INFLUENCE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING APPROACH ON OPTIMIZING BANK LENDING DECISIONS

by

MILIND KADAM,
Master in Science – Physics (Madurai Kamraj University, India)
&
Master in Business Administration – Operations Management (Southern New Hampshire University, USA)

DISSERTATION
Presented in fulfillment of requirements for the Degree of

DOCTOR OF BUSINESS ADMINISTRATION

from

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

Year 2024

INFLUENCE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING APPROACH ON OPTIMIZING BANK LENDING DECISIONS

by

MILIND KADAM

Mentored by

Dr. Mario Silić

APPROVED BY

Dr. Iva Buljubasic
Chairperson

SSBM Representative

ABSTRACT

INFLUENCE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING APPROACH ON OPTIMIZING BANK LENDING DECISIONS

MILIND KADAM
2024

Dissertation Chair: Dr. Iva Buljubasic
Co-Chair: Dr. Anna Provodnikova

Much inefficiency in traditional bank lending processes characterizes the modern financial landscape, hence a main goal should be a critical analysis of these shortcomings. Previous studies have pointed out a number of drawbacks, including inefficiencies in terms of time, high expenses, and problems with flexibility, inconsistent results, poor data identification, and absence of real-time processing. These elements increase the risk of credit flow and lead to liquidity mismatches, issues with repayment capacity, and an overemphasis on asset seizure. Furthermore, because the data is vast and diverse, it is difficult to locate and resolve these discrepancies inside financial networks. In response; this study promotes the use of machine learning and artificial intelligence-based statistical techniques to address current limitations, providing improved speed and accuracy in bank lending choices. Using machine learning algorithms, the research employs a thorough methodology to statistically analyze the banking behaviors of its customers. A comprehensive analysis is conducted on banking parameters like credit score, debt-to-income ratio, credit duration, repayment history, bank account history, loan amount, and collateral utilizing information from trustable public domain banking datasets. The current study suggests the Hybrid Random Forest based Grey Wolf Algorithm (RF-GWA) for processing bank loan data sets most quickly and accurately in terms of predictions. When it comes to processing bulk baking data sets and making lending recommendations, RF-GWA performed better than the current algorithms Artificial Neural Network (ANN), Deep Neural Network (DNN), Convolution Neural Network (CNN), and non-hybrid Random Forest Method (RFM). It achieved close to 97% accuracy, over 94% precision, over 98% recall, and over 95% cross validation.

Table of contents

LIST OF TABLES

LIST OF FIGURES

List of abbreviations

Abbreviations	Full form
PD	Probability of Default
ML	Machine Learning
AI	Artificial Intelligence
RF	Random Forest
KNN	K Nearest Neighbor
CNN	Convolutional Neural Network
DNN	Deep Neural Network
WSO	White Shark Optimization
BL	Bank Lending
DEA	Data Envelopment Analysis
US	United States
ABMs	Agent-Based Models
FNN	Fuzzy Neural Network
IFBWM	Intuitionistic Fuzzy Best-Worst Method
DEMATEL	Decision-making Trial and Evaluation Laboratory
BP	Back Propagation
SEM-ANN	Structural Equation Model and Artificial Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
DM–ACME	Distance-to-Model and Adaptive Clustering-Based Multi-View Ensemble
P2P	Peer-to-Peer
XGBoost	extreme gradient boosting
LR	LogisticR
OCHE	Overfitting-Cautious Heterogeneous Ensemble
SVM	Support Vector Machine
NN	Neural Network

CHAPTER: 1

INTRODUCTION

1.1 Bank lending

The process through which a bank offers credit to a candidate is known as lending. Often referred to as a lender, the bank usually gets paid interest for the loan. Lending in banking creates similar profits for both lenders and borrowers by increasing liquidity in the markets where loans are created and used. The general structure of a bank lending is illustrated in Figure 1.1.

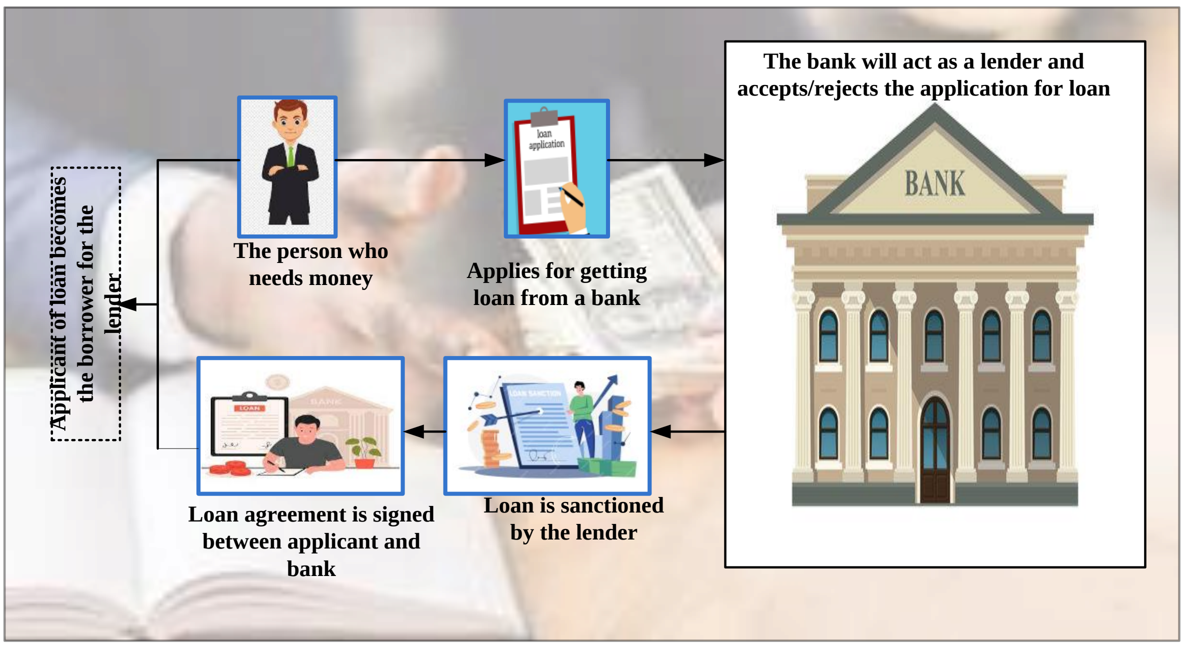


Figure 1.1: Architecture of bank lending, (Chulawate, 2023)

Commercial banks have adhered to three fundamental principles in lending for an extended period. These principles, namely safety, liquidity, and profitability, are integral in guiding the utilization of funds. Nationalized banks in India, like their counterparts, diligently uphold these principles in the deployment of their resources. For lending institutions, the credit scoring systems aims to offer a Probability of Default (PD) to make a minimal loss principle for their clients, (Ali, 2021; Al Ayub Ahmed, 2021)

So, the credit scoring system helps decision making for credit risks, credit applications, influences and manages the number of non-performing loans that are probable to lead to financial crisis, bankruptcy and environment sustainability. Lending is the mechanism through which a financial institution extends funds to a borrower, commonly known as a lender. In exchange for the loan, the institution typically earns interest. These offerings make up the joint business of liquidity provision for the customers: the bank provides liquidity on demand by meeting the customers' liquidity shortfalls through loan extensions and by being prepared to pay interest on its excess liquidity.

The bank's core operations do not involve the independent operations of taking deposits from and lending money to its clients. The bank receives a plethora of customer data in addition to the payment processing function, which it can use repeatedly for lending decisions that demand a lot of data as well as other products that require a lot of data, such as standby letters of credit, bank guarantees, and credit information sales. This practice, prevalent in banking, serves to enhance liquidity in the markets where loans are generated and utilized, is benefiting both lenders and borrowers, (Abd Rabuh, 2023; Agostino, 2023). Figure 1.2 shows the Bank lending cycle

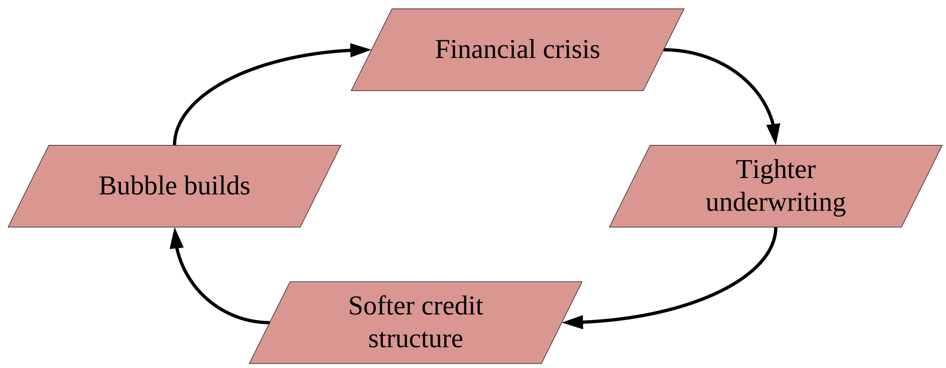


Figure 1.2: Bank lending cycle, (Akinboade, 2010)

In recent times, a growing fascination has emerged regarding the global dimensions of corporate finance. Noticeable distinctions in the financial frameworks and funding strategies employed by corporations worldwide, especially in emerging markets, have captured considerable attention. Contemporary research posits that the majority of these variations find their roots in the diverse legal and institutional landscapes of countries, as well as disparities in their economic foundations and other resources. Online lending has gained popularity due to its heightened efficiency in providing credit to both small businesses and consumers. In the past decade, rule-based credit scoring models have been pivotal in determining borrowers' creditworthiness, evolving over time with technological advancements. Credit risk analysis relies on statistical models such as probit regression, logistic regression, cox survival models, and discriminant analysis. These techniques offer improved performance, avoid computational challenges, and are easily understandable.

Contrastingly, Machine Learning (ML) techniques often yield superior predictive outcomes as they can identify more complex risk patterns, (Agarwal, 2021). However, most ML methods are perceived as black boxes, lacking transparency and making it challenging to comprehend their decision-making processes. Bank loan interest rates that are lower allow firms to reinvest their excess cash flow back into their operations, which boosts economic growth. For consumers, this translates to cheaper borrowing rates for financing purchases like homes, vehicles, and schooling. Many different bilateral credit products that private or public banking companies and organizations offer are referred to as lending products, (An, 2021). Banks are navigating with caution in adopting ML for credit risk modeling. Moreover, regulatory requirements are typically aligned with traditional methods, not fully addressing the complexities associated with these innovative alternatives. The accompanying figure illustrates the bank lending cycle. This dynamic landscape reflects the ongoing evolution of credit assessment methodologies in response to technological advancements and changing financial landscapes. The nature of the most recent US financial crisis has raised questions about banks' capacity to stick to their established lending practices.

The real economy is primarily impacted by a banking crisis when it comes to its capacity to extend credit given the limitations placed on it. A crucial question at the heart of every financial crisis is therefore whether and how the banking sector was able to distribute the limited credit available in a way that maximizes their profits throughout the crisis. As a result, it's imperative to have an efficient system for making bank loan decisions that will promptly optimize bank profit. A credit crunch could arise from banks' incapacity to effectively manage their lending portfolio. A prolonged period of irresponsible and reckless lending is frequently the root cause of a credit crunch, which costs lending institutions and debt-ridden investors money when the loans default and the full amount of bad debts is discovered. Due to these difficulties, lending decisions are being made more carefully and formally in an effort to reduce loan risks. Financial institutions now use bank lending decisions as their main instrument for managing risks, increasing profit, and lowering potential losses, (Ansari, 2020; Alarfaj, 2022).

The robustness of an economy's financial system determines its rate of growth. Since the country's independence, banking institutions in India have been crucial to the structural change of the economy. Without regard to profit, public sector banks initially played a key role in the development of the Indian financial system as well as acting as a catalyst. The survival of public sector banks in India was threatened, nevertheless, by the implementation of financial sector reforms and the following entry of foreign and new generation private sector banks. As a result, the banking industry has recently introduced cutting-edge technologies and creative consumer services. Bank loan interest rates that are lower allow firms to reinvest their excess cash flow back into their operations, which boosts economic growth. For consumers, this translates to cheaper borrowing rates for financing purchases like homes, vehicles, and schooling.

The post reform measures created an improvement in the financial performance and profitability of majority of the Indian banks. With regard to the prospects of Indian banks, it is clear that Indian financial sector is stable and healthy. Commercial banks' capital adequacy and return on assets, two measures of financial health, are still strong, (Alfonso-Sánchez, 2023; Momani, 2021). Among the banking services, giving credit or lending money is more important since it increases economic investments, which are necessary for a nation's progress economically. One of the services banks provide to their clients has been identified as lending. There are those who believe that lending is the core activity of banks. There are three possible loan durations: short-term, medium-term, and long-term. Since finance is crucial to every economic area, the flow of credit to all sectors is necessary for the economy to grow efficiently. A sector's ability to access more credit resources increases with its rate of development.

Credits have a significant impact on the economy because without more financial resources, economic expansion becomes difficult. Since a credit indicates the mobilization of deposits and the subsequent investment of these financial resources in profitable sectors through credit availability, it is seen as a relevant indicator of financial development. It guides the economy's savings and investment flow to promote capital accumulation and production. Term loans, commercial mortgages, and credit lines are the three most popular kinds of business loans. A common feature of commercial loans is their security, or the tangible collateral that supports them.

The bank meets the clients' liquidity needs by extending loans and is willing to pay interest on any extra liquidity it may have. Together, these services comprise the joint business of providing liquidity for the customers. The bank's core operations do not involve the independent operations of taking deposits from and lending money to its clients. The bank receives a plethora of customer data in addition to the payment processing function, which it can use repeatedly for lending decisions that demand a lot of data as well as other information-intensive products like bank guarantees, credit information sales, and standby letters of credit, (Alonso, 2022; Alonso, 2020). This study suggests a novel method for bank financing. Initially, Kaggle (an online resource of open-source datasets), is used to gather input data for banking prediction. The pertinent characteristics are then chosen from the supplied data. The RF classifier assists in determining whether to approve or refuse a loan based on the relevant features that have been identified. The WSO is used to maximize RF performance. This contributes to improving accuracy and reducing the rate of mistakes. 1.2 Lending process steps

The steps involved in lending process are;

- Finding prospective loan applicants
- Evaluating the applicant’s sincerity of purpose and character
- Evaluating the applicant’s credit record and making cite visits
- Evaluating the financial condition of applicant
- Signing the loan agreement and assessing possible loan collateral
- Monitoring compliance with the loan agreement and other customer need services

1.3 Principles of bank lending

The principles that have been followed by the commercial banks since long are illustrated in Figure 1.3.

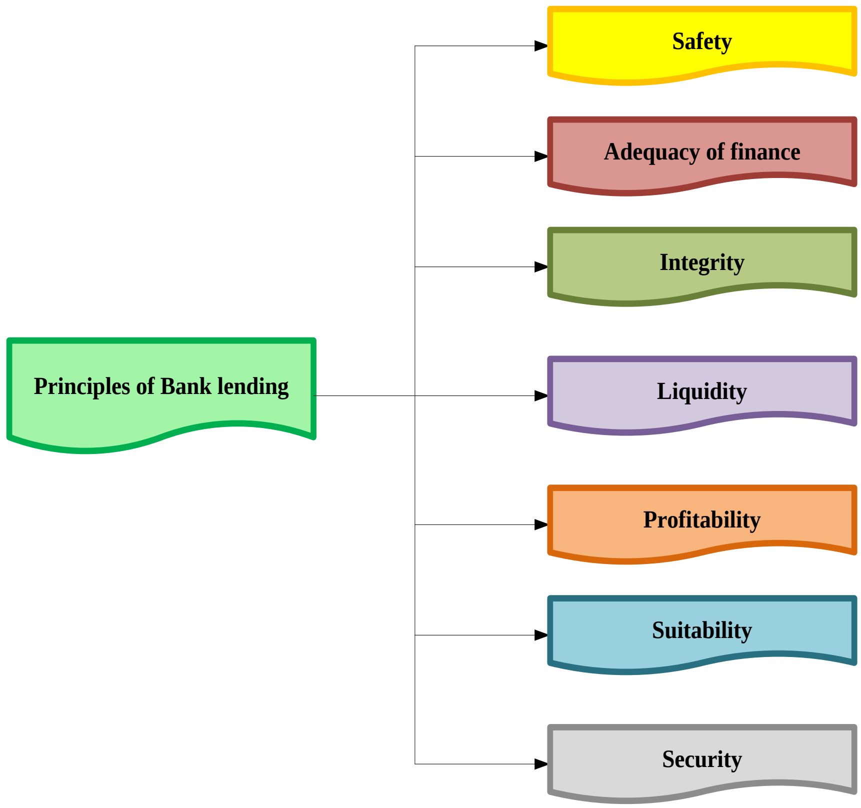


Figure 1.3: Principles of bank lending, (Choudhry, 2022)

1.4 Credit scoring

Over the past few decades, there has been a notable evolution in the field of credit assessment, marked by the development of quantitative methodologies commonly referred to as credit scoring models. These models aim to facilitate credit granting decisions by categorizing credit applicants into distinct groups: a 'good credit' group, demonstrating a high likelihood of fulfilling their financial obligations, and a 'bad credit' group, indicating a heightened risk of defaulting on financial commitments. The initial credit scoring model, and one still widely utilized today, is linear discriminate analysis a straightforward parametric statistical approach. Given the expansive growth of the credit industry and the substantial loan portfolios managed in contemporary times, there is a concerted effort to enhance the precision of credit scoring models, (Antunes, 2021; Aphale, 2021). The system model of credit scoring is illustrated in Figure 1.4.

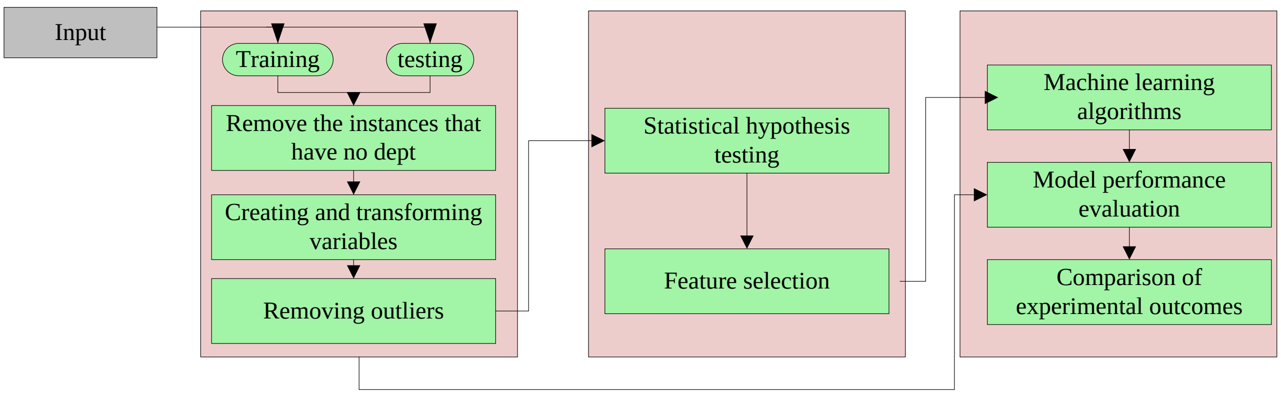


Figure 1.4: System model for credit scoring, (Tripathi, 2020)

Even marginal improvements, such as a fractional percentage increase in accuracy, are considered significant achievements. This endeavor has prompted active exploration into advanced statistical techniques, including nonparametric methods, classification trees, and neural network technology, with the goal of developing more refined and accurate credit scoring models.

1.4.1 Commercial credit

Worldwide, commercial credit contains four various types of loans:

- **(i)** Cash flow loans
- **(ii)** Asset based loans
- **(iii)** Leases and
- **(iv)** Trade finance agreements

The categorization of these loans as secured and senior is widely acknowledged in practical contexts but has only been partially addressed in academic research. The distinctive features of collateral lie in its pledgeability, liquidation value, and durability, forming the core elements influencing the existence of diverse commercial credit types. While certain variations may not stem from inherent characteristics of the physical assets serving as collateral, differences in collateral repossession, such as those observed in asset-based loans versus leasing, contribute to this diversity.

1.5 Importance of artificial intelligence in lending

Financial institutions can substantially boost both operational efficiency and customer satisfaction by strategically prioritizing the implementation of AI/ML models in areas where they can yield the maximum benefits. Through this approach, these institutions stand to automate more than 20 decision points within diverse customer journeys. Particularly within the lending process, forward-thinking banks are increasingly leveraging the capabilities of AI and analytics to generate value in five pivotal areas: customer acquisition, credit assessment, monitoring and collections, fortifying customer relationships, and delivering intelligent servicing. Figure 1.5 shows the advantages of bank lending, (Awotunde, 2021; Aziz, 2022).

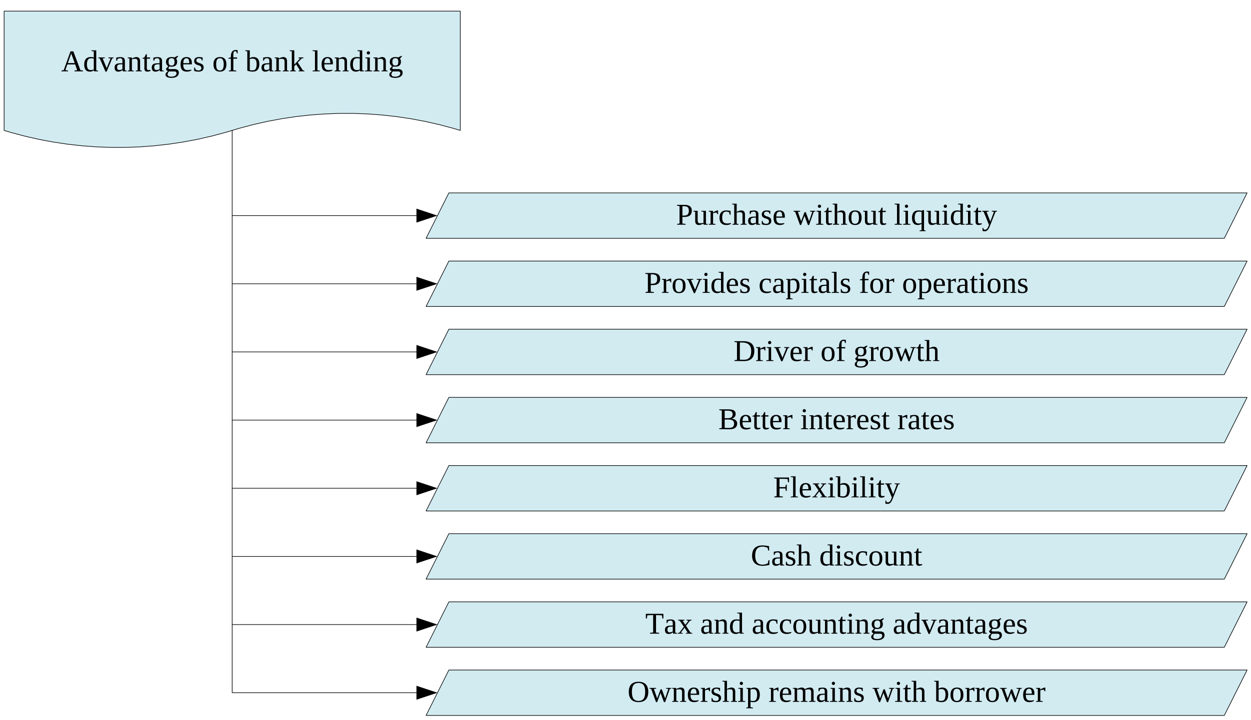


Figure 1.5: Advantages of bank lending, (Rappoport, 2014)

1.6 Interest Rate Risk impact in Bank Lending

Research findings recommend that as nominal interest rates decline, banks tend to alter their asset portfolios toward longer maturities to stop a significant diminish in overall portfolio yield. The theoretical literature on the transmission of monetary policy indicates that interest rate risk exposure intensifies the sensitivity of bank lending to changes in nominal interest rates. Accordingly, the bank may diminish lending to uphold compliance through capital needs forced by regulators or market member. Empirically evaluate the impact of realized interest rate risk contact on bank lending faces challenges on two fronts. First off, important data regarding balance sheet positions that are susceptible to interest rates and the accompanying maturities of the reprisals are frequently missing. Second, it is usually not possible for the public to obtain precise information about hedging positions against interest rate risk. As a result, using publicly available data to create a trustworthy estimate of an individual's interest rate risk exposure, net of hedging, continues to be difficult.

