St. Mary's University

Department of Computer Science

Predicting Financial Credit Risks of Banks Using Machine Learning Algorithm

By: SenaitTeklemarkos

Advisor: Dr. Million M.

February, 2025

Addis Ababa, Ethiopia

ACCEPTANCE

Predicting financial credit risks of Banks using machine learning algorithm

By

SenaitTeklemarkos

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science

Thesis Examination Committee:

Name	**Signture**	**Date**
____________________	__________________	________________
Internal Examiner		
____________________	__________________	________________
External Examiner		

Dean, Faculty of Informatics

Name	**Signture**	**Date**
____________________	__________________	________________

December, 2024

Declaration

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been duly acknowledged.

Name: SenaitTeklemarkos

Signature: ____________________________

Date: ____________________________

This thesis has been submitted for examination with my approval as the advisor, **Dr. Million M.**

Signature: ____million____________________

Date: ___January 02, 2025___________________

Addis Ababa,

Ethiopia

December, 2024

Abstract

Credit risk is an important factor influencing bank financial performance, and the capacity to foresee it enables institutions such as Banks to manage potential risks and maintain their profitability. Accurate credit risk prediction allows banks to make informed decisions by identifying customers who are likely to default in advance. In this study, multiple machine learning approaches are used on Awash Bank customer data to create a prediction model capable of predicting credit risk. Missing values in numerical features are filled using the mean, while categorical features are filled with the mode. Categorical features are encoded using Label Encoding, except the 'Branch' variable, which, due to its high cardinality of 124 unique values, is encoded using the Hasher function, a method suggested for features of this type. The dataset is split into training, testing, and validation sets using an 80:20:10 ratio, where 10% of the training set is reserved for validation. Key characteristics are identified by applying a correlation analysis and the ExtraTreesClassifier, and class imbalance is handled using the SMOTE oversampling approach to avoid bias against the majority class. Five machine learning models—XGBoost, CatBoost, Random Forest, Support Vector Machine (SVM), and Deep Neural Networks (DNN)—are trained on the dataset and tested for accuracy, precision, recall, and F1 score. Hyperparameter tuning is performed using RandomizedSearchCV() to optimize the performance of each selected model. The results show that the XGBoost algorithm outperformed the others, with an accuracy of 92.2%, followed by CatBoost and Random Forest.This study contributes to the limited research on credit risk prediction in the Ethiopian banking sector by utilizing real data from Awash Bank and demonstrating the potential for machine learning, particularly ensemble methods such as XGBoost, to improve credit risk management in the banking industry. However, a major limitation of this study is the reliance on a limited dataset focused exclusively on loans, which may not fully represent the diverse customer base of Awash Bank, particularly those seeking other types of credit products. Future research could address this limitation by incorporating additional data sources or conducting longitudinal studies to enhance predictive accuracy and generalizability.

Keywords: Banking *industry, Credit risk prediction, Ensemble Machine Learning, Deep Neural Networks*

Acknowledgment

Foremost, I would like to thank the almighty GOD for giving me the ability, and Strength to learn, understand, and conclude my thesis report with his mercy and grace.

This thesis is dedicated to my parents, to whom I am deeply grateful to my mother, TsehayDegene, for her endless love and encouragement, and to my father, Teklemarkos Haile, for his unwavering belief in my potential.

First, I sincerely thank my advisor, Dr. Million.M, for his exceptional guidance, valuable advice, and constant support throughout this thesis work. I am also grateful to all the postgraduate coordinators at St. Mary's University for their availability and continued advice.

It is a great pleasure to acknowledge my deepest thanks and gratitude to Awash International Bank, and My Dearest Friend EdemealmDesalegn for their support and guidance. They supported me during data collection and understanding of the domain.

Finally, thank you! All my friends and the defense committee.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
AIB	Awash International Bank
CatBoost	Categorical Boosting
CSV	Comma-separated value
CV	Cross-Validation
DM	Data Mining
FS	Feature selection
KNN	K-Nearest Nehbour
MFI	Micro Finance Institution
ML	Machine Learning
NPL	Non-Performing Loan
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
RF	Random Forest
KNN	K- nearest neighbor
XGB	Extrem Gradient Boosting
TP	True Positive

CHAPTER ONE

Introduction

1.1. Background

In the realm of machine learning (ML) and artificial intelligence (AI), data is considered the foundation and even the backbone of these technologies [1]. A vast chunk of data is created from numerous sources worldwide, such as social media and financial institutions. This data has huge potential, and nowadays almost all companies are focusing on the benefits of data to gain profitability, reduce risks, and improve performance using AI-based solutions [2].

Facing credit risk prediction is a known challenge for financial institutions worldwide [3]. In emerging markets, it is crucial to have credit risk management to ensure the stability and growth of both the banking sector and the country as a whole. Different studies suggest various approaches to enhance credit risk management in financial institutions. A proactive credit management plan is essential, focusing on risk identification, evaluation, and strategic measures to mitigate potential losses while optimizing cash flow and ensuring timely payments[4]. The integration of automated systems for real-time credit risk monitoring is emphasized, allowing organizations to dynamically track changes in credit risk and streamline processes such as credit applications[5]. Furthermore, leveraging data analytics and machine learning models can significantly improve credit risk assessment by identifying patterns and predicting defaults more accurately, which includes handling missing values and addressing class imbalances in the data[5]. In Ethiopia, the financial system has been dealing with challenges in managing its credit portfolio [6].

The issue of credit risk in countries like Ethiopia is exacerbated by a lack of data and limited access to advanced credit risk assessment models [6]. Reliant on traditional methods, like manual and rule-based approaches, have proven insufficient in predicting the creditworthiness of borrowers, particularly those with limited financial backgrounds or minimal exposure to formal credit channels [7].

Machine learning technologies have made great strides recently and promise to address the problem of credit risk [8]. For example, they are valuable tools used in analyzing complex data sets that allow financial institutions to design more accurate and data-driven credit risk assessment models. These advanced techniques are invaluable for analyzing complex data sets, enabling financial institutions to design more accurate and data-driven credit risk assessment models. While traditional machine learning methods have proven effective, deep learning approaches further enhance predictive accuracy by automatically learning hierarchical representations from large amounts of data, making them particularly suitable for capturing intricate patterns in credit risk.

This research uses supervised machine learning techniques and deep learning to construct a predictive model for credit risk assessment using data collected from one of the well-known Ethiopian banks, Awash International Bank (AIB). This study proposes the use of advanced data analytics and modeling techniques aimed at improving credit risk management practices in banks which possibly motivates other financial institutions in the country as well as international ones. The study uses Deep Neural Networks, RandomForest, CatBoost, XGBoost, and Support Vector Machine algorithms to build a robust credit risk prediction model.

The successful development and implementation of an operative machine learning-based credit risk prediction model can have far-reaching impacts on the stability and growth of Ethiopia's banking sector as well as its wider economic development. Through this research, therefore, a significant opportunity presents itself for dealing with this critical challenge facing the Ethiopian financial system and opening up avenues for sounder institutional frameworks in emerging economies concerning prudent but workable credit risk management systems.

1.2. Statement of the Problem

The lack of effective techniques and models for credit risk assessment and management poses a major challenge for developing countries, leading to prominent economic problems due to financial crises. In Ethiopia, financial institutions often face bankruptcy due to poor credit scoring management, as they depend on the individual skills and knowledge of banks rather than robust, data-driven methods. For instance, AIB, one of the oldest and most impactful banks in the country, has faced significant issues with its credit portfolio [6].

Emerging technologies like artificial intelligence, machine learning, and blockchain have significant impacts on IT management since they introduce new capabilities and efficiency.

However, the integration of these technologies emphasizes the importance of individual skills and expertise inside organizations. This problem is not solely Ethiopian, as most developing countries have suffered similar economic challenges. For instance, the proportion of non-performing loans in several African countries ranged from 5% to 20% of total loan portfolios [9].

Different works are done to apply machine learning algorithms for constructing predictive models. Bizuwork et al. [10] have attempted to develop automated prediction models using machine-learning techniques for loan risk analysis. TsegaAsres et al. [11] investigated credit risk assessment for predicting loan defaulters in Ethiopian banking industry. Also, Walfaanaa M Ejeta[12], tried to apply artificial intelligence-enhanced credit risk assessment to the Commercial Bank of Ethiopia. Tamiru et al.[13] assessed Performance analysis of deep and machine learning algorithms for loan evaluation model.

Despite the advancements in loan risk analysis, there remain significant gaps in the existing research, particularly in the application of machine learning models. Many studies, such as those by TsegaAsres et al. and Walfaanaa M. Ejeta, have focused on specific aspects of credit risk assessment but often lack comprehensive real-world data and diverse research methodologies. Furthermore, while TamiruMelese et al. have conducted performance analyses of various algorithms, there is still a limited exploration of how these machine-learning approaches can be effectively integrated into risk management practices in Ethiopian financial institutions. These issues underscore the pressing need for advanced risk management tools at institutions like Banks to predict customer repayment behavior accurately and mitigate credit risks. This study, therefore, investigates and constructs a machine learning model for credit risk predictions, thereby improving the quality of credit portfolios in Ethiopia with a potentialto extend it to other developing countries facing similar challenges.

1.3. Research questions

To solve the problem mentioned above, this study investigates and answers the following research questions:

- What are the key features for predicting credit risk within Bank?
- Which machine learning algorithm is the most suitable for constructing optimal credit risk predictive modeling?

- What is the performance of the proposed predictive model in credit risk predictions?

1.4. Objective of the study

The study outlines the following general and specific objectives.

1.4.1. General Objective

The general objective of this study is to construct machine learning models that can predict credit risk of Banks, thereby enhancing the quality of their credit portfolio.

1.4.2. Specific Objectives

To achieve the general objective of the study, the following specific objectives are formulated.

- To review literature and related works to identify techniques and algorithms.
- To compile and arrange a comprehensive collection of loan records including details about borrowers and their repayment histories.
- To prepare the data collected for experimentation
- To evaluate several machine learning techniques, to determine the most reliable and accurate one for credit risk prediction.
- To implement the chosen machine learning algorithms, evaluate their efficacy, and select the best-performing model for improving the bank's loan portfolio
- To evaluate the performance of the proposed model

1.5. Scope of the study

The research is conducted within the financial services industry, specifically on the credit analysis and loan approval processes in banks.This focus is due to practical considerations, including the availability of data, the willingness of the organization to provide access, and resource constraints that limit the scope of the study. The research aims at investigating and identifying patterns that help predict the probability of default for a given loan applicant and classify whether to accept or reject the loan application using the borrowers' historical data. This work encompasses the entire process, from preprocessing the loan and borrowers' historical data to their proper prediction and classification of the loan applicants.

The research attempts to pull the loan and borrowers' data collected from AIB to serve as the dataset for learning, prediction, and classification purposes. This dataset spans a time coverage from 2017 to 2024, allowing for a comprehensive evaluation of patterns over several years. Evaluating the performance of the machine learning models against loan portfolios used by AIB for credit analysis and loan approval decisions.

1.6. Significance of the study

Ethiopian banks will find boundless worth in this study. The stability and extension of Ethiopia's banking industry as well as the state's economy as a whole are significantly impacted by credit risk management approaches.

This study aims to solve a main issue that the Ethiopian banking sector and other emerging nations face: the absence of efficient credit risk evaluation. Explicitly, it intends to construct an efficient machine-learning model to boost credit risk assessment in banks. If the strategy is successfully implemented, it can improve the quality of the credit portfolio, lower the risk of financial crises, and support sustainable economic growth in Ethiopia and probably other developing countries dealing with comparable problems.

Although this study primarily utilizes data from AIB, the largest private commercial bank in Ethiopia, its findings and recommendations hold broader significance for the entire Ethiopian banking sector. Given AIB's leading market position, improvements in credit risk management practices at this institution can serve as a model for other banks in Ethiopia. By demonstrating the effectiveness of data-driven, machine learning-based strategies, this research can encourage other financial institutions both in Ethiopia and in other developing countries to adopt similar advanced approaches to credit risk management. Ultimately, this study aims to contribute to a more robust banking environment in Ethiopia, fostering financial stability and sustainable growth.

The findings of this research enhance credit risk assessment and loan decision-making processes at Banks, leading to more informed and efficient lending practices. The implementation of machine learning models can improve risk management strategies, reducing default rates and enhancing profitability. Policymakers may benefit from insights into effective credit risk management practices, which can inform regulations and guidelines that promote financial stability within the banking sector. Researchers can use this study as a foundation for further exploration into data-driven methodologies in credit risk assessment, fostering innovation in the field. Additionally, customers stand to gain from improved lending practices, as more accurate risk assessments may lead to fairer loan terms and increased access to credit. Overall, this study aims to create a positive ripple effect across the Ethiopian banking landscape and beyond.

1.7. Methodology of the Study

1.7.1. Research design

This research follows experimental research. Experimental research is a scientific method that involves the use of controlled experiments or empirical observations to investigate a specific research question or hypothesis. This type of research design lets the researcher establish causal relationships between the manipulated variables and the observed outcomes[14]. The controlled nature, causal inference, and hypothesis-testing abilities of experimental research make it a priceless tool for advancing scientific knowledge and driving progress across a wide range of academic and applied disciplines.

To conduct an extensive experiment, the study undertakes data collection and preparation, constructing an optimal model using machine learning algorithms, and evaluating the proposed model to measure its prediction performance.

1.7.2. Data collection and preparation

Data is collected from AIB,to evaluate the quality of past loan data including application details, credit profiles, approval decisions, and loan performance.The dataset contains a set of features that is used to predict whether to approve or deny a loan application.The data covers more than 13 thousand customers who joined the bank between 2014 and 2024 to access credit services.

Dtata preparation then takes place, handling missing values, removing outliers, performing feature engineering, encoding categorical variables, and normalizing the data; after that transforming the data into the required format and structure for running machine learning algorithms to create credit risk prediction model.

1.7.3. Implementation tool

For implementation purposes, Python programming language is used. Python is a broadly used and versatile programming language that is best suited for building predictive models and data-driven applications. Python is an outstanding choice for building classification models due to[15]:

- Extensive Library Ecosystem: Python has a vast collection of robust and well-maintained machine-learning libraries, such as Panda, Numpy, Scikit-learn, TensorFlow, and Keras, which provide a wide range of classification algorithms and utilities.
- Ease of Use and Readability: Python's syntax is known for its simplicity and readability, making it easy for beginners to understand and write code.
- Integration with Other Tools: Python can integrate with other tools and technologies, such as databases, web frameworks, and visualization libraries, without a glitch.
- Flexibility and Versatility: Python is a general-purpose programming language, which implies it can be used for a wide variety of tasks, including data preprocessing, feature engineering, model training, and deployment.

These features make Python a brilliant choice for building classification models, as it delivers a robust, flexible, and user-friendly environment like the Pandas and Numpy, Scikit-learn (sklearn) and Matplotlib and Seaborn for data scientists and machine learning engineers, as well as academic researchers, to develop, test, and deploy their solutions.

- Pandas :- this data manipulation and analysis libraries are essential for preprocessing and preparing credit risk data for modeling.
- Matplotlib and Seaborn:- these data visualization libraries are used to explore and visualize the credit risk data, which can provide valuable insights during the modeling process.
- Scikit-learn (sklearn):- provides a wide range of algorithms and tools for classification, regression, and clustering tasks, including those relevant to credit risk prediction. It also provides evaluation metrics that can be used to evaluate the performance of various machine learning models. It also provides functionality for splitting datasets into training and testing sets.

1.7.4. Evaluation methods

Evaluation is necessary for ensuring that machine learning models are capable of making accurate predictions on new, unseen data, which is crucial for their successful deployment in real-world applications. Evaluation metrics help to realize the strengths and weaknesses of the credit risk prediction model and lead to improvements. In this study accuracy, recall, precision and F-score are used for measuring effectiveness model based on confusion matrix.

Confusion matrix: provides a breakdown of the model's predictions compared to the actual outcomes, allowing us to assess key performance metrics like accuracy, precision, recall, and F1-score.

Accuracy measures the overall correctness of the model in classifying the data set into its possible classes. **Precision** measures the proportion of true positives classified among the predicted positive instances. **Recall** measures the proportion of true positives that are correctly identified from the data set. **F1-Score** finds harmonic mean by Combining precision and recall of the model into a single metric.

CHAPTER TWO

Review of Literature AND RELATED WORKS

2.1. Overview

This chapter offers a detailed review of the existing literature on the application of machine learning techniques for credit risk prediction, concentrating specifically on the banking sector in Ethiopia. It starts by discussing the critical importance of effective credit risk management for financial institutions. The review highlights the challenges banks face in properly assessing the creditworthiness of borrowers and the potential financial instability caused by loan defaults.

The chapter further digs into the growing research on using machine learning algorithms for credit risk prediction, observing supervised learning, and unsupervised learning techniques. Additionally, the literature review explores the unique characteristics of the Ethiopian banking sector, including data availability, regulatory environment, and socio-economic factors that may affect credit risk. It discusses existing studies on machine learning applications for credit risk assessment in the Ethiopian financial market, identifying gaps and limitations in current research. The chapter concludes by synthesizing insights from the reviewed literature and outlining this thesis's potential contributions, which aim to develop and evaluate robust machine-learning models tailored to the specific needs and challenges faced by Awash Bank S.C.

2.2. Introducing Credit Risk

Credit risk refers to the risk of financial losses suffered by lenders when borrowers fail to meet their repayment duties [16]. In the financial sector, credit risk is a chief concern, as it can have significant implications for the stability and profitability of banks, lending institutions, and the overall financial system.

Credit risk can arise from various sources. Some of the risks related to borrowers, financial institutions, and country are described as follows [17][18]:

- Borrower default risk:- This refers to the risk that a borrower, whether an individual, a business, or a government entity, may fail to make scheduled payments on a loan or other debt obligation [18].

- Concentration risk:- This is the risk that arises when a financial institution's credit exposures are concentrated in a particular sector, geographic region, or group of connected borrowers [17].
- Collateral risk:- This refers to the risk that the value of the collateral provided by the borrower may not be sufficient to cover the outstanding loan amount in the event of default [18].
- Counterparty risk:- This is the risk that a counterparty to a financial transaction, such as a derivative contract or a repurchase agreement, may fail to fulfill its contractual obligations.
- Country risk:- This encompasses the risks associated with the political, economic, and social conditions of the country in which the financial institution operates.

Effective credit risk management is crucial for the stability and profitability of financial institutions, as it involves robust credit analysis, diversification of the loan portfolio, collateral management, and the implementation of prudent credit risk policies and procedures[19].

AIB's approach to credit risk management is primarily based on manual processes rather than comprehensive or advanced technology. The bank conducts a thorough loan underwriting and approval procedure, evaluating applicants' creditworthiness, collateral, and repayment capacity. Although the credit risk management team performs careful investigations, the approval process primarily relies on multiple layers of review to ensure loan quality. The bank has implemented a system for monitoring and reporting credit risk, but it heavily depends on manual assessments to detect early indicators of potential issues in the loan portfolio. It generates frequent reports and has basic early warning systems, but lacks the advanced AI-powered technologies that could enhance proactive risk management.

This reliance on manual processes could result in mistakes and errors. Currently, AIB relies on traditional procedures rather than AI-powered solutions for proactive risk mitigation, which could restrict the effectiveness of its credit risk management. While the bank's practices follow to regulatory requirements and industry standards, its reliance on manual systems highlights the need for additional research into the efficacy of its credit risk management strategies, particularly in light of Ethiopia's evolving banking environment.

2.2.1. Importance of Credit Risk Management

Credit risk management is vital in the financial industry for numerous reasons. As noted by Scott, Amajuoyi, Adeusi, Kudirat[3], credit risk management has the following benefits:

- Protecting Profitability and Solvency:- Operative credit risk practices help banks avoid NPL and minimize losses, safeguarding their profitability and long-term sustainability.
- Maintaining Financial Stability:- Robust credit risk management mitigates the risk of systemic financial crises, conserving confidence in the overall financial system.
- Compliance with Regulations:- Devotion to credit risk management requirements demonstrates an institution's responsible lending practices and helps maintain regulatory compliance.
- Improved Lending Decisions:- Effective credit risk assessment enables banks to make more informed, data-driven lending choices, optimizing their capital allocation and loan pricing.
- Enhanced Customer Relationships:- Liable credit risk management can build trust and strengthen long-term relationships with customers.

2.2.2. Challenges in Assessing Creditworthiness

The key challenge in evaluating creditworthiness stems from incomplete or asymmetric information. Borrowers might not disclose all relevant financial details, making it difficult for lenders to accurately assess their true creditworthiness. Additionally, fluctuations in macroeconomic factors, such as interest rates and market volatility, can significantly impact a borrower's ability to repay loans, requiring lenders to constantly adapt their credit risk assessment models to account for these dynamic conditions [20].

Further complicating the matter, traditional credit scoring models may have limitations in capturing the full complexity of a borrower's financial situation [20]. These models may overlook alternative data sources or fail to accurately predict default risk for certain borrower segments. Lenders must also contend with the threat of fraud and misrepresentation, necessitating robust fraud detection mechanisms. Additionally, the rapidly evolving lending landscape, with the emergence of new models like peer-to-peer platforms, presents challenges in adapting credit risk assessment methodologies. Finally, financial institutions must balance regulatory requirements with the need for innovative credit risk management approaches such that stakeholders can digitally monitor all business activities of other stakeholders in the network, so that it is more transparent and honest,, adding further complexity to the process.

