 11449–11455, 2016, doi: 10.15680/IJIRCCE.2016.

[60] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016, doi: 10.21037/atm.2016.03.37.

[61] E. Rosenzweig, "Successful User Experience :," 2015.

[62] M. Sun, "Support Vector Machine Models for Classification," *Encycl. Bus. Anal. Optim.*, pp. 2395–2409, 2014, doi: 10.4018/978-1-4666-5202-6.ch215.

[63] A. Thakur, "Fundamentals of Neural Networks," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. VIII, pp. 407–426, 2021, doi: 10.22214/ijraset.2021.37362.

[64] H. Amnur, "Customer relationship management and machine learning technology for identifying the customer," *Int. J. Informatics Vis.*, vol. 1, no. 1, pp. 12–15, 2017, doi: 10.30630/joiv.1.1.10.

[65] N. Singh, P. Singh, and M. Gupta, "An inclusive survey on machine learning for CRM: a paradigm shift," *Decision*, vol. 47, no. 4, pp. 447–457, 2020, doi: 10.1007/s40622-020-00261-7.

[66] B. Bhattacharjee, "A COMPREHENSIVE STUDY OF MACHINE LEARNING APPROACHES FOR CUSTOMER," vol. 06, no. October, 2024, doi: 10.37547/tajet/Volume06Issue10-11.

[67] V. Anandi and M. Ramesh, "Analysis of customer relationship management using Machine Learning," *Turkish J. Comput. Math. …*, vol. 12, no. 14, pp. 5505–5512, 2021.

[68] A. Khaliq, S. Ajaz, A. Ali, D. Shakir, and K. Baig, "From Data to Decisions : Predictive Machine Learning Models for Customer Retention in Banking," vol. 4, no. 3, 2024.

[69] O. Id *et al.*, "Утримання клієнтів комерційного банку як задача класифікації у машинному навчанні," vol. 0259, pp. 179–190, 2024.

[70] Z. Revesai, M. Dzinomwa, B. Ndlovu, P. Nyoni, and S. Dube, "Customer Churn Analytics using Classical Machine Learning Algorithms and Deep Neural Networks: A Case of Zimbabwe Banks," no. October, 2023, doi: 10.46254/ap04.20230040.

[71] A. Todupunuri, "Develop Machine Learning Models to Predict Customer Lifetime Value for Banking Customers , Helping Banks Optimize Services," no. October, 2024.

[72] Y. W. Wardana and I. R. Pratama, "The Effect of Internet Banking , Product Advantages , and Customer Relationship Management on Customer Loyalty at Bank Syariah Indonesia Semarang City," vol. 6, no. 2, pp. 146–161, 2024.

[73] M. Rahman and V. Kumar, "Machine Learning Based Customer Churn Prediction in Banking," *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, no. October, pp. 1196–1201, 2020, doi: 10.1109/ICECA49313.2020.9297529.

[74] P. Verma, "Churn prediction for savings bank customers: A machine learning approach," *J. Stat. Appl. Probab.*, vol. 9, no. 3, pp. 535–547, 2020, doi: 10.18576/JSAP/090310.

[75] Ibrahim Adedeji Adeniran, Christianah Pelumi Efunniyi, Olajide Soji Osundare, and Angela Omozele 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, doi: 10.51594/csitrj.v5i8.1489.

[76] R. Author *et al.*, "This content downloaded from 129.252.86.83 on Wed," *Source MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.

[77] J. vom Brocke, A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, "Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research," *Commun. Assoc. Inf. Syst.*, vol. 37, pp. 205–224, 2015, doi: 10.17705/1cais.03709.

[78] K. Peffers *et al.*, "A Model for Producing and Presenting Information Systems Research," *Comuter Sci. Res.*, p. 24, 2007, [Online]. Available: h

[79] Y. Roh, G. Heo, and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1328–1347, Apr. 2021, doi: 10.1109/TKDE.2019.2946162.

[80] A. Amine and S. Gatfaoui, "Temporarily vulnerable consumers in a bank services setting," *J. Serv. Mark.*, vol. 33, no. 5, pp. 602–614, 2019, doi: 10.1108/JSM-05-2018-0154.

[81] A. Astorino, E. Gorgone, M. Gaudioso, and D. Pallaschke, "Data preprocessing in semi-supervised SVM classification," *Optimization*, vol. 60, no. 1–2, pp. 143–151, Jan. 2011, doi: 10.1080/02331931003692557.

[82] D. Z. Abidin, S. Nurmaini, R. Firsandava Malik, Erwin, E. Rasywir, and Y. Pratama, "RSSI Data Preparation for Machine Learning," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Nov. 2020, pp. 284–289. doi: 10.1109/ICIMCIS51567.2020.9354273.