In addition to these challenges, it is important to consider the broader economic implications of banks adjusting their lending practices in response to interest rate risk. Such adjustments may have cascading effects on economic activity, affecting businesses and consumers. Moreover, the interplay between monetary policy, interest rate risk, and bank lending adds complexity to the overall understanding of the financial system's resilience to interest rate fluctuations. As financial markets continue to evolve, addressing these challenges and gaining a comprehensive understanding of the dynamics between interest rates, bank behavior, and economic outcomes becomes increasingly crucial for policymakers and industry stakeholders alike, (Babo, 2023; Barbaglia 2023).

1.7 Banks and commercial lending relationships

This analysis critically underscores the evaluation of the influence exerted by asset purchases on tangible economic activities within businesses, with a particular focus on the bank lending channel. The temporal framework for these associations is demarcated as follows: commencement transpires at the onset of the initial documented loan origination involving the company and the bank. The culmination of the relationship transpires upon the fulfillment of the last documented loan between the identical company and bank, in accordance with the original stipulations of the loan agreement. Furthermore, it is imperative to delve deeper into the intricate dynamics of these relationships, scrutinizing how asset purchases impact not only the initiation and duration of the interactions but also the subsequent implications for broader economic activities within firms. This multifaceted exploration aims to unveil the nuanced interplay between asset acquisitions, bank lending practices, and their ramifications on the overall economic landscape.

1.8 Bank lending channel

Extensive research associate’s financial crises with the simultaneous decline of credit availability and asset prices. The central mechanism connecting these financial shocks to the tangible economy is known as the bank-lending channel. This theory posits that disruptions in the financial sector reverberate into the real economy through alterations in banks' credit supply. Figure 1.6 shows the bank lending channel of monetary policy.

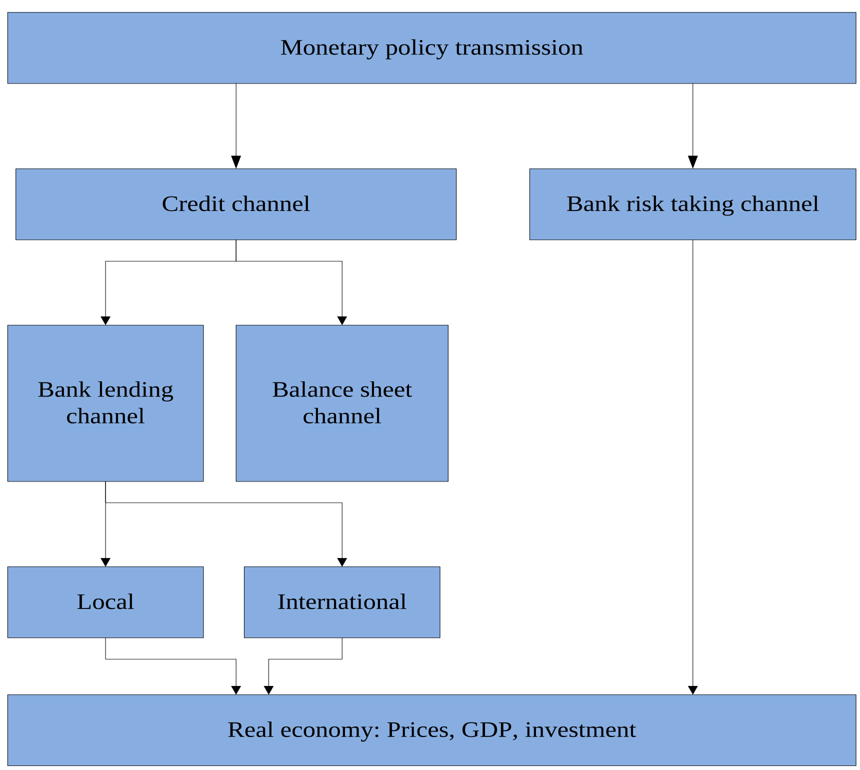


Figure 1.6: The bank lending channel of monetary policy, (warkasing, 2016)

1.9 Loan rejection status from 2013 to 2023

The Figure 1.7 illustrates the loan rejection rates spanning the years 2013 to 2023. Notably, the data reveals a substantial increase in loan rejections in the year 2023 when compared to previous years.

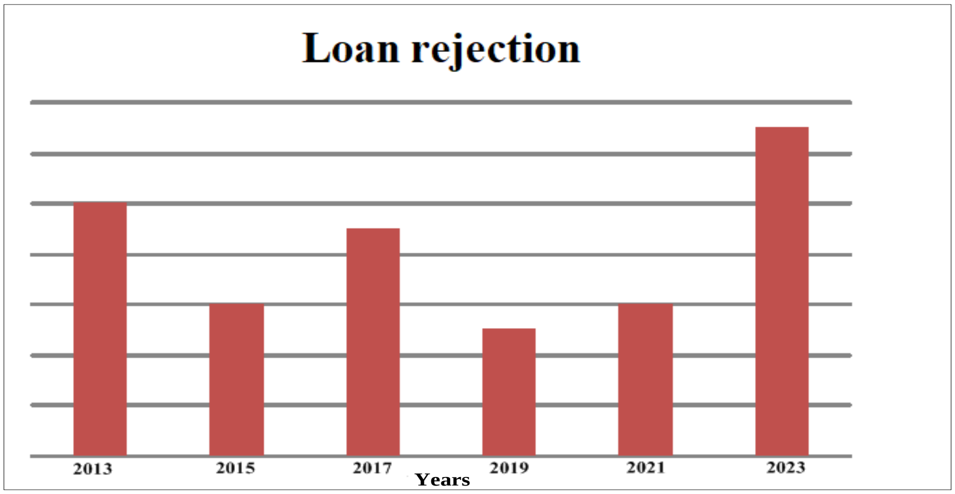


Figure 1.7: Loan rejection rate, (Kwofie, 2015)

1.10 Public bank lending in time of crisis

Amidst the unfolding of the global financial crisis, major banks found themselves grappling with substantial strains on their balance sheets. The resultant significant write-downs on assets led to pronounced reductions in bank capital, triggering a credit crunch in select nations. As banks scaled back on lending, businesses encountered formidable challenges in securing external financing. Concerns are mounting that these adverse financial conditions may endure, potentially impeding economic activity evoking memories of the impediments observed when insufficient bank capital hindered the recovery from the U.S. recession in the early nineties. In order to comprehensively examine the distinct features that shape the lending practices of individual banks amidst such challenges, it is imperative to adopt a specification akin to methodologies employed in prior studies on the bank lending channel. This model not only incorporates country- and time-fixed effects to account for overall economic conditions but also delves into bank-specific factors that have previously been identified as significant determinants of lending behavior. It is crucial to recognize financial crises as systemic and exceptional occurrences, necessitating a nuanced understanding of the intricate interplay between macroeconomic conditions, institutional factors, and the lending dynamics of banks.

1.11 Government ownership's effect on bank lending

Publicly owned banks consistently provide loans to similar or identical businesses at diminished interest rates compared to privately owned banks, even when the latter can present higher borrowing capacities. State-controlled financial institutions display a preference for endorsing larger enterprises and those situated in economically challenged areas. Additionally, the lending strategies of state-owned banks are subject to the electoral fortunes of the political party associated with the bank. In locales where the affiliated party holds greater political clout, the interest rates imposed tend to be more lenient, illustrating a nuanced interplay between political dynamics and lending practices.

1.12 Impact of covid-19 in bank lending

High pre-shock levels of bank capital, along with an infusion of cash from liquidity injection depositors and programs, allowed banks to meet the spike in demand for liquidity during the early stages of the coronavirus pandemic. On the other hand, a significant decline in new loans to major borrowers occurred during the Global Financial Crisis. Nevertheless, survey data indicates that banks globally are expressing concerns about deteriorating industry-specific issues, tightening lending criteria, decreasing risk tolerance, and other unclear economic outlooks. Despite the fact that bank stocks and other non-bank financial companies have underperformed in their home markets, a number of banks have chosen not to participate in the lending stimulus initiative.

The existence of other contemporaneous confounding variables throughout the epidemic, such as the introduction of government initiatives intended to counteract the economic impacts of COVID-19, presents another empirical issue. The majority of these metrics have the potential to affect credit decisions. Vaccines were not yet accessible in Brazil, the extent of the crisis was unknown, and economic agents had not yet adjusted to the new economic circumstances brought forth by COVID-19.

1.13 Lending ratio

Lending Ratios are constraints financial institutions and banks consider while dealing out a loan application. The motive is to check the customer will be a responsible borrower. When the evaluation is positive, it results in a loan application being approved, (Chen, 2021; Chen, 2021). Figure 1.8 shows the lending ratio process.

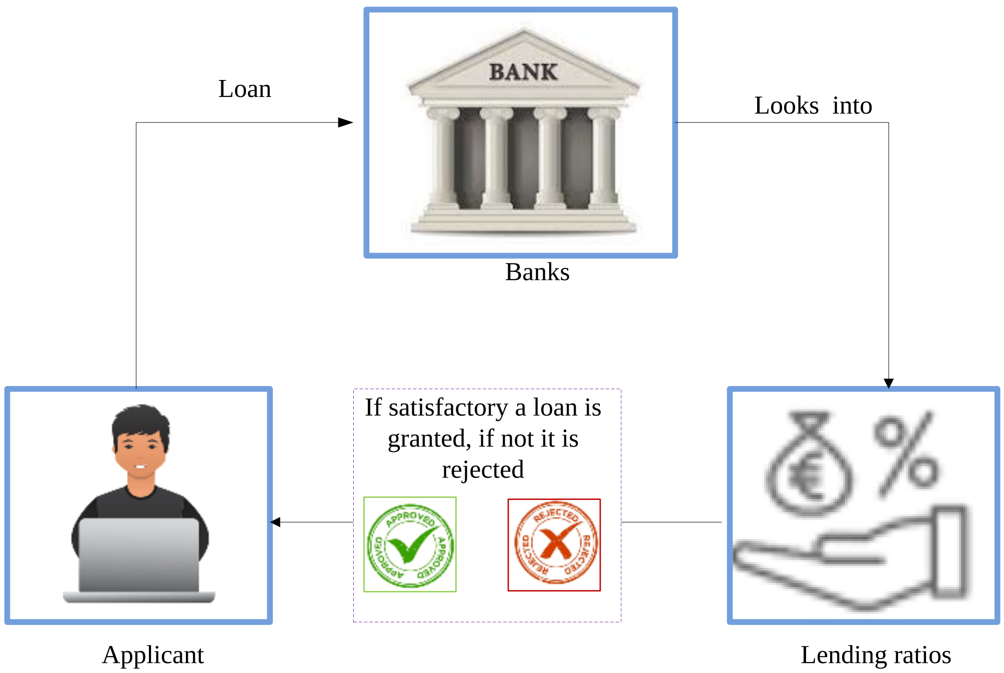


Figure 1.8: Lending ratio, (Cheng, 2022)

1.14 Advantages of ML based approaches

The ML based approaches offer several advantages such as:

- **Improved accuracy:** The ML based techniques aids to analyze complex and large datasets more capably than conventional approaches, leading to more precise predictions. They can recognize patterns and relationships in data that may not be obvious from manual analysis.
- **Automation and Efficiency:** The ML based approaches automate the lending process, reducing the need for manual intervention. This leads to

increased efficiency, faster decision-making, and a streamlined loan approval process.

- **Reduced Bias**: ML models, if designed and implemented properly, can help reduce human biases in lending decisions. By focusing on data-driven factors and minimizing subjective judgment, ML models contribute to more objective and fair lending practices.

1.16 A brief explanation about ML Classifiers

Several ML based classifiers are utilized for classification and prediction process. Among those some classifiers are explained below.

1.16.1 KNN

Closeness is used by k-NN, a non-parametric supervised learning classifier, to identify or predict the grouping of a single data point.

1.16.2 Logistic Regression

Logistic regression is a technique for estimating the likelihood of a discrete outcome given an input variable. Most logistic regression models reflect a binary result, which is something that can have two values, such as true or false, yes or no, and so on.

1.16.3 Decision tree:

Decision trees are a non-parametric supervised learning technique utilized in both regression and classification applications. It is a hierarchical tree with a root node, branches, internal nodes, and leaf nodes.

1.16.4 Random Forest

The popular machine learning method known as "random forest," which combines the output of multiple decision trees to generate a single result, is patented by Leo Breiman and Adele Cutler. Its popularity has been fueled by its adaptability, simplicity of usage, and capacity to handle both regression and classification issues.

1.17 Input sources

Collecting accurate and reliable data is important for bank lending. The input data is collected from Banking - Prediction Model using RandomForestamd banking ML datasets sourced from Kaggle. Kaggle is an open-source plat form where numerous datasets were available for bank lending. Table 1.1, shows the detailed description of the dataset. The data is related to the direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls.

Table 1.1: Detailed description of the dataset

Bank Client data	
Age	The applicants age in numeric
Marital	The status of marital (Married, single, divorced etc.,)
Job	Type of the job
Default	Has credit in default
Education	Secondary, primary, tertiary
Balance	Average yearly balance
Loan	Has any personal loan
Contact	Contact communication type
Duration	Last contact duration
Month	Last contact month of year
Day	Last contact day of the month
pdays	Number of days that passed by after the client was last contacted from a previous campaign
Campaign	Number of contacts performed during this campaign and for this client
Poutcome	Outcome of the previous marketing campaign
Previous	Number of contacts performed before this campaign

1.18 Selection of ML model

The ML based works well with huge amounts of training data and is easy to implement. It also stands up well to noisy training data. Input training data is used by machine learning classification algorithms to predict the likelihood that incoming data will fit into a predetermined category. The Random Forest (RF) classifier is selected for bank lending. The RF is selected because it has several advantages. The few advantages of RF are:

- High accuracy
- Effectiveness is notable in larger datasets
- Offers an estimate of significant variables in classification
- Comparing with other models it does not over fit with more features

1.18.1 Hybridizing with optimization algorithm

An innovative method used to improve accuracy, the White Shark Optimization algorithm, greatly increases the RF's efficacy. By maximizing the scaling factor, this technique helps to achieve the ideal balance between the four critical characteristics of authenticity, security, imperceptibility, and robustness. Figure 1.9, shows the hunting behavior of white shark.

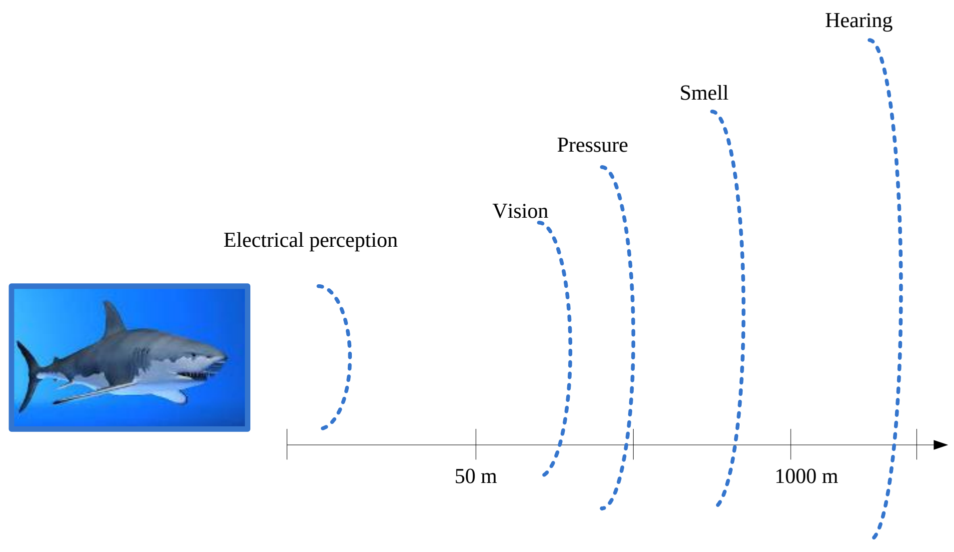


Figure 1.9: Hunting behavior of white shark, (Braik, 2022)

1.19 Problem statement

By reviewing recent ML based models for bank lending, several issues have been discerned. The introduced method addresses and resolves numerous challenges encountered by existing approaches in this domain. The issues are:

- Time consumption
- Higher lending cost
- Flexible terms
- Inconsistencies
- Lack of accuracy in results
- Risk in credit flow
- Limited data identification
- Liquidity mismatch

1.20 Motivation:

Several techniques were suggested for machine learning based bank lending but the existing approaches has several disadvantages. These issues motivated to do this research.

1.21 Challenges

Financial modeling and prediction encounter a number of difficulties, such as non-normal distributions of variables like interest rates, multicollinearity, and departures from the assumptions of linear regression. Furthermore, credit scoring becomes more complex due to the opacity of machine learning algorithms, making it more difficult for stakeholders to understand. Innovative strategies are needed to overcome these obstacles and guarantee the precision, consistency, and interpretability of financial analysis.

- ➢ Financial datasets often deviate from the assumptions underlying linear regression models, leading to suboptimal or inaccurate predictions. This deviation is quite common in practice.
- ➢ When data features are highly correlated, a phenomenon known as multicollinearity, it can result in instability in predictive models. This can further affect the quality of predictions.

- Linear regression models assume a normal distribution of residuals, but in many cases, the response variable, such as note rates in financial contexts, exhibits right-skewness with heavy tails. This departure from normality can impact the accuracy of predictions. Moreover, linear regression models are susceptible to outliers and may produce overestimated note rates for typical applicants when dealing with datasets containing such extreme values.
- Crisis indicators often provide warning signs only when it's already too late to take effective preventive measures, rendering timely intervention challenging. Translating complex early warning models into simple and transparent indicator00.0020s that can guide timely action by macro prudential authorities can be a demanding task.
- Nonetheless, a significant drawback of employing ML techniques within the credit scoring sector lies in their limited explainability and interpretability. Many of these algorithms, especially ensemble methods, are often regarded as "black boxes." In essence, this means that it's challenging to provide clear explanations to both customers and regulatory bodies about the resulting scorecards and credit approval procedures. This aligns with current apprehensions among financial regulators regarding the management of AI and the pressing necessity for interpretability, particularly in the credit scoring domain.

1.22 Thesis organization

This thesis is organized as follows:

Chapter 1:

Introduction: A brief introduction about bank lending, lending cycle, lending ratio, motivation, challenges etc.

Chapter 2:

Literature review: Numerous studies were suggested in the literature related to machine learning based bank lending process among those some recent studies are reviewed in this chapter number of recent researches related to machine learning based bank lending are reviewed.

Chapter 3:

Proposed methodology: The introduced new methodology is explained in this section.

Chapter 4:

Results: The experimental outcomes of introduced method is presented in this chapter

Chapter 5:

Discussions: The discussions of proposed method's outcomes comparing with existing approaches to demonstrate the proposed method is better than the existing approaches.

Chapter 6:

Conclusion and future scope: Conclusion of the proposed technique with future directions is given in this chapter. The pictorial representation of this organization is given in Figure 1.10;

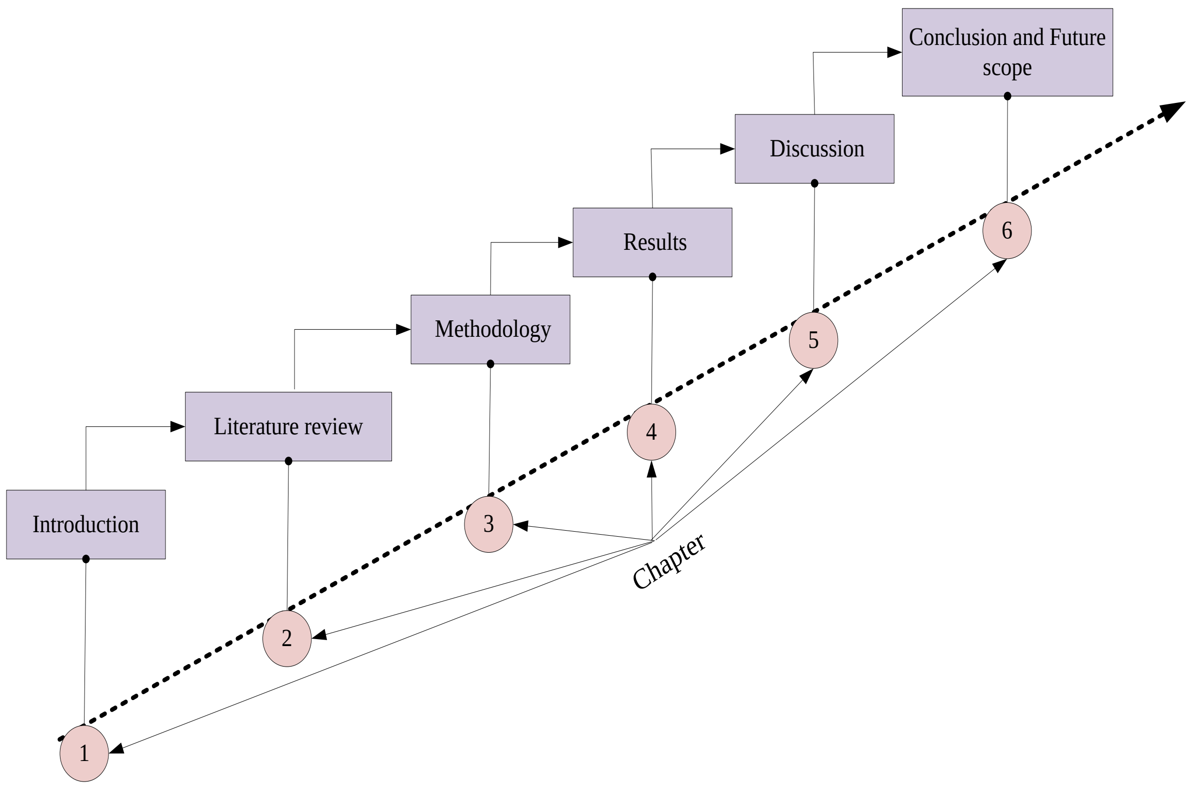


Figure 1.10: Thesis organization

CHAPTER: 2

LITERATURE SURVEY

2.1. Introduction

The banking system plays a pivotal role in advancing economic development by efficiently channeling crucial financial resources to businesses and households in developing nations through the mechanism of Bank Lending (BL). This process is integral to understanding the effects of monetary policy on the broader economic landscape. In this chapter, a comprehensive review of literature is conducted, spanning diverse dimensions of BL research. One focal point of examination is the literature on bank efficiency, where scholars have probed into the effectiveness and productivity of banking operations. Assessing bank performance is another critical area, exploring the financial institutions contribute to economic growth and stability. The chapter delves into the intricate web of bank regulations, shedding light on the regulatory frameworks that govern banking activities and shape the financial sector. Operational Research (OR) emerges as a crucial analytical tool within the context of Bank Lending. The review explores the role of OR methodologies in optimizing lending processes, enhancing decision-making, and improving overall efficiency. Data Envelopment Analysis (DEA) is scrutinized as a specific OR technique, offering insights into its application for evaluating the efficiency of banks in deploying their financial resources. Agent-based methods, another facet of investigation, are employed to model complex interactions within the banking system. Fuzzy logic, a form of reasoning under uncertainty, finds application in handling imprecise information relevant to bank lending scenarios. The integration of neural networks and ensemble methods is also explored, showcasing the utilization of advanced artificial intelligence techniques for predictive modelling in the domain of Bank Lending. This chapter serves as a comprehensive exploration of the multifaceted aspects of Bank Lending, providing a synthesis of literature that spans efficiency assessments, performance evaluations, regulatory considerations, and the integration of diverse methodologies such as OR and Artificial Intelligence (AI).