2.3. Machine Learning

Machine learning (ML) focuses on developing systems and algorithms capable of learning and improving from data, without being explicitly programmed [8]. The diverse range of ML techniques allows software applications to enhance their performance and capabilities over time, adapting to new information and patterns in the data.

At the core of machine learning are algorithms that are trained on historical data to identify relationships, discover hidden patterns, and uncover insights. These models can then be used to make predictions, classify information, group data points into meaningful clusters, and reduce the dimensionality of complex datasets [21].

The flexibility of machine learning enables its application across a wide spectrum of industries and domains, from computer vision and natural language processing to predictive analytics, personalized recommendations, and autonomous systems[22]. As the field continues to evolve, the integration of machine learning with other emerging technologies is expected to unlock new possibilities and drive further innovations that transform the way we interact with and leverage data-driven solutions. The majority of organizations are either directly adopting and integrating machine learning-based solutions into their operations or embracing machine learning indirectly through the use of ML-infused products and services provided by third-party vendors and technology providers, reflecting the growing recognition of the transformative potential of this technology [8]. Nowadays machine learning discipline is extended to deep learning that can bring a breakthrough in problem-solving and decision-making.

2.3.1. Deep learning

Deep learning is a subfield of machine learning that works with artificial neural networks with multiple hidden layers to learn and make predictions from data in a highly effective manner [14]. Deep neural networks have become relevant for the more general field of reinforcement learning, where there is no supervising teacher [4]

The breakthrough in deep learning came around 2012 [15] when researchers showed that deep neural networks could achieve remarkable results in image recognition tasks, far surpassing previous methods. This success was largely due to the increased computational power of modern computers, improved training techniques, and access to large datasets [23]. Since then, deep learning has advanced rapidly, influencing many areas of machine learning. In particular, its application to reinforcement learning, where systems learn by interacting with their environment, has opened up new possibilities for creating AI systems that can learn and make decisions without direct supervision and minimal human intervention [24].

2.4. Types of machine learning

Machine learning involves feeding a large volume of data to a machine so that it can learn and make predictions, find patterns, or classify data. The three basic machine learning types are supervised, unsupervised, and reinforcement learning as shown in figure 2.1 below.

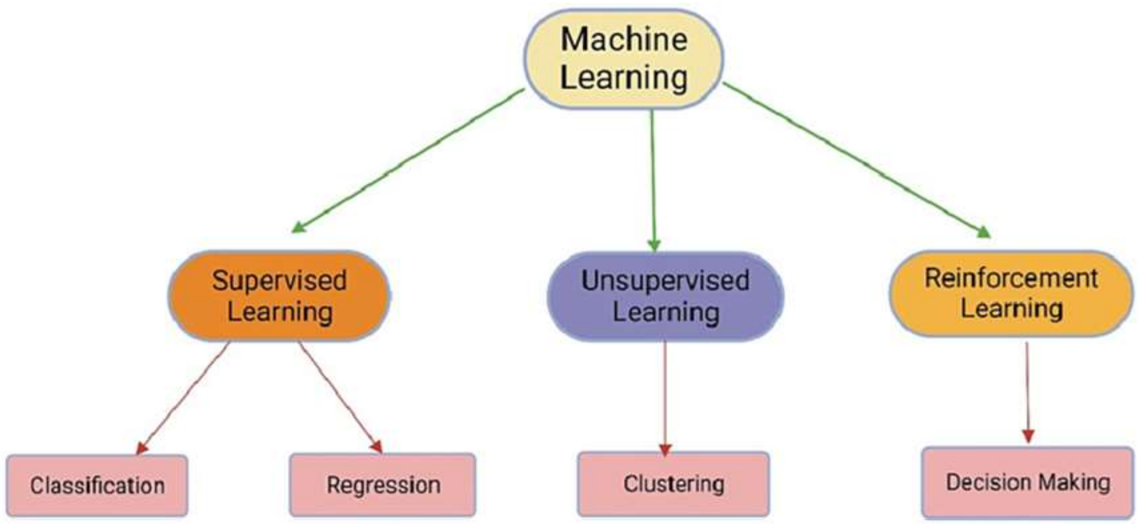


Figure 2.1. Types of machine learning approaches[25]

2.4.1. Supervised Machine Learning

Supervised machine learning is a type of machine learning where the algorithm is trained on a labeled dataset, implying the input data has corresponding output or target variables. Supervised learning aims to learn a mapping function from the input data to the output labels, allowing the model to make accurate predictions on new, unseen data.

Some common examples of supervised machine learning algorithms include linear regression for predicting continuous values, logistic regression for binary classification, decision trees for both regression and classification tasks, and support vector machines for high-dimensional classification problems. These algorithms are widely used in a variety of applications, such as spam detection, credit risk assessment, medical diagnosis, and image recognition[26].

2.4.2. Unsupervised Machine Learning

Unsupervised machine learning is a type of machine learning where the algorithm provided with unlabeled data without any corresponding output variables. The goal of unsupervised learning is to discover hidden patterns, structures, or groupings within the data without any prior information about the expected outputs. Some common examples of unsupervised machine learning algorithms include k-means clustering for grouping data points into clusters, principal component analysis (PCA) for dimensionality reduction, and anomaly detection algorithms for identifying outliers or unusual data points. These techniques are widely used in customer segmentation, image compression, recommendation systems, and fraud detection [26].

2.4.3. Reinforcement Machine Learning

Reinforcement learning is a bit different from supervised and unsupervised learning. In reinforcement learning, the model learns from the consequences of its actions. The model receives feedback on its performance and uses that information to adjust its actions and improve its performance over time[27]

This study aims to use supervised machine learning algorithms to analyze credit risk and extract patterns, which has a significant customer base in Ethiopia. The study evaluate various algorithms, including both ensemble and deep learning techniques, as well as traditional classification methods.

2.5. Supervised Machine Learning Algorithms

The current study explores Deep Neural Networks (DNNs) from the deep learning category, CatBoost, Random Forest, and XGBoost from the ensemble learning category, and Support Vector Machines (SVM) from machine learning classification algorithms. An overview of each of them is discussed as follows.

2.5.1. Deep Neural Networks (DNNs)

Deep Neural Networks (DNNs) are leading machine learning algorithms globally, especially in fields of credit risk prediction and other challenging problem-solving. DNNs are excellent at analyzing large volumes of data and seeing complex patterns that conventional algorithms frequently miss. Financial institutions, including those in Ethiopia, may learn from a variety of information, including transaction histories and client attributes, thanks to their capacity to model nonlinear relationships. This eventually results in improved risk management [28].

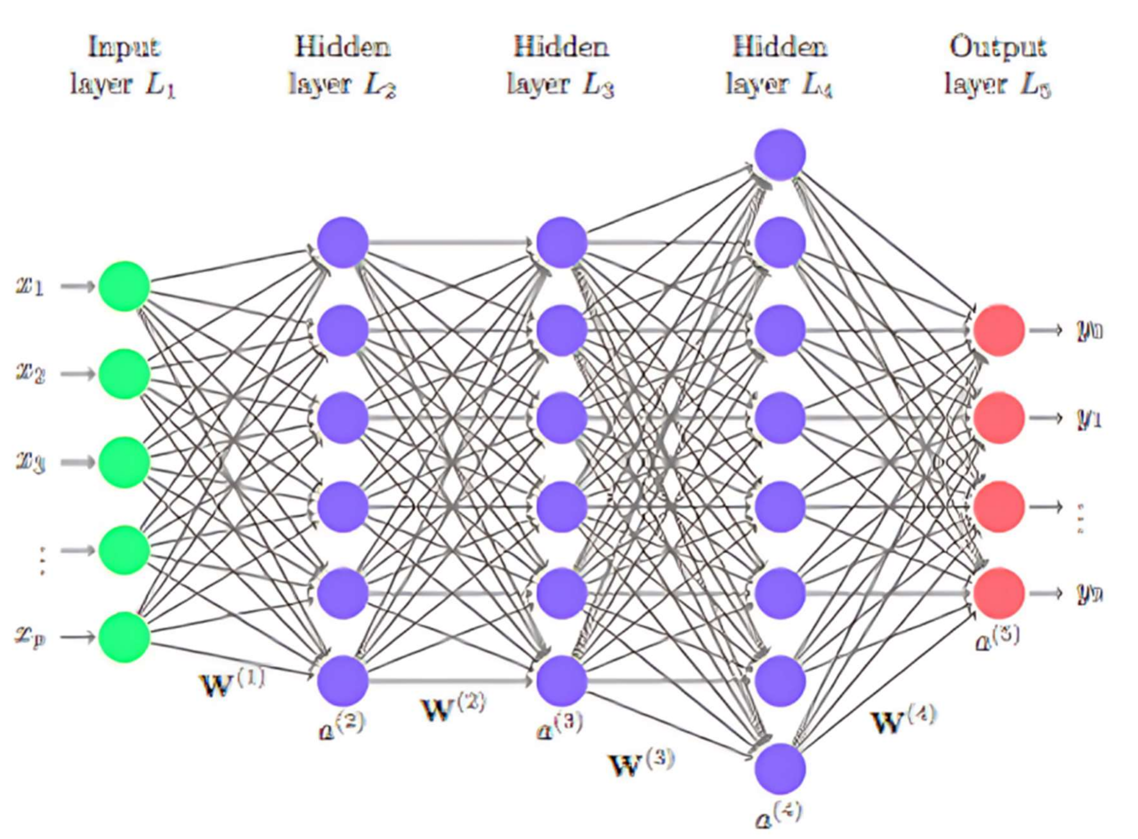


Figure 2.2: How Deep Learning Actually Works[29]

Deep learning processes and learns from massive datasets using artificial neural networks, as shown in the figure above. These networks are made up of layers of linked nodes (neurons), with each layer changing the data and extracting spatterns. As data travels across the layers, the network modifies its connections to increase forecast precision.

2.5.2. Random Forest

Random forest is another method of classification based on decision trees, a collection of tree predictions. It is a machine learning technique that is used to solve regression and classification problems. It uses ensemble learning, a technique that combines more than one classifier to solve complex problems. Random forest algorithms include many decision trees. It has more advantages than other classification algorithms for better performance on outlier data. The Random forest also provides regression functionality with training and testing datasets[30].

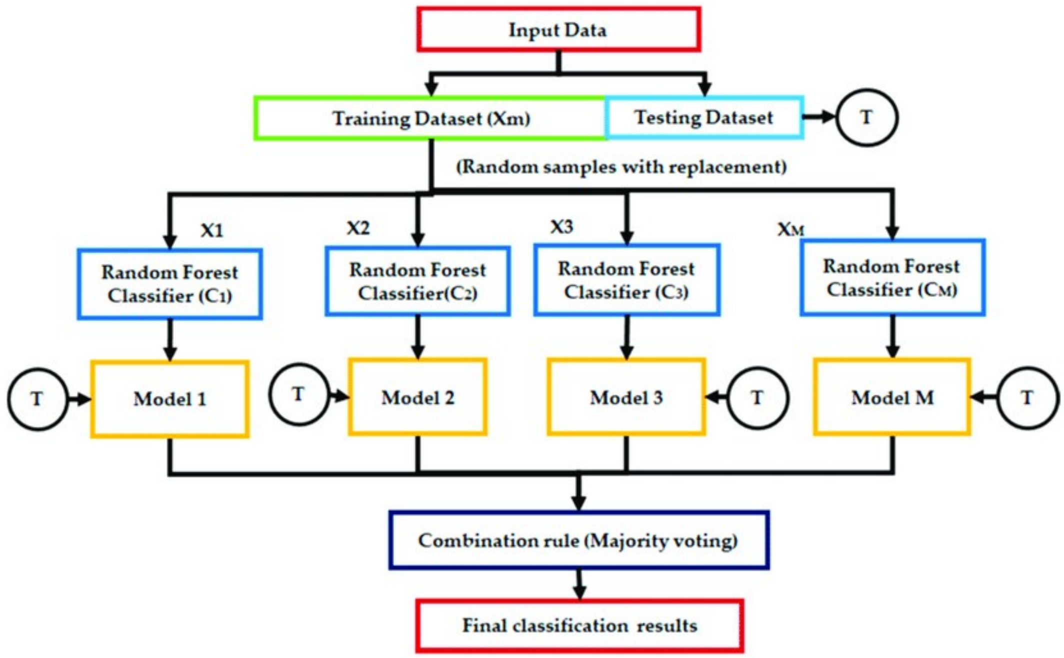


Figure 2.3: Workflow of the Random Forest Algorithm[31]

Figure 2.3 illustrates how the Random Forest method works. During training, it builds many decision trees, each containing a random subset of the data and features. Each tree provides a forecast, which the algorithm combines using a majority vote. Compared to single decision trees, this ensemble method reduces overfitting while enhancing prediction reliability and accuracy. Random Forest produces a more trustworthy and generalized model by averaging the predictions of numerous trees, making it ideal for complicated datasets with various properties.

2.5.3. Categorical Boosting (CatBoost)

One element that sets CatBoost apart is how well it handles categorical features, which are frequently found in financial datasets. Due to the algorithm's ability to function well without requiring a lot of preparation, Ethiopian banks would find it ideal as their data frequently contains categorical factors such as income source and employment. CatBoost enhances predictive performance by reducing overfitting by ordered boosting, which is a major benefit for credit risk modeling[32].

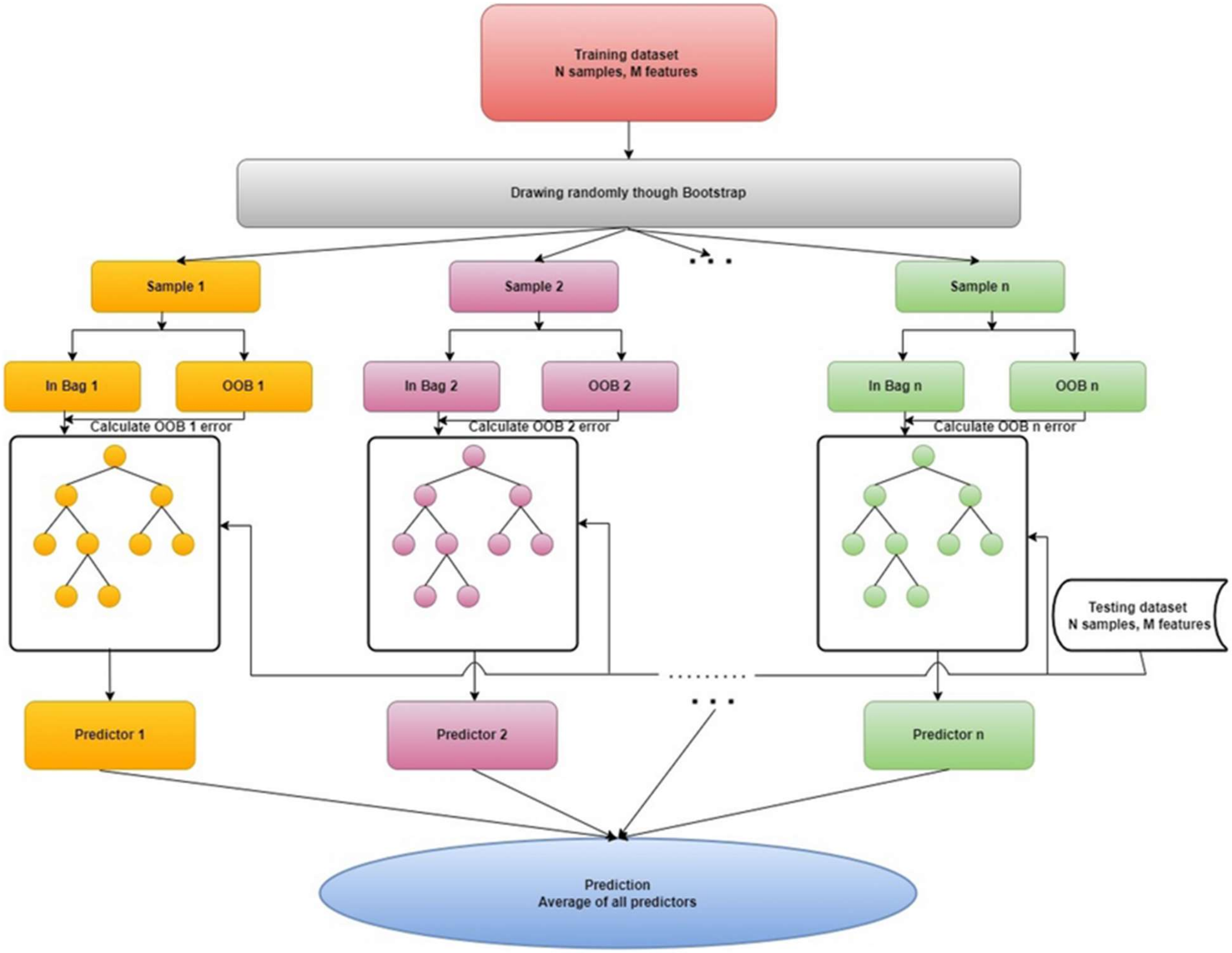


Figure 2.4: CatBoost Bootstrapping and Evaluation[33]

Figure 2.4 shows how the CatBoost algorithm works. During training, CatBoost builds a series of decision trees in order, with each tree indicating to fix the errors of the preceding ones. During the bootstrapping process, each tree is trained on a subset of the data known as the in-bag samples, while the remaining data, known as out-of-bag samples, is used to evaluate the model's performance. This method enables CatBoost to enhance predictions using gradient boosting, reducing the loss function, and adding categorical data without requiring significant preprocessing. CatBoost significantly reduces overfitting, improves model accuracy, and provides robust performance across complex data sets by combining in-bag and out-of-bag samples.

2.5.4. Extreem Gradient Boost (XGBoost)

XGBoost is well-known throughout the world for its predictive analytics speed and accuracy. Because of its ability to process massive datasets quickly, financial organizations find it to be an indispensable tool. In addition to offering accurate forecasts, XGBoost gives comprehensible insights into the key variables influencing credit risk decisions in the context of Ethiopian banks. Based on data-driven information, this capacity helps banks create more strategic lending practices [34].

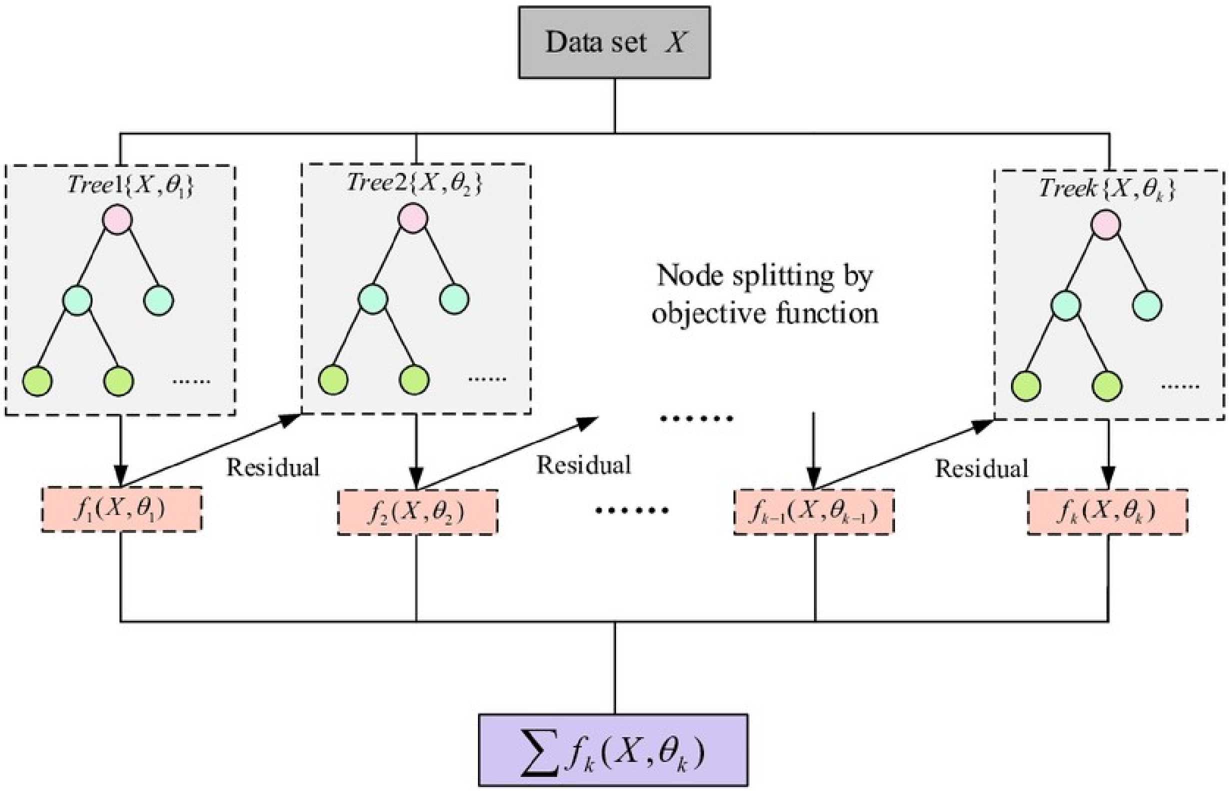


Figure 2.5: Workflow of Extreme Gradient Boosting (XGBoost)[35]

Figure 2.5 shows how Extreme Gradient Boosting (XGBoost) works. XGBoost builds an ensemble of decision trees repeatedly, beginning with Tree1(X,θ), which predicts values based on input attributes X and parameters θ. The initial tree's residuals (errors) are computed and used to train subsequent trees, indicated as ∑fk(X,θk), which correct the residuals from prior trees. This repeated procedure improves the model by reducing the loss function and improving the predictions. XGBoost focuses on residual correction, as opposed to CatBoost, which depends on gradient boosting but is designed for categorical data processing and reduces overfitting using permutations. Random Forest, on the other hand, creates many decision trees separately on random subsets of data and features, then combines their predictions by majority voting without sequential correction, depending on ensemble averaging rather than boosting.