[83] S. Swain, S. Chakravarty, B. Paikaray, and H. Bhoyar, "Heart Disease Detection Using Machine Learning Techniques," *Lect. Notes Electr. Eng.*, vol. 990 LNEE, pp. 273–284, 2023, doi: 10.1007/978-981-19-9090-8_24.

[84] F. Of, T. O. N. The, and N. Of, "SETH CARNAHAN University of Illinois," *Business*, pp. 1–43.

[85] V. Sowmya and N. Kayarvizhy, "An Efficient Missing Data Imputation Model on Numerical Data," in *2021 2nd Global Conference for Advancement in Technology (GCAT)*, 2021, pp. 1–8. doi: 10.1109/GCAT52182.2021.9587886.

[86] J. S. Murray and J. P. Reiter, "Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models With Local Dependence," *J. Am. Stat. Assoc.*, vol. 111, no. 516, pp. 1466–1479, Oct. 2016, doi: 10.1080/01621459.2016.1174132.

[87] E. Shaaban, Y. Helmy, and A. Khedr, "A Proposed Churn Prediction Model," *Mona Nasr / Int. J. Eng. Res.*

Appl., vol. 2, no. 4, pp. 693–697, 2012, [Online]. Available: www.ijera.com

[88] X. Guo, L. Wang, L. Tian, and X. Li, "Elevation-dependent reductions in wind speed over and around the Tibetan Plateau," *Int. J. Climatol.*, vol. 37, no. 2, pp. 1117–1126, 2017, doi: 10.1002/joc.4727.

[89] Ö. Gür Ali and U. Arıtürk, "Dynamic churn prediction framework with more effective use of rare event data: The case of private banking," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7889–7903, 2014, doi: https://doi.org/10.1016/j.eswa.2014.06.018.

[90] A. Hevner and S. Chatterjee, "Design Science Research in Information Systems BT - Design Research in Information Systems: Theory and Practice," A. Hevner and S. Chatterjee, Eds., Boston, MA: Springer US, 2010, pp. 9–22. doi: 10.1007/978-1-4419-5653-8_2.

[91] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: A proposal and a discussion," *Requir. Eng.*, vol. 11, pp. 102–107, Mar. 2006, doi: 10.1007/s00766-005-0021-6.

[92] J. vom Brocke, A. Hevner, and A. Maedche, "Introduction to Design Science Research BT - Design Science Research. Cases," J. vom Brocke, A. Hevner, and A. Maedche, Eds., Cham: Springer International Publishing, 2020, pp. 1–13. doi: 10.1007/978-3-030-46781-4_1.

[93] D. Callaghan, J. Burger, and A. K. Mishra, "A machine learning approach to radar sea clutter suppression," *2017 IEEE Radar Conf. RadarConf 2017*, pp. 1222–1227, 2017, doi: 10.1109/RADAR.2017.7944391.

[94] J. Patterson and A. Gibson, *Deep learning: A practitioner's approach*. " O'Reilly Media, Inc.," 2017.

[95] G. Schaefer, "Strategies for imbalanced pattern classification for digital pathology," in *2017 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT)*, 2017, pp. 1–4. doi: 10.1109/ICIEV.2017.8338552.

[96] M. I. C. Rachmatullah, J. Santoso, and K. Surendro, "A Novel Approach in Determining Neural Networks Architecture to Classify Data With Large Number of Attributes," *IEEE Access*, vol. 8, pp. 204728–204743, 2020, doi: 10.1109/ACCESS.2020.3036853.

Deep Neural Network Python Code

```
import pandas as pd
import numpy as np
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, accuracy_score
X = df3.drop(['CUSTOMER_LEVEL'], axis=1)
y = df3['CUSTOMER_LEVEL']
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
from imblearn.over_sampling import SMOTE
oversample = SMOTE()
X_balance, y_balance = oversample.fit_resample(X_train, y_train)
scaler = StandardScaler()
X_balance = scaler.fit_transform(X_balance)
X_test = scaler.transform(X_test)
model = Sequential()
model.add(Dense(128, input_dim=X_balance.shape[1], activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(64, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(16, activation='relu'))
num_classes = len(np.unique(y_balance))
model.add(Dense(num_classes, activation='softmax'))
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
history = model.fit(X_balance, y_balance, epochs=50, batch_size=32, validation_split=0.2, verbose=1)
y_pred = np.argmax(model.predict(X_test), axis=1)
print("Classification Report:\n", classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred)
```