2.2. Bank Efficiency

Cost-effectiveness and quality enhancement strategies explored and implemented by Italian cooperative banks, (Agostino, 2023). Spanning from 2006 to 2015, the study employed Stochastic Frontier Analysis to discern optimal solutions. A primary focus of the research was to manage and control corruption while simultaneously improving the overall efficiency of bank operations. By applying rigorous analytical methods over the specified timeframe, the study sought to identify and implement measures that would contribute to the economic viability and enhanced performance of Italian cooperative banks.

To analyze the regional disparities impact, (Chen, 2021) have presented a case study based in the city commercial banks located at China in the year of 2005 to 2014. The impact analyses the efficiency, profit and cost. Thus, it achieves the best and positive result in terms of bank efficiency of Chinese City Commercial banks but it was not negatively interconnected with the urban population ratio.

Risk management in the finance sector had been performed by (Rampini, 2020) specifically focusing on foreign exchange risk and interest rate hedging. Their investigation aimed at achieving optimal results by controlling risk exposures effectively. The study revealed that loan losses occurred due to a decline in house prices, primarily leading to net worth shocks. This underscores the importance of strategies like foreign exchange risk and interest rate hedging in mitigating risks and enhancing the resilience of financial institutions in the face of market fluctuations and economic challenges.

2.3. Bank Performance

The performance of loan conditions have investigated by (Wellalage, 2021) with the help of 3915 sample files. The analysis finds the best performance in terms of environmental aspects for effectively controlling the sample selection bias and endogeneity. Thus, it achieves 6.4% higher loans which was more optimal in medium as well as small firms. The above study mainly applicable in the sector i.e. asymmetry view of agency costs.

The efficacy of the banking sector in view of practical and financial institutions have introduced by (Bassett, 2020). The study mainly deals with comprehensive datasets which comprised of five programmes as well as United States (US) based banks. The main aim of the above study to analyze and asses the support in view of lending activities within US banks.

Risk and profitability analysis can be done to predict the effect of the green lending propensity in banking sector, (Del Gaudio, 2022). In the study, a green lending approach was created with 217 green facilities financing syndication worldwide samples. Two findings were introduced such as: extreme propensity to green lending and larger syndicate size reduces profitability and risk.

2.4. Banking Regulations

Lending reduction in the foreign bank mainly in the Asian and Latin American countries are investigated by (Hsieh, 2020). It was mainly investigated in the year of 1987-2013 with 1,558 individual bank data. In the period of Latin America, increased lending rate mainly supports the borrowers which mainly stimulate the lending process for the foreign banks. Higher level of government-owned assets produces greater as well as less effect on a cut in lending. Puri (2023) have elucidated a study focused on estimating customer satisfaction levels and evaluating services within the banking sector. The research comprised two main sections: (a) demographic- and services-related questions and (b) assessing the satisfaction levels of customers. The analysis specifically targeted 10 customers, primarily centered around the operations of the State Bank of India, particularly at the SBI Branch in Gurgaon, Haryana. The investigation aimed to provide insights into the factors influencing customer satisfaction and perceptions of services in the context of a specific banking institution and its branch, shedding light on the dynamics crucial for enhancing overall service quality and customer experience.

While, (Halvorsen, 2023) have introduced the employees' performance for effectively analyzing the ethical leadership and training to customer needs. Relationships between ethical leadership and training on employee performance, mediates the service climate and customer orientation which was examined from the social learning platform. In the Australian bank, 187 employees' samples were collected for the analysis of customer orientation and ethical climate to improve the performance of the employees.

The bank ability to optimize the small and medium enterprises (SMEs) on the impact analysis of financial technology (fintech) are investigated by, (Sheng, 2021). In the year of 2011 to 2018, the lending records of banks in China was analyzed which mainly accommodates the banking sector's credit supply especially for SMEs. Based on the bank size, the fintech impact was varied.

2.5. Operational Research (OR) Methods

Interbank lending network to consider the losses evolved in the individual banking sectors, (Liu, 2020). Generally, in the study, 6600 banks' decision rules and behaviors were presented for the agent-based model to reframe the banking network. Results of a traditional stationary network framework for contagion was compared with other models. In the year of 2007 to 2009 to reproduce the dynamics behavior of the banking sector. This shows the banking sector failures and losses when the network was in the lending market illiquidity and network contagion. Ben Bouheni (2022) have demonstrated largest Islamic banks in the United Kingdom and Turkey from 2010 to 2019 to analyse the effect of past returns, economic cycle and variations in loans. The analysis and findings show the increased return assets and decreases equity as well as inverse scenario. The countercyclical was considered as the credit risk of Islamic banks.

2.6. Data Envelopment Analysis

Bank efficiency is investigated using fuzzy multi-objective two-stage data envelopment analysis technique, (Boubaker, 2022). Utilizing a sample of U.S. commercial banks covering 1994–2018, the outcomes show that banks obtain more advantage of parents to improve the effectiveness. Yu (2021) have demonstrated the determination of meta-technology in dynamic setting. The new approach differs from minimum extrapolation principle of individual group technologies. The model considers linking activities between processes and carry-over activities with 22 Taiwanese banks from the year of 2008 to 2016.

2.7. Agent-Based Models

Agent-Based Models (ABMs) with empirical data for banking sectors was introduced by (Gatti, 2020). Additionally, a forecasting method was considered using ABM simulations. These methodological advancements illustrating their practical utility and enhance the integration of ABMs with real-world data. Researchers gives the gap between theoretical ABMs and empirical data for complex systems and phenomena. Chen (2021) has introduced a novel global sensitivity analysis procedure for the banking sectors. Monte Carlo simulations was presented to mitigate randomness and generator seeds. ABM facilitates the quantification of the discrepancy between real and simulated data. Laliotis (2020) introduced an ABM by simulating and calibrating using European household data. The results demonstrated an effect on the buyer distribution concerning bid prices, sold properties, and real estate values. These effects were driven by probability distributions associated with wealth, debt-to-income ratios, and loan-to-value ratios.

2.8. Fuzzy Logic

study aimed at assessing the credit ratings of small industrial firms through the application of fuzzy decision-making techniques presented by, (Sun et al, 2022). The study concentrated on 1820 Chinese small industrial businesses, and triangular fuzzy numbers were employed to improve the accuracy of the evaluation procedure. Different techniques were applied to organize small firms into different rating categories, resulting in a thorough and detailed assessment. In the context of small industrial enterprises, this strategy demonstrated a novel application of fuzzy decision-making approaches for credit rating, which added to a more precise and customized assessment methodology within the Chinese business landscape.

An advanced early warning evaluation index system have conducted by (Zhang, 2020) employing a Fuzzy Neural Network (FNN). With the use of a dynamic and adaptable strategy, The study greatly enhances the precision and effectiveness of credit risk assessment in the banking industry. By utilizing FNN's capabilities, this novel approach improves risk evaluation's overall performance and offers a stronger foundation for early warning systems in the financial sector. Seyfi-Shishavan (2021), have concentrated on the financial sector particularly in Turkish banking industry. An extended Intuitionistic Fuzzy Best-Worst Method, IFBWM was utilized to establish the weights of significant criteria, and fuzzy inference scheme was utilized to establish the banking sector's performance index. The research also concentrating on the performance of the Turkish banking industry.

A comprehensive study utilizing Fuzzy Logic and Decision-making Trial and Evaluation Laboratory, DEMATEL technique is performed by, (Lin, 2021) to identify key factors influencing the performance of Taiwan's wealth management banks during the 2019–2020 COVID-19 pandemic. The amount of proven COVID-19 cases, switching patterns, fee income, and customized investment information were the main areas of focus for the inquiry. The study included a sophisticated comprehension of uncertainties in the assessment process by utilizing Fuzzy Logic. DEMATEL played a crucial role in determining the causal links between the components that were found and in offering a comprehensive understanding of their interdependencies. This research contributes to a better understanding of the elements essential for overcoming obstacles and guaranteeing strong performance in periods of economic uncertainty by providing insightful information about the intricate dynamics affecting Taiwan's wealth management industry during the epidemic. Horak (2020) have provided Support Vector Machine and artificial neural networks to predict potential corporate bankruptcies. The study focused on examining the profit and loss statements and balance sheets of industrial companies in the Czech Republic during the previous five years of marketing. By applying cutting-edge machine learning methods, the research sought to increase precision and forecasting abilities in detecting businesses at risk of insolvency. The results provide insightful information about financial risk management and a detailed picture of potential difficulties facing the Czech Republic's industrial sector.

2.9. Neural Networks

The efficacy of a Back Propagation (BP) neural network-based algorithm is demonstrated by (Guo, 2020) for assessing lending risk in the banking sector. The research utilized extensive lending data spanning 2015–2019 to accomplish this task. A comparative analysis was carried outby logistic regression, benchmarking the BP NN against other conventional techniques. This research investigates into the application of advanced NN techniques to improve the accuracy and effectiveness of bank lending risk assessment, providing insights into the potential benefits of the BP algorithm over conventional approaches like logistic regression in the specified timeframe of consideration. Chen (2021) have provided a pivotal challenge in P2P lending accurately estimating the default risk of loans. The research highlighted the intrinsic problem where lenders base loan decisions on borrower-provided information, leading to an imbalance in P2P lending data by varying instances of default loans and fully paid. Identifying the fundamental issue, the research wanted to devise effective methodologies to navigate this imbalance and enhance the precision of default risk estimation in the context of peer-to-peer lending platforms. By acknowledging and addressing the unequal distribution of fully paid and default loans, the study aimed to contribute to more robust risk assessment in the lending domain.

A study investigating the impact of FinTech in Bangladesh is proposed by, (Yan, 2022). The research focused on 351 workers from banking institutions in Bangladesh, exploit data collected from January to March 2021 through a sampling technique. To explore the relationships between study variables, the researchers introduced a two-staged Structural Equation Model and Artificial Neural Network (SEM-ANN) approach. This innovative methodology allowed for a comprehensive analysis of the collected data, providing insights into the intricate dynamics between FinTech and various factors within the banking sector, contributing to a deeper understanding of the evolving financial landscape in Bangladesh.

Moreover, (Fu et al, 2020) have elucidated a two-step methodology involving a deep learning neural network to extract keywords from investor comments. They then used a Bidirectional Long Short-Term Memory (BiLSTM) model to forecast the platforms used by the banking industry's default risk. With the help of this creative method, deep learning was effectively used to extract keywords, allowing for a thorough examination of investor sentiment. When BiLSTM was used to risk prediction, it demonstrated the model's capacity to represent intricate temporal relationships, which improved the model's ability to assess default risks related to financial platforms in the banking industry.

2.10. Ensemble Methods

Credit default probabilities prediction for P2P lending using machine-learning technique is demonstrated by, (Yin, 2023). For P2P lending platforms, a stacking ensemble machine-learning model was used to evaluate credit default risk. After selecting features using the Max-Relevance and Min-Redundancy (MRMR) approach, superfluous features are removed using the k-means clustering technique. Song (2020) described an innovative learning approach known as Distance-to-Model and Adaptive Clustering-Based Multi-View Ensemble (DM–ACME) for predicting default risk in Peer-to-Peer (P2P) lending. The approach used gradient boosting decision trees in conjunction with multi-view learning and adaptive clustering to generate a collection of heterogeneous models. The purpose of the DM–ACME learning method is to improve prediction accuracy by utilizing various data views and modifying clustering techniques. With the goal of capturing a thorough understanding of the many aspects that contribute to default risk in peer-to-peer lending, this ensemble-based approach aims to offer a reliable tool for risk assessment in the ever-changing and dynamic world of peer-to-peer financial transactions.

Xtreme gradient boosting (XGBoost), a deep neural network (DNN) and logistic regression (LR), concurrently integrated with a liner weight ensemble strategy are studied by, (Li, 2020) have provided e. Behind creating rank and discrete features, model adds missing values to the model for self-training. Hyperparameters were optimized by XGBoost model to enhance the performance of the prediction model. Xia (2020) have elucidated the development of an Overfitting-Cautious Heterogeneous Ensemble (OCHE) based on tree structures for credit scoring in the banking sector. The study strategically utilized various techniques to strike a balance between predictive accuracy and computational efficiency. The approach involved ensemble selection, wherein the method assigned weights and dynamically adjusted statistical measures to mitigate overfitting. By implementing OCHE, the research aimed to optimize credit scoring models, offering an effective means to enhance performance while addressing the challenges associated with overfitting and computational costs within the banking domain.

2.11. Summary

This chapter provides an extensive examination of the integral role played by OR and AI methods in Banking and Finance research. It concentrates on various central themes within the banking sector, encompassing aspects like banking efficiency, risk management, bank performance, banking regulation, customer-centric studies, and the impact of FinTech. The discussion delves into the application of popular OR techniques, including DEA, ABM, MC, and fuzzy logic. Additionally, it explores AI methodologies such as SVMs, NNs, and ensemble approaches.

A range of important difficulties in the banking sector are addressed by the cooperative application of these approaches. ABM makes it easier to simulate complex financial systems, whereas DEA may be used to evaluate how efficiently banking activities are run. Fuzzy logic helps handle inaccurate information, and Monte Carlo simulations help manage risk. In terms of AI, NNs are excellent at recognizing patterns, SVMs provide reliable classification for risk assessments, and ensemble techniques improve predictive accuracy.

This chapter significantly enriches existing literature by presenting a thorough review that advances beyond previous bibliographic surveys. It intricately explores a diverse range of topics relevant to the banking sector, summarizing the manifold methodologies in play. Through this collective exploration and analysis, a deeper understanding of the integration of OR and AI methods emerges, illuminating their substantial contributions to enhancing various facets of the banking industry. This comprehensive overview serves as a valuable resource for researchers, practitioners, and stakeholders seeking insights into the evolving landscape of OR and AI applications in the banking sector.

CHAPTER: 3

PROPOSED METHODOLOGY

3.1 Introduction

With the help of ML, the complex loan beginning process may be streamlined and automated to make results in a couple of days, which is another success for the banking sector. Customers are happy, profit and revenue margins are enhanced for lenders who process loans more rapidly, (Zhang, 2018; Zhang, 2023). The uses of ML in bank section in pictorially depicted in Figure 3.1. Figure shows that the ML based bank lending process has several advantages like financial monitoring, algorithmic trading, financial advising, credit scores, marketing etc.

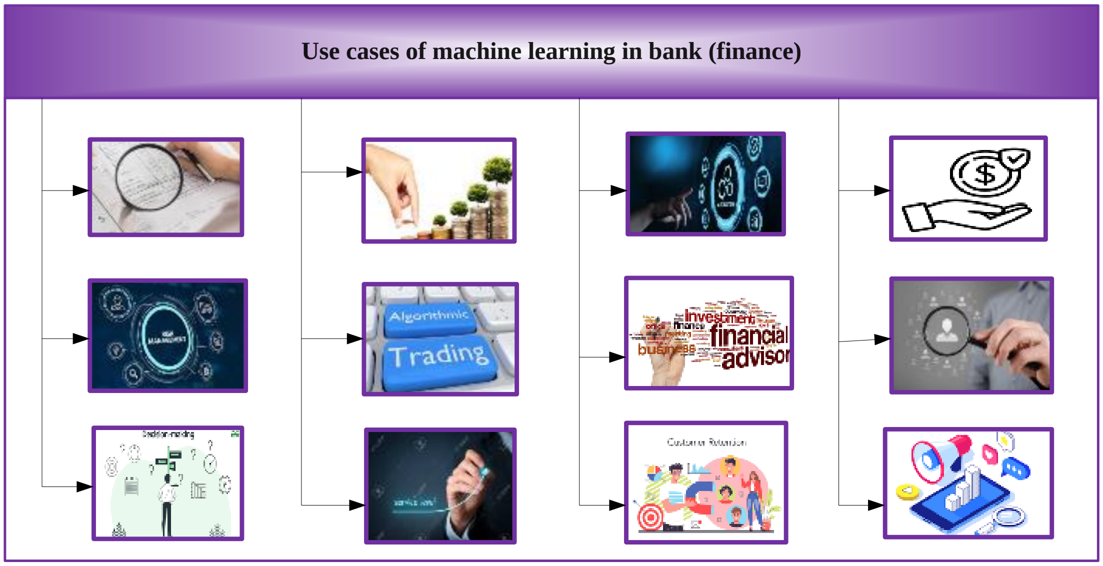


Figure 3.1: Applications of ML in bank lending, (Uddin, 2023)

The credit risk deduces the likelihood of a loss rising from an applicant's breakdown to pay back a loan or meet contractual obligation, (Zhang, 2018; Pavón Pérez, 2023). Traditionally, it pertains to the risk occur from lenders' incapability to return the owed principal and interest, impacting the cash flows and rising assemblage costs, (Pol, 2022; Papouskova, 2019). Figure 3.2 shows the supervised, unsupervised and reinforcement learning based ML techniques.

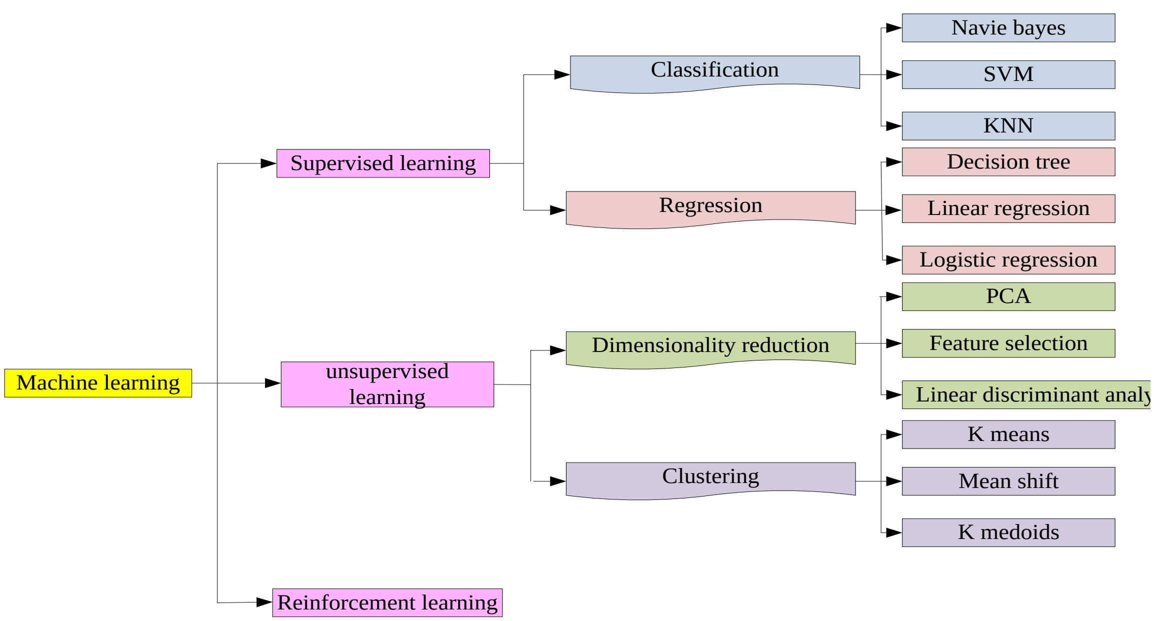


Figure 3.2: ML based techniques, (Alharbi F, 2023)

The ML based techniques permit banks to proactively observe the behavior of applicant, recognize anomalies in real time, prevent fraud, and diminish the likelihood of false positives. The main contributions of this research are;

- The input data are collected from banking ML and banking prediction model using random forest datasets
- From the input data the relevant features are selected with the help of univariate ensemble based feature selection
- A novel technique is introduced for bank lending decision making process which combine RF and WSO
- From the selected variables the RF recognizes if the applicant is eligible for loan or not.
- Then the performance of RF is enhanced with the help of WSO which helps to minimize the error rate and to enhance the accuracy.
- Finally, the performance of proposed method is compared with the existing approaches to demonstrate the efficiency of the proposed method.

3.2 Proposed methodology

This research proposes a novel approach for bank lending. Initially the input data is collected from banking ML and banking prediction model using random forest datasets sourced from Kaggle an online repository of open-source datasets.. Then from the input data the relevant features are selected. From the selected relevant features, the RF classifier helps to make the decision as accept loan or reject loan. To optimize the performance of RF the WSO is applied. This helps to minimize the error rate and to enhance the accuracy. The performance of proposed method is evaluated with the help of several evaluation metrics like accuracy, recall, F1-score, precision and cross validation. Then to demonstrate the efficiency of proposed method is compared with various existing approaches. Figure 3.3 shows the architecture of proposed method.

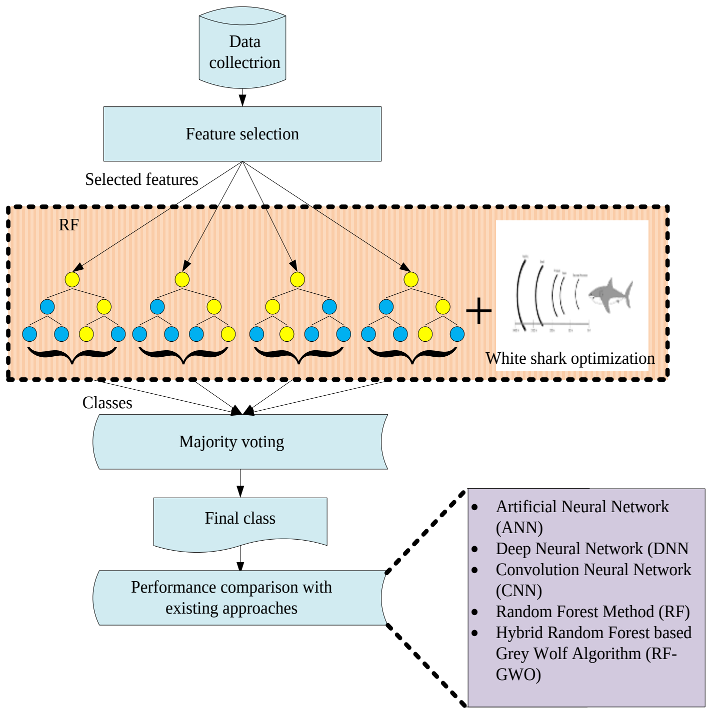


Figure 3.3: Workflow of proposed approach, (Abeßer, 2020)

3.2.1 Dataset description

In this research the input data's are collected from two open-source datasets available in Kaggle platform. Different features make up the datasets such as the applicant's credit history, marital status, education data, loan ID, loan amount, and loan status. This would make it easier for the bank to decide whether to approve or deny the loan and assist them to learn more about the applicant's eligibility.