2.5.5. Support Vector Machines (SVM)

Support Vector Machines (SVMs) are commonly utilized in international finance due to their accuracy in the categorization of jobs. Their ability to handle high-dimensional data makes them especially valuable in credit risk assessment. SVMs can precisely identify between varied risk levels for Ethiopian financial institutions using complicated customer data. The capacity to implement non-linear decision boundaries through kernel functions improves adaptability in varied financial conditions [36].

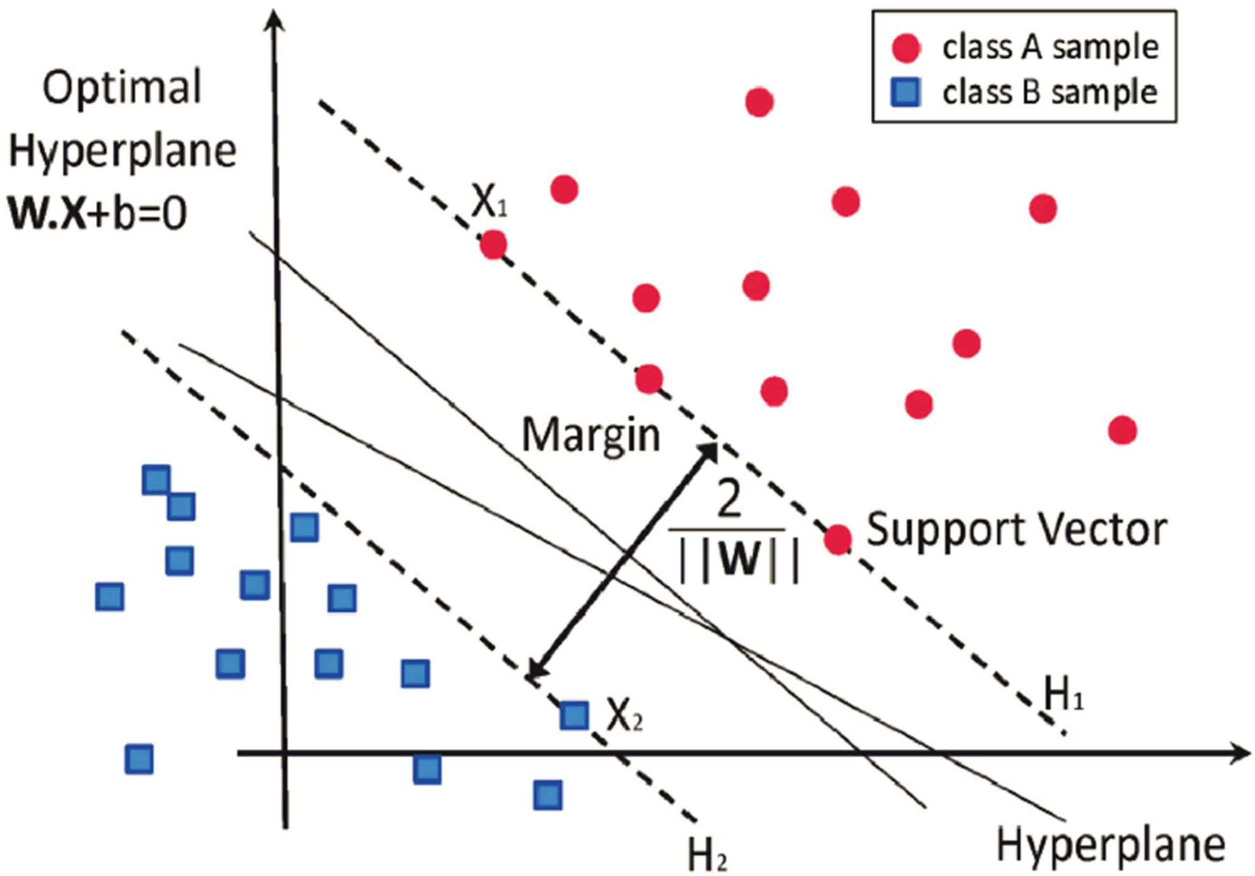


Figure 2.6: Support Vector Machine Margin Calculation[37]

SVM aims to find the optimal hyperplane defined by the equation W·X+b=0, where W is the weight vector and b is the bias term. This hyperplane separates data into different classes, with parallel lines called H_1 and H_2 representing the margins on either side of the hyperplane. The margin is calculated as $\frac{2}{||W||}$, which is the distance between the hyperplane and the support vectors, the closest data points to the hyperplane. SVM maximizes this margin to ensure the best separation between classes, and if the data is not linearly separable, kernel functions are used to map the data into a higher-dimensional space where a linear separation can be achieved.

2.6. Feature selection

Feature selection is a critical step in the machine learning process, as it helps to detect the most relevant and informative features from the available data to be used in model development. This is particularly important when dealing with high-dimensional datasets, where the presence of irrelevant or redundant features can significantly impact the model's performance and generalization capabilities [38]. Figure 2.7 below shows feature selection approaches.

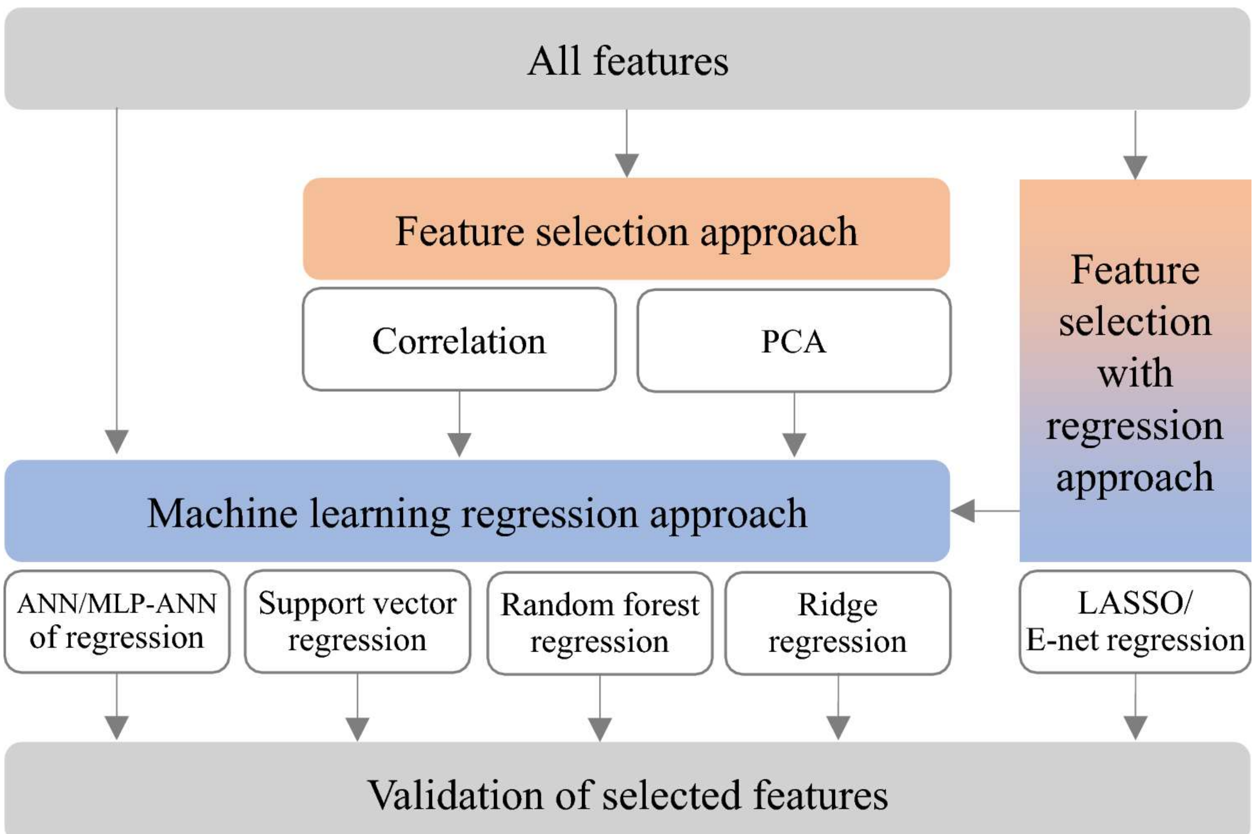


Figure 2.7: Feature Selection Methods [39]

This study uses the correlation approach to identify features that have a strong relationship with the target variable, allowing it to filter out irrelevant or less useful data. Furthermore, the Extra Trees Classifier is used to rank features according to their contribution to model accuracy, offering a reliable estimate of feature relevance. This approach combination guarantees that both statistically significant and model-relevant characteristics are picked, which improves the model's overall performance and interpretability[40].

Types of Feature Selection Methods in machine learning are Filter Methods, Wrapper Methods and Embedded Methods[41].

2.6.1. Filter methods

Filter methods, which evaluate and rank the features based on their statistical properties, such as correlation, mutual information, or chi-square test, without considering the model itself[42]. Some Examples of filter methods include Pearson correlation, ANOVA, and mutual information-gain based techniques. These methods are computationally efficient and can provide a good initial understanding of the feature importance, but they do not take into account the complex interactions between features and the target variable.

2.6.2. Wrapper methods

Wrapper methods, on the other hand, use a machine learning model to evaluate the feature subsets and select the ones that provide the best performance for that specific model[15]. This approach can capture the non-linear relationships and feature interactions, but it can be computationally expensive, especially when dealing with a large number of features. Recursive Feature Elimination (RFE) and Sequential Feature Selection are examples of wrapper methods, where the model is trained iteratively, and the least important features are progressively removed.

2.6.3. Embedded methods

Embedded methods combine the advantages of both filter and wrapper methods, where the feature selection is performed as part of the model training process[42]. Lasso regression, for instance, is an embedded method that can effectively shrink the less important feature coefficients to zero, effectively performing feature selection during the model training. Tree-based models, such as Random Forests, also provide feature importance scores that can be used for feature selection.

2.7. Application of Machine Learning

Machine learning has found numerous applications in Ethiopia, contributing to the country's development and addressing encounters[36] . Here are a few examples:

Agriculture:- Machine learning techniques have been engaged to improve agricultural productivity in Ethiopia. studies used satellite imagery and machine learning algorithms to predict crop yields, facilitating better resource allocation and planning for farmers.

Healthcare:- Machine learning models have been developed to assist in disease diagnosis and prediction in Ethiopia. to predict the risk of disease among people, facilitating early intervention and prevention strategies.

Energy:- Machine learning has been used to optimize energy systems in Ethiopia. to forecast solar energy generation, aiding in the integration of renewable energy sources into the national grid.

Transportation:- Machine learning techniques have been used to improve traffic management and transportation planning in Ethiopia. to predict traffic congestion, enabling more efficient traffic control and routing strategies.

Finance: These demonstrate the growing applications of machine learning in various sectors of the Ethiopian economy, contributing to improved decision-making, resource optimization, and problem-solving.

2.8. Challenges of Machine Learning

Machine learning is an extremely effective tool for problem-solving, streamlining corporate processes, and automating chores, but it's also a highly challenging technology that calls for a substantial investment of time and deep experience[43]. Common machine learning challenges are discussed in detail below:-

2.8.1. Data Accessibility and Quality

The availability and quality of data in Ethiopia is a major barrier that affects how successful machine learning (ML) models are. Sparse, inconsistent, and unstructured data is a problem for many financial institutions and can reduce the precision of machine learning predictions. Robust models require high-quality data to be trained; yet, inadequate or incorrect data might lead to erroneous or misleading outcomes in the models, which can be harmful when making financial decisions.

2.8.2. Infrastructure Limitations

Infrastructure constraints impede the application of ML models in Ethiopian financial institutions. Many institutions lack the high-performance computing capabilities they need, including dependable internet connectivity, powerful servers, and cloud services. The efficiency with which complex machine learning algorithms may be run and big datasets processed can be hampered by this poor infrastructure. Advanced machine learning (ML) solutions are difficult to develop and frequently unfeasible without the proper technological infrastructure.

2.8.3. Gaps in Talent and Skill

The lack of qualified personnel in Ethiopian financial institutions is a major obstacle to the effective application of machine learning. Trained data scientists and machine learning specialists with the requisite domain knowledge and programming abilities are hard to come by. Institutions find it challenging to effectively create, implement, and maintain machine learning models due to this talent imbalance. The overall performance of ML initiatives might be impacted by a lack of trained staff, which can result in dependency on outside consultants or insufficient internal capabilities.

2.8.4. Cost and Investment

Financial institutions in Ethiopia may find it difficult to use ML solutions since they need a large financial commitment. Developing infrastructure, hiring expertise, and purchasing technology may be extremely expensive, especially for organizations with tight budgets. Institutions may be discouraged from pursuing these technologies or find it difficult to fully utilize ML's potential due to the hefty implementation costs.

2.8.5. Integration

A difficult task is integrating machine learning models with current legacy systems. Many financial organizations are still using antiquated systems that can't handle the sophisticated machine-learning algorithms of today. Integrating ML with these legacy systems can be a labor-intensive and technically challenging task. Because of the potential for operational disruptions and a decrease in the overall efficacy of ML projects, this integration problem must be properly planned and managed.

2.8.6. Risks Associated with Cybersecurity

New cybersecurity vulnerabilities are introduced by the incorporation of ML. Adversarial assaults, in which nefarious actors alter input data to trick the model, can affect machine learning models. Furthermore, there is a greater chance of cyber assaults due to the growing dependence on digital data and infrastructure. To ensure that the advantages of machine learning (ML) are not outweighed by security risks, financial institutions need to invest in strong cybersecurity measures to guard against data breaches and other cyber threats.

2.9. Related Works

There are different studies conducted on credit risk analysis and prediction. Here under we present some of the works done by foreign and local scholars.

The study by Melese et al., [44] provided an in-depth exploration of the applicability of hybrid machine learning approaches in credit risk prediction. The authors present an original hybrid CNN-SVM/RF/DT model, and they provide some analysis to conclude how effective it was. The CNN model did not perform as well as the hybrid models in the performance metrics considered in the paper, so the authors might need to make some modifications to this model or benchmark it against another model. Overall, the scientific technique and data analysis in the paper will need to be improved in several ways. For example, it may improve the reproducibility of models by offering more architectural components. There should also be a fuller statistical analysis and a broader range of performance evaluation metrics used to make fewer claims about the value of comparisons among the performance of the models. Additionally, the paper did not elaborate on what specific future work should be encouraged by this work or talk thoroughly about the possible explanations for the findings as limitations or future work.

The work of Berhane et al., [7] evaluated the effectiveness of hybrid machine learning models for credit risk prediction in P2P (peer-to-peer) lending, with a focus on handling inherent data challenges. To address these challenges, the paper seeks to improve prediction models for credit risk. The research utilizes an appropriate methodology by first preprocessing imbalanced lending data using SMOTE sampling, CNN for feature extraction, and integration of the LR, GBDT, and kNN models through hybridization. The performance evaluation involved an empirical experiment conducted on an authentic P2P lending dataset using traditional metrics like accuracy, recall, and AUC. According to the results, the CNN-kNN hybrid model outperformed the other hybrid models in CNN-LR and CNN-GBDT in all evaluation metrics. This indicates that kNN, along with the features learned from CNN layers, was the most effective algorithm for categorizing credit risk. To explain the effective hybrid model, the authors provide a detailed and consistent interpretation of the results. They offer an insightful explanation of the algorithmic effectiveness of this type of problem. In summary, the paper critically evaluates the application of hybrid machine learning models to an important and relevant problem. The methodology, experiments, and statistical explanation are technically sound and offer new insights that could be used by P2P lending organizations to improve their credit risk assessment models. The paper meets high standards of scientific rigor and scientific scholarship.

The Alagöz & Çanakoğlu, [45] conducted research on improving credit risk prediction models for commercial banks, focused on overcoming obstacles associated with unbalanced datasets. Traditional statistical approaches sometimes fall inadequate in cases when one class dominates the data. Lukas' research creates a credit default prediction model by combining several credit-related datasets and using Min-Max normalization to normalize feature values. To overcome data imbalance, the study uses both undersampling and oversampling strategies and assesses their effects on model performance. The Gradient Boosted Decision Tree approach, paired with K-means SMOTE oversampling, resulted in considerable accuracy gains. Accuracy improves from 66.9% to 89% for Taiwan, 70.7% to 84.6% for South Germany, and 65% to 87.1% for Belgian datasets.A one-way ANOVA validates the statistical significance of these results ($p < 0.001$). Lukas' study demonstrates the efficacy of balanced datasets and advanced resampling approaches by offering an interpretable model that stakeholders can access online to better manage credit risk.

The MarteyAddo[46] presented an insightful comparative study of machine learning and deep learning models for credit risk analysis. The authors build binary classifiers using real-world data and examine the stability of the models by testing them on separate datasets. Their outcomes indicate that tree-based models exhibit greater stability compared to multilayer neural networks, raising valid questions about the intensive deployment of deep learning systems in enterprise settings, where model reliability and interpretability are crucial. The methods taken by the researchers are rigorous, leveraging well-established techniques from the machine learning and credit risk domains. The selection and testing of the top ten important features provides valuable insights into the key drivers of credit risk. While the study is focused on the financial sector, the broader implications around algorithm transparency and the ethical use of predictive models are extremely relevant across many industries. Overall, this work contributes meaningfully to the ongoing discourse around responsible development and deployment of progressive analytics capabilities.

Shan & Nilsson, [47] studied the efficacy of several classification algorithms for credit risk assessments using a large peer-to-peer lending dataset. The study looked at models including Support Vector Machine (SVM), Decision Trees, Logistic Regression, Multilayer Perceptron Neural Network (MLP), Probabilistic Neural Network (PNN), and Deep Learning. Sultan discovered that SVM offered the greatest classification accuracy (97%), exceeding other models. Decision Trees followed with an accuracy of roughly 95.6%, demonstrating good performance similar to earlier studies. Logistic Regression and MLP performed similarly, with accuracies of 93.7% and 93.6%, respectively, adding to the dispute about whether MLP has benefits over Logistic Regression. PNN had a lower accuracy of roughly 92.9%, which contradicted prior studies indicating greater performance in smaller datasets.in comparison to previous studies, Deep Learning models performed poorly, most likely due to the fundamental technique utilized in this investigation. The study also identified difficulties with data imbalance, with the majority of non-default situations and a high rate of false negatives, creating dangers to stakeholders. Sultan's findings highlight the need for more research into advanced deep learning algorithms and tactics for addressing data imbalance and improving classification accuracy.

Bizuwork et al., [10] presented research on using machine learning algorithms to predict loan risk in the context of microfinance institutions (MFIs) in Ethiopia. The author applies several machine learning techniques, including K-Nearest Neighbors (KNN), Logistic Regression, Naive Bayes, and Support Vector Machines (SVM), to a real-world dataset from various Ethiopian MFIs. The aim is to identify the most effective algorithm for accurately predicting loan default risk, which is a serious problem for the financial sustainability of MFIs. The author compares the performance of the different algorithms using metrics such as accuracy, precision, recall, and F1-score. The results show that the SVM model outperforms the other algorithms, accomplishing the highest overall predictive performance. The paper provides valuable insights into the application of machine learning for loan risk assessment in the microfinance sector, which can help MFIs make more informed lending decisions and improve their financial stability. However, the paper does not explore more advanced machine learning techniques, such as ensemble methods or hyperparameter tuning, which could potentially improve the predictive accuracy of the models. It focuses primarily on the technical aspects of the modeling process and does not provide much discussion on the practical implications of the findings or recommendations for microfinance institutions to implement the proposed approach, limiting the real-world applicability of the study.

Chhetri et al., [48] aimed to conduct a comparative analysis of ensemble machine learning algorithms for credit risk prediction in the financial sector. The researchers explore the use of algorithms such as Logistic Regression, Random Forest, Naive Bayes, AdaBoost, and XGBoost to build robust credit scoring models. They aim to evaluate the performance of these ensemble techniques in terms of accuracy, precision, recall, and F1-score, identify the most significant parameters influencing bank loan repayment, and develop an optimal hybrid machine learning model for credit risk prediction. The study utilizes a real-world dataset of 13,600 credit applications from a cooperative financial institution, with 70% of the data used for training the models and 30% reserved for testing. The performance of the individual machine learning algorithms is evaluated and compared, and feature importance analysis is conducted to determine the key factors affecting loan default. The researchers then propose a hybrid model by combining the strengths of multiple algorithms to achieve superior predictive capabilities. The findings indicate that the ensemble algorithms, particularly Random Forest and XGBoost, outperformed traditional techniques like Logistic Regression, and factors such as occupation, income ratio, credit history, and savings account balance were identified as the most influential in predicting loan defaults. The study provides valuable insights for financial institutions to enhance their credit risk assessment and decision-making processes, ultimately leading to better risk management and increased profitability.

Recently, Lev & Linda, [43] thoroughly examined the strengths and weaknesses of applying machine learning (ML) techniques in credit risk prediction. The strengths include ML's ability to process extensive data volumes and enhance prediction accuracy, particularly through models like XGBoost and LightGBM. Additionally, the use of evaluation metrics such as AUC and precision is well-supported, and the importance of feature selection is emphasized. However, the paper also highlights significant weaknesses, such as the "black box" nature of many ML models, which limits their interpretability, and the challenge of dataset representativeness, with many studies relying on public datasets that may not reflect real-world conditions. The review calls for more robust models capable of handling imbalanced data and improving transparency through Explainable Artificial Intelligence (XAI) techniques, suggesting a need for further research to address these limitations.

Emmanuel et al., [49] presented a robust approach to credit risk prediction by leveraging a stacked classifier model and a filter-based feature selection (FS) method. The study's strength lies in its innovative use of a multi-level ensemble method, combining Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) to outperform traditional models in accuracy, F1-Score, and Area Under the Curve (AUC) metrics. Additionally, the use of a filter-based FS technique based on Information Gain ensures that the most relevant features are selected, enhancing the model's performance. However, the paper's reliance on accuracy as the primary performance metric, while incorporating F1-Score and AUC, might not fully capture the model's effectiveness in handling the class imbalance inherent in credit risk datasets. Despite these concerns, the proposed stacked model's sequential architecture demonstrates significant improvements over individual estimators and other existing methods, marking it as a valuable contribution to credit risk prediction.

The study conducted on machine learning-based credit risk assessment for predicting loan defaulters in the Ethiopian banking industry, authored by TsegaAsresaMengistu et al. [11], presents a robust analysis of various classification algorithms to enhance risk management in banking. The strengths of the study lie in its comprehensive approach, employing multiple machine learning techniques such as Random Forest, Decision Tree, Gradient Boosting (GB), XGBoost, and Multi-Layer Perceptron (MLP) to classify borrowers effectively. The results indicate that the XGBoost algorithm outperformed others, achieving a training accuracy of 97.8% and testing accuracy of 98.0%, showcasing its effectiveness in distinguishing between defaulters and non-defaulters. However, the study has some weaknesses, including the reliance on a single dataset, which may limit the generalizability of the findings, and the lack of detailed exploration of the interpretability of the models, which is crucial in the banking context for regulatory compliance and trust. Additionally, while the study mentions preprocessing and feature selection processes, it could benefit from a more detailed explanation of these methodologies. Overall, the research highlights the potential of machine learning in improving credit risk assessment but calls for further validation across diverse datasets and deeper insights into model interpretability.

The study conducted on Artificial Intelligence-Enhanced Credit Risk Assessment at the Commercial Bank of Ethiopia, Ejeta[12] aims to develop a machine learning model to predict loan approval status using a dataset of 32,285 applicants. The research employs various supervised machine learning techniques, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Deep Neural Network, to evaluate performance based on accuracy and key metrics such as recall, precision, and F1-Score. The strengths of the study include a comprehensive dataset and the effective application of advanced machine learning methods, with the Random Forest classifier achieving the highest accuracy of 93.62%, indicating its robustness in capturing the complexities of credit risk. Additionally, the use of techniques like SMOTE for dataset balancing and correlation analysis for feature selection enhances the model's effectiveness. However, the study has limitations, such as a potential lack of diversity in the dataset and the need for further exploration of additional variables that could improve predictive capabilities. Overall, the findings underscore the importance of leveraging AI in financial institutions to enhance decision-making processes and reduce default rates, while also highlighting opportunities for future research to refine and expand upon the credit risk assessment models developed.

2.9.1. Summary of related works

Hereunder table 2.1 presents summary of related works done to enhance their credit risk assessment and analysis.

Author	Problem explored	Approach	Results	Limitations / Gaps
Melese, et al. (2023) [44]	Applicability of hybrid ML models in credit risk prediction.	CNN-SVM/RF/ DT hybrid model	98%	Need for better statistical analysis and model benchmarking.
Berhane et al. (2024) [7]	Evaluated hybrid ML models for P2P lending	CNN-LR, CNN-GBDT, CNN-kNN	91.87%	Limited implications discussion
Alagöz&Çan akoğlu (2021) [45]	Credit risk prediction for commercial banks, focusing on unbalanced datasets.	Gradient Boosted Decision Tree with K-means SMOTE	Accuracy improved from 66.9% to 89% for Taiwan dataset	More exploration needed on resampling strategies, limitations in real-world application due to specific dataset use.
Shan & Nilsson (2018) [47]	Efficacy of classification algorithms for credit risk assessments.	SVM, Decision Trees, Logistic Regression, MLP, PNN, DL	SVM achieved highest accuracy of 90%,	Issues with data imbalance and need for advanced DL techniques for better performance.
Bizuwork et al. (2019) [10]	Loan risk prediction for microfinance institutions (MFIs) in Ethiopia.	KNN, Logistic Regression, Naive Bayes, SVM	KNN outperformed others with accuracy of 99.91%	Lacks exploration of advanced techniques like ensemble methods or hyperparameter tuning
Chhetri, et al. (2023) [48]	Comparative analysis of ensemble ML algorithms for credit risk prediction.	Logistic Regression, Random Forest, AdaBoost, XGBoost	XGBoostoutpe rfomedothers with accuracy of 93.7%	Lacks deeper exploration of hybrid model structures
Emmanuel et al. (2024)	Robust credit risk prediction using	Random Forest, Gradient	94.4% F1-Score on the	Reliance on accuracy as a primary metric may not fully

[49]	stacked classifier models.	Boosting, XGBoost	German dataset, and AUC with stacked models	address class imbalance, lack of exploration into handling complex real-world datasets.
TsegaAsresa Mengistu(2023) [11]	Predicting loan defaulters	Applied ML algorithms (Random Forest, Decision Tree, etc.)	Testing accuracy: 97.8%	Limited context applicability and potential biases.
WalfanaMagarsaEjeta(2024) [12]	Identifying credit risk and predicting loan status	Developed a ML model using classifiers (LR, DT, RF, SVM, DNN) on a dataset of 32,285 applicants with 10 attributes.	RF registers best accuracy of 93.62%	Limited exploration of external factors affecting credit risk.

Table 2.1 Summary of selected related works

This study aimed to address several gaps in the literature on credit risk prediction using machine learning. To fill the gaps highlighted in the related works summary table, this research implemented Deep Neural Networks (DNN), Random Forest, Categorical Boosting, XGBoost, and Support Vector Machines (SVM). These proposed algorithms offered advantages over previous techniques. Specifically:

- The deep neural network tuned parameters for optimized model performance.
- Missing data addressed using imputation techniques.
- SMOTE employed as an oversampling technique to focus on data balancing.
- An Extra Trees Classifierused to assess feature importance, enhancing the model’s interpretability and accuracy.

CHAPTER THREE

PROPOSED ARCHITECTURE AND METHODS

3.1. Overview

This chapter outlines the key materials and methods used to develop the proposed risk prediction system. The study employed an experimental research approach to systematically test and evaluate various models and methods. The literature review examined diverse studies to establish a strong foundation for the system design by providing a deep understanding of the problem and potential solutions. The data collection strategies are detailed, including the sources, formats, and preprocessing techniques used to gather information for the model. The design phase is explored, highlighting the specific techniques and tools leveraged to experiment the proposed architecture.

3.2. Proposed architecture

This study employed an experimental research design, leveraging multiple algorithms.. First, I undertook an extensive literature review to understand the problem and formulate research questions, followed by data collection. A solution is then designed and proposed to address the credit risk problem, which is subsequently evaluated by implementing the solution and analyzing the results. The high-level architecture of the proposed design, which includes data preprocessing steps such as handling missing values, data transformation, feature selection, and data balancing, as well as various machine learning models like Deep Neural Network, CatBoost, XGBoost, Random Forest, and Support Vector Machine, is depicted in Figure 3.1 below.

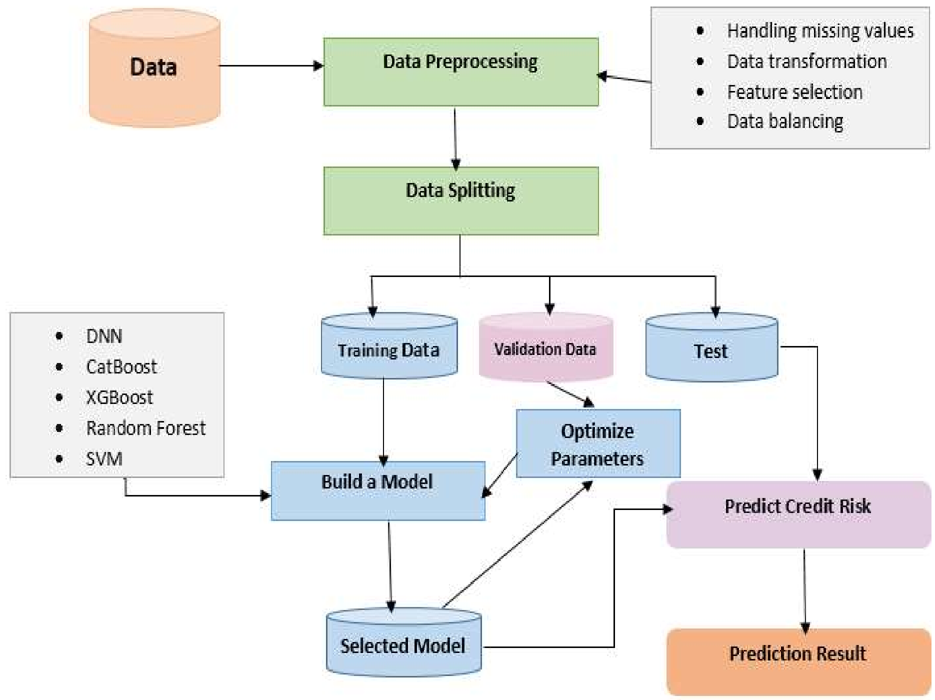


Figure 3.1 The proposed architecture for predicting credit risk

3.3. Data collection and Preparation

Data preparation involves collecting data from the identified source, cleaning it to address errors and missing values, and transforming it into a suitable format. This process also includes integrating data from multiple sources, reducing the dataset by selecting essential features, and engineering features to enhance model performance. These steps ensure that the data is accurate and ready for analysis.

3.3.1. Data collection

Awash International Bank (AIB) is one of the largest customer base and oldest private commercial banks in Ethiopia, established in 1995 GC [8]. It's headquartered in Addis Ababa and has a wide network of over 400 branches nationwide. Awash Bank held around 12% of the total banking sector assets, 13% of total deposits, and 14% of total loans [6]. In terms of its financial performance, as of June 30, 2021, AIB ensured a total asset base of Birr 143.6 billion (around $3 billion) and a total loan and advances portfolio of Birr 71.2 billion (over $1.5 billion)[8]. AIB has faced challenges with non-performing loans. In 2021, the bank's NPL ratio stood at around 3.2%, which is relatively low compared to the industry average in Ethiopia, but this issue is still a problem for their economy [6].

The data for this study is collected from Awash Bank, a leading financial institution in Ethiopia. Awash Bank is one of Ethiopia's Largest private or non-government banks, with a vast customer base and comprehensive financial data set, making it an ideal source for this credit risk prediction study. The dataset includes 22 features or attributes related to customer credit risk, such as demographic information (age, gender, marital status, etc.), financial history (income, assets, liabilities, etc.), transaction data (payment patterns, account activity, etc.), and credit performance indicators (payment history, defaults, credit utilization, etc.). The data covers more than 13 thousand customers who joined the bank between 2014 and 2024 to access credit services. In the table 3.1 belowfeatures collected for credit risk prediction are listed with description.

Table 3.1 : - Credit Risk Features and their description

Feature	Data type	Description
Segment	Nominal	Borrower category (e.g., individual, business)
Branch	Nominal	Branch of the lending institution
Region	Nominal	Geographic area of the borrower
Loan Type	Nominal	Type of loan (e.g., personal, auto)
Product Description	Nominal	Details of the loan product
Amount Granted	Numeric	Total loan amount given
Date Granted	Date	Loan issuance date
Due Date	Date	Date by which the loan should be repaid
Grace Period	Numeric	Period after due date before payment is overdue
Mode of Repayment	Nominal	Method of payment (e.g., monthly)
Repayment Amount	Numeric	Amount to be paid periodically
Expected Collection to Date	Date	Total expected collection by a specific date
Amount Collected to Date	Numeric	Actual amount collected by a specific date
Arrears Amount	Numeric	Amount overdue
Principal	Numeric	Original loan amount
Interest	Numeric	Cost of borrowing, excluding principal
O/S Balance	Numeric	Remaining amount to be paid (principal + interest)

Interest Rate	Numeric	Percentage charged on the principal
Next Repayment Date	Date	Date of the next payment due
Overdue Date	Date	Date when payment is considered overdue
Days Overdue	Date	Number of days payment is overdue
Status	Nominal	Risk level of the loan: low, medium, high

However, not all characteristics are important to this study's goal of predicting credit risk; hence, a feature selection method is required to refine the dataset. The variable "Region" was specifically removed as it has the same value across all records. Following this reduction, the remaining 21 variables passes through additional investigation via data preprocessing.

3.3.2. Data preprocessing

Data preprocessing is crucial for building an effective credit risk prediction model. This includes handling missing values, addressing outliers, and transforming the data. Missing data imputation and outlier treatment help ensure the dataset is complete and accurate. Data transformation, such as normalization and standardization, creates a more uniform distribution and reduces variability. Feature engineering is then used to identify independent variables that have a strong relationship with the target credit risk variable, which is essential for developing a high-performing model. By meticulously preparing the data, machine learning algorithms can operate on high-quality inputs, leading to more accurate and interpretable results for credit risk prediction.

Several preprocessing procedures are performed to create the final dataset for the credit risk prediction model, which are implemented using data from Awash Bank. This Data Preprocessing step included several key tasks, such as cleaning missing values, encoding non-numeric data, normalizing (scaling) the data, feature selection, sampling, and finally splitting the dataset for training and testing purposes. One variable in the Awash Bank customer dataset, entitled "Region," contained a single unique value: "West Addis Ababa Region." Because this column did not give any important variability or information for analysis, it is deemed useless for modeling purposes. As a result, the column is droped out from the dataset with the Python drop command, ensuring that the dataset contained only important properties for effective credit risk. On the other hand, the product description column have many unique values. However, these values are often divided into two categories: standard loans and premium loans. Standard loans are defined as any entries in the product description that includ terms such as personal, housing, education, health, staff, transportation, building, construction, and murabaha. All other entries are categorized as premium loans. This segmentation not only simplifies the analysis, but also allows for more targeted risk management methods based on the kind of loan. The following data preprocessing techniques are employed to prepare the data set in this credit risk prediction study:

3.3.2.1. Handling Missing Values

Several factors, including incomplete client records or mistakes in data gathering, could lead to missing values in the dataset. Using the proper methods, the missing values are imputed to resolve this problem. To maintain the general distribution of the data, missing values for numerical features are imputed using the mean of the corresponding feature. Since this preserves the natural links between the categories, missing values for categorical features are imputed using the corresponding feature's mode.

The majority of attributes in the collected data are complete and do not have missing values, however, as shown in table 3.2 there are attributes containing missing values. So resolving these missing values is critical to enhance the model accuracy. Since all attributes are with missing values less than 50%, it is possible to fill them with approximate new values.

Table 3.2. Attributes with missing values and imputation methods used

Attribute	Missing values (in %)	Data type	Imputation method
Amount Granted	3.24%	Numeric	Mean
Date Granted	5.40%	Date	Mode
Due Date	3.24%	Nominal	Mode
Grace Period	3.73%	Numeric	Mean
Mode of Repayment	3.76%	Nominal	Mode
Repayment Amount	4.06%	Numeric	Mean
Expected Collection to Date	9.22%	Numeric	Mean
Amount Collected to Date	9.22%	Numeric	Mean
Arrears Amount	9.23%	Numeric	Mean
Interest Rate	5.47%	Numeric	Mean
Next Repayment Date	15.07%	Date	Mode

3.3.2.2. Data Transformation

To ensure a common scale between the features and to mitigate the impact of outliers, the dataset undergoes Min-Max normalization. This technique linearly scales the features to a common range, typically between 0 and 1, without affecting the underlying data distribution. By normalizing the features, the model is able to learn the patterns more effectively, as it is not be biased by the differences in magnitude or scale between the features[50].

In this study, numerical variables such as Amount Granted, Repayment Amount, Expected Collection to Date, Amount Collected to Date, Arrears Amount, Principal, Interest, O/S Balance, and Interest Rate are normalized or standardized. Also, as shown in table 3.3, Date-related variables, such as Date Granted, Due Date, Next Repayment Date, and Overdue Date, are converted to datetime format and essential properties extracted. By undergoing these transformations, the dataset is more suitable for analysis and has a better predictive ability for evaluating credit risk.

Table 3.3. List of attributes transformed into new values

Attribute	Original value	New value
Date Granted	2022-07-21	2022-07-21
Due Date	2031-04-05	2031-04-05
Next Repayment Date	26/07/2024	26/07/2024
Overdue Date	26/07/2024	26/07/2024

3.3.2.3. Data Encoding

Categorical variables are prevalent in bank data and require specific handling in machine learning. Certain algorithms, like CatBoost, have built-in functions for managing categorical data without the need for manual encoding. In contrast, other methods necessitate transforming categorical data into numeric values using techniques such as Label Encoding. In this dataset, there are six attributes with categorical values, including the 'Branch' variable, which featured a significant number of unique values. To address this high cardinality, the Hasher function is applied, while the remaining categorical variables are converted into numerical representations using scikit-learn'sLabelEncoder. Table 3.4 below presents the attributes that are encoded from categorical to numeric values.

Encoding is essential for converting categorical variables into numerical formats suitable for analysis, as most machine learning algorithms rely on numerical input for mathematical computations. In this study, categorical variables are encoded using both Label Encoding and the Hasher function. Label Encoding assigns distinct numeric values to each category, while the Hasher function efficiently handles high-cardinality variables by hashing them into a fixed number of features. For the 'Branch' variable, which have 124 unique values, a n_features value of 10 was selected based on the square root rule for unique values. This approach added ten new columns for the 'Branch' variable without significantly increasing the number of columns for other variables. The LabelEncoder is used for five categorical variables ('Segment,' 'Loan Type,' 'Product Description,' 'Mode of Repayment,' and 'Status'), while the Hasher function is applied exclusively to the 'Branch' variable, facilitating effective modeling and analysis.

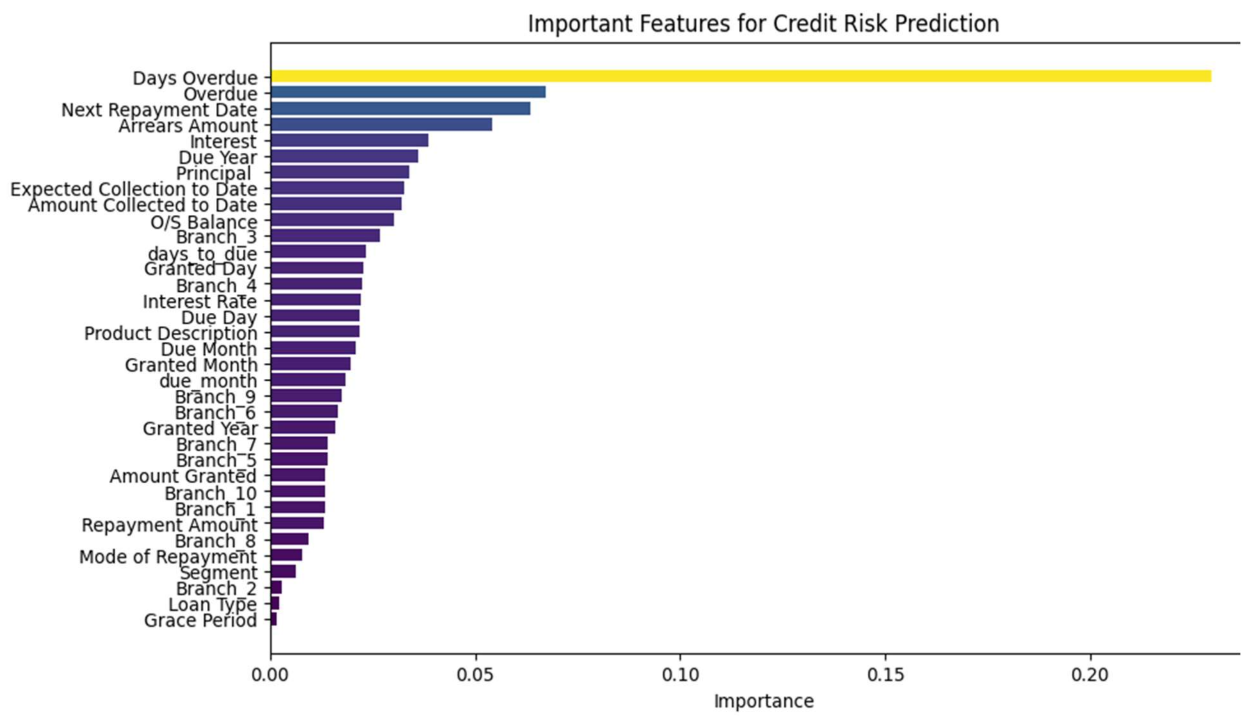
Table 3.4. Attributes encoded to numeric values

Attributes	Categorical values	New encoded values
Branch	DejazmachBalchaSafo, Kolfe Branch,Jemo Branch,Merkato Branch, Addis Ketema ….	hashed_0, hashed_1, hashed_2, hashed_3, ….
Segment	Business, Corporate,GOV, IFB Segment	0,1,2,3…
Loan Type	Advanced Against Export Bills, Advanced Against Murabaha, Overdraft loan, Qard Financing…	0,1,2,3….
Product Description	Premium Loan, Standard Loan	0,1
Mode of Repayment	Anually, Annually, Monthly, Quarterly	0,1,2,3
Status	Low, medium, High	0,1,2

3.3.2.4. Feature Selection

The dataset may contain numerous features, and not all of them are equally important for predicting credit risk. To identify the most relevant features, the Extra Trees Classifier algorithm ranks the features based on their importance, Extra Trees Classifier provides a reliable feature importance ranking. After the ranking, the top features are selected for further model development, ensuring that subsequent machine learning models focus on the most informative and predictive elements, thereby enhancing overall model performance and interpretability.