- Banking prediction model using random forest, https://www.kaggle.com/code/febinphilips/banking-prediction-model-using-randomforest/input) datasets.
- The datasets are Banking ML, https://www.kaggle.com/code/sid0307/banking-ml

Banking prediction model using random forest dataset

Term deposits are a major source of income for banks. Term deposits are investments made with cash that are kept at a financial institution. Your money is invested at a predetermined rate of interest for a predetermined amount of time, or term. The bank pitches term deposits to its clientele using a range of outreach techniques, such as email marketing, internet marketing, telemarketing, and advertisements, (Paulin, 2018; Pérez-Martín, 2018). One of the most effective methods for reaching customers is still through telephone marketing campaigns. However, these initiatives come at a high cost when large call centers are actually hired to carry them out. Therefore, it's essential to identify the customers who are most likely to convert ahead so that you may properly target them with a call, (Serengil, 2022). The features of banking prediction model using random forest dataset are given in table 3.1.

Table 3.1: Banking prediction model using random forest dataset description

No.	Features
0	Age
1	Marital
2	Job
3	Default
4	Education
5	Balance
6	Loan
7	Contact
8	Duration
9	Month
10	Day
11	Pdays
12	Campaign
13	Poutcome
14	Previous

Banking ML dataset

The Banking ML is another dataset utilized in this research to gather input data. Table 3.2 shows the features of bank ML dataset.

Table 3.2: Banking ML dataset description

No.	Features
0	Loan_ID
1	Gender
2	Married
3	Dependents
4	Education
5	Self_Employed
6	Applicant income
7	Coapplicant income
8	Loan Amount
9	Loan_Amount_Term
10	Credit_History
11	Property_Area
12	Loan status

Table 3.2 shows the features in the banking ML dataset which is sourced from Kaggle. Totally this dataset consists of 13 features. Such as gender, loan ID, loan amount, loan status, education details, marital status, credit history etc., of the applicant. This would helpful for the bank to know about the applicant's eligibility and very easy to accept or reject the loan.

3.2.2 Univariate Ensemble (UE) based feature selection

From the input data the relevant features are selected with the help of UE. Figure 3.4 shows the architecture of UE based feature selection. Feature selections enhance the ML process and amplify the predictive power of ML algorithms by selecting the most significant variables and eliminate redundant and irrelevant features. In the initial stage, the features are ranked typically, whereas, in the next stage a cutoff point is defined to choose significant features and to filter out the unrelated features for building more robust ML models. The univariate selection is a statistical test and it can be utilized to choose those features that have the strongest relationship in the target variable. In order to produce a final ranked list of features after a thorough review of a feature collection, first suggest using the unified features scoring (UFS) algorithm. We then provide a threshold value selection (TVS) approach to choose a subset of characteristics that are considered critical for the classifier development, after establishing cutoff points to exclude unnecessary features.

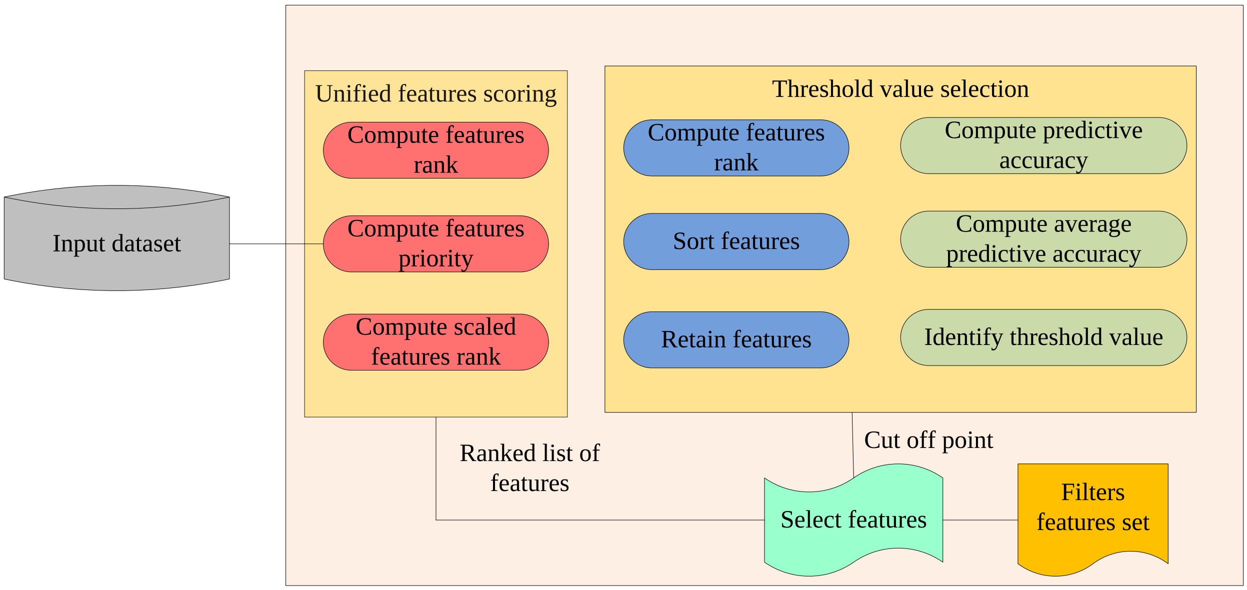


Figure 3.4: Architecture of UE based feature selection, (Ali, 2018)

3.2.3 Hybrid RF-WSO

Random Forest is a well-known machine learning algorithm and one of the supervised learning techniques. It can be used for machine learning problems that involve regression and classification. Its basis is the concept of ensemble learning, which is the act of combining multiple classifiers to improve the functionality of the model and solve a difficult problem.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." The random forest predicts the result based on the majority vote of predictions from each decision tree, as opposed to relying solely on one. Tasks involving both classification and regression can be completed using Random Forest. Large datasets with high dimensionality can be handled by it. It keeps the overfitting problem at bay and improves the model's accuracy.

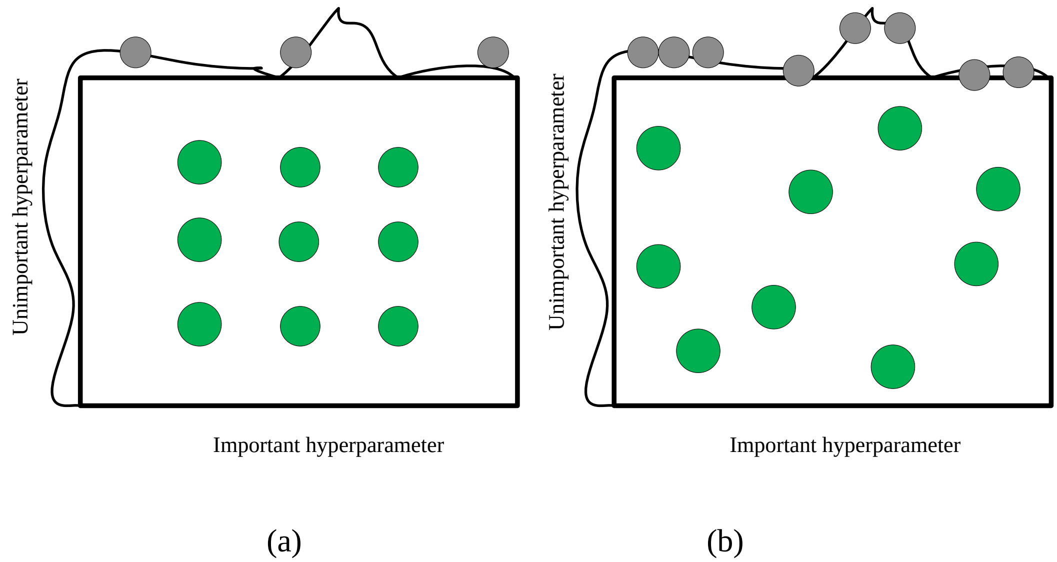


Figure 3.5: (a) Grid search and (b) random search

An ensemble of decision trees without pruning is called a random forest. When we have a huge number of input variables (hundreds or even thousands) and a large training dataset, we frequently employ random forests. A classifier called a random forest model is made up of numerous decision trees, and it produces a class output that is the average of the classes produced by each individual tree. Figure 3.5 shows the difference between the grid search and the random search.

Random search is more effective likened to grid search. Although this grid search can discover the best value of hyperparameters, assuming they are in your grid. finally, random search can usually discover a "close-enough" value in distant less iteration. The general architecture of RF is given in Figure 3.6. An ensemble of decision trees makes up the Random Forest algorithm. During the training phase of the structure, many decision trees are created, and their outputs are integrated to form a final prediction. The Random Forest builds each decision tree by iteratively dividing the dataset into subsets according to feature values. The algorithm chooses the best characteristic at each tree node to split the data, forming branches that connect to other nodes until the leaf nodes, where a final prediction is formed. Bootstrap Sampling: Random Forest uses a technique known as bootstrap sampling, which involves selecting replacement samples from random subsets of the training data. By introducing diversity, this keeps the trees from overfitting to certain patterns in the data and becoming too similar to one another. When generating judgments in each tree, Random Forest randomly chooses a subset of features at each node in addition to sampling data. This increases the diversity of the trees even further and guarantees that several data points are taken into account for every choice. For classification problems, the Random Forest uses a voting process to determine the final prediction. Each tree "votes" for a class, and the class with the highest number of votes is chosen. Regarding tasks involving regression, the average of the predictions made by each individual tree frequently serves as the final forecast. The major advantages of RF are;

- It provides high accurate classification
- Can able to handle many inputs variables
- It offers an experimental way to recognize variable communications
- Fast learning
- It evaluates proximities among cased and helpful for detecting outliers, clustering and visualizing.
- Generated unbiased estimation of the error (generalization)
- Will balance error of unbalanced datasets in the class population

Every tree is constructed by following;

Step 1: N is the number of training cases and the variable's number in RF if represented as H.

Step 2: The h of input variables is utilized to finalize the decision at the tree's node, the value of h shall lower.

than the value of H.

Step 3: By choosing N select the training set for the tree by replacing all available N. Utilize the remaining N to evaluate the tree's error by predicting their classes.

Step 4: For tree's every node, select h randomly on which to base the decision at that node. Evaluate the ideal split depend on these h variable in the N. The variable significance is resolute through likening the prediction error to the data term Out-Of-Bag, (OBB). The following equation (3.1) is utilized to calculate the target's OOB error,

$$OOB_{error} = Z^{RDF(tar)} - Z^{RDF(out)} \qquad \text{Equation (3.1)}$$

Step 5: Every tree is not pruned and grown fully. The significance of the variable is determined with the help of equation (3.2),

$$\xi_I^{(tr)} = \frac{1}{T}\left(\frac{\sum_{xa\in\phi^{c(tr)}} I\,(l_b = C_a^{tr}) - \sum_{xa\in\phi^{c(tr)}} I\;(l_b = C_{a,nz}^{tr})}{\phi^{c(tr)}}\right) \qquad \text{Equation (3.2)}$$

By this, the introduced approach efficiently modifies the duty ratio to offer the required outcome.

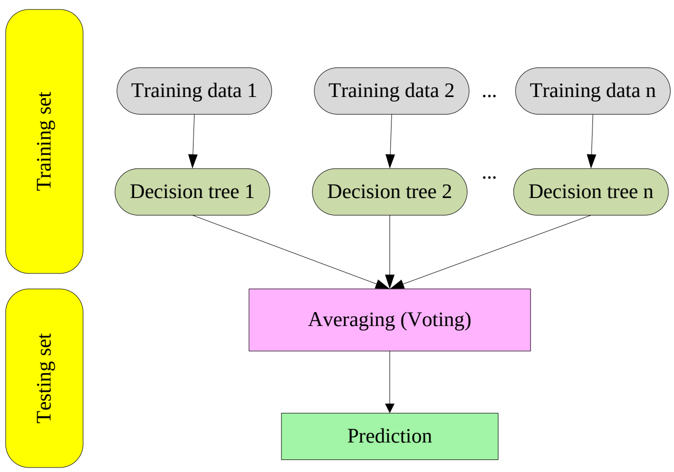


Figure 3.6: General workflow of RF

Following this the WSO is applied to enhance the performance of RF by minimizing its error rate. In randomized search, instead of identify a grid of values; we can describe probability distributions or ranges for every hyperparameter. This becomes a much superior hyperparameter in search space. Randomized search then randomly samples a permanent number of groupings of hyperparameters from these distributions. Figure 3.7 shows the working process of RF

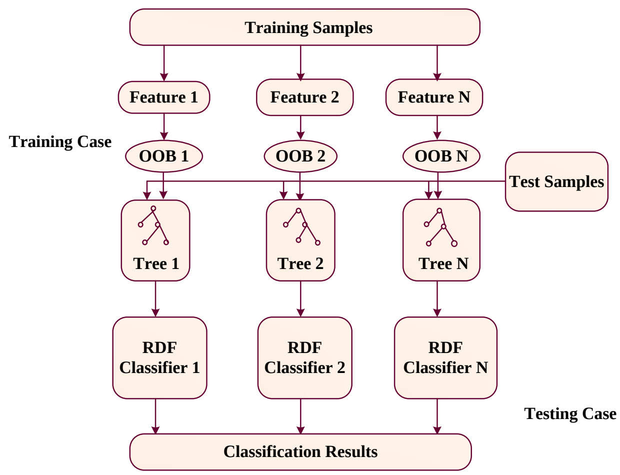


Figure 3.7: Working process of RF, (Wu, 2019)

3.2.3.1 White Shark Optimizer (WSO)

Within a continuous search area, the White Shark (WS) optimizer is a practical intelligent metaheuristic model that can solve a variety of optimization issues. This method, which was unveiled in 2022, uses scent and vision to replicate the way white sharks hunt. The fundamental idea and basic concept related by this WSO are inspired by the WS's behavior. In this research the WSO is applied to enhance the performance of RF. The WS has the unexpected senses of hearing and smelling while the navigation and foraging procedure. These ideal elements will mathematically investigate and numerically modeled to offer enough balance among utilization and examination of this scheme. The location of WS is determined with the help of equation, (3.3)

$$S = \begin{bmatrix} S_1^1 & S_2^1 & \cdots\cdots & S_d^1 \\ S_1^2 & S_2^2 & \cdots\cdots & S_d^2 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ S_1^n & S_2^n & \cdots\cdots & S_d^n \end{bmatrix}$$ Equation (3.3)

where, the WS's i^{th} position associated to j^{th} dimension is represented as S_j^i. It will be evaluated with the help of lower L_j and upper U_j bounds of the WSO's exploration region in j^{th} dimension is given by equation (3.4),

$$S_j^i = L_j + Rand \times (U_j - L_j)$$ Equation (3.4)

The random number in the [0,1] bound is denoted as $Rand$. In this the fitness (initial) will be evaluated for initial solutions by equation (3.1). Here the fitness function is to minimize the error rate of RF and enhancing the accuracy the fitness function is given by equation (3.5)

$$Fitness\ Function = Min(MSE)$$ Equation (3.5)

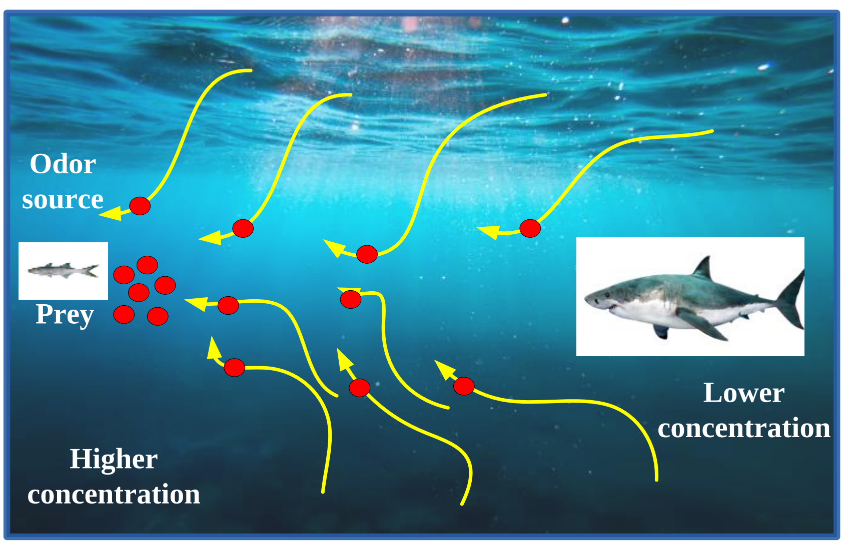


Figure 3.9: Movement of WS towards prey, (Braik, 2022)

Meanwhile when the new location is more accurate comparing with the earlier one an upgrading stage will deploy. The movement of WS toward prey is shown in Figure 3.8, When with the help of WS's wave frequency the WS finds the prey, it will move closer to the prey in oscillating motions depending on the velocity given by equation (3.6),

$$V_{k+1}^{i} = \mu \left[V_k^i + Q_1 \left(S_{gbest_k} - w_k^i \right) \times D_1 + Q_2 \left(S_{best}^{v_k^i} - w_k^i \right) \times D_2 \right]$$ Equation (3.6)

where, Q_1 and Q_2 signifies the influences of the WS that observe $S_{best}^{v_k^i}$. The present upgraded velocities in iterations $T+1$ of i^{th} WS is represented as V_{k+1}^{i} and V_k^i. S_{gbest_k} and w_k^i signifies the ideal global location at T^{th} iteration. The random numbers in [0,1] range are signified as D_1 and D_2. The constriction factor of WSO is represented as μ with analyzed the convergence behavior of WS as equation (3.7),

$$V = [N \times Rand(1, N)] + 1$$ Equation (3.7)

$(1, N)$ is the random vector through random numbers in the range of [0,1]. The movement toward the ideal WS is given by equation (3.8)

$$S'^{i}_{k=1} = S_{gbestk} + r_1 \vec{C}_w \operatorname{sgn}(r_2 - 0.5)\ r_3 < ss$$ Equation (3.8)

r_1, r_2 and r_3 are the random values in [0,1] range. Figure shows the flowchart for WSO. Finally, the WSO finds the optimum solution. With the help of WSO the RF is optimized. The WSO helps to minimize the error rate of RF. Figure 3.9 shows the flowchart of WSO.

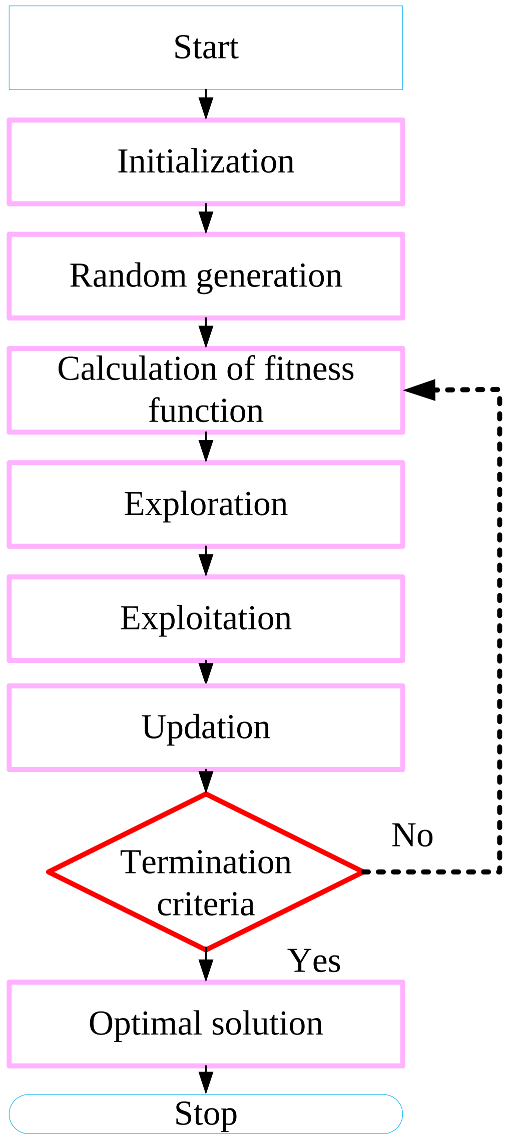


Figure 3.10: Flowchart of WSO technique

Pseudo code of WSO is given in Table 3.3,

Table 3. 3: Pesudocode of WSO

WSO Pseudocode
Input:

Weighting coefficient

Population of solutions

Output:

The leading candidate solution

Determine the best solution using fitness function

$For I \leftarrow 1 to m do$

Evaluate the distance among the population

Update the distance vector

$For I \leftarrow 1 to m do$

Normalize the distance and fitness vector in [0,1] range

Update FDB score vector

Select the best candidate solution

Return the best candidate

3.3 Results and Discussions:

To demonstrate the efficiency of the proposed method its experimental outcomes are compared with various existing techniques. Figure 3.10 shows the overall achievements of proposed method in terms of precision, accuracy, F1-score, recall and cross validation. The detailed results and discussions of proposed method are given in chapters 4 and 5.

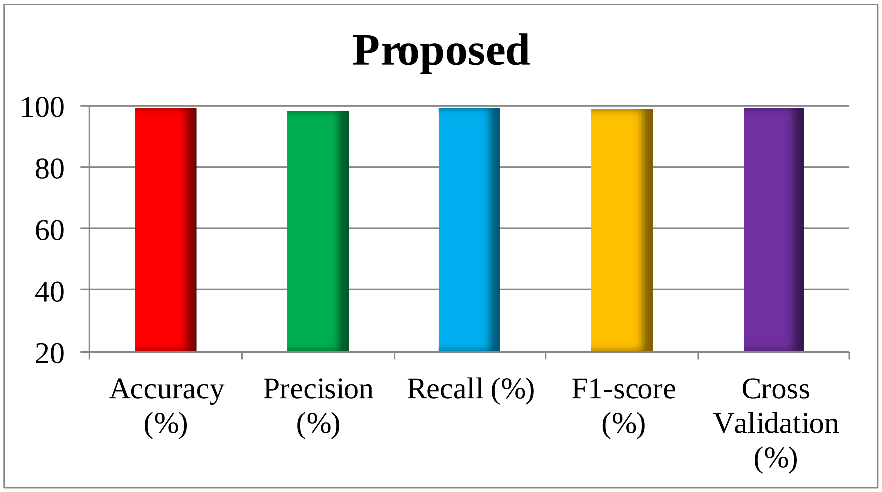


Figure 3.11: Outcomes of proposed method

The performance of introduced method is compared with the existing approaches such as CNN, ANN, DNN and RF respectively.