Additionally, correlation analysis further improves model performance in credit risk prediction. In high-dimensional financial datasets, correlation analysis is crucial for filtering out unnecessary features by identifying variables that strongly correlate with loan default risk. Once this initial filtering is done, the Extra Trees Classifier evaluates feature importance while considering non-linear correlations and interactions. This technique enhances both interpretability and prediction accuracy, leading to well-informed lending decisions in credit risk assessment. By systematically reviewing and ranking features based on their contribution to predictive power, this method not only improves model performance but also provides valuable insights into the factors influencing credit risk. A graph of feature importance is shown below in figure 4.8:



The feature importance graph shows that "Days Overdue" is the most important predictor of the target variable, emphasizing its critical role in assessing credit risk. Following that, " Overdue" emerges as a significant component, indicating a link between higher rates and increased default risk. Furthermore, "Next Repayment Date" and "Arrears Amount" are notable elements, showing their importance in determining repayment behavior. In contrast, the 'Grace Period' is considered less significant, suggesting that a history of timely payments may not be a strong indicator of future credit risk. Overall, these findings highlight the need of prioritizing the most important elements in credit risk evaluations.

3.3.2.5. Data Balancing

Credit risk datasets often exhibit a class imbalance, where the majority of instances belong to the "low-risk" class, and the minority instances belong to the "high-risk" class. This imbalance can lead to biased model predictions, where the model tends to perform better on the majority class but struggles with the minority class.

In machine learning-based credit risk prediction, SMOTE (Synthetic Minority Over-sampling Technique) successfully tackles class imbalance by creating synthetic samples for the minority class, which is often loan defaults. The procedure begins by finding each minority instance's nearest neighbors, such as borrowers who default on loans. SMOTE then generates new synthetic samples by interpolating between these instances and their neighbors and using a random scaling factor to generate new data points[51]. This procedure is continued until the dataset has a more equal representation of defaulting and non-defaulting debtors. By expanding the minority class, SMOTE improves the model's capacity to learn from varied experiences, eventually enhancing the accuracy and dependability of credit risk forecasts.

Table 3.5:- Target Variable Counts

Status		
Low	Medium	High
7856	1025	371

Figure 3.2 below presents the class distribution of instances before applying SMOTE oversampling.

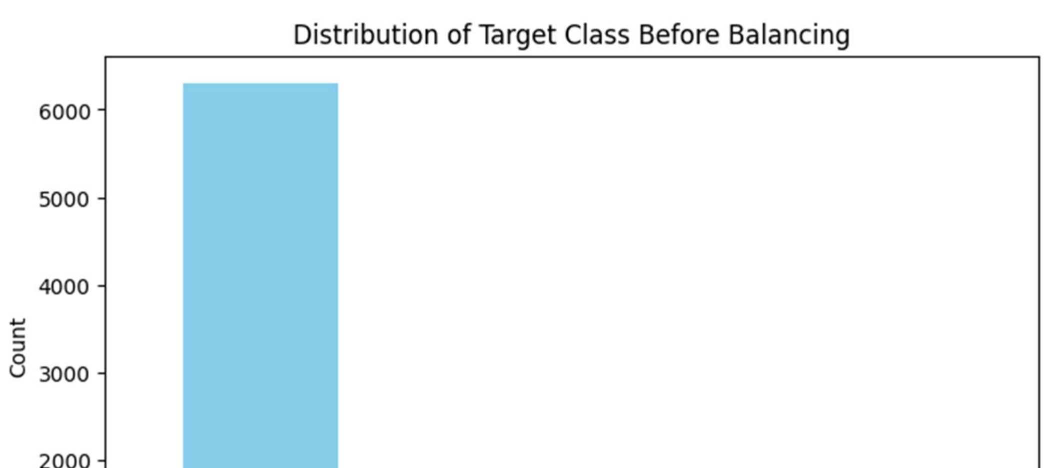


Figure 3.3: Class Distribution Before SMOTE

Figure 3.3 below shows the improved class distribution after applying SMOTE oversampling, highlighting a more balanced representation of both low-risk, medium risk and high-risk classes.

Figure 3.4: Class Distribution After SMOTE

The Synthetic Minority Over-Sampling Technique (SMOTE) is commonly used to address class imbalance in datasets. As noted in prior research, SMOTE generates synthetic samples by interpolating between existing minority class instances[52]. This involves selecting a minority instance and its k-nearest neighbors, creating new instances along the line segments connecting them. By enriching the dataset with these synthetic examples, SMOTE improves class representation, enhancing the performance of machine learning models.

3.4. Machine Learning Models

The preprocessed dataset is used to train and evaluate various supervised machine learning models for credit risk prediction. Algorithms such as Random Forest, XGBoost, CatBoost, support vector machines, and deep neural networks are comparatively assessed. These modeling techniques are commonly applied in the financial sector for predictive analytics tasks like credit risk assessment and loan default forecasting. By benchmarking the performance of these diverse algorithms on the credit risk data, the study aims to identify the optimal model that can deliver the most accurate and reliable predictions. The comparative evaluation provides insights into the strengths and weaknesses of each modeling approach in the context of credit risk prediction, allowing the selection of the best-performing model for further refinement and deployment.

3.4.1. Random Forest Classfier

Random Forest Classifiers are suitable for credit risk prediction because of the ability to handle large datasets with various features and their resistance to overfitting. In the context of credit risk assessment, a Random Forest model is used to look into the complex correlations between borrower attributes and loan outcomes, which are essential for effectively predicting the likelihood of default. During training, the Random Forest algorithm generates a large number of decision trees (see figure 2.3 in section 2.5.2). Each tree is built using a random portion of the training data and a random selection of features, a technique known as "bagging" (bootstrap aggregation). This randomness serves two purposes: it decreases the possibility of overfitting by allowing individual trees to collect noise unique to their data subset, and it improves the model's capacity to generalize to new, previously unknown data.

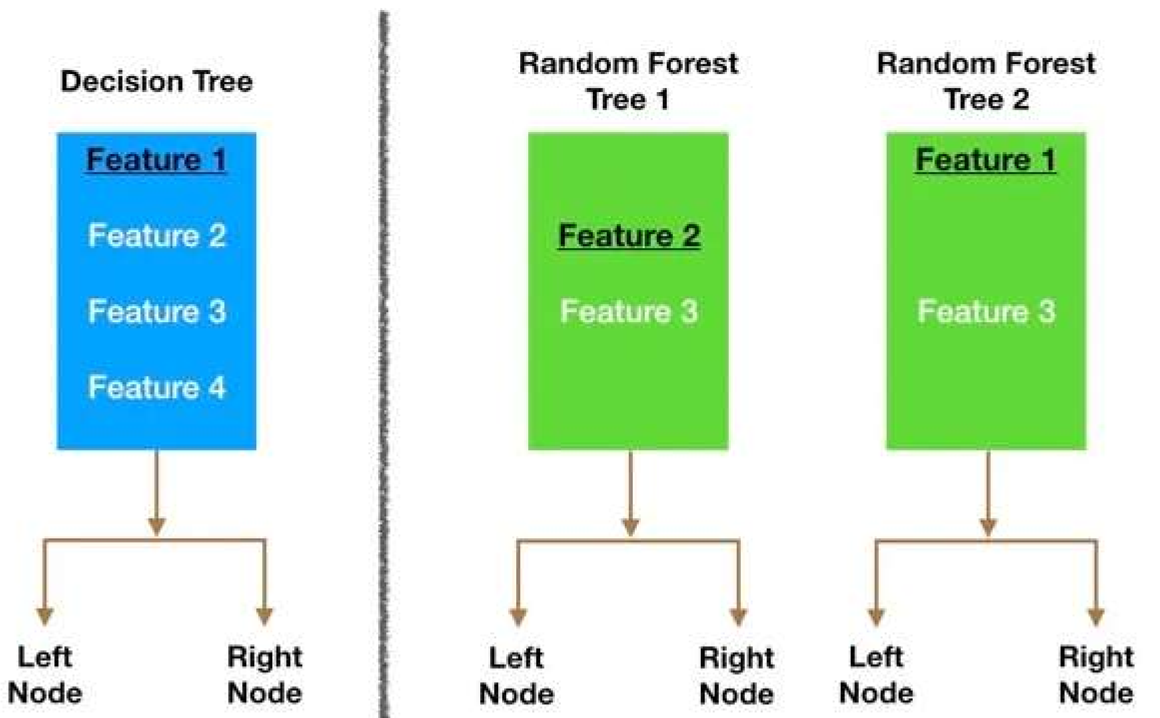


Figure 3.5: Illustration of the Random Forest Algorithm[53]

Once trained, the Random Forest model predicts credit risk by combining forecasts from all individual trees, which classify borrowers as "low risk," "medium risk," or "high risk." The ultimate predicted is decided by a majority vote across all trees. This classifier efficiently examines a variety of criteria, including credit history, loan amount, repayment behavior, and income level, allowing it to find trends and make predictions. Furthermore, the Random Forest algorithm gives insights into feature relevance, allowing stakeholders to determine which aspects have an important effect on credit risk. This openness enables financial institutions to make more informed lending decisions and improve risk management measures, eventually boosting overall lending performance and decreasing the possibility of defaults.

3.4.2. XGBoost

XGBoost (Extreme Gradient Boosting) is an advanced algorithm that has grown in popularity for its ability to predict credit risk due to its effectiveness and versatility in dealing with complex data sets. This approach uses a boosting strategy in which a number of weak learners, often decision trees, are trained successively (see figure 2.5 in section 2.5.4). Each new tree is especially designed to repair the faults of the previous trees, allowing the model to repeatedly improve its prediction performance. In terms of credit risk prediction, XGBoost excels at assessing numerous borrower variables such as credit history, loan amount, repayment behavior, and income, successfully capturing deep linkages and patterns in the data.

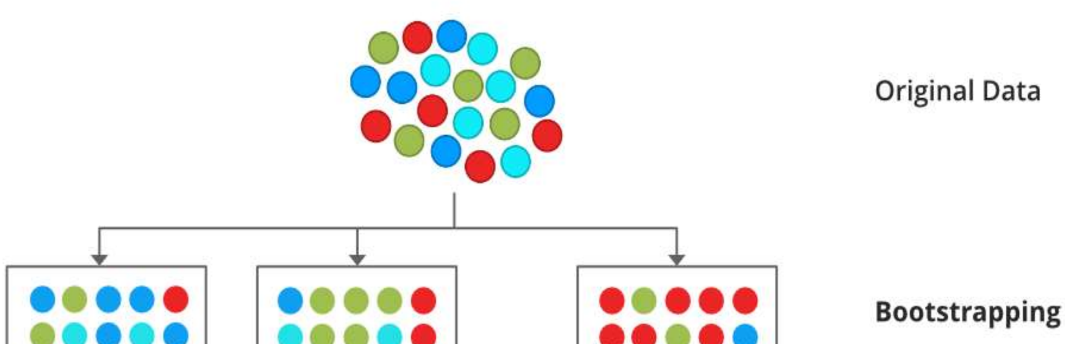


Figure 3.6: Illustration of the XGBoost Algorithm[54]

One of the key advantages of XGBoost is the use of regularization techniques, which reduce the risk of overfitting, making it especially useful for datasets with a large number of features. Furthermore, XGBoost has a built-in technique for addressing missing values, which improves its flexibility and real-world application. The algorithm also gives insights into feature importance, helping users to identify the features that have a major effect on credit risk. XGBoost's excellent accuracy and computational efficiency enable financial organizations to make accurate analyses of customer risk, resulting in better lending decisions and risk management techniques. Overall, XGBoost's effectiveness and flexibility make it a top choice for predicting credit risk, which leads to reduced default rates and enhanced credit effectiveness.

3.4.3. CatBoost

A gradient-boosting technique called category Boosting, or CatBoost, is designed to build credit risk prediction models efficiently, especially when working with datasets that contain a large amount of category features. CatBoost is unique in that it does all of this automatically, removing the need for a lot of preprocessing, including encoding and transformation of these category data. This functionality is especially helpful in credit risk scenarios when borrower factors like loan type, job status, and payback history may be included in the data. The method uses an ordered boosting strategy to build a more resilient model by using dataset changes to lower the chance of overfitting. CatBoost builds trees one after the other (see figure 2.4 in section 2.5.3), focusing on the mistakes made by earlier models. This enables it to repeatedly improve its predictions and capture intricate feature correlations.

CatBoost works by transforming category features to numerical representations while preserving their inbuilt associations. It achieves this using a technique known as "ordered target statistics," which computes the statistics of a target variable (such as the chance of default) while taking into account the order of the data. This strategy reduces the influence of target leakage and improves model accuracy. Once trained, CatBoost makes predictions based on the weighted contributions of each tree in the ensemble. Hyperparameters are adjusted to achieve optimal results on the credit risk prediction task, allowing the model to make use on its ability to detect trends in customer behaviour. By successfully handling categorical data and utilizing complex boosting algorithms, CatBoost helps financial institutions to make taught lending choices, eventually leading to fewer default.

3.4.4. Support Vector Machine (SVM)

A Support Vector Machine (SVM) model is implemented to classify customers into high-risk and low-risk categories. SVMs are known for their ability to handle non-linear relationships and their robustness to outliers, which make them a suitable choice for credit risk prediction.

The SVM model is trained on the preprocessed data, and the hyperparameters, such as the kernel function and the regularization parameter, are tuned to optimize the model's performance on the credit risk prediction task.

Support Vector Machine (SVM) is a strong credit risk prediction algorithm that accurately classifies borrowers as "low risk," "medium risk," or "high risk" based on their financial specifications. It finds the best hyperplane for separating multiple groups by maximizing the margin between the nearest data points, known as support vectors (see figure 2.6 in section 2.5.5). When the data is not linearly separable, SVM uses kernel functions to translate it to a higher-dimensional space, allowing for more accurate classification. Furthermore, SVM has a regularization parameter (C) that balances model complexity and training accuracy, hence boosting generalization to new data. SVM improves the accuracy of credit risk assessments by assessing several borrower factors such as credit history, repayment behavior, and income levels, allowing financial institutions to make better financing choices and lower default rates.

3.4.5. Deep Neural Network (DNN)

A Deep Neural Network (DNN) is developed to detect complex patterns in data and provide credit risk predictions. The DNN's architecture will have an input layer that represents the selected features from the feature selection process, as well as many hidden layers designed to acquire hierarchical data representations (see figure 2.2 in section 2.5.1). The model's architecture specify the number of hidden layers, the number of neurons in each layer, and the activation functions used. In addition, regularization techniques such as dropout layers are used to reduce the risk of overfitting.

The Sequential API is used in this credit risk prediction study because it is a simple and efficient way for building Deep Neural Networks (DNNs) layer by layer. This technique begins with an input layer that corresponds to significant parameters such as credit history, income level, and repayment behavior, followed by hidden layers made up of neurons initiated by functions such as ReLU to capture nonlinear correlations in the data. The last layer uses a softmax or sigmoid activation function to provide probability scores indicating whether borrowers are categorized as low, medium, or high risk. The Sequential model's simplicity allows for simple modifications to layers and hyperparameters, resulting in optimal training and performance refinement. Finally, this approach improves the accuracy of credit risk assessments, enabling financial institutions to make more accurate lending decisions.

3.5. Evaluation methods

The performance of the machine learning models (RF, XGBoost, CatBoost, SVM, and DNN) are evaluated using the following metrics. To simplify evaluation, confusion matrix is constructed, which helps assess classification model performance in machine learning by comparing predicted values against actual values for a dataset. Table 3.5 below shows confusion matrix for two class problem.

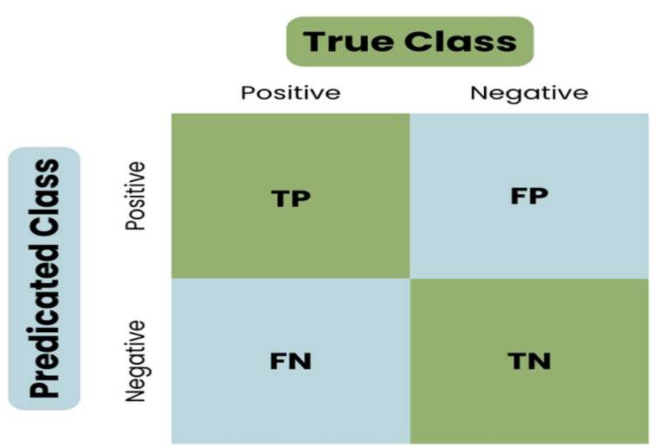


Table 3.5. Confusion matrix

Confusion matrix provides a breakdown of the model's predictions compared to the actual outcomes, allowing us to assess key performance metrics like accuracy, precision, recall, and F1-score.

True Positives (TP): These are the credits that are correctly predicted as being approved.

True Negatives (TN): These are the credits that are correctly predicted as being rejected.

False Positives (FP): These are the credits that are incorrectly predicted as being approved when they are actually rejected.

False Negatives (FN): These are the credits that are incorrectly predicted as being rejected when they are actually approved.

Based on these four values (TP, TN, FP, FN), the following effectiveness evaluation metrices can be calculated:

Accuracy: The ratio of correctly predicted instances to the total number of instances. This metric provides an overall measure of the model's performance in correctly classifying customers as high-risk or low-risk.

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

Recall: The ratio of correctly predicted high-risk instances to the total number of actual high-risk instances. This metric is particularly important in the context of credit risk prediction, as it measures the model's ability to correctly identify high-risk custom ers, which is crucial for effective risk management.

$$\text{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}}$$

Precision: The ratio of correctly predicted high-risk instances to the total number of instances predicted as high-risk. Precision measures the model's ability to avoid false positive predictions, which is essential to minimize the risk of incorrectly identifying low-risk customers as high-risk.

$$\text{Precsion} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

F1-score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance. The F1-score considers both the model's ability to correctly identify high-risk customers (recall) and its ability to avoid false positive predictions (precision).

$$\mathrm{F1-score} = 2 * \frac{\mathrm{precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

3.6. Summary

In this chapter, the overall processes followed and methods used to answer the research questions and achieve the objective of the study have been discussed. Some of them include the data collection and preprocessing approaches, implementation of the proposed method, and evaluation methods.

CHAPTER FOUR

EXPERIMENTATION AND RESULT

4.1. Overview

This section outlines how experiments are conducted based on the steps mentioned in Chapter Three. It covers all aspects of data preprocessing and the construction of the machine learning models.

The research employs an experimental research design, which primarily includes data collection and preparation (described in Chapter Three), the design and implementation of the proposed solution, and the execution of experiments to create an optimal model and report evaluation results. Details are presented in the following sections.

4.2. Dataset used for experimentation

The final dataset obtained after addressing missing values, normalizing the data, selecting features, and applying sample techniques is split into training, validation andtesting sets. The generated dataset is shown in table 4.1 below.

Table 4.1:- Final Dataset Description

Variable Name	Descriptions	Data Types
Segment	Identifier for the customer segment	Numeric
Branch	Identifier for the bank branch	Numeric
Product Description	Identifier for the product description	Numeric
Mode of Repayment	Identifier for the repayment mode	Numeric
Repayment Amount	Amount to be repaid in installments	Numeric
Expected Collection to Date	Expected amount to be collected by date	Numeric
Amount Collected to Date	Amount of money collected to date	Numeric
Arrears Amount	Amount in arrears	Numeric
Principal	Principal amount of the loan	Numeric
Interest	Interest amount on the loan	Numeric
Interest Rate	Interest rate of the loan	Numeric

Status	Current status of the loan (encoded)	Numeric
days_overdue	Days overdue for the payment	Numeric
on_time_payment	Indicator of on-time payment (1 = Yes, 0 = No)	Numeric
due_day_of_week	Day of the week when payment is due	Numeric

For modeling purposes, the final dataset is divided into training, validation, and testing sets, with 70% of the data for the training set,10% for the validation set, and 20% for the testing. This 70/10/20 split is often used because it guarantees that the model has enough data to discover underlying linkages and patterns while retaining enough data for a separate assessment[55]. A bigger training set (70%) ensures the model does not overfit. Additionally, a validation set is used during the selection process to provide an intermediate check on model performance and guide hyperparameter optimization for the model tuning. Finally, a test set (20%) offers an objective evaluation of the model's performance on untested data, guaranteeing its robustness and predictive ability. This split finds the perfect compromise, providing the model with a wealth of samples to learn from without jeopardizing the validity of the assessment procedure.

4.3. Statistical analysis

Hereunder we present data visualization of some of the key variables to have detailed information about each attributes and their effect on loan risk status.

4.3.1. Loan Type

The loan type variable is a categorical feature that takes five different values:-**term loans, overdraft loans, advanced against export bills, Qard Financing, and advanced against murabaha**. Each of these categories reflects a distinct financial product offered by Awash Bank to meet the diverse demands of its customers. Understanding these loan kinds is critical for assessing lending patterns and customer preferences since they represent the many possibilities accessible in the financial market. This knowledge can be useful in analyzing risk, improving loan offers, and personalizing financial services to individual customer needs.

As shown in Figure 4.1, term loans are the most frequent loan preference among customers, followed by overdraft loans. According to banking experts, term loans provide a predictable repayment structure with fixed installments over an extended period, which helps customers manage their budgets effectively. Additionally, term loans usually have lower interest rates than other products like overdrafts, making them a more cost-effective option.

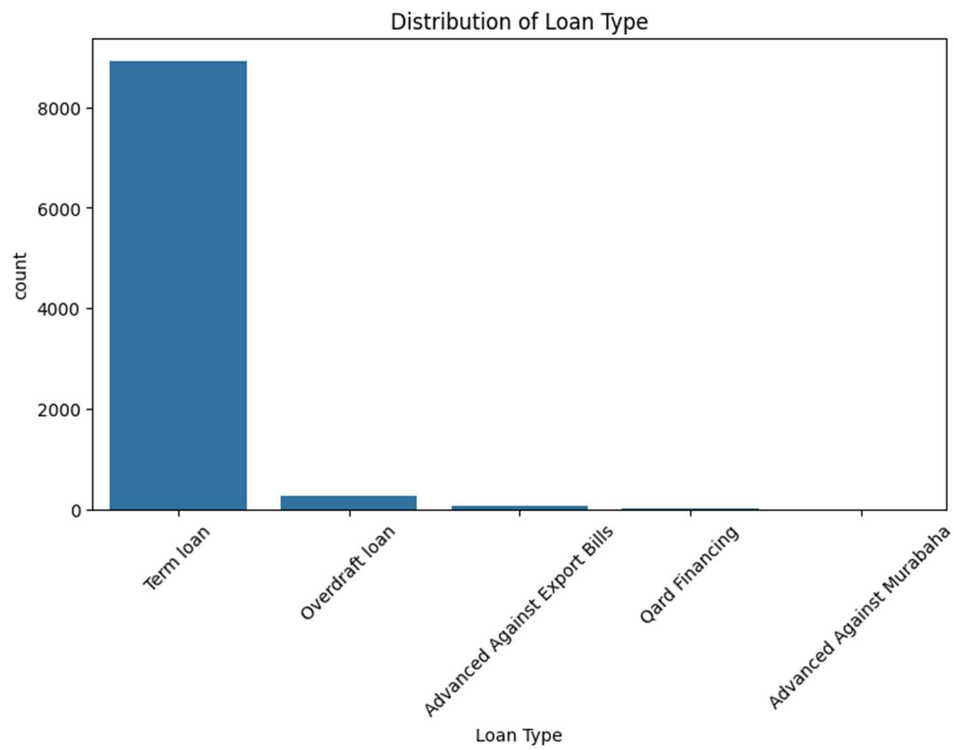


Figure 4.1:- Loan Type Variable Graph

4.3.2. Segement

The segment variable is a categorical feature with nine distinct values: retail, SME, corporate, business, government, IFB, IFB section, institutional, IFBCRM. It classifies businesses based on factors including industry type, size, revenue, and risk profile. Understanding these categories allows Awash Bank to personalize its lending approach as well as risk assessments, as various segments have different levels of risk, that can influence default risk. Notably, the data indicates an equal distribution across these segments, implying identical representation for each group.

As presented in figure 4.2 below all segments are almost with equal distribution. This means that the bank provides equal service for all segments. This balanced approach could indicate the bank's strategic focus on serving a wide range of customer segments, ensuring equitable access to its financial services.This inclusive approach has a great contribution to overall stablity and resilence of the bank in the long run.

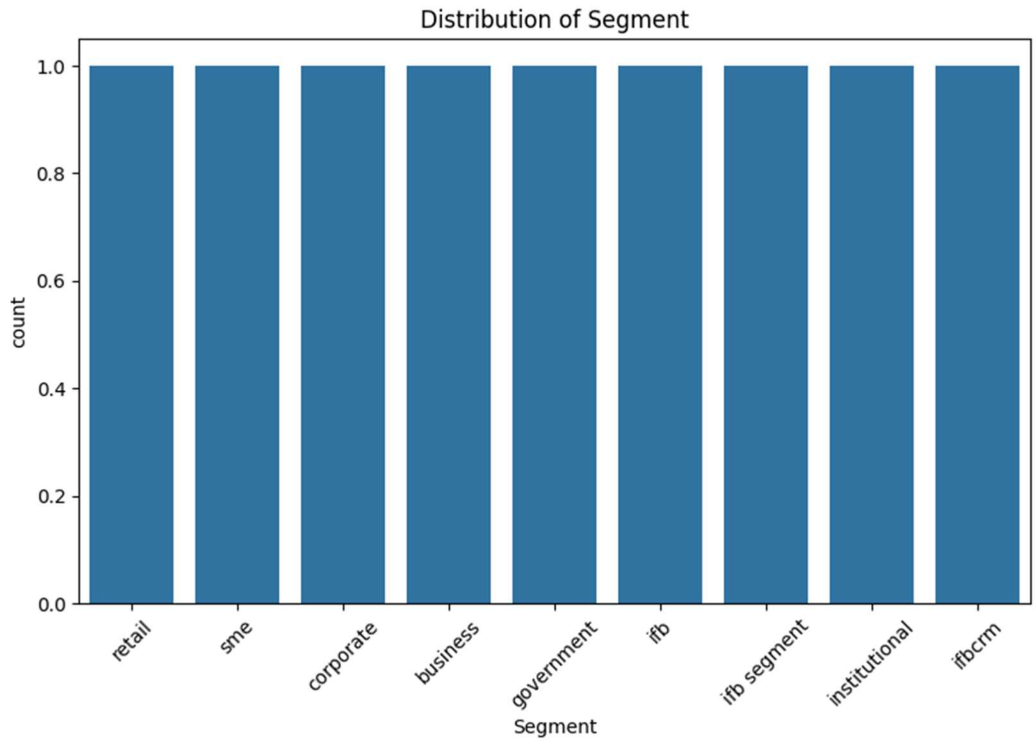


Figure 4.2:- Segment Variable Graph

4.3.3. Amount Granted and Arrears Amount

The Amount Granted and Arrears Amount variables obtained from Awash Bank serve as key indicators of the bank's borrowing activities and credit risk management strategies. Amount Granted indicates the overall amount distributed to borrowers, showing the bank's dedication to enabling loan access across all segments, including individuals as well as business organizations.

In contrast, Arrears Amount indicates the total amount of payments that are late and is an important indicator of credit risk. A smaller arrears figure indicates excellent credit management and a solid payback culture among borrowers, whereas a greater figure may signal financial instability for the bank. Overall, these factors give critical insights into the bank's lending activity and borrower repayment patterns, which help in the assessment of its credit risk measures.

Figure 4.3 below shows a large difference between the total Amount Granted and the Arrears Amount, with the Arrears Amount representing only 0.41% of the Total Amount Granted. This highlights the efficacy of the bank's credit evaluation systems and the general financial health of its borrower base.

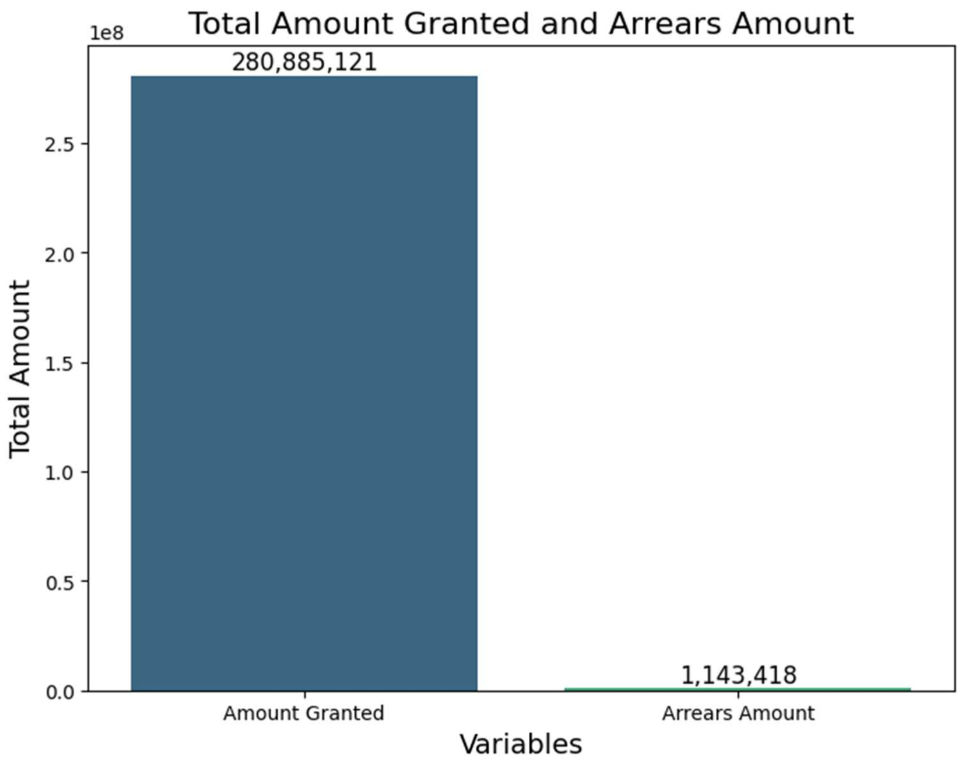


Figure 4.3: Amount Granted and Arrears Amount Variable Graph

4.3.4. Interest Rate

The histogramshown in Figure 4.4below illustrates the distribution of interest rates and days overdue in the loan portfolio.

The left histogram shows the distribution of interest rates. It has a bimodal form, with a large peak about 0 and a smaller peak around 10. This shows that the dataset comprises a large number of loans with extremely low interest rates, as well as a smaller number of loans with higher rates around 10. The skewed distribution suggests that most loans have low interest rates, with a smaller group having higher rates.

The right histogram shows the distribution of overdue days. It has a highly right-skewed shape, with a huge peak about 0 days overdue and a lengthy tail that extends to higher values.This means that the majority of loans are only a few days overdue, however just a low number are considerably overdue, perhaps indicating increased risk or late payments.

These distributions give insightful information that can help the lender with risk management, underwriting, and portfolio optimization. The bimodal interest rate pattern and right-skewed days overdue data provide a more detailed knowledge of the loan portfolio's characteristics, allowing the institution to make better decisions.

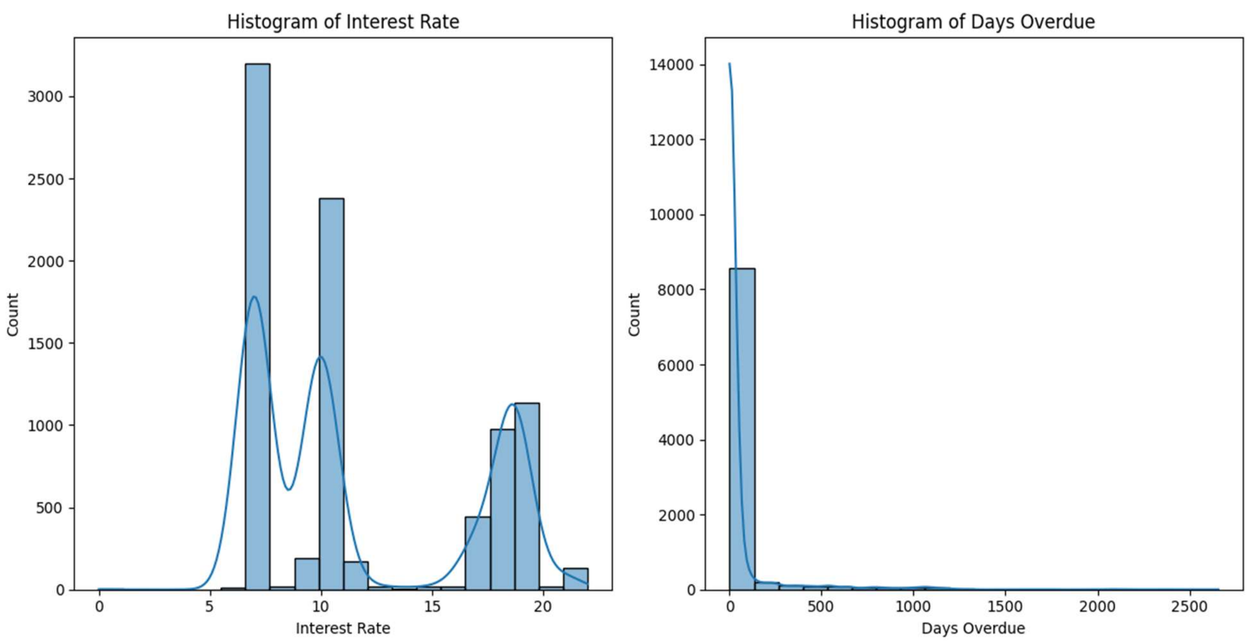


Figure 4.4: Interest Rate and Days Overdue Variable Graph

4.3.5. Branch

The Branch variable shows the addresses of 124 different branches from which loans are issued. Figure 4.5 below shows the top 10 branches with the highest loan counts, providing insights into the lending activities and dynamics at the branches generating the most loans. Analyzing this variable is critical for understanding regional variations in lending patterns. Understanding branch-level lending dynamics not only helps with risk assessment, but it also informs strategic decisions for optimizing future lending efforts, ensuring that credit services are spread evenly throughout regions.

Figure 4.5 shows the top ten branches by loan count. The 18 Mazoria and Addis Ketema branches are the most visible, demonstrating its importance in lending activity. Other branches, such as DAFRIQUE BRANCH and IFB-Addis Ketema Branch, are also heavily involved, although IFB-Lideta, Betel, and IFB-Sidamo Tera have more balanced loan distributions. Loan distribution disparities may be impacted by local economic conditions, business activities, branch capacity, and borrower demographics. Understanding these patterns is critical for determining loan access equity and improving lending tactics across geographies.

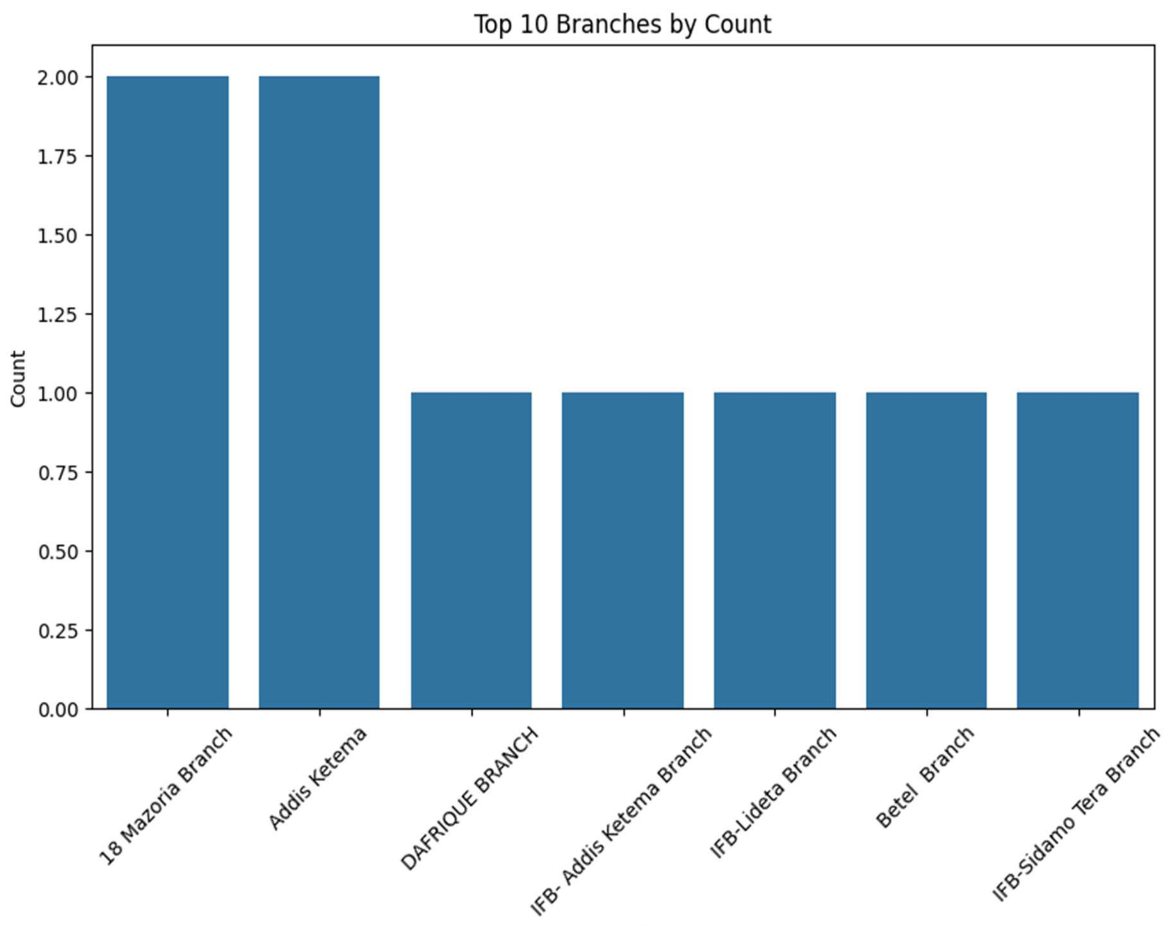


Figure 4.5: Distribution of Customer Transactions Across Top Branches

4.4. Descriptive Statistics

In-depth summaries of the dataset are offered by descriptive statistics, which display important metrics including count, standard deviation (STD), mean, minimum and maximum values. Researchers measure and characterize the essential features of the data with the help of these statistics. Analytical metrics, such as count, mean, standard deviation, and measures of central tendency, are computed and are shown in Table 4.2in order to produce a more informative summery.

Figure 4.6: Descriptive Statistics of Customer Data

The Summary Statistics Heatmap presents a comprehensive overview of the bank's lending portfolio, showcasing a wide range of loan characteristics across different metrics. The data reveals a diverse set of borrowers accessing financing, with loan amounts ranging from under $10,000 to over $2.5 billion, grace periods varying from less than a day to 72 days, and repayment amounts spanning nearly $90 million. Additionally, the heatmap highlights the bank's loan collection performance, outstanding balances, interest rates, and days overdue, providing valuable insights into its credit risk management practices. This detailed data can inform strategic decisions to optimize the bank's lending operations, manage risks, and ensure equitable access to its financial services across the customer base.

4.5. Correlation Analysis

Correlation analysis is used to investigate the correlations between the independent and dependent variables in the context of credit risk prediction. This research assists in identifying potential multicollinearity among independent characteristics, which might have an impact on predictive model performance. Given that the dataset contains both continuous and categorical variables, the Spearman correlation method is selected as the best strategy for this research.

The Spearman correlation is especially appropriate for this study since it determines how well the relationship between two variables can be captured by a monotonic function. This means it can successfully capture non-linear correlations, making it appropriate for credit risk analysis where the interactions between features and the target variable are complex. Furthermore, Spearman correlation is robust to outliers and does not assume a normal distribution, consistent with the diverse character of financial data.

Figure 4.6 below depicts a correlation heatmap created to visualize these associations. This heatmap clearly shows the degree and direction of connections, allowing for a better understanding of how different variables relate to credit risk outcomes.