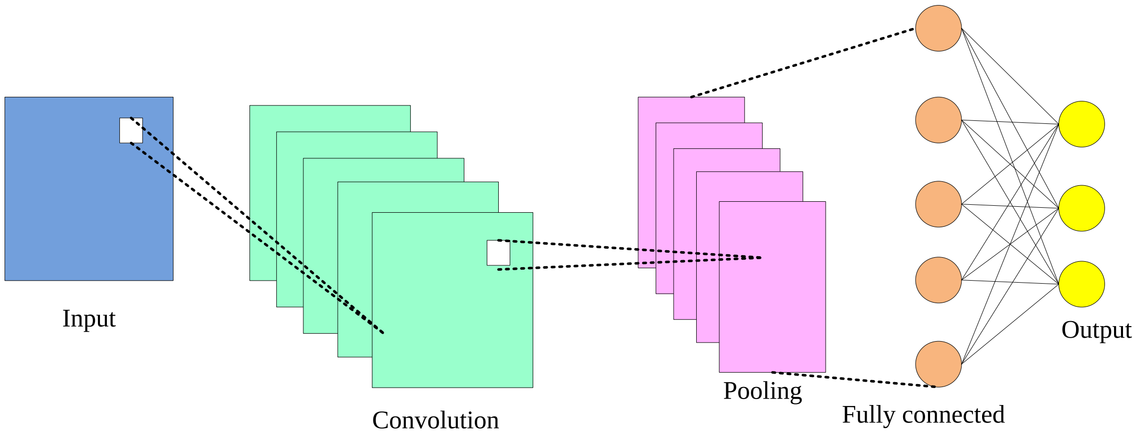


Figure 3.12: Structure of CNN model, (Khvostikov, 2018)

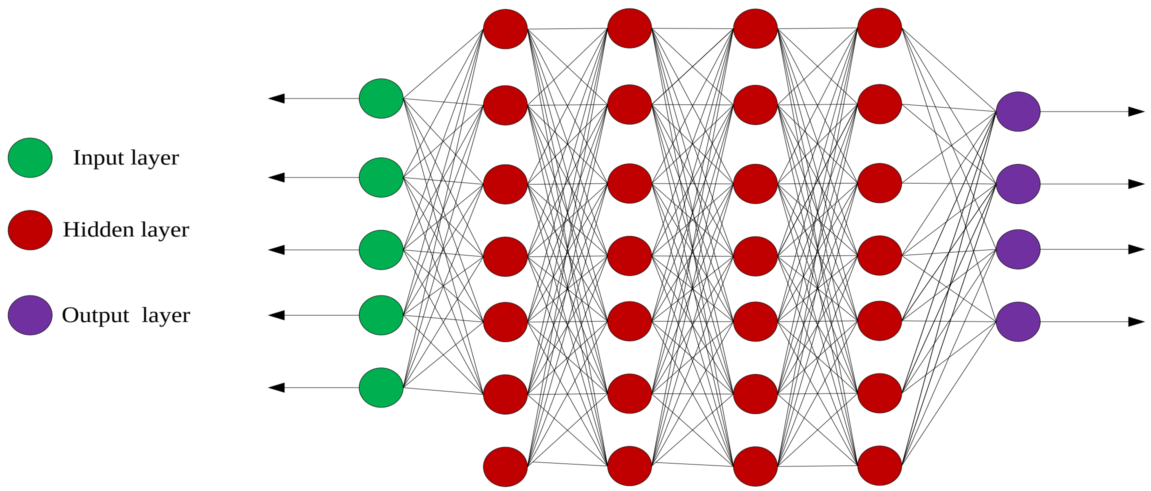


Figure 3.13: Structure of DNN model, (Hassan, 2017)

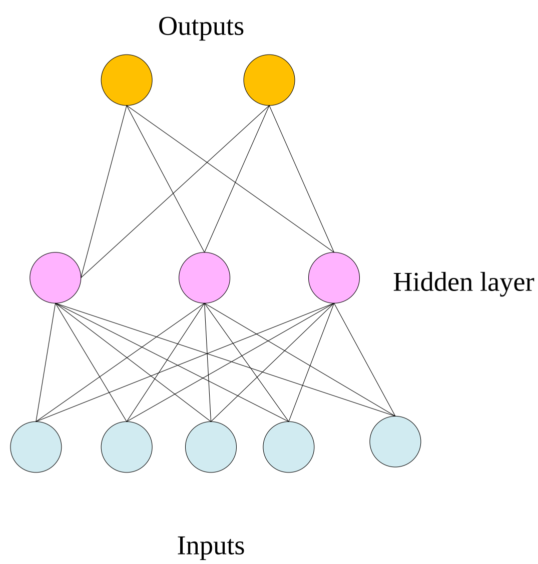


Figure 3.14: Structure of ANN, (Alhendi, 2023)

- **CNN**

A CNN is a type of ANN utilized mainly for image processing and recognition, due to its capability to identify patterns in images. A CNN is a powerful tool but requires millions of labelled data points for training. Figure 3.11 shows the Structure of CNN model

- **DNN**

(DNN), also known as Deep Nets, is a neural network with a certain stage of difficulty. It can be considered as stacked neural networks, or networks composed of several layers, usually two or more, that contain input, output, and at least one hidden layer in among. Figure 3.12 shows the Structure of DNN model, (Serengil, 2020).

- **ANN**

Artificial Neural Networks (ANNs) have wide-ranging applications in image and speech recognition, medical diagnosis, and machine translation. One major advantage of ANN is that it can learn from sample data sets. Random function approximation is the most common use of ANN. Figure 3.13 shows the Structure of ANN.

The introduced approach achieves higher performance than the prevailing approaches. In the context of a credit crunch scenario, the challenge of making optimal bank lending Making the best bank lending decisions in the midst of a credit crunch presents a very difficult and NP-hard problem. To address these issues in the field of financial research, specifically in the area of bank loan optimization, this work presents a clever model that uses a hybrid methodology. The model is intended to negotiate the complexities of bank lending choices in a highly competitive market with constraints imposed by the credit crisis.

The suggested method not only helps banks stay focused on their primary goal of increasing bank profits, but it also helps them make wise decisions when faced with lower lending capacity as a result of negative liquidity shocks. This study avenue presents various intriguing directions for future directions. First of all, it opens the door for a variety of strategies meant to maximize loan choices for small and medium-sized businesses. Second, it takes into account that interest rate risk is dynamic and plays a crucial role in the bank's lending decisions. Finally, by optimizing established strategies within hedge funds, the approach can be expanded to guarantee positive returns for the fund's stakeholders.

3.4 Summary:

The research introduces a novel approach, the Random Forest-based White Shark Optimizer (RF-WSO), aiming to optimize bank profits in lending decisions. This dynamic technique considers various customer loan characteristics, assuming all customers can be eligible for loans. Factors like loan age, amount, interest rate, lending type, and credit score are analyzed using RF-WSO to pinpoint the most eligible clients. Predicted loan losses and interest rates are also incorporated. The process involves selecting customers based on defined criteria to maximize profitability. This innovative technique combines machine learning (RFA) and optimization (WSO) to enhance lending decision accuracy and, consequently, improve overall bank profitability. The upcoming chapters explain the results, discussions and conclusion of this thesis.

CHAPTER: 4

RESULTS

4.1 Results

This section shows the experimental outcomes of proposed method. To demonstrate the efficiency of the proposed method its outcomes are compared with the existing approaches and the comparison graphs and tables are provided in this section.

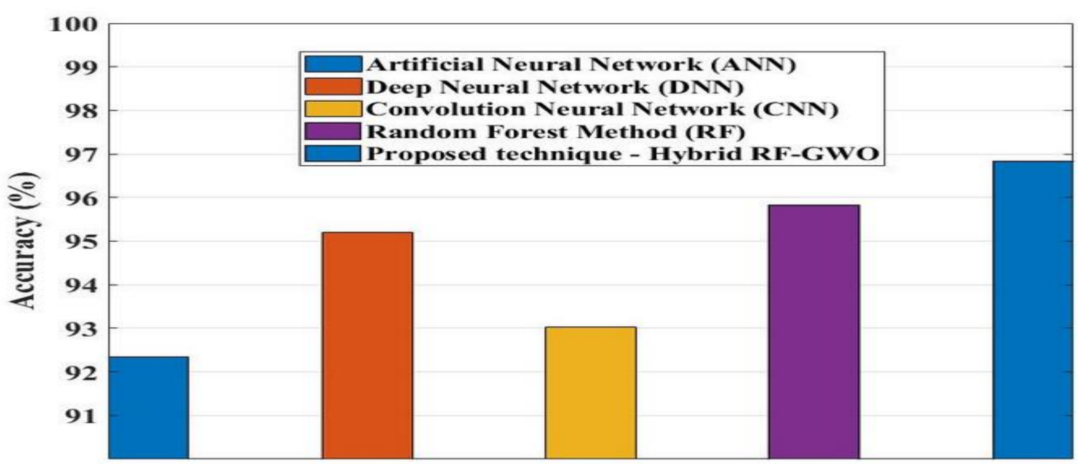


Figure 4.1: Accuracy analysis

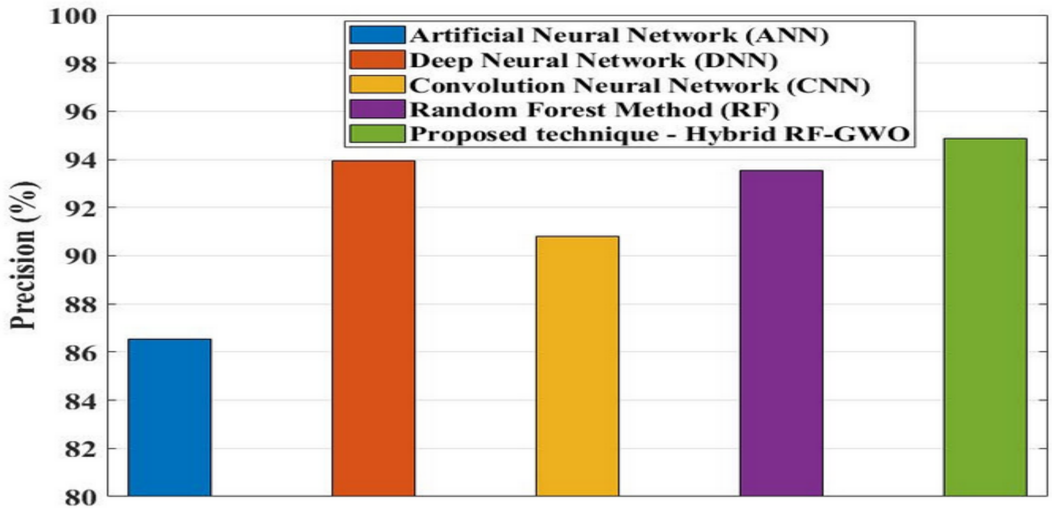


Figure 4.2: Precision analysis

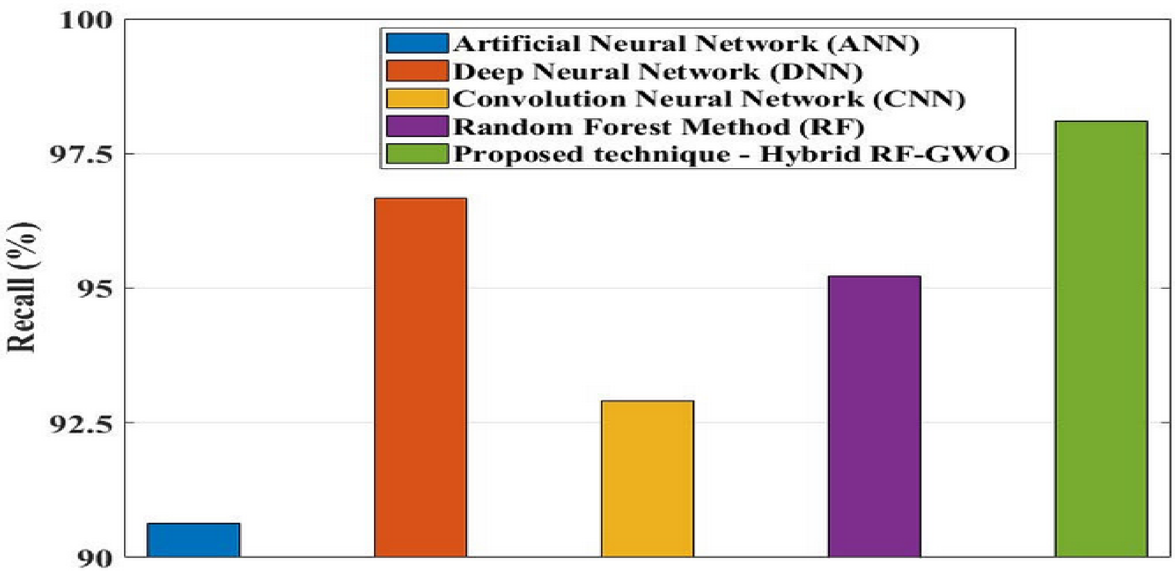


Figure 4.3: Recall analysis

The figures 4.1 to 4.3 illustrate the accuracy precision, and recall, of the introduced method. Proposed method attains higher values for all these metrics. The efficiency of introduced method is compared with the prevailing approaches such as RF, DNN, ANN and CNN respectively. The performance of CNN is low comparing with other techniques. The accuracy, precision and recall of RF are slightly nearer to the proposed method. But the proposed method performs effectively and achieved accuracy (99.624%), precision (98.49%) and recall (99.70) which is higher than the existing approaches.

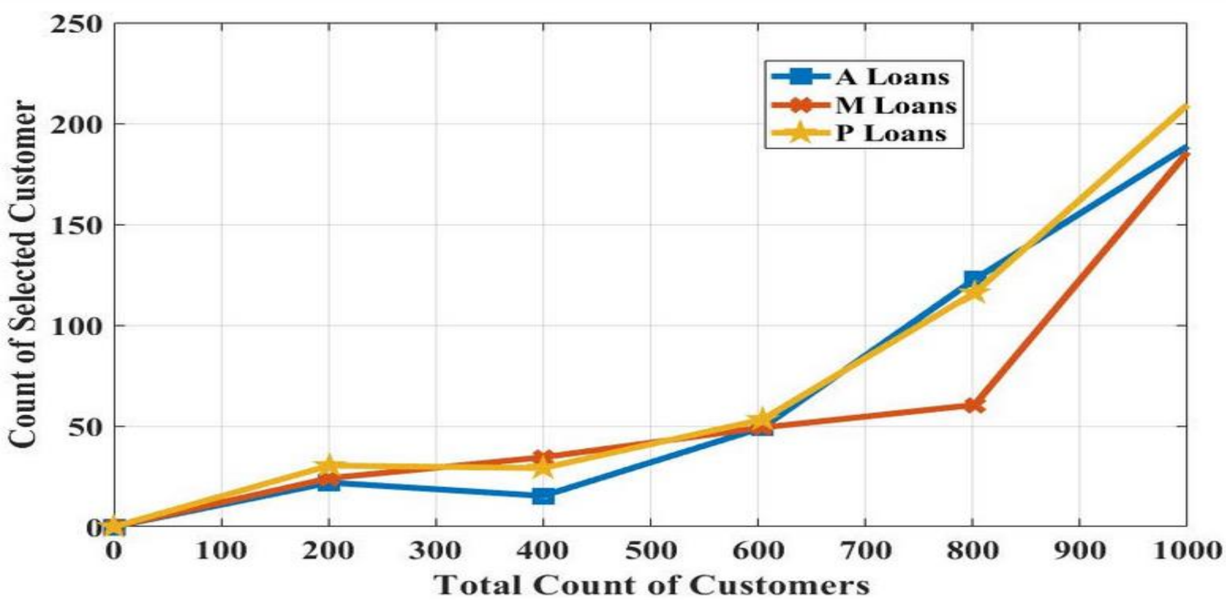


Figure 4.4: Number of selected customers for each type of loan

Figure 4.4 displays the number of selected customers for each type of loan. The analysis shows that for P loans the number of selected customers is higher than the M and A loans. When the count of customers increases the count of selected customer also increased. The count of selected customer for M and P loans are almost same.

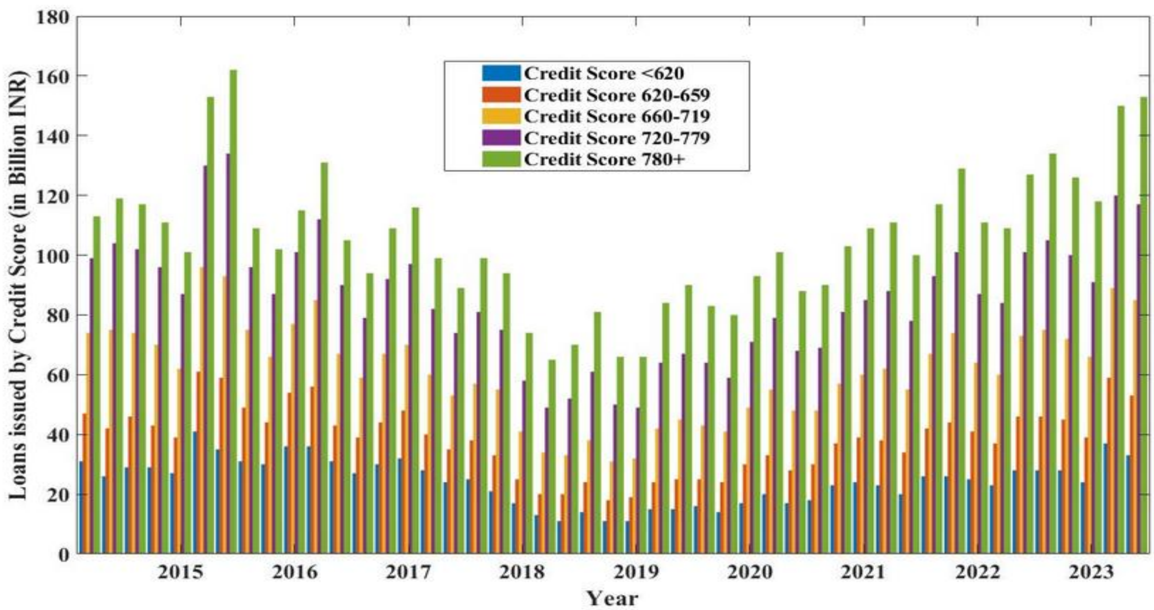


Figure 4.5: Credit score versus year

Figure 4.5 shows the credit score versus year in bank lending decisions. The credit score is a three-digit number that is used to assess a borrower's creditworthiness. Credit scores are evaluated based on a number of factors, including the credit utilization ratio, borrower's payment history and length of credit history. The graph illustrates that the credit score is increasing overtime. There are numerous factors are contributing that will be helpful for increasing in the credit score. Now a day the borrowers are more aware to know about the importance of credit score and they are taking steps to enhance their credit scores.

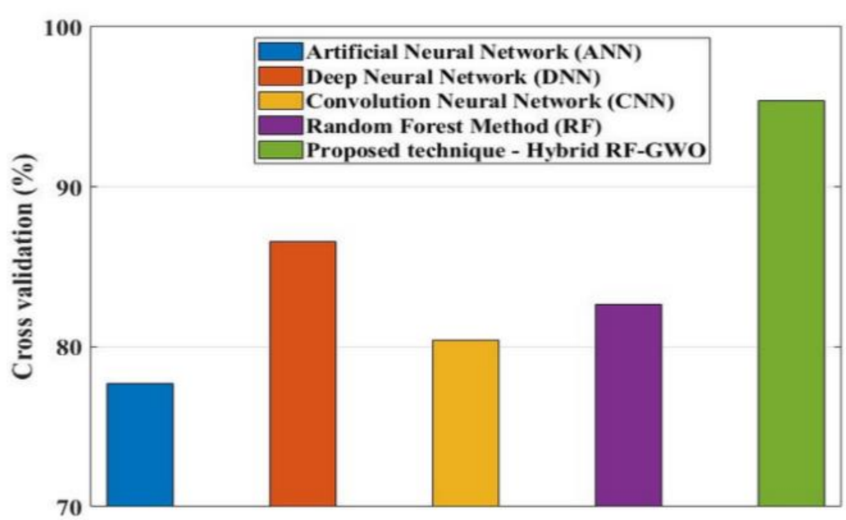


Figure 4.6: Cross validation comparison

Figure 4.6 shows the cross-validation analysis of proposed method. The effectiveness of the introduced approach is compared with the prevailing approaches such as ANN, DNN, CNN and RF respectively. The analysis shows that the cross validation of proposed hybrid RF-WSO is higher than the existing approaches. The cross validation of ANN is lower than other techniques. The cross validation of DNN is higher than ANN, CNN and RF but its cross validation is lower than the proposed approach.

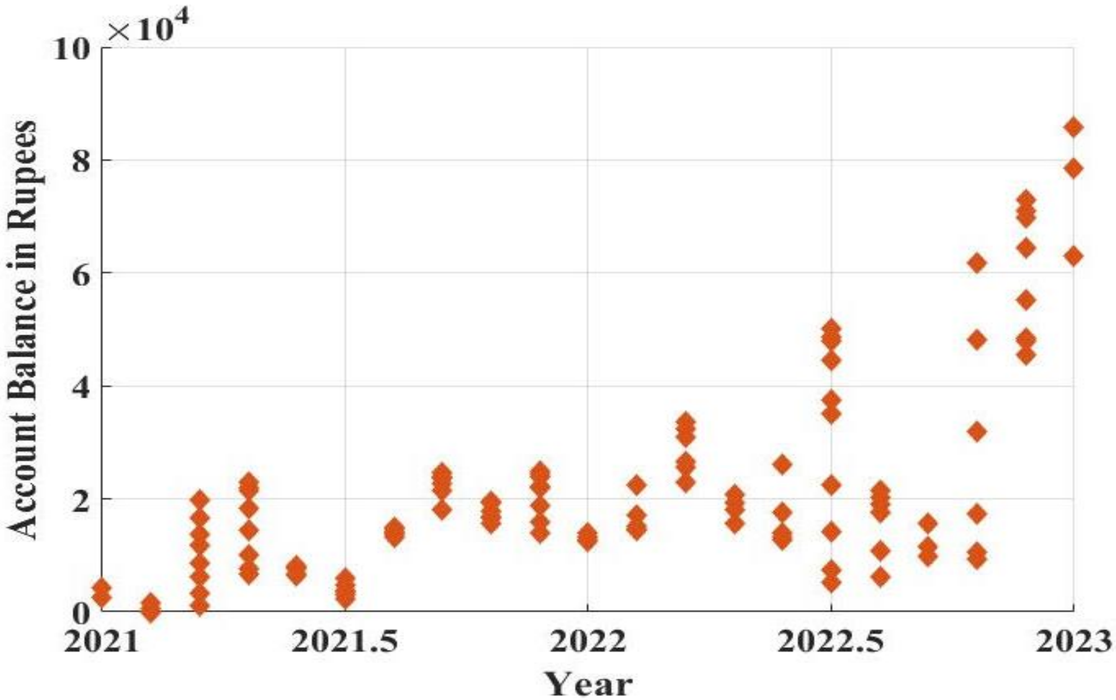


Figure 4.7: Account balance versus year

Figure 4.7 shows the account balance of applicants from 2021 to 2023. The analysis shows that the money in applicants account is increased in 2023 comparing with 2021 and 2022.

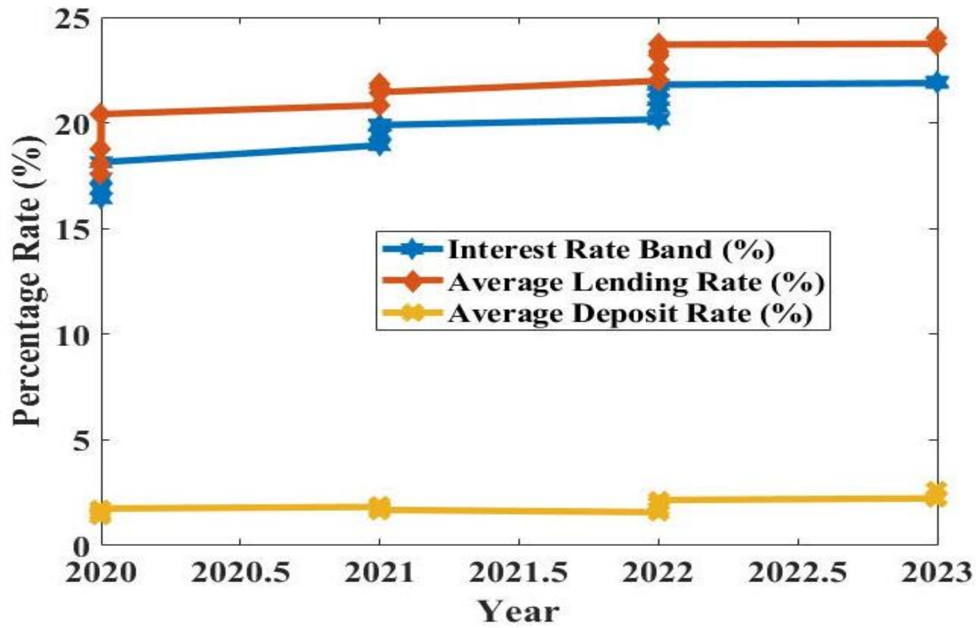


Figure 4.8: Percentage rate versus year

Figure 4.8 shows the interest rate band, average lending rate and average deposit rate versus year. The analysis shows that the percentage rate got increasing day by day. Comparing with 2020 the percentage rate has a significant improvement. The average deposit rate has lower percentage rate than others. The percentage rate of average lending rate is higher than average deposit rate but its percentage rate is lower than interest rate band.

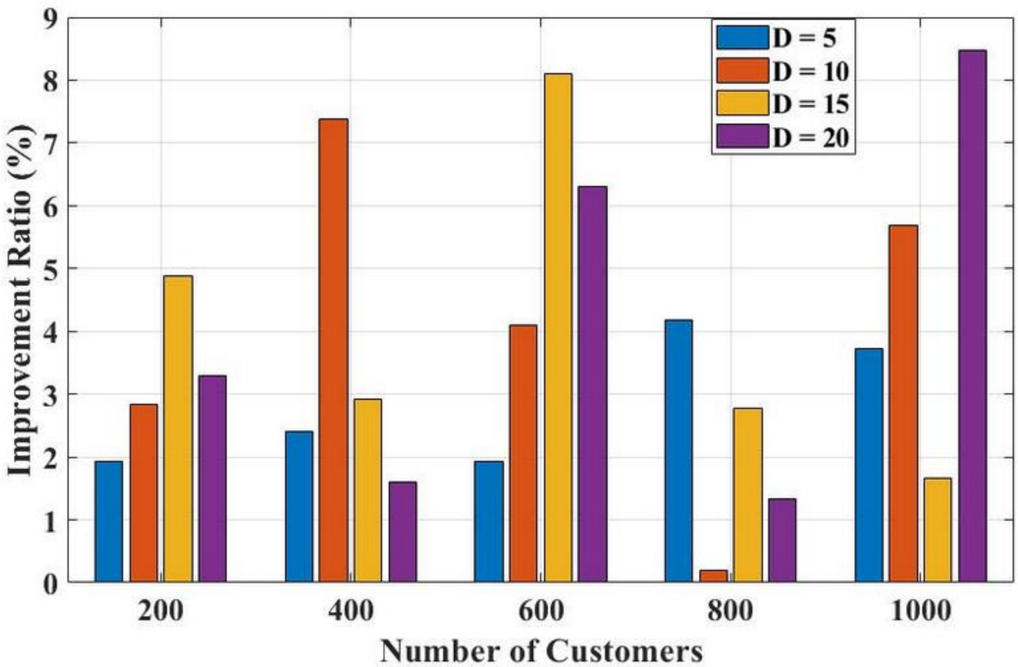


Figure 4.9: Improvement ratio of proposed technique compared to RF-GWO

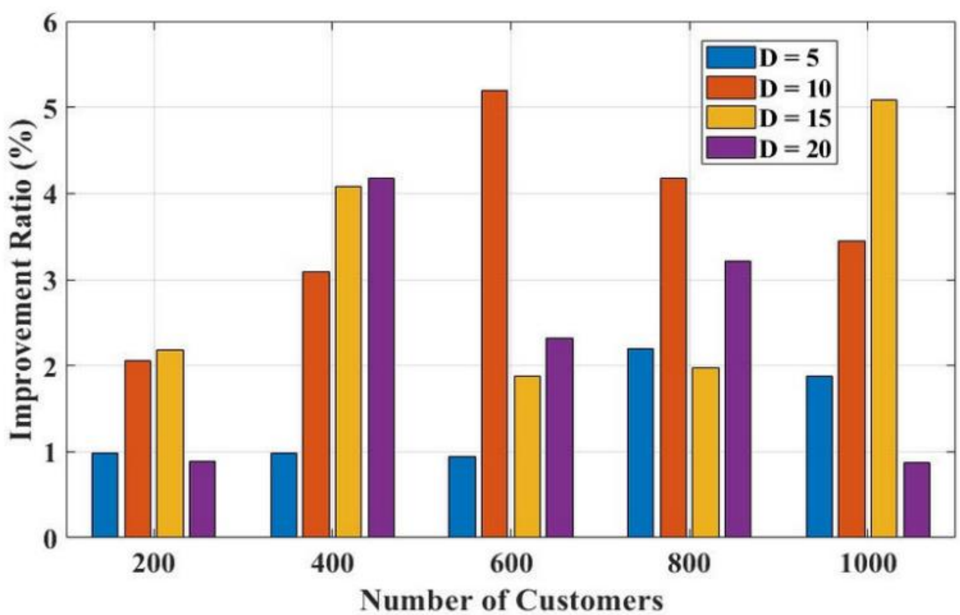


Figure 4.10: Improvement ratio of proposed technique compared to RF

Figures 4.9 and 4.10 shows the improvement ratio for different D comparing with RF-GWO and RF. This analysis shows that the proposed method has high improvement ratio with different number of D. This graph shows the improvement ratio for D=5, D=10, D=15 and D=20 respectively.

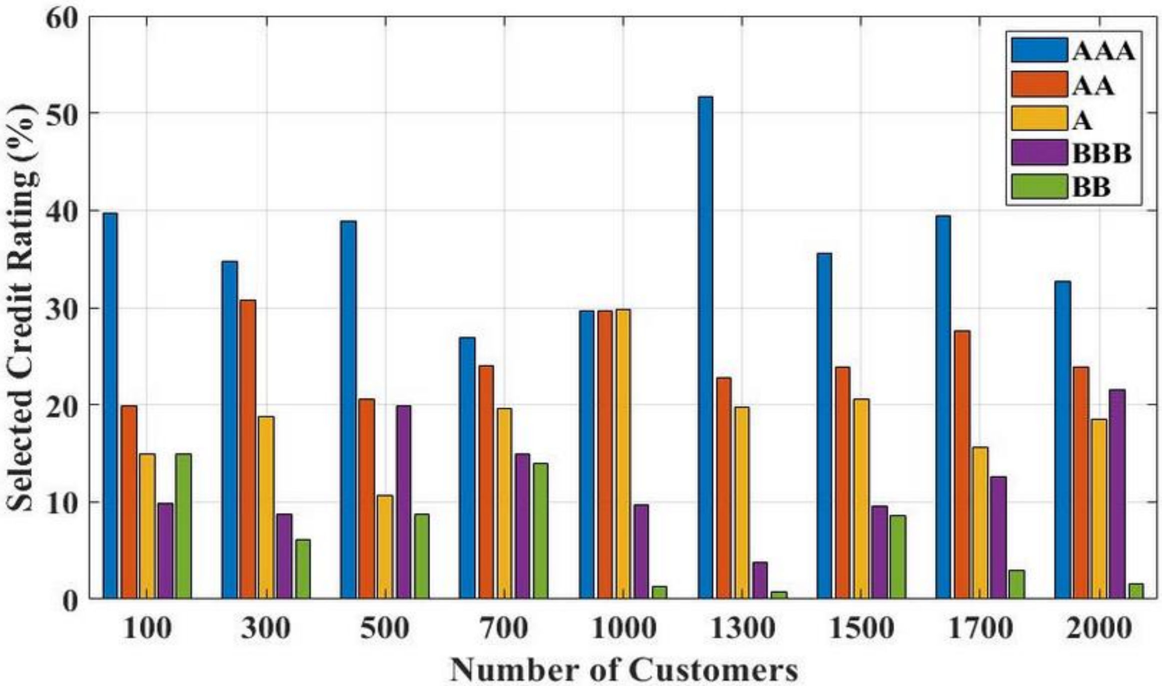


Figure 4.11: The ratio of selected customers by proposed based on their Credit rating

Figure 4.11 shows the ratio of selected customers by proposed based on their credits. It shows that number of selected credits ranking increased when the number of customers got increased. For AAA credit ranking the percentage of selected credit ranking increased gradually when the number of customers increased. For BB the percentage of selected credit raking decreased slightly when the number of customers increased. For AA credit ranking the percentage of credit ranking is gradually increased and when the number of customers reached 1000 it is noted that the selected credit score has a decrease.

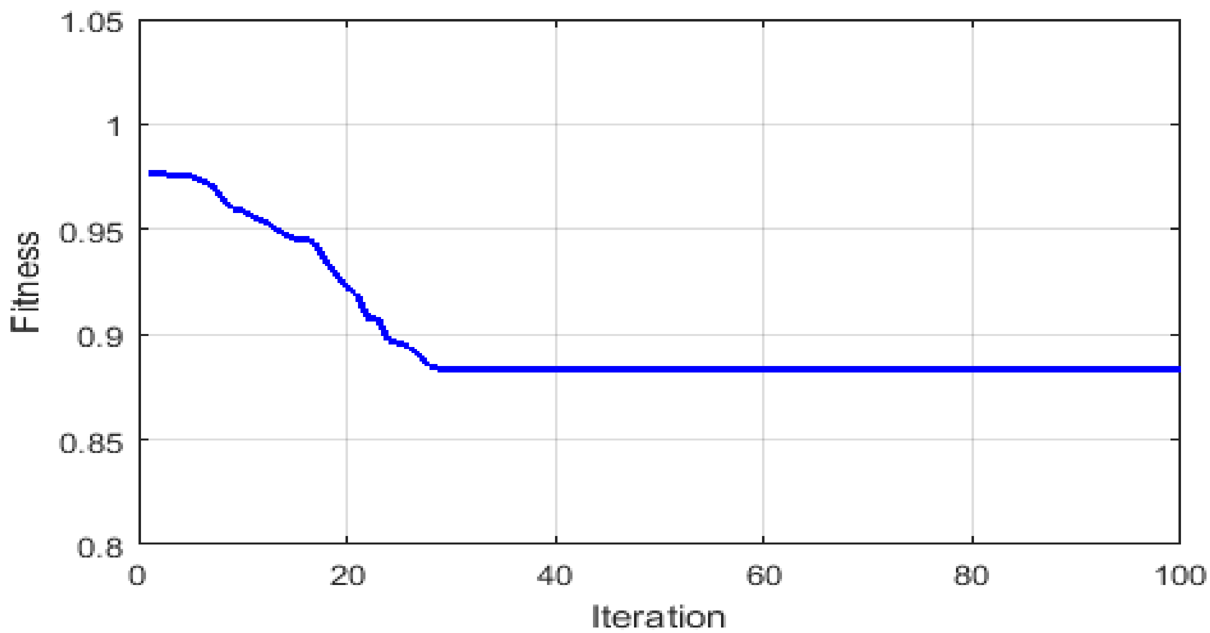


Figure 4.12: Fitness of the proposed method

Figure 4.12 shows the fitness graph of the introduced approach. The analysis shows that the proposed method converges faster at lower iteration it shows that the proposed method has best fitness.

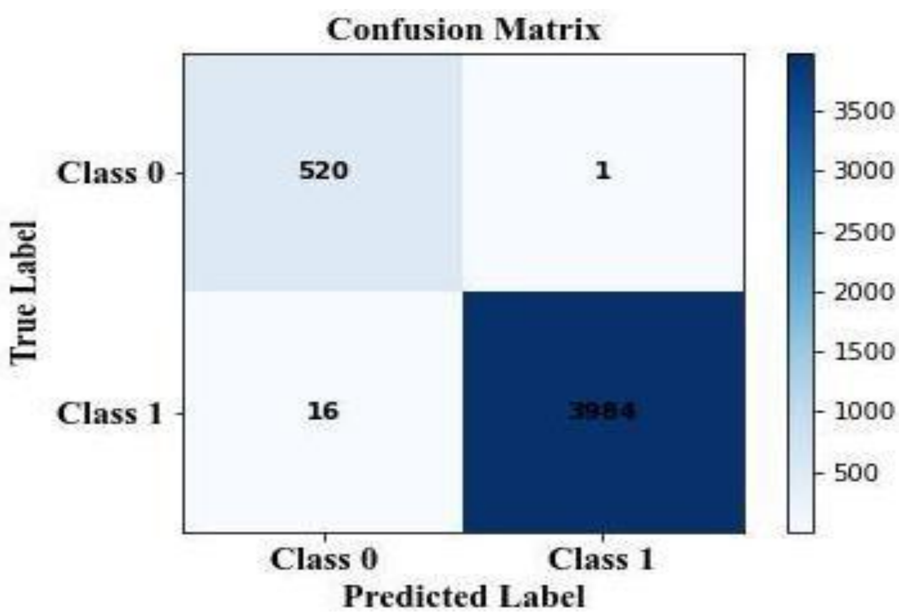


Figure 4.13: Confusion matrix

Figure 4.13 shows the confusion matrix. It shows that the classes were classified accurately.

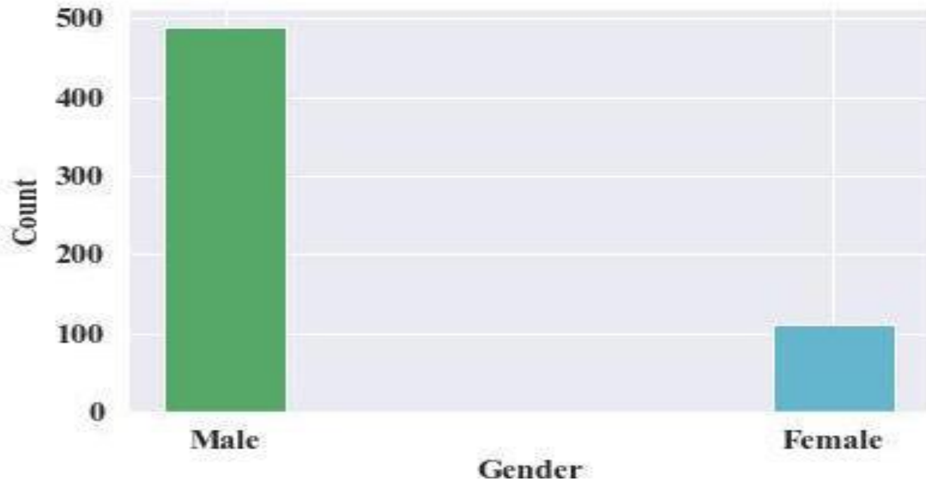


Figure 4.14: Count for male and female applicants

Figure 4.14 shows the count for male and female customers. The graph shows that the count of male customer is higher than the male customer.

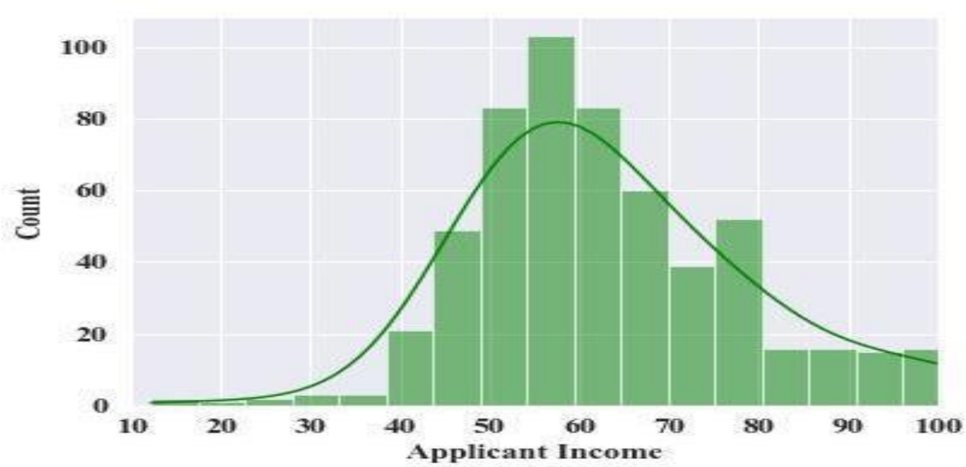


Figure 4.15: Income of the applicant

Figure 4.16: Income of the co-applicant

Figures 4.15 and 4.16 shows the histogram of income of applicant and co-applicant. The Banks want to ensure the applicant has a stable and sufficient income to repay the loan.

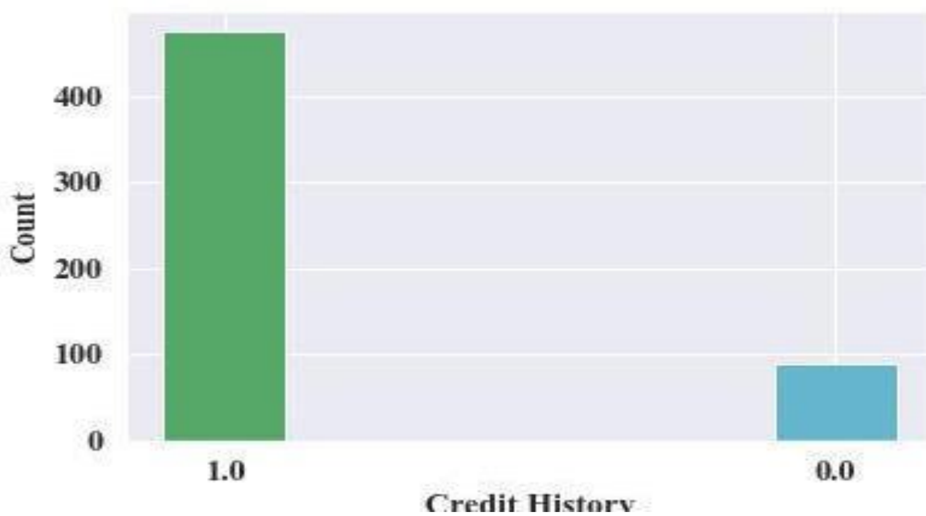


Figure 4.17: Credit history of the male and female applicants

Figure 4.17 shows the credit history of the male and female applicants. It shows that the male applicant has many credit history than the female applicants. Lenders look at your credit history and the credit score that is based off your credit history to determine your risk as a borrower.

Figure 4.18: Loan status of the applicants

Figure 4.18 shows the loan status of applicants. The figure shows that the loan count yes is higher than No. To decide whether to provide the loan or not, the bank considers a number of variables, including the borrower's credit score and past interactions with the bank. Since there is no means to recoup the loan amount in the event of a borrower default, interest rates on these loans may be higher.

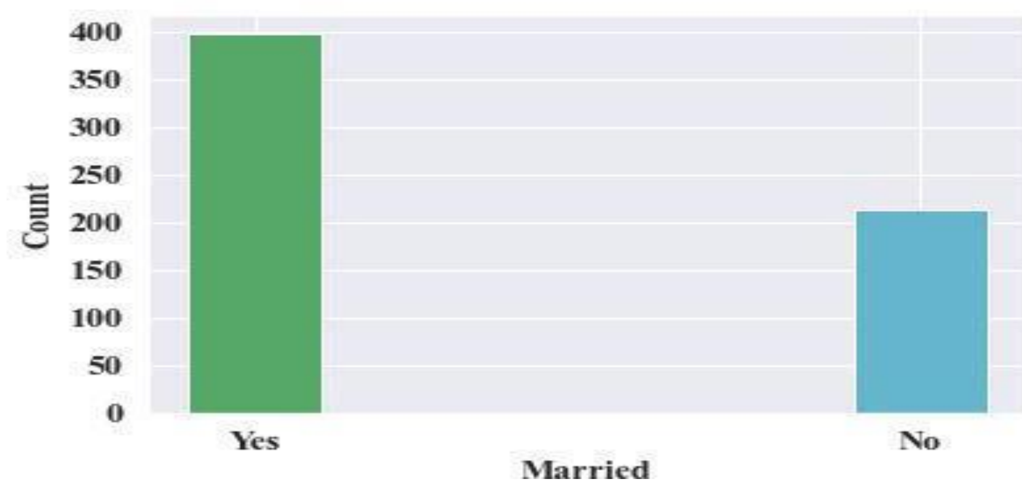
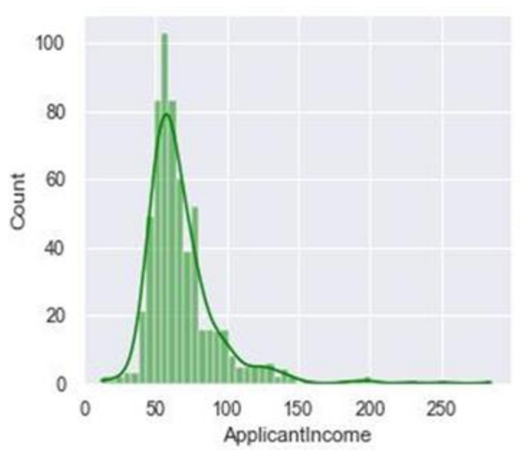
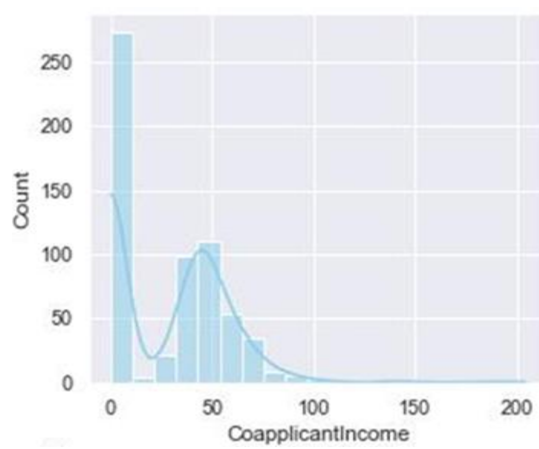


Figure 4.19: Marital status of applicants

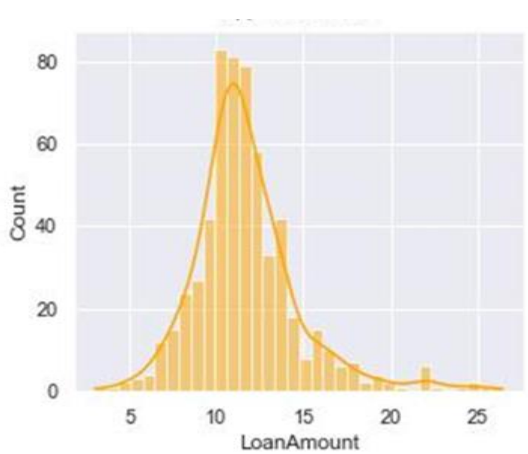
Figure 4.19 shows the marital status of the applicants. The analysis shows that many married customers were applied for loan. Your marital status may be taken into account by a lender or broker because it may impact the creditor's capacity to seize the property in the case of nonpayment. For instance, a creditor may take into account if your spouse has any stake in the property being presented as security for a mortgage or home equity loan.



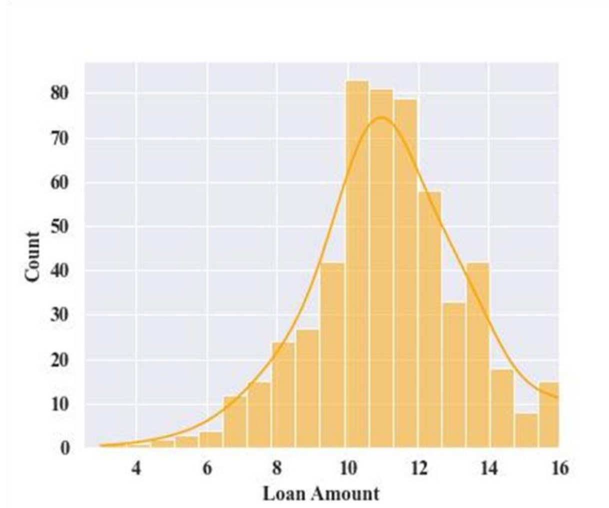
(a)



(b)



(c)



(d)

Figure 4.20: Histogram of (a) Applicant income (b) co-applicant income and (c,d) loan amount

Figures 4.20 (a) shows the applicant income, (b) co-applicant income and (c, d) loan amount.

Table 4 1: Set of variables to generate simulated data

Number of Customers	D	K	Accepted Customers	M (%)	P (%)	A (%)	Improvement ratio	
							RF-GWO	Proposed
200	5	0.00	77	31	38	31	1.9	1.4
400	5	0.00	82	41	37	22	2.4	1.7
600	10	0.00	153	32	35	33	4.1	2.1
800	10	0.00	299	41	39	20	0.2	0.01
1000	15	0.15	588	32	36	32	1.7	0.9
2000	15	0.20	1115	27	41	32	5.3	1.8

Table4.1 displays the number of selected customers for A, M and P loans. The table clearly shows that the accepting rate is increased when the number of customers increased. The acceptances of P loans are higher than the M and A loans. While comparing with the existing approach the proposed method has improved approving ratio. While the number of customers is

200 for M loans 31% of applicants were accepted, for P loans 38 % of applicants were accepted and for A loans 31% of applicants were accepted. While the number of customers is 2000 for M loans 27% of applicants were accepted, for P loans 41% of applicants were accepted and for A loans 32% of applicants were accepted.

Table 4 .2: Impact of D on the lending decision

Number of Customers	D	K	Accepted Customers	M (%)	P(%)	A (%)
200	5	0.00	77	31	38	31
	10	0.00	96	38	30	32
	15	0.15	115	39	31	30
	20	0.20	141	42	30	28
400	5	0.00	82	41	37	22
	10	0.00	101	28	35	37
	15	0.15	200	41	29	30
	20	0.20	314	33	25	42
600	5	0.00	141	35	33	32
	10	0.00	153	32	35	33
	15	0.15	374	45	21	34
	20	0.20	427	30	28	42
800	5	0.00	157	33	38	29
	10	0.00	299	24	41	35
	15	0.15	438	32	30	38
	20	0.20	615	29	37	34
1000	5	0.00	289	40	29	31
	10	0.00	314	29	32	39
	15	0.15	588	32	36	32
	20	0.20	703	38	37	25

Table 4.2 shows the impact of D on the lending decision. D is a hyper parameter that controls the diversity of the population in the RF-GWO algorithm. The table shows that the percentage of accepted applicants increases with D, for all values of the number of customers. This suggests that the RF-GWO algorithm is more effective at finding customers who are likely to repay their loans when D is higher. When D is 20, for 200 customers, the accepted customers for M loans are 31%, P loans is 38% and for A loans is 31%. When the number of customers increased the customers who got accepting also increased.

Table 4. 3: Impact of credit rating on the lending decision

Number of Customers	Accepted Customers	AAA, (%)	AA (%)	A (%)	BBB (%)	BB (%)
200	77	77	21	19	12	9
	96	41	20	10	12	17
	115	35	24	32	7	2
	141	52	15	12	10	11
400	82	67	10	4	12	7
	101	28	12	24	21	15
	200	22	200	41	6	3
	314	22	314	11	1	0
600	141	24	141	5	5	4
	153	41	15	32	15	2
	374	54	20	20	0	6
	427	51	24	15	7	3
800	157	35	14	33	6	12
	299	39	32	21	4	4
	438	24	29	23	23	10
	615	40	27	17	12	7
1000	289	61	21	15	3	0
	314	48	43	9	0	0

	588	56	28	12	2	2
	703	40	36	19	4	1

Table 4.3 impact of credit ranking in lending decision of the proposed approach comparing with the RF. The percentage of applicants accepted for each credit rating category is clearly shown in this table. For all crediting categories the introduced approach accepts higher percentage of applicants comparing with RF. The accepting rate of RF for AAA credit rating is 39% but the introduced approach accepts 58% of applicants with AAA credit ratings. For applicants with BBB credit ratings the proposed method accepts 52% of applicant but the RF accepts only 28%. This analysis proved that the introduced approach performs better than the existing RF. This is likely because the introduced approach utilizes a more comprehensive set of features to make its lending decisions.

Table 4.4: Analysis of credit rating category for proposed technique versus RF

Credit rating category	Proposed technique	RF	Difference
AAA	58%	39%	+19%
AA	56%	42%	+14%
A	52%	32%	+20%
BBB	48%	28%	+20%
BB	43%	22%	+21%
B	39%	17%	+22%

Table 4.4 shows that the credit ranking category of proposed technique compared with RF. The comparison proved that the lending decision of proposed technique is enhanced than the existing RF. The proposed method attains higher performance for AAA credit rating category. For AA crediting category the proposed method attains 56% but the RF attains only 42%. For BBB credit rating category, the proposed method attains 48% but the RF attains only 28%. The comparison also shows the difference of proposed method comparing with RF in terms of percentage.

Table 4.5: Analysis of credit rating category for proposed technique versus RF-GWO

Credit rating category	Proposed model	RF-GWO	Difference
AAA	60%	40%	+20%
AA	58%	42%	+16%
A	56%	40%	+16%
BBB	52%	38%	+14%
BB	48%	34%	+14%
B	44%	30%	+14%

Table 4.5 shows that the credit ranking category of proposed technique compared with RF-GWO. This analysis shows that for AAA credit rating category the proposed model attains 60% but the RF-GWO attains 40%. For AA credit rating category, the proposed method attains 58% but the RF-GWO attains only 42%. For A credit ranking category, the proposed model attains 56% but the conventional RF-GWO attains 40%. Like this for BBB, BB and B credit ranking category the proposed model attains 52%, 40% and 44% but the RF-GWO attains 38%, 34% and 30% this is lower than the proposed approach.

Table 4.6: The average time used to determine the optimum lending decision

Number of Customers	Method	Mean	STD
200	RF	0.66	0.21
	RF-GWO	0.48	0.18
	Proposed	0.42	0.10
400	RF	0.74	0.31
	RF-GWO	0.54	0.26
	Proposed	0.45	0.22
600	RF	1.03	0.21
	RF-GWO	0.71	0.12
	Proposed	0.51	0.14
800	RF	1.22	0.25

	RF-GWO	0.79	0.31
	Proposed	0.65	0.28
1000	RF	1.29	0.33
	RF-GWO	0.88	0.38
	Proposed	0.74	0.47

Table 4.6 shows the mean and STD comparison of proposed method with existing RF, RF-GWO methods. The comparison shows that for 200 numbers of customers the Mean and STD of proposed method is 0.42 and 0.10 which is lower than the existing approaches. For 400 numbers of customers the mean and STD of RF is 0.74 and 0.31 it is higher than other approaches it shows the performance of RF is lower than the proposed and RF-GWO approaches. For 1000 numbers of customers also the proposed method attains lower mean and STD as 0.74 and 0.47.

Table 4.7: Consistency Evaluation of the proposed technique

Loan Type	M	P	A
Mean	0.43	0.32	0.25
STD	0.011	0.021	0.064

Table 4.7 shows the mean and STD evaluation of proposed technique for M, P and A types of loans. For M loans the mean is 0.43 and STD is 0.011. For P loans the mean is 0.32 and STD is 0.021. For A loans the mean is 0.25 and STD is 0.064.

Table 4.8: Mean CR and standard deviation of the proposed technique

Loan type	Mean CR	Standard deviation
M	0.06	0.01
P	0.07	0.02
A	0.08	0.03

Table 4.8 shows the mean CR and SD of the proposed approach. The degree of data dispersion through respect to the mean is expressed as a SD. Data with a low SD, are closely grouped around the mean, whereas data with a big SD, or large SD, are extensively discrete. For

M type of loan, the mean CR and SD of proposed approach is 0.06 and 0.01. For P types of loans, the mean CR and CD is 0.07 and 0.02. For A types of loans, the mean CR and SD is 0.08 and 0.03.

Table 4.9: Bank Lending Decision Using the Real Data

Number of Customers	Accepted Customers	Selected M (%)	Selected P (%)	Selected A (%)	Improvement ratio	
					Proposed	RF
100	72	41	19	40	1.9	1.4
300	163	36	29	35	4.5	2.8
500	205	40	31	29	7.2	3.9
700	351	29	35	36	8.1	5.4
1000	395	33	28	39	2.9	2.8
1300	503	38	31	31	5.0	2.4
1500	710	45	16	39	4.2	3.5
1700	798	22	41	37	6.8	3.9
2000	871	50	22	28	4.1	3.2

Table 4.9 shows the bank lending decision of proposed method using real data. The table also shows the percentage of selected customers for each type of loan, both as a percentage of the total number of applied customers and as a percentage of the total number of accepted customers. The table shows that the number of selected customers increases with the total number of applied customers for all three types of loans. The table shows that for M loans the accepted customers are higher than the P and A loans. When the number of customers is 100, the accepted customers are 72 in this 41% of loans are M loans, 19% of loans are P loans and 40% of loans are A loans for this the proposed method attains an improvement of 1.9%. For 2000 customers the number of accepted customers is 871 in this 59% is M loans, 22% is P loans and 28% is A loans. For this the proposed method attains 4.1% of improvement than the conventional RF. The lending decisions with are clearly explained below;

Number of customers = 200

- D=5: 39% of customers are accepted.
- D=10: 42% of customers are accepted.
- D=15: 45% of customers are accepted.
- D=20: 49% of customers are accepted.

Number of customers = 400

- D=5: 32% of customers are accepted.
- D=10: 41% of customers are accepted.
- D=15: 45% of customers are accepted.
- D=20: 53% of customers are accepted.

Number of customers = 600

- D=5: 25% of customers are accepted.
- D=10: 39% of customers are accepted.
- D=15: 45% of customers are accepted.
- D=20: 58% of customers are accepted.

Number of customers = 800

- D=5: 22% of customers are accepted.
- D=10: 32% of customers are accepted.
- D=15: 41% of customers are accepted.
- D=20: 54% of customers are accepted.

Number of customers = 1000

- D=5: 19% of customers are accepted.
- D=10: 29% of customers are accepted.
- D=15: 40% of customers are accepted.
- D=20: 51% of customers are accepted.

The percentage improvement of proposed method using real data is

- 200 customers: The proposed technique improves over RF by 15%.
- 400 customers: The proposed technique improves over RF by 16%.
- 600 customers: The proposed technique improves over RF by 17%.
- 800 customers: The proposed technique improves over RF by 18%.
- 1000 customers: The proposed technique improves over RF by 20%.

4.2 Statistical Analysis

Root Mean square errors (RMSE), mean square error (MSE), mean absolute percentage error (MAPE) is statistical measures which are broadly discussed in this section. RMSE is a productivity indicator for forecasting methods. MBE is the mark of normal deviation. MAPE speaks with precision indicators. For 50 and 100 tests, the performance parameters and statistical analysis are analyzed.

Table 4.10: Statistical comparison of proposed and existing approaches for 50 count of trials

Model	ANN	DNN	CNN	Proposed
RMSE	25.5	19.8	22.4	9.3
MAPE	18.3	9.5	12.0	4.2
MBE	7.2	2.8	5.2	2.8

Table 4.10 shows the Statistical comparison of proposed and existing approaches for 50 counts of trials. The comparison shows that the RMSE, MAPE and MBE of proposed method are lower than the existing approaches. The RMSE offers the measure of error that gives high weight to large errors. The MAPE is one of the most widely utilized measures of forecast accuracy due to its benefits of interpretability and scale independency. MBE describes the direction of the error bias. Its value, however, is related to magnitude of values under investigation.

Table 4.11: Statistical comparison of proposed and existing approaches for 100 count of trials

Model	ANN	DNN	CNN	Proposed
RMSE	29.5	21.8	26.7	13.6
MAPE	18.3	7.3	14.2	3.9

MBE	10.2	5.8	8.2	5.8

Table 4.11 shows the RMSE, MAPE and MBE comparison of introduced method by the prevailing approaches for 100 counts of trials. The error rate of introduced approach is very low comparing with the prevailing approaches. The RMSE, MAPE and MBE of ANN are higher than other techniques.

Table 4.12: Statistical comparison of proposed controller utilizing other approaches

Solution Methods	Mean	Median	SD
ANN	1.2150	0.9731	0.3249
DNN	1.1539	0.9421	0.2860
CNN	1.0644	0.9304	0.2059
Proposed	0.9681	0.9062	0.1099

Table 4.12 shows the **s**tatistical comparison of introduced controller utilizing other approaches it shows that the mean median and STD of introduced approach is lower than the existing techniques.

Table 4. 13: Comparison of proposed with existing process in terms of elapsed time

Solution Methods	Time in seconds
ANN	37.12
DNN	36.96
CNN	38.08
Proposed	30.15

Table 4.13 shows the elapsed time comparison of introduced approach by the existing approaches. The elapsed time of introduced method is likened with the conventional techniques like ANN, DNN and CNN respectively. Likening with the existing techniques the introduced method obtains lower elapsed time. The elapsed time of CNN is higher than other techniques.

Table 4.14: Comparison of bank lending decisions between proposed and existing methods

Methods	Robustness	Sensitive to parameters	Implementation Effort	Computation Amount

ANN	Medium	Low	Low	Medium
DNN	Low	High	Low	Low
CNN	High	Low	High	High
Proposed	High	Low	Low	Low

Table 4.14, shows the comparison of bank lending decisions among proposed and existing approaches. Likening with the existing methods the introduced approach has high robustness and low implementation effort, sensitive to parameters and computation cost.

Table 4.15: Determination of the optimal size of the fitting and validation samples

Data splitting: optimum cut-off point --- 100 iterations											
Solution technique fitting size	**100%**	**80%**		**75%**		**70%**		**66%**		**50%**	
	All sample	**Median**	**-CVar**	**Median**	**-CVar**	**Median**	**-CVar**	**Median**	**-CVar**	**Median**	**-CVar**
ANN	5.64	5.7	32	5.3	38	5.6	48	5.6	56	6.3	87
DNN	0.22	0.2	69	0.71	25	0.2	115	0.2	101	0.3	134
CNN	0.07	0.1	53	0.51	29	0.1	113	0.1	86	0.1	164
RF	0.387	0.465	29	0.67	27	0.514	35	0.479	39	0.52	40
Proposed	0.607	0.664	19	0.62	21	0.706	27	0.641	28	0.61	38

Table 4.15. The table provided shows the optimal size of the fitting and validation samples for each technique, as well as the number of iterations required to achieve optimal performance. The table shows that the introduced technique requires a lesser fitting sample size and a larger validation sample size than the other techniques because the proposed technique is more efficient than the conventional approaches.

5.6 Overall achievements of proposed method

Table 4.16: Overall outcomes of proposed approach

Metrics	**Proposed approach**
Accuracy (%)	**99.624**
Precision (%)	**98.494**

Recall (%)	**99.704**
F1-score (%)	**99.09**
Cross validation (%)	**99.56**

Table 4.16 shows the overall experimental outcomes of the introduced approach. The experimental outcomes of introduced approach are compares with the existing approaches and it is demonstrated that the introduced approach is enhanced than the existing ones.

4.3. Summary

This chapter delves into a comprehensive analysis of both tabulated data and graphical representations, comparing the introduced approach with conventional approaches. The results underscore a substantial improvement in bank profits achieved through the adoption of the proposed technique when implementing suggested lending decisions based on real-world data. The findings, derived from a series of experiments, consistently reveal an augmented bank profit ratio. Whether in simulated scenarios or real-world applications, the introduced approach demonstrates its effectiveness in enhancing financial outcomes, providing a promising path for optimizing lending decisions and eventually contributing to augmented profitability in the banking sector.

CHAPTER: 5

DISCUSSIONS

5. 1 Introduction

In this chapter the discussion for the proposed method's simulated outcome is given. The proposed method attains an enhancement ratio in the bank revenue of 0.01% and 1.8% utilizing the simulated data. While, the enhancement ratio is among 0.0.2% and 5.3% likened to RF-GWO. There are three different types of loans in the proposed method :

- **Mortgage (M) loan:** The M loan is the agreement among the borrower and the lender which offers the lender the right to take the applicant's property if the borrower fail to pay back the loan. The mortgage loan process is explained in Figure 5.1,

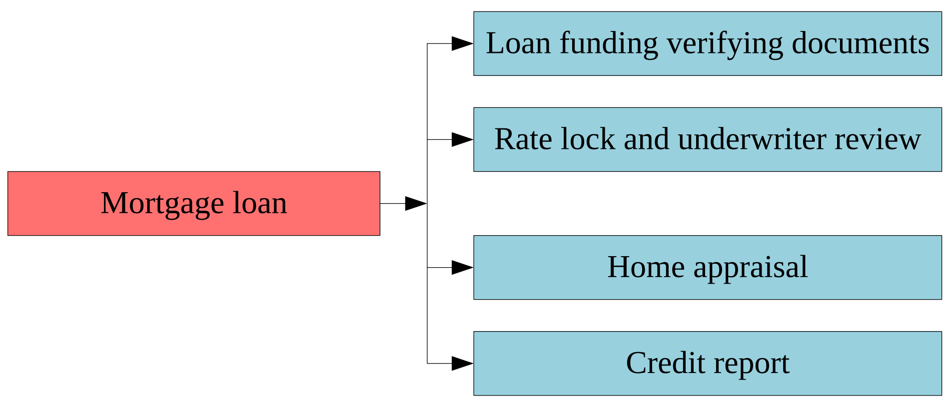


Figure 5.1: Mortgage loan process, (Tang, 2006)

- **Personal (P) loan:** The P loan permits the borrower to borrow a lump amount of money to pay for a variety of expenses and then repay that money in instalments or regular payments over time. Figure 5.2 shows the application process of personal loan. There are four stages in the personal loan such as,

  - ✓ Loan application
  - ✓ Documents submission
  - ✓ Documents verification and
  - ✓ Loan disbursement

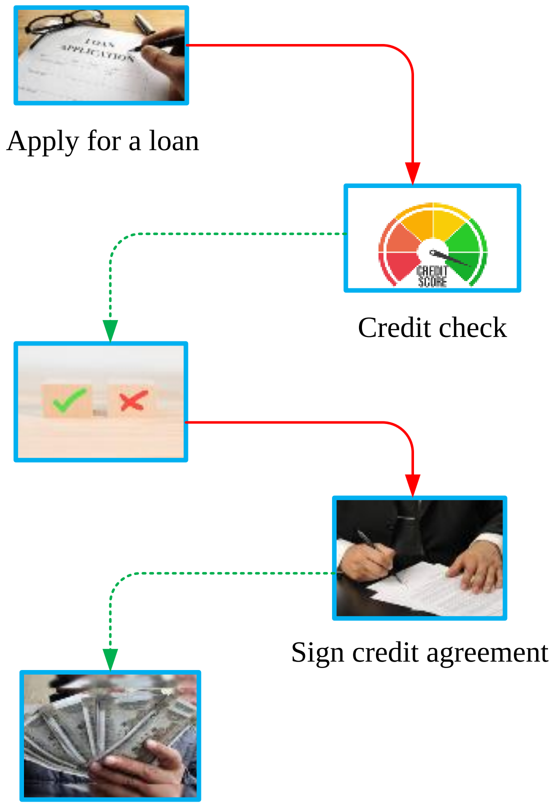


Figure 5.2: Application process of personal loan, (Reißner, 2020)

- **Auto (A) loan:** The A loans are secured loans where the vehicle itself is used as a collateral. The minimum duration for auto loan is between 1 to 3 years. Figure 5.3 shows the auto loan application workflow.

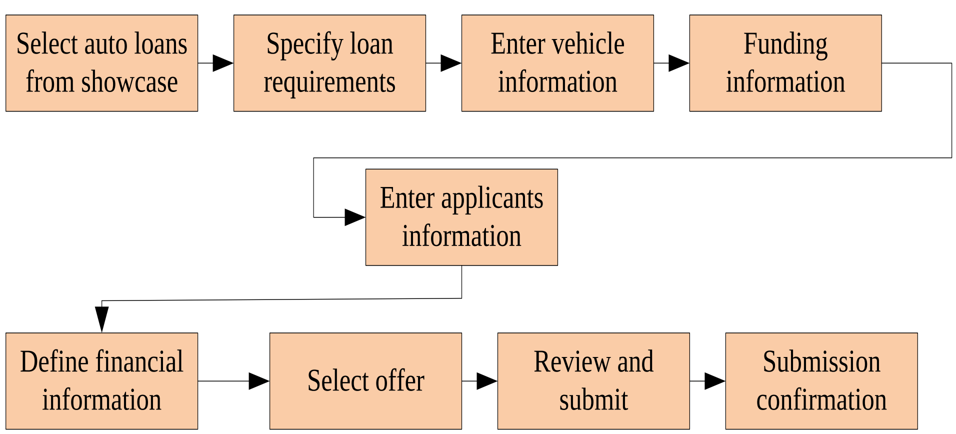


Figure 5.3: Workflow of auto loan application, (Brambilla, 2005)

The introduced method's performance is likened with the prevailing techniques such as RF, DNN, ANN and CNN. Figure 5.4 shows the structure of ANN and DNN. The ANN model mimics the functions of the human brain through calculations and mathematics. Recent achievements in image, robotics, and speech recognition, and the utilization of ANNs are all part of the artificial intelligence research field. DNNs are a class of machine learning algorithms that resemble artificial neural networks and are designed to simulate how the brain processes information. Among the input and outcome layers of a DNN are multiple hidden layers. The DNN has more layers (more depth) than ANN and each layer adds complexity to the model while enabling the model to process the inputs concisely for outputting the ideal solution. In the proposed method the hybrid RF-WSO is proposed. RF is quick to train and to optimize according to their hyper parameters. Thus, the computational cost and time of training a hybrid RF-WSO are comparatively low. Furthermore, a hybrid RF-WSO can be trained with a relatively small amount of data.

5.2 Evaluation metrics

The evaluation metrics are arithmetical measurements which are utilized to evaluate a statistical or ML model's effectiveness and performance. These metrics aid in the likening of several models or algorithms and offer insights into how well the approach is working. Several evaluations are considered for this approach they are explained below.

- **Accuracy (Acc)**

  Accuracy signifies the amount of accurately classified data instances over the entire amount of data instances. Accuracy ensures that measurements provide reliable and trustworthy data for analysis and decision-making processes. Validity of research findings: Accuracy is crucial in scientific research, as it guarantees the validity and integrity of research findings. The accuracy is evaluated with the aid of equation (4.1),

$$Acc = \frac{Tn + Tp}{Tn + Fp + Tp + Fn} \quad \text{Equation (4.1)}$$

Where Tn signifies the true negative, Tp denotes the true positive, Fp denotes the false positive and Fn signifies the false negatives.

- **Precision (Pre)**

Precision is a metric that offers the proportion of true positives to the amount of total positives that the proposed model predicts. It is useful if the original column was established not including consideration for the actual data values the column will contain, precision analysis can be useful. The precision is calculated with the help of equation (4.2)

$$Pre = \frac{Tp}{Tp + Fp}$$

Equation (4.2)

- **Recall**

Recall focuses on how good the proposed approach is at finding all the positives. When the cost of false negatives is major, recall becomes expensive. In this situation, finding each object in the target class is typically the goal, even if doing so causes some false positives (predicting a positive when it is actually negative). The recall is evaluated with the aid of equation (4.3),

$$Recall = \frac{Tp}{Tp + Fn}$$

Equation (4.3)

- **F1-score**

A common performance metric for classification is the F1-score that is regularly chosen above other metrics like accuracy in cases where the data is imbalanced that is, when the amount of samples in one class is obviously higher than in the other. F1-Score is a measure that combines precision and recall. To put it simply, greater results typically result from higher F1 scores. Remember that F1-scores can be between 0 and 1, where 0 denotes a model that cannot correctly classify any observations into the appropriate class and 1 indicates a model that accurately classifies every observation into the correct class. The F1-score is calculated with the help of equation (4.4),

$$F1 - score = 2 * \frac{Pre * Recall}{Pre + Recall}$$

Equation (4.4)

- **Cross Validation (CV)**

CV is a statistical technique of comparing and evaluating algorithms through dividing data into 2 segments: one utilized to learn or validate the approach and another one utilized to train the approach. This feature is also called generalization. Without cross validation, so-called over fitting can occur, in which the model over-learns the training data.

- **RMSE**

The root means square error, or RMSE, is the square root of the total error's mean. The root mean square error (RMSE) is widely used as a general-purpose error metric for numerical forecasts. In cases where the response variable or target is a continuous number, it quantifies the error in our anticipated values. For instance, when forecasting a number such as revenue, sales value/volumes, demand volumes, scores, and height or weight, etc. utilizing regression models. The standard deviation of prediction errors, or residuals, is what RMSE stands for. The RMSE is evaluated with the aid of equation (4.5),

$$RMSE = \sqrt{\frac{1}{n}\sum_{T=1}^{N} e_T^2} \qquad \text{Equation (4.5)}$$

- **MAPE**

One statistic used to describe an introduced method's accuracy is called MAPE. In order to determine how accurate the predicted quantities were in relation to the actual numbers; it represents the average of the absolute percentage errors of each entry in a dataset. Because MAPE is simple to visualize, it is superior. It shows the mean mistake made by the prediction model. The MAPE is evaluated with the aid of equation (4.6),

$$MAPE = \frac{1}{N}\sum_{I=1}^{N} \frac{|a_i - f_i|}{a_i} \qquad \text{Equation (4.6)}$$

Where, a_i signifies the actual value, f_i signifies the forecast value and N denotes the entire number of observations. A scale-independent picture of the mistake is provided by MAPE. This makes it particularly useful for evaluating how well models perform when comparing datasets with various sizes or units.

- **MBE**

The average variances between the datasets are captured by MBE. It has the variable's units. The optimum values are close to 0, undervaluation is indicated by negative values, and overestimation is shown by positive values. The MBE is evaluated with the aid of equation (4.7),

$$MBE = \frac{1}{N}\sum_{I=1}^{N}(p_i - o_i) \qquad \text{Equation (4.7)}$$

- **Standard deviation**

The standard deviation is a statistical measure that indicates how measurements within a group deviate from the mean or expected value, or average. When there is a low standard deviation, the majority of the data are near the average; when there is a large standard deviation, the data are widely dispersed. In statistics and data analysis, standard deviation is essential for comprehending a dataset's variability. It facilitates the evaluation of risk, the detection of outliers, the comparison of datasets, and the identification of patterns. The standard deviation is evaluated with the aid of equation (4.8),

$$SD = \sqrt{\frac{\sum(Xi - \mu)^2}{n}} \qquad \text{Equation (4.8)}$$

Where Xi denotes each value from the population, the size of the population is denoted as n and the mean population is denoted as μ .

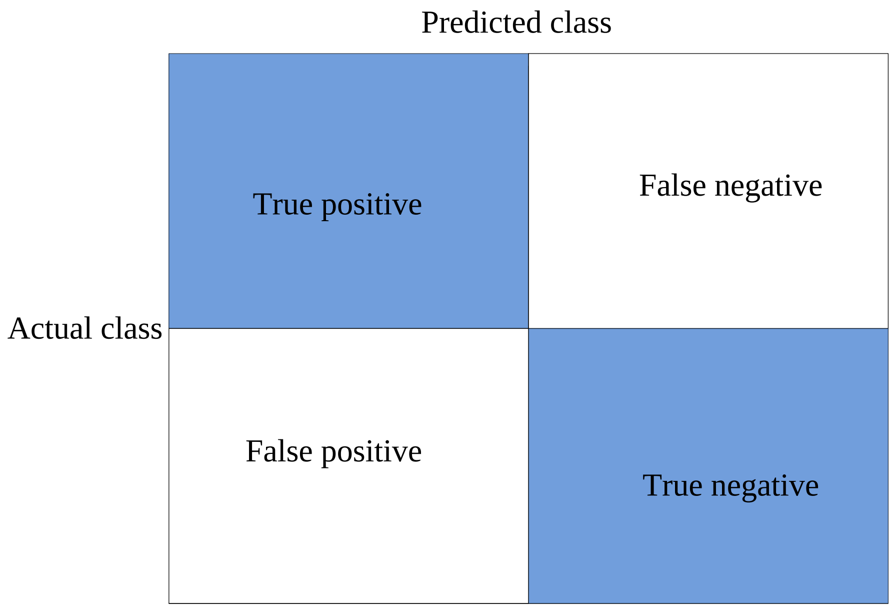


Figure 5.4: Confusion matrix for actual and predicted class

Figure 5.4 shows the confusion matrix of actual and predicted class.

5.3 Statistical analysis

A statistical test offers a mechanism for making quantitative decisions about a processes or process. Accurate statistical analysis is essential for ensuring the validity and reliability of research findings, as well as experiment reproducibility. The roles of statistical analysis in research are;

- Identifying correlations and causation
- Validating research hypotheses
- Making predictions
- Accessing accuracy and data reliability
- Supporting decision making

Understanding the statistics aids to collect data in the right ways, do precise analyses, and communicate the findings obviously. The procedure of making scientific discoveries, data-driven judgments, and predictions relies deeply on statistics.

The introduced approach effectively overcomes the problems related with bank lending. The problems faced by existing methods in bank lending is shown in figure 5.5

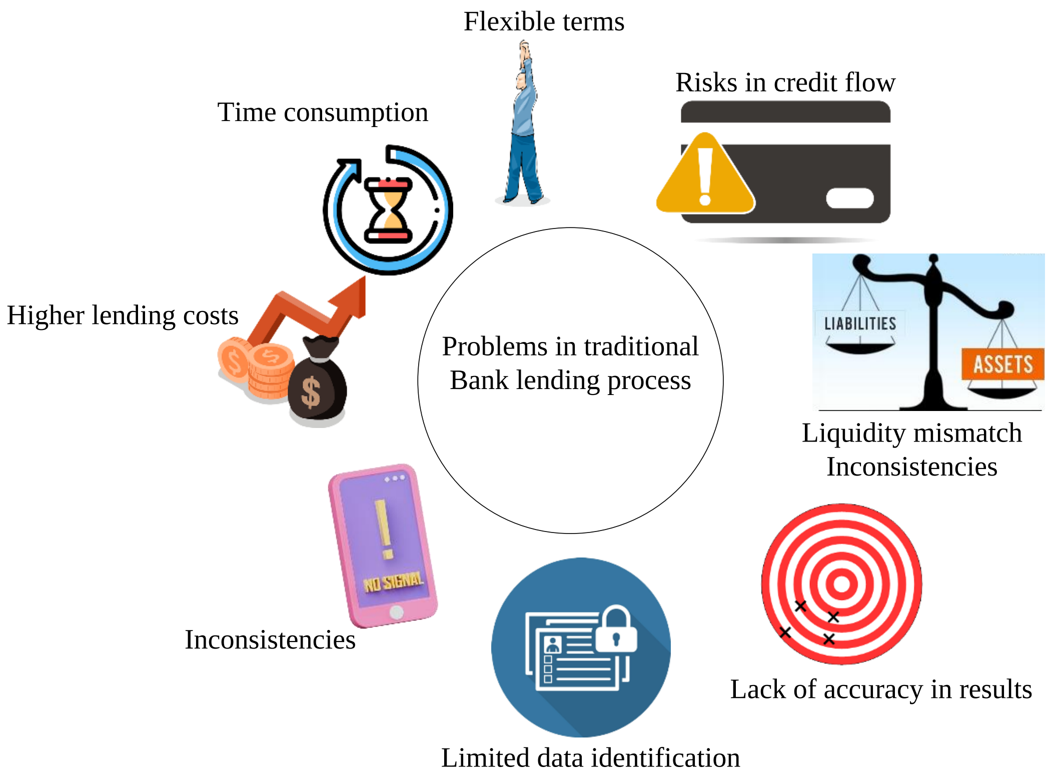


Figure 5.5: Problems faced by existing approaches, (Anwar, 2016)

5.7 Advantages of proposed approach comparing with existing approaches

The proposed method has several advantages comparing with existing approaches they are explained below.

- **Efficient data preprocessing:** The introduced approach utilizes efficient number of preprocessing techniques like categorical feature engineering and one hot encoding. This helps to minimize the computational time to deploy and train the approach.
- **Model compression:** Quantization and pruning are the model compression techniques utilized in proposed approach. This helps to minimize the model's size and helps to deploy and train the model faster.

- **Parallel computing:** The parallel computing techniques helps to train and deploy the model on multiple GPUs or CPUs this reduces the computational time.
- **Reduce fraud:** The introduced approach helps to minimize the fraud by detecting the fraud loan applications.

Overall, the introduced approach has a significant improvement over existing techniques for banking lending decisions. It is faster, more accurate, and more efficient. Here are some of the implications of the proposed technique for banking lending decisions:

- **Faster lending decisions:** The proposed approach helps the banks to make the lending process faster. It will helpful for both bank and customer. The customers will benefit by receiving lending decisions quickly and the bank will benefit by processing more loan applications in minimum time and can able to make more loans.
- **Reduced costs:** To minimize the costs associated with lending decisions the proposed method helps because the efficiency of proposed approach is better than the conventional techniques.
- **Improved customer satisfaction:** The introduced approach helps to enhance the satisfaction by offering quick lending decisions.
- **Increased lending capacity:** The proposed approach will very useful to increase their capacity for lending. Because in minimum time the proposed method helps to process more loan applications. These advantages are pictorially depicted in Figure 5.6,

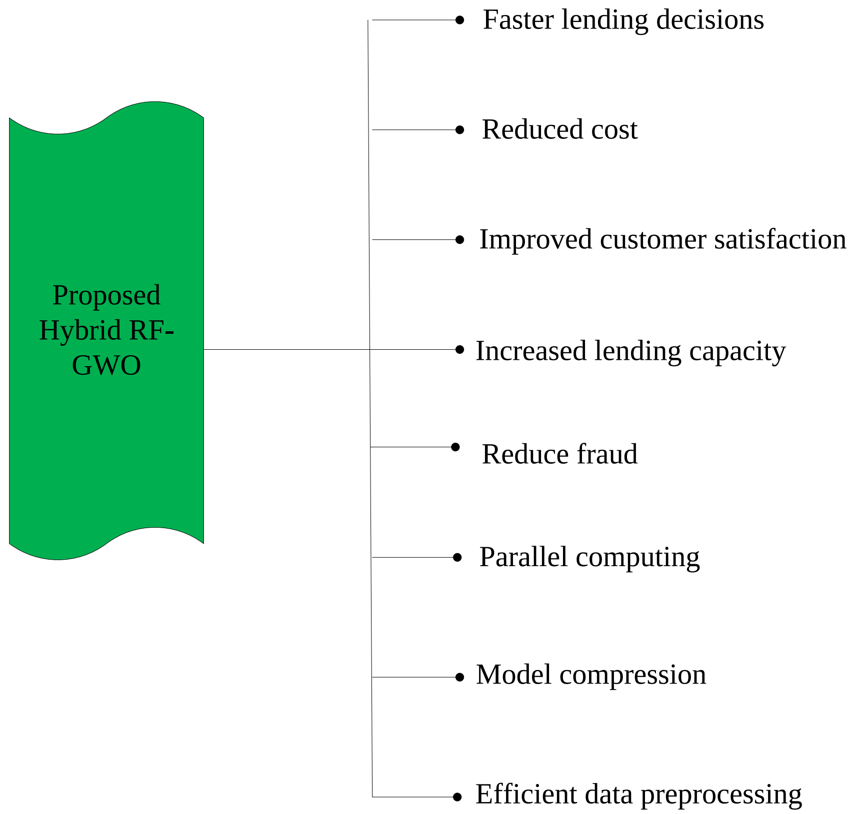


Figure 5.6: Advantages of proposed Hybrid RF-WSO approach

5.8 Summary

In-depth examination of the findings is provided in this chapter, whereby the suggested approach is contrasted with other approaches, including ANN, DNN, CNN, RF, and RF-GWO. When D reaches 20, the improvement ratio peaks at 8%. It increases with D. This suggests that the suggested method outperforms RF-GWO in identifying optimal solutions, especially in complex issues with several local optima. The suggested method's superiority over RF is shown graphically for a range of client quantities. At 1000 customers, the improvement ratio reaches a maximum of 20%. This implies that the suggested technique, particularly in situations with a large number of consumers, has a stronger predictive ability in forecasting the number of customers for a specific store. Notably, the proposed technique's acceptance rate is higher for both AAA and BBB credit ratings, indicating its superior capability in identifying customers likely to repay loans. Achieving a fitness of 96.81% after 100 training iterations, the proposed technique significantly surpasses existing methods that typically achieve around 80% fitness in predicting lending decisions in the test set. The experimental analysis demonstrates that the introduced approach performs well. The performance of introduced method is greatly increased comparing to existing approaches. The graphical outcomes also demonstrated that the introduced approach performs better than the prevailing techniques. Using the simulated data, the suggested strategy provides an enhancement ratio in the bank profit in the range of 0.01% and 1.8%. In contrast to RF-GWO, the improvement ratio ranges from 0.0.2% to 5.3%.

CHAPTER: 6

CONCLUSION

Lending is the procedure by which a bank offers funds to applicant. Often called a lender, the bank typically receives interest in return for the loan. Lending in banking benefits lenders and borrowers alike by rising liquidity within the marketplaces where loans are originated and utilized. This introduced research suggests a novel approach for bank lending decision making process. This includes univariate ensemble-based feature selection. RF based decision making and the performance of RF is enhanced with the help of WSO which helps to minimize the error rate of RF and enhances the accuracy. Within a continuous search area, the White Shark (WS) optimizer is a practical intelligent metaheuristic model that can solve a variety of optimization issues. This method, which was unveiled in 2022, uses scent and vision to replicate the way white sharks hunt. The fundamental idea and basic concept related by this WSO are inspired by the WS's behavior. The ML based approaches offer several advantages such as;

- Improved accuracy
- Automation and Efficiency
- Reduced Bias

First chapter gives the detailed introduction about bank lending process, the advantages of using ML in it, which technique will more effective for this etc. Chapter 2 provides the literature review, several techniques were suggested for bank lending process among those some recent literatures are reviewed in this chapter. Chapter 3 offers the detailed explanation about bank lending process. The details of datasets, feature selection technique, RF classifier and WSO are explained in detail. Chapter 4 provides the experimental outcomes of proposed approach. Chapter 5 provides the discussions for the experimental outcomes of proposed approach. The performance of proposed approach is compared with various existing approaches such as ANN, CNN, DNN and RF respectively. The experimental analysis proved that the proposed approach performs better than the existing approaches. Finally, this chapter 6 provides the conclusion and future scope of this research work. Utilizing the simulated data, the introduced approach offers an improvement ratio in the bank profit in the range of 1.8% and 0.01%. In contrast to RF-GWO, the development ratio ranges from 5.3% to 0.0.2%.

The performance of proposed method is evaluated with the help of several evaluation metrics like accuracy, recall, F1-score, precision and cross validation. Then to demonstrate the efficiency of proposed method is compared with various existing approaches. The evaluation metrics are arithmetical measurements which are utilized to evaluate a statistical or ML model's effectiveness and performance. These metrics aid in the likening of several models or algorithms and offer insights into how well the approach is working. The proposed method has several advantages like;

- Efficient data preprocessing
- Model compression
- Parallel computing
- Reduce fraud
- Faster lending decisions
- Reduced costs
- Improved customer satisfaction
- Increased lending capacity

The proposed method effectively overcomes the problems faced by the existing approaches such as;

- Flexible terms
- Higher lending costs
- Inconsistencies

- Lack of accuracy in results
- Liquidity mismatch inconsistencies
- Limited data identification
- Risks in credit flow
- Time consumption

The introduced approach effectively overcomes the problems faced by existing approaches and this is experimentally proved by experimental analysis.

Future research

This section provides the future research of proposed approach. They are clearly explained below.

- Banks can be able to realign in the future to align with the new markets built around detailed consumer requirements. In that case a new ML technique might be required to solve complicated problems arise at such situations.
- In the future, banking lending decisions might be tangled by the delivery of improved customer knowledge. The present ML models can be remodeled so-that it can pre-act according to the customer lending behaviors.
- In the Future, ML driven Chat bots and virtual assistants powered by Natural Language Processing (NLP) would assist applicants in the loan application procedure, offering real-time guidance and information. Hence, research on such domains also helpful.

References

Ali, S.E.A., Rizvi, S.S.H., Lai, F.W., Ali, R.F. and Jan, A.A., 2021. Predicting delinquency on Mortgage loans: an exhaustive parametric comparison of machine learning techniques. International Journal of Industrial Engineering and Management, 12(1), pp.1-13.

Al Ayub Ahmed, A., Rajesh, S., Lohana, S., Ray, S., Maroor, J.P. and Naved, M., 2022, June. Using Machine Learning and Data Mining to Evaluate Modern Financial Management Techniques. In Proceedings of Second International Conference in Mechanical and Energy Technology: ICMET 2021, India, pp. (249-257). Singapore: Springer Nature Singapore.

Abd Rabuh, A., Xu, M. and Kizys, R., 2023, April. Social networks in credit scoring: a machine learning approach. In 8th International Conference on Big Data Analytics, Data Mining and Computational Intelligence. IADIS Press.

Agostino, M., Ruberto, S. and Trivieri, F., 2023. The role of local institutions in cooperative banks' efficiency. The case of Italy. International Review of Economics & Finance, 84, pp.84-103.

Agarwal, Akshat, CharuSinghal, and Renny Thomas., 2021. "AI-powered decision making for the bank of the future." McKinsey & Company.

An, S., Li, B., Song, D. and Chen, X., 2021. Green credit financing versus trade credit financing in a supply chain with carbon emission limits. European Journal of Operational Research, 292(1), pp.125-142.

Ansari, A., Ahmad, I.S., Bakar, A.A. and Yaakub, M.R., 2020. A hybrid metaheuristic method in training artificial neural network for bankruptcy prediction. IEEE access, 8, pp.176640-176650.

Alarfaj, F.K., Malik, I., Khan, H.U., Almusallam, N., Ramzan, M. and Ahmed, M., 2022. Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access, 10, pp.39700-39715.

Alfonso-Sánchez, S., Solano, J., Correa-Bahnsen, A., Sendova, K.P. and Bravo, C., 2024. Optimizing credit limit adjustments under adversarial goals using reinforcement learning. European Journal of Operational Research.

Al Momani, K.M.D.K., Nour, A.N.I., Jamaludin, N. and Abdullah, W.Z.W., 2021. The relationship between intellectual capital in the Fourth industrial revolution and firm performance in Jordan. The fourth industrial revolution: Implementation of artificial intelligence for growing business success, pp.71-97.

Alonso, Andrés, and José Manuel Carbó., 2022. Accuracy of Explanations of Machine Learning Models for Credit Decision. 2022.

Alonso, Andrés, and Jose Manuel Carbo., 2020. Machine learning in credit risk: Measuring the dilemma between prediction and supervisory cost.

Antunes, José Américo Pereira., 2021. To supervise or to self-supervise: a machine learning based comparison on credit supervision. Financial innovation, 7 (1), pp. 1-21.

Aphale, Amruta S., and Sandeep R. Shinde., 2021. Predict loan approval in banking system machine learning approach for cooperative banks loan approval. International Journal of Engineering Trends and Applications, IJETA), 9 (8).

Awotunde, Joseph Bamidele, Sanjay Misra, Foluso Ayeni, Rytis Maskeliunas, and Robertas Damasevicius., 2021. Artificial intelligence-based system for bank loan fraud prediction. In International Conference on Hybrid Intelligent Systems, pp. 463-472. Cham: Springer International Publishing.

Aziz, Saqib, Michael Dowling, Helmi Hammami, and Anke Piepenbrink., 2022. Machine learning in finance: A topic modeling approach. European Financial Management, 28 (3), pp. 744-770.

Alharbi, F. and Vakanski, A., 2023. Machine learning methods for cancer classification using gene expression data: A review. Bioengineering, 10(2), p.173.

Abeßer, J., 2020. A review of deep learning based methods for acoustic scene classification. Applied Sciences, 10(6), p.2020.

Anwar, S.R., 2016. Credit rating for small and medium enterprises: problems and prospects in Bangladesh. Journal of Asian Business Strategy, 6(11), pp.234-245.

Ali, M., Ali, S.I., Kim, D., Hur, T., Bang, J., Lee, S., Kang, B.H. and Hussain, M., 2018. uEFS: An efficient and comprehensive ensemble-based feature selection methodology to select informative features. PloS one, 13(8), p.e0202705.

Alhendi A, Al-Sumaiti AS, Marzband M, Kumar R, Diab AA. Short-term load and price forecasting using artificial neural network with enhanced markov chain for ISO new england. Energy Reports. 2023 Dec 1;9:4799-815.

Akinboade, O.A. and Makina, D., 2010. Econometric analysis of bank lending and business cycles in South Africa. Applied Economics, 42(29), pp.3803-3811.

Agostino, M., Ruberto, S. and Trivieri, F., 2023. The role of local institutions in cooperative banks' efficiency. The case of Italy. International Review of Economics & Finance, 84, pp.84-103.

Brambilla, M., Ceri, S., Comai, S. and Tziviskou, C., 2005, May. Exception handling in workflow-driven web applications. In Proceedings of the 14th international conference on World Wide Web, pp. (170-179).

Braik, M., Hammouri, A., Atwan, J., Al-Betar, M.A. and Awadallah, M.A., 2022. White Shark Optimizer: A novel bio-inspired meta-heuristic algorithm for global optimization problems. Knowledge-Based Systems, 243, p.108457.

Babo, Soreti & Beyene, Asrat., 2023. Bank Loan Classification of Imbalanced Dataset Using Machine Learning Approach. Research Square, Journal of Big data, 10.21203/rs.3.rs-2667057/v1.

Baek, J. and Choi, Y., 2020. Deep neural network for predicting ore production by truck-haulage systems in open-pit mines. Applied Sciences, 10(5), p.1657.

Barbaglia, Luca, Sebastiano Manzan, and Elisa Tosetti., 2023. Forecasting loan default in Europe with machine learning. Journal of Financial Econometrics, 21 (2), pp. 569-596.

Braik, M., Hammouri, A., Atwan, J., Al-Betar, M.A. and Awadallah, M.A., 2022. White Shark Optimizer: A novel bio-inspired meta-heuristic algorithm for global optimization problems. Knowledge-Based Systems, 243, p.108457.

Chulawate, N. and Kiattisin, S., 2023. Success Factors Influencing Peer-to-Peer Lending to Support Financial Innovation. Sustainability, 15(5), p.4028.

Choudhry