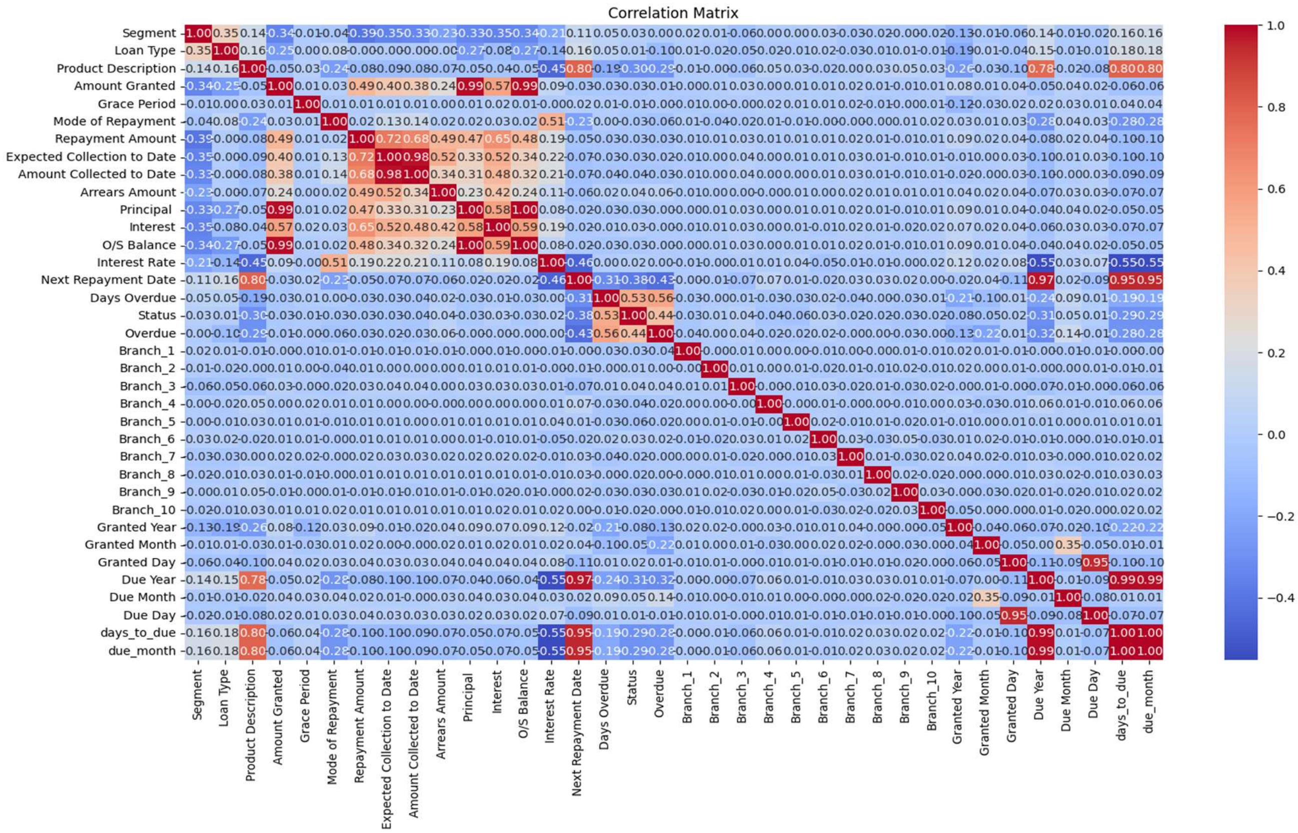


Figure 4.7: Correlation Heatmap of the Variables

In this study, several variables displaying significant multicollinearity were identified as shown above in the diagram, including 'Amount Granted','O/S Balance','Due Year','Due Day', 'days_to_due','due_month', and 'Next Repayment Date'. Multicollinearity may increase error rates, limit model interpretability, and bias prediction, making it difficult to evaluate the individual contributions of related variables. To improve the robustness and reliability of the prediction models, these multicollinear variables are dropped from the dataset. This removal approach is to enhance model performance and allow for a better contributions of the existing variables to credit risk prediction, resulting in improved creditworthiness assessment prediction. Below figure 4.7 is the heatmap of correlations after discarding multicollinear variables.

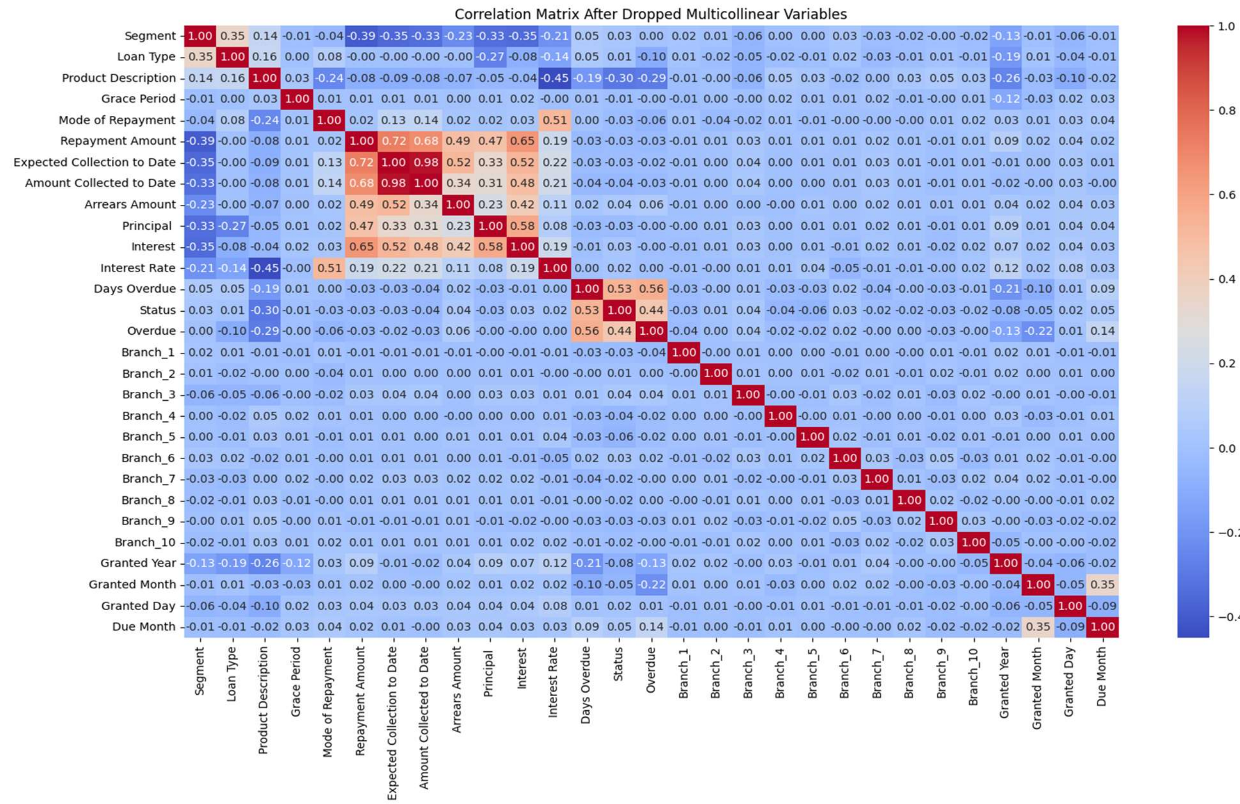


Figure 4.8: Correlation Heatmap After Removing Multicollinearity

4.6. Hyper-parameter tuning

Machine learning employs numerous parameters to produce results based on the user's requirements. However, selecting parameters to fit the model is a complicated task. Hyperparameter tuning is critical for overcoming the obstacles of selecting parameters that generate better results or reduce learning errors. In many cases, Bayesian optimization is recommended, especially when the overall dataset exceeds 100,000 samples and comprises more than 100 features[56]. This approach effectively explores the parameter space by learning from prior assessments, making it appropriate for large and complex datasets.

However, because there are fewer attributes used for the purpose of credit risk prediction, Random Search is selected. With less computational time than complex techniques, Random Search provides an easy-to-use and effective tool for exploring the hyperparameter space. Despite evaluating each possible combination of parameters, Grid Search can be computationally expensive and result in lengthy processing times, particularly when dealing with huge datasets[57]. The user experience is a major factor in Manual Search, however, it can add bias and not necessarily produce the best results. Random Search offers a balanced solution in this situation, allowing for quick iterations while efficiently optimizing model performance without adding unnecessary effort.

4.7. Modeling

In this section, five machine learning algorithms have been evaluated to determine how effective they are in predicting credit risk. The goal is to find the model that makes the most accurate and reliable prediction.

4.7.1. Extreme Gradient Boosting

The XGBoost algorithm is written in Python using the xgboost module, which is a fast and scalable gradient boosting method. Instead of utilizing default settings for credit risk prediction, the RandomizedSearchCV() method is used in this study to optimize hyperparameter performance. After tuning, the following optimal parameters are selected:

Table 4.2: Tuned XGBoost Parameters

Parameter	Value
subsample	0.6
reg_lambda	5
reg_alpha	0.1
n_estimators	1000
min_child_weight	4
max_depth	4
learning_rate	0.01
gamma	0.1
colsample_bytree	0.6

This study uses an XGBoost model with the optimal parameters found through tuning to predict credit risk of customers using the data collected from Awash Bank. The model is built up with these best parameters, and the performance results are provided below.

Table 4.3: XGBoost Training Accuracy Evaluation Results

Measures	Accuracy
Accuracy	92.4%
Precision	93.4%
Recall	92.4%
F1 score	92.7%

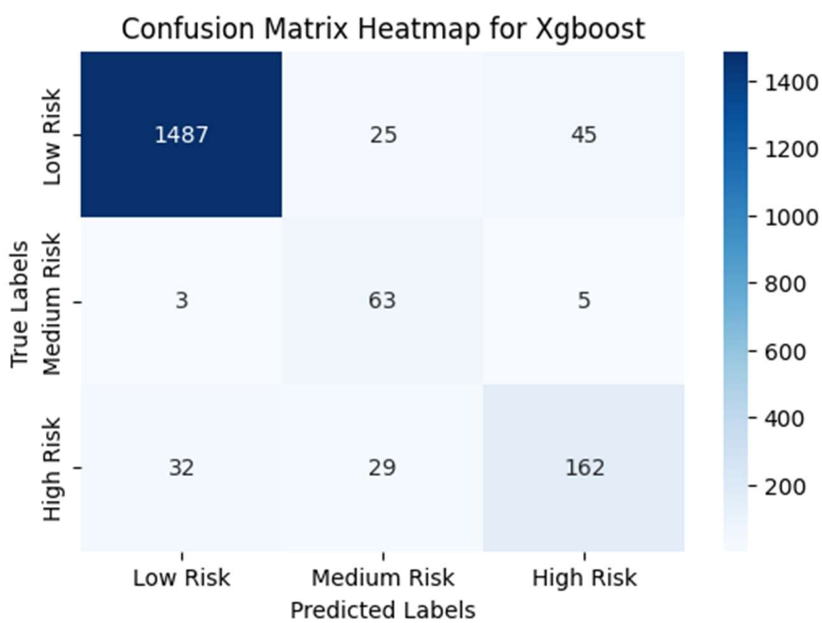


Figure 4.9: Confusion Matrix Heatmap for XGBoost

4.7.2. Categorical Boosting

The Python catboost package is used to develop the CatBoost model for credit risk prediction. The RandomizedSearchCV() technique is used to tune the hyperparameters of the model in order to optimize its performance. After tuning, the following optimal parameters are selected:

Table 4.4: Tuned CatBoost Parameters

Parameter	Value
learning_rate	0.2
Depth	8

l2_leaf_reg	5
iterations	500
bagging_temperature	1
border_count	64

The CatBoost algorithm for credit risk prediction produced the following results based on the optimal parameters selected throughout the tuning process:

Table 4.5: CatBoost Training Accuracy Evaluation Results

Measures	Values
Accuracy	91.19%
Precision	91.15%
Recall	91.19%
F1 score	91.34%

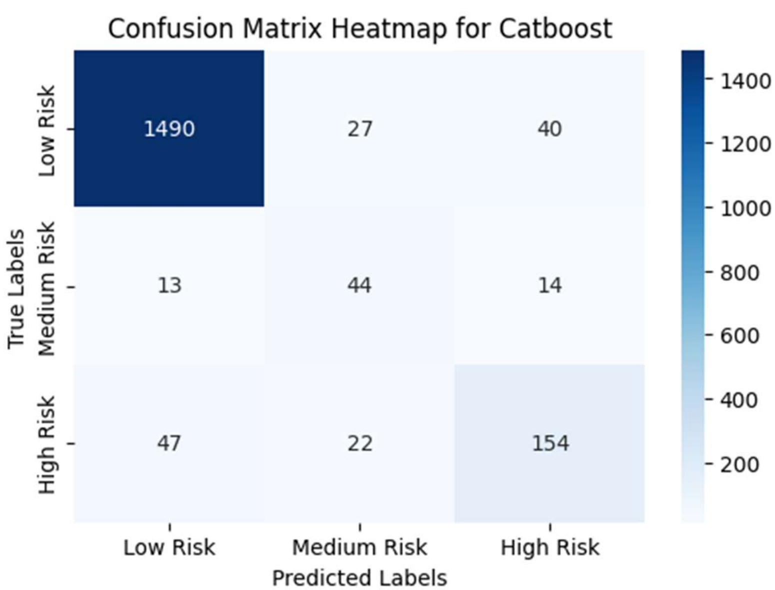


Figure 4.10: Confusion Matrix Heatmap for CatBoost

4.7.3. Support Vector Machine

The SVM function from the Python sklearn.svm package is utilized for building the support vector machine (SVM) model. In order to improve the model for credit risk prediction in this study, a linear kernel with C = 10, max_iter = 2000 and tol = 0.0001 are used using the RandomizedSearchCV() technique. The SVM algorithm's kernel trick is essential since it establishes the classification's decision boundary. The linear kernel used in this study makes it possible to simply dividing the credit risk groups using a single line. When compared to more complicated kernels, the linear kernel is faster in processing time which is its main benefit [58]

Table 4.6: Tuned SVM Parameters

Parameter	Value
C	10
max_iter	2000
tol	0.0001

The following table 4.10shows the result obtained from the support vector machine in credit risk prediction.

Table 4.7: SVM Training Accuracy Evaluation Results

Measures	**Values**
Accuracy	63.9%
Precision	4.20%
Recall	50.4%
F1 score	39.3%

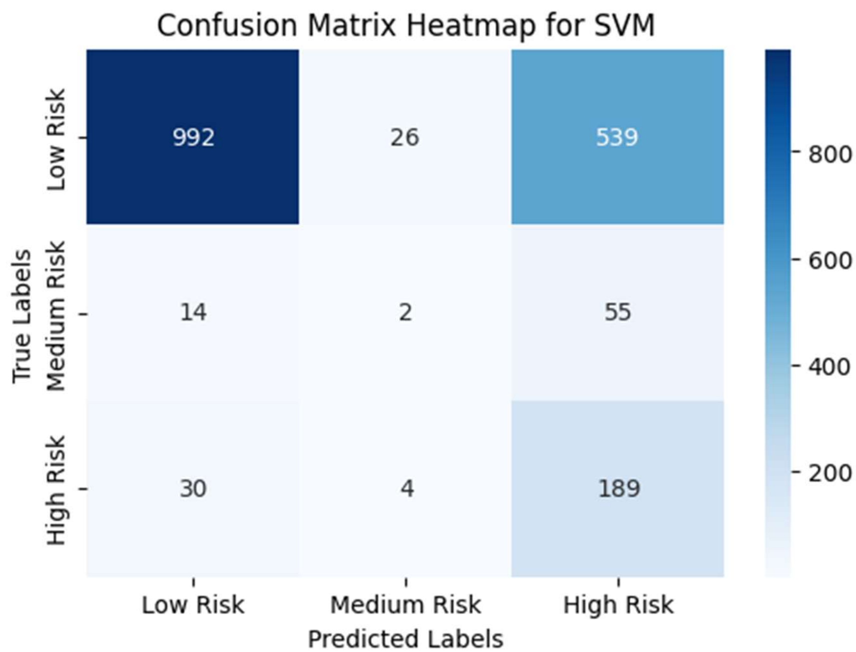


Figure 4.11: Confusion Matrix Heatmap for SVM

4.7.4. Deep Neural Network

Deep neural networks are a type of artificial neural network with two or more hidden layers. The advantage of deep neural networks lies in their ability to learn complex features and handle more intensive tasks by executing multiple complex operations simultaneously[59]. In this study, a deep neural network algorithm is used to predict the credit risk level of customers at Awash Bank, comparing its performance with other traditional algorithms on the same dataset. Parameter tuning is performed to select optimal parameters for this deep learning model, using a method called RandomizedSearchCV(). This approach is chosen due to its efficiency in finding the right parameters with fewer iterations. After tuning, the following optimal parameters are selected:

Table 4.8: Tuned Deep Neural Network Parameters

Parameter	Value
Optimizer	Adamax

Neurons	128
learning_rate	0.001
Epochs	22
batch_size	64
Activation	Linear

This study used a Deep Neural Networkmodel with the optimal parameters found through tuning to predict credit risk of Bank customers. The model is built up with these best parameters, and the performance results have are provided below in table 4.13.

Table 4.9: Deep Neural Network Training Accuracy Evaluation Results

Measures	**Values**
Accuracy	83.6%
Precision	74.2%
Recall	83.6%
F1 score	77.2%

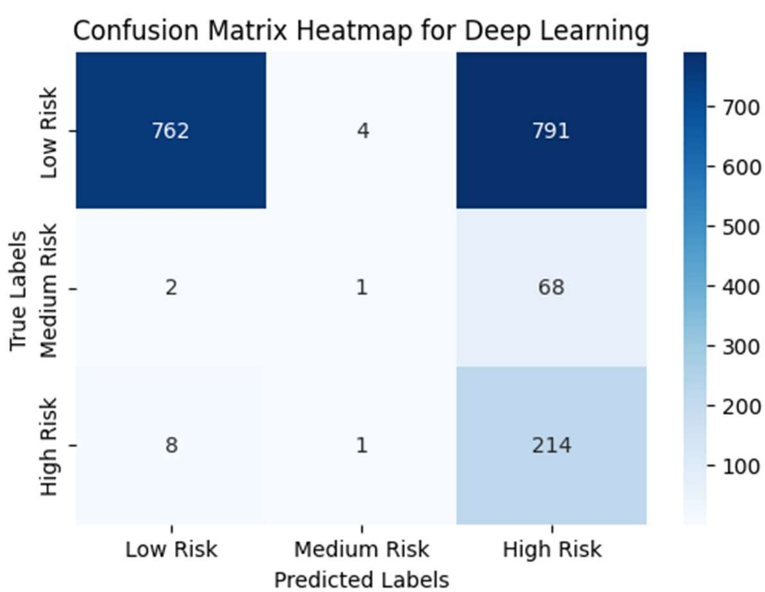


Figure 4.12: Confusion Matrix Heatmap for DNN

4.7.5. Random Forest

The Random Forest algorithm is implemented in Python using the sklearn.ensemble class. This algorithm is based on ensemble learning, which combines multiple classifiers to improve predictions on a dataset. The RandomizedSearchCV() function is used to select the optimal parameters for the Random Forest classifier. After tuning, the following optimal parameters are selected:

Table 4.10: Tuned RandomForest Parameters

Parameter	Value
n_estimators	200
min_samples_split	10
min_samples_leaf	1
max_features	log2
max_leaf_nodes	20
Criterion	Entropy
max_depth	10

This study used a Random Forest model with the optimal parameters found through tuning to predict credit risk of Bank customers. Table 4.16below provids the performance results of random forest algorithm.

Table 4.11: RandomForest Training Accuracy Evaluation Results

Measures	**Values**
Accuracy	90.4%

Precision	71.5%
Recall	80.6%
F1 score	74.04%

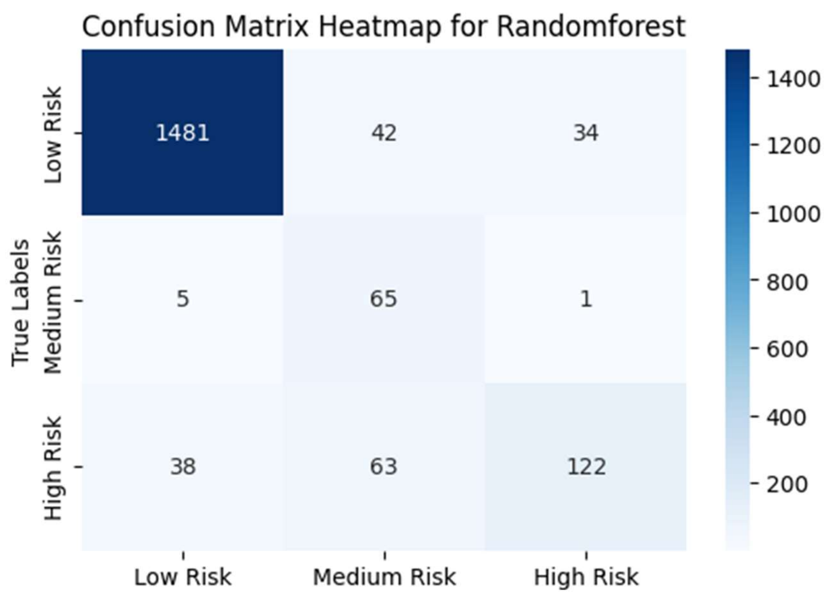


Figure 4.13: Confusion Matrix Heatmap for Randomforest

4.8. Comparison of models

An experiment is carried out to determine the most effective machine learning model for predicting credit risk using the Awash Bank dataset. To address the class imbalance in the dataset, the SMOTE sampling approach is used to balance the data prior to model building.

The machine learning algorithms are evaluated based on key evaluation metrics such as accuracy, precision, recall, and F1 score. In the experiments, across all evaluation metrics, the XGBoost algorithm outperformed Random Forest, SVM, CatBoost, and Deep Neural Network. The graph below provides a comparative analysis of the accuracy of the machine learning algorithms used in this study Deep Neural Network, Random Forest, SVM, CatBoost, and XGBoost.

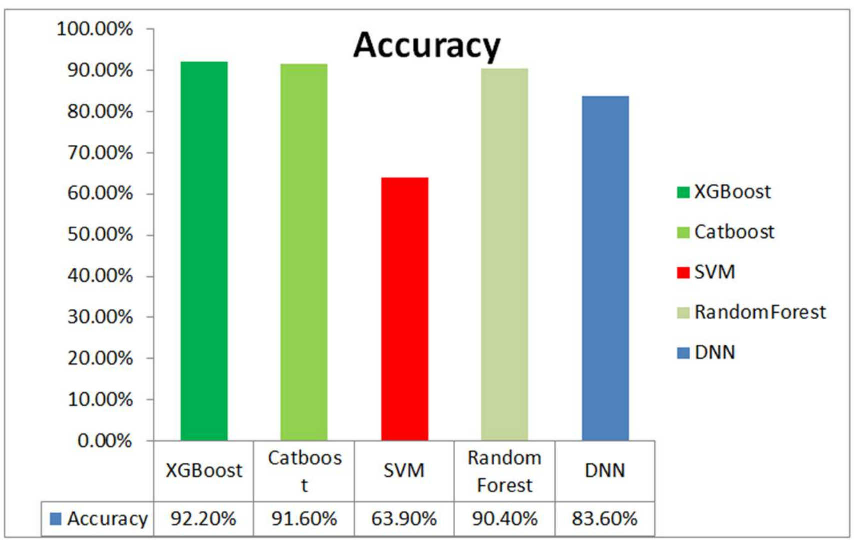


Figure 4.14:- Accuracy Comparison graph

Here in this above plot, it could be seen that the Accuracy is 92.2% for the XGBoost Algorithm which comparativily implies that this model is quite outperformed other algorithms, and hence we suggest the use of XGBoost in predicting the credit risk in banks.

4.9. Discussion of result

This study successfully solves the research topics presented, providing useful insights on credit risk prediction in the banking sector. The identification of nineteen essential variables required for predicting credit risk reflects a comprehensive examination of the available data and literature. Using approaches such as ExtraTreesClassifier and correlation matrices, the study not only identified these features but also underlined their relevance in improving prediction accuracy, resulting in improved lending decisions.

In assessing multiple machine learning models, the study successfully compares techniques such as XGBoost, CatBoost, SVM, Random Forest, and DNN. The findings show that the XGBoost model beats the others, with remarkable accuracy. This demonstrates XGBoost's capacity to handle complicated data interactions, which makes it a better for credit risk prediction.

This research answers each research question that are raised for investigation in the first chapter.

The first research question asks to identify the key features for predicting credit risk within theBank. Accordingly, the research being conducted identified important features to improve the prediction accuracy of the credit risk model, based on previous research and a comprehensive review of the available data. The ExtraTreesClassifier and correlation matrix have been applied to identify the most important and correlated attributes. This research identified nineteen key characteristics for predicting credit risk. These variables are: segment, loan type, product description, grace period, mode of repayment, repayment amount, expected collection to date, amount collected to date, arrears amount, principal, interest, interest rate, days overdue, status, overdue, branch, granted year, granted month, granted day, and due month. By focusing on these key features, the study aims to improve the accuracy of credit risk predictions and contribute to better decision-making in lending processes.

The second research question concerned about suggesting the most effective machine learning model for credit risk prediction, **The study tries to** evaluate several algorithms such as XGBoost, CatBoost, SVM, Random Forest, and DNN are experimented to identify an optimal model for bank credit risk prediction. .

Finally, the performance of the proposed predictive model is tested in credit risk predictions.The performance of the proposed XGBoost model in credit risk prediction is evaluated, in which the model obtained an accuracy of 92.2%, demonstrating its ability to really predict events of credit risk. Furthermore, it demonstrated a precision of 93.2%, indicating its ability to identify true positive instances while reducing excessive false positives. The recall also gained around 92.2%, demonstrating the model's capacity to capture actual positive cases. Furthermore, the F1 score of

92.5% demonstrates a balanced performance between precision and recall, validating the model's durability. The result shows that XGBoost is the most effective one than other ML algorithms. Its effective handling of complex relationships within the data has given it the power to effectively perform on the given Bank dataset. These findings therefore highlight the XGBoost model's significance to allow banks for more informed decisions regarding loans and improved credit risk management techniques.

Chapter Five

CONCLUSION AND RECOMMENDATIONS

5.1. Overview

This chapter presents concluding remarks to show the strength of the study and on the attempts made to construct an optimal model for credit risk prediction. Based on the findings of the study, practical recommendations and the way forwards are suggested for improving credit risk assessment methods in financial institutions.

5.2. Conclusion

This study demonstrates the importance of credit risk prediction in improving profitability for banking institutions. In the banking business, accurately predicting credit risk is critical to reduce defaults and increasing income. This work used an experimental method for building a prediction model for predicting credit risk of bank's customers. Data is collected using a stratified random sampling approach, and preprocessing steps are performed to verify that the dataset is suitable for machine learning. To assess the performance of the model's, the data is divided into training and test: 80% for training and 20% for testing. Feature engineering and transformation strategies are used to improve the model's prediction capabilities.

A significant issue in this work is dealing with class imbalance and dealing with the limitations of deep learning models due to the small dataset. To resolve class imbalance, the SMOTE approach is used, which helped to balance the training dataset. Despite these attempts, the Deep Neural Network model struggled with overfitting and underperformance, most likely due to the small size of dataset. To confirm the reliability of the results, five distinct machine learning algorithms are experimented: XGBoost, DNN, CatBoost, SVM, and Random Forest. To maximuize the performance of each model, among others, tuning techniques using RandomizedSearchCV is used to tune hyperparameters.

All Selected algorithms are evaluated based on metrics like accuracy, precision, recall, and F1 score. The XGBoost model outperformed than the other algorithms, with the maximum accuracy of 92.2%. This finding demonstrates the efficacy of ensemble learning approaches like XGBoost in credit risk prediction, especially when applied to complex financial data. The findings imply that XGBoost is a very reliable method for predicting credit risk at Awash Bank, and its use might greatly enhance financial decision-making processes.

The primary strength of this study is its use of state-of-the-art ensemble algorithms and deep learning approaches for accuratly predicting credit risk using data from Awash Bank. The findings show that Extreme Gradient Boosting (XGBoost) outperformed than other models, such as Random Forest, SVM, CatBoost, and DNN, in predicting credit risk based on customer data. Furthermore, this study uses original, unprocessed data obtained directly from the bank, giving a solid foundation for analysis.

This study also addresses a significant gap in the existing literature, since previous research has mostly focused on microfinance institutions (MFIs) and a small subset of banking. Using actual customer data from Awash Bank, this study adds useful insights into credit risk prediction in a banking sector where such research has been limited. In addition, the utilization of parameter tuning approaches, especially Randomized Search and correlation analysis, to address multicollinearity in the dataset strengthens the findings. This methodological approach emphasizes the trustworthiness and validity of the findings stated in this study.

The significant limitation of this study is Ethiopia's diversified financial landscape, which comprises over 28 banking institutions. Due to time and resource constraints, it is not feasible to include data from all banks in this study. As a result, the research focused on a small sample of data of customers from Awash Bank, which may not correctly reflect the institution's overall demographics of banking sectors. This limitation causes a significant difficulty, because, despite the banks providing similar services and having comparable customer data management and payment systems, the findings may lack generalizability to the broader range of financial institutions in the entire country.

One of the limitation is the underperformance of DNN algorithm's, which is further exacerbated by the limited amount of training data. The DNN model tended to overfit the training data, generating less reliable predictions. Contributions and Impact

This study contributes significantly to the area of credit risk prediction by focusing on data from Awash Bank, where little research has been conducted. Most previous research in Ethiopia focused on microfinance institutions, the tax authority, or other datasets, but little attention has been given to predicting credit risk in big financial institutions such as Awash Bank. Using actual customer data from Awash Bank, this study gives significant insights into the literature. The study's findings can be used to provide the groundwork for future research on credit risk prediction utilizing data from additional Ethiopian banks, therefore helping to address a knowledge gap in the area of study.

In addition, this study provides the use of machine learning techniques in the financial services sector, particularly in Ethiopia. While machine learning methods are commonly used in many sectors, deep learning's use in credit risk prediction is limited. This study demonstrates the potential of deep learning, with certain difficulties such as overfitting because of limited data. Furthermore, the study underlines the effectiveness of ensemble algorithms like XGBoost, CatBoost, and Random Forest in predicting credit risk. XGBoost, in particular, outperforms the other models by attaining 92.2% accuracy, demonstrating that it is the well-performed algorithm for credit risk prediction in the Awash Bank dataset. From a business viewpoint, Awash Bank can proactively evaluate customer risk due to these predictive models, which enhance decision-making and may boost revenue through improved risk management techniques.

5.3. Future Work and Recommendations

The dataset must be enlarged for future research in order to increase the accuracy of credit risk prediction models. While the data used for this study is obtained from Awash Bank, a deeper examination might be possible if a larger and more varied collection of customer data are collected. Incorporating data from other Ethiopian financial organizations or branches might result in more comprehensive models that more accurately represent the banking industry as a whole. The data obtained from bank has only contained customers live in Addis ababa, so for the future the variable “Region” should conatin other areas. Additionally, by including outside variables like market dynamics and macroeconomic trends, the model's predictive power would be enhanced, enabling the bank to foresee wider economic impacts on credit risk.

Another significant area for future research is resolving the limitations faced by deep learning models in this study. Due to data limitation, the DNN model underperformed, demonstrating the need for a bigger dataset to fully realize its capabilities. Future research should focus on collecting additional data and use advanced methods like cross-validation and regularization to prevent overfitting and increase model generalization. Additionally, investigating different machine learning techniques, such as LightGBM other ensemble methods, might improve credit risk prediction at Awash Bank. Implementing and refining these methodologies allow the bank to make more accurate, data-driven decisions, resulting in better risk management and financial performance.

Such research can further be extended to other organizations offering credit to their customers thereby to measure and manage credit risk they may encounter in their business activities.

References

[1] D. Ohia and C. Daniel, “Advancements and Challenges in Machine Learning and Artificial Intelligence: Shaping the Future of Technology,” Jan. 2024.

[2] B. Chaudhary, *Importance of Data (A Term Paper)*. 2023. doi: 10.13140/RG.2.2.23837.69602.

[3] A. Scott, P. Amajuoyi, and K. Adeusi, “Effective credit risk mitigation strategies: Solutions for reducing exposure in financial institutions,” *Magna Sci. Adv. Res. Rev.*, vol. 12, pp. 198–211, May 2024, doi: 10.30574/msarr.2024.11.1.0084.

[4] Dileep Kumar, “Credit Management Control.” [Online]. Available: https://www.highradius.com/resources/Blog/strategies-involved-in-credit-risk-management/

[5] A. S. Deng, L. Rono, and J. Sang, “Credit Risk Management and the Performance of Financial Institutions in South Sudan,” *Mod. Econ.*, vol. 11, no. 11, pp. 1919–1928, 2020, doi: 10.4236/me.2020.1111128.

[6] National Bank of Ethiopia (NEB), “Quarterly Bulletin Second Quarter 2017/18 Fiscal Year Series,” vol. 34, no. 2, p. 154, 2018.

[7] T. Berhane, T. Melese, and A. M. Seid, “Performance Evaluation of Hybrid Machine Learning Algorithms for Online Lending Credit Risk Prediction,” *Appl. Artif. Intell.*, vol. 38, no. 1, pp. 1–26, 2024, doi: 10.1080/08839514.2024.2358661.

[8] Rackspace Technology, “The 2023 AI and Machine Learning Research Report,” *Rocksp. Technol.*, 2023.

[9] K. Alnabulsi, E. Kozarević, and A. Hakimi, “Assessing the determinants of non-performing loans under financial crisis and health crisis: evidence from the MENA banks,” *Cogent Econ. Financ.*, vol. 10, no. 1, pp. 1–23, 2022, doi: 10.1080/23322039.2022.2124665.

[10] D. Bizuwork, H. Advisor, and G. Semeon, “Loan Risk Prediction Using Machine Learning Algorithms:-The Case of Ethiopia’s Micro-Finance Institution’s,” no. June,

2019.

[11] T. Mengistu, T. Dargaso, and M. Demessie, *Machine Learning-Based Credit Risk Assessment for Predicting Loan Defaulters in Ethiopian Banking industry*. 2023. doi: 10.21203/rs.3.rs-3800266/v1.

[12] W. Magarsa, E. Stream, D. S. Advisor, and N. E. Benti, "College of Natural and Computational Sciences Computational Data Science program Artificial Intelligence-Enhanced Credit Risk Assessment on Commercial Bank of Ethiopia," 2024.

[13] T. Melese, T. Berhane, A. M. Seid, A. Walelgn, and B. Dar, "Performance analysis of deep and machine learning algorithms for loan evaluation model," vol. 16, no. June, pp. 101–114, 2023.

[14] K. Martin, B. Hailey, and L. Knoll, "Psychology 303: Research Methods in Psychology," no. Straub 180, 2012.

[15] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive text mining and database fusion approach," *Environ. Health Perspect.*, vol. 127, no. 9, pp. 2825–2830, 2019, doi: 10.1289/EHP4713.

[16] T. Zewdu, "ST. MARY'SUNIVERSITY DEPARTMENTS OFMASTERS OF BUSINESS ADMINISTRATION Assessment of Credit Risk Management Practice of Awash Bank," 2021.

[17] T. AYALEW, "the Effect of Credit Risk on the Financial Performance of Commercial Banks: Evidence From Selected Private Commercial Banks in Ethiopia," *Corp. Bus.*, vol. 4, no. 3, pp. 18–25, 2022.

[18] K. Tegene and B. Venkataram, "THE EFFECT OF CREDIT RISK MANAGEMENT ON FINANCIAL PERFORMANCE OF COMMERCIAL BANKS IN ETHIOPIA," vol. 42, pp. 471–482, Feb. 2023, doi: 10.17605/OSF.IO/GZNHP.

[19] J. Witzany, "Rating and Scoring Systems. In Credit Risk Management," *Springer Cham*, pp. 19–116, 2017, [Online]. Available: https://www.bis.org/publ/bcbsc125.pdf

[20] A. Khandani, A. Kim, and A. Lo, "Consumer Credit-Risk Models Via Machine-Learning

Algorithms," *J. Bank. Financ.*, vol. 34, pp. 2767–2787, Nov. 2010, doi: 10.1016/j.jbankfin.2010.06.001.

[21] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.," 2022.

[22] A. Bin Rashid and M. A. K. Kausik, "AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications," *Hybrid Adv.*, vol. 7, no. August, p. 100277, 2024, doi: 10.1016/j.hybadv.2024.100277.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[24] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.

[25] S. Singh, R. Kumar, S. Payra, and S. Kumar Singh, "Artificial Intelligence and Machine Learning in Pharmacological Research: Bridging the Gap Between Data and Drug Discovery," *Cureus*, vol. 15, Aug. 2023, doi: 10.7759/cureus.44359.

[26] I. Goodfellow, Y. Bengio, and A. Courville, "Example Notation for Deep Learning," *Raw.Githubusercontent.Com*, 2016.

[27] Houssem Ben Salem, "Supervised VS Unsupervised VS Reinforcement learning." [Online]. Available: https://medium.com/@bensalemh300/supervised-vs-unsupervised-vs-reinforcement-learning-a3e7bcf1dd23

[28] N. Hordri, S. Yuhaniz, and S. M. Shamsuddin, *Deep Learning and Its Applications: A Review*. 2016.

[29] IAN Editorial Team, "Deep learning implementation in enterprise." [Online]. Available: https://www.intelligentautomation.network/intelligent-automation-ia-rpa/articles/deep-learning-implementation-in-enterprise

[30] S. Aljahdali and S. N. Hussain, "Comparative prediction performance with support vector machine and random forest classification techniques," *Int. J. Comput. Appl.*, vol. 69, no. 11, 2013.

[31] D. Tewari, U. D. Dwivedi, and S. Biswas, “A Novel Application of Ensemble Methods with Data Resampling Techniques for Drill Bit Selection in Oil and Gas Industry,” *Energies*, Jan. 2021, doi: 10.3390/en14020432.

[32] R. Yang, P. Wang, and J. Qi, “A novel SSA-CatBoost machine learning model for credit rating,” *J. Intell. Fuzzy Syst.*, vol. 44, pp. 1–16, Oct. 2022, doi: 10.3233/JIFS-221652.

[33] S. Paul, P. Das, A. Kashem, and N. Islam, “Sustainable of rice husk ash concrete compressive strength prediction utilizing artificial intelligence techniques,” *Asian J. Civ. Eng.*, vol. 25, pp. 1–16, Aug. 2023, doi: 10.1007/s42107-023-00847-3.

[34] W. Wang, X. Zuo, and D. Han, “Predict credit risk with XGBoost,” *Appl. Comput. Eng.*, vol. 74, pp. 164–177, Jul. 2024, doi: 10.54254/2755-2721/74/20240462.

[35] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, and D. Gao, “Degradation state recognition of piston pump based on ICEEMDAN and XGBoost,” *Appl. Sci.*, vol. 10, p. 6593, Sep. 2020, doi: 10.3390/app10186593.

[36] B. Boyle and T. Serdar, *Support Vector Machines: Data Analysis, Machine Learning and Applications*. 2011.

[37] I. Setiawan *et al.*, “Credit Risk Management Prediction Using the Support Vector Machine (SVM) Algorithm,” 2022, pp. 195–206. doi: 10.2991/978-94-6463-084-8_18.

[38] K. P. Singh, N. Basant, and S. Gupta, “Support vector machines in water quality management,” *Anal. Chim. Acta*, vol. 703, no. 2, pp. 152–162, 2011, doi: 10.1016/j.aca.2011.07.027.

[39] S. Jomthanachai, W. P. Wong, and K. W. Khaw, “An application of machine learning regression to feature selection: a study of logistics performance and economic attribute,” *Neural Comput. Appl.*, vol. 34, no. 18, pp. 15781–15805, 2022, doi: 10.1007/s00521-022-07266-6.

[40] “ML | Extra Tree Classifier for Feature Selection.” [Online]. Available: https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-selection/

[41] N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, *Filter Methods for*

Feature Selection – A Comparative Study. 2007. doi: 10.1007/978-3-540-77226-2_19.

[42] A. Peña Yañez, “El anillo esofágico inferior.,” *Rev. Esp. Enferm. Apar. Dig.*, vol. 26, no. 4, pp. 505–516, 1967.

[43] C. Lev and T. Linda, “Application areas of Machine learning.” [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

[44] T. Melese, T. Berhane, A. Mohammed, and A. Walelgn, “Credit-Risk Prediction Model Using Hybrid Deep—Machine-Learning Based Algorithms,” *Sci. Program.*, vol. 2023, pp. 1–13, 2023, doi: 10.1155/2023/6675425.

[45] G. Alagöz and E. Çanakoğlu, “Credit Risk Analysis Using Machine-Learning Algorithms,” in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, 2021, pp. 1–4. doi: 10.1109/SIU53274.2021.9477873.

[46] P. Addo *et al.*, “Credit Risk Analysis using Machine and Deep Learning models HAL Id : halshs-01719983 Centre d ’ Economie de la Sorbonne Documents de Travail du,” 2018.

[47] Q. Shan and M. Nilsson, “Credit risk analysis with machine learning techniques in peer-to-peer lending market,” p. 40, 2018.

[48] V. Kanaparthi, “Credit Risk Prediction using Ensemble Machine Learning Algorithms,” *6th Int. Conf. Inven. Comput. Technol. ICICT 2023 - Proc.*, vol. 147, no. 1, pp. 41–47, 2023, doi: 10.1109/ICICT57646.2023.10134486.

[49] I. Emmanuel, Y. Sun, and Z. Wang, “A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method,” *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00882-0.

[50] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, “Survey:Time-series data preprocessing: A survey and an empirical analysis,” *J. Eng. Res.*, 2024, doi: https://doi.org/10.1016/j.jer.2024.02.018.

[51] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[52] D. Elreedy and A. F. Atiya, “A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance,” *Inf. Sci. (Ny).*, vol. 505, pp. 32–64, 2019.

[53] Tony Yiu, “Understanding Random Forest.” [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[54] “XGBoost.” [Online]. Available: https://www.geeksforgeeks.org/xgboost/

[55] S. Ray, K. Alshouiliy, and D. Agrawal, “Dimensionality Reduction for Human Activity Recognition Using Google Colab,” *Information*, vol. 12, p. 6, Dec. 2020, doi: 10.3390/info12010006.

[56] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian optimization of machine learning algorithms,” *Adv. Neural Inf. Process. Syst.*, vol. 4, pp. 2951–2959, 2012.

[57] K. Khadka, J. Chandrasekaran, Y. Lei, R. N. Kacker, and D. R. Kuhn, “A Combinatorial approach to hyperparameter optimization,” *Proc. - 2024 IEEE/ACM 3rd Int. Conf. AI Eng. - Softw. Eng. AI, CAIN 2024*, pp. 140–149, 2024, doi: 10.1145/3644815.3644941.

[58] O. F.Y, A. J.E.T, A. O, H. J. O, O. O, and A. J, “Supervised Machine Learning Algorithms: Classification and Comparison,” *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, 2017, doi: 10.14445/22312803/ijctt-v48p126.

[59] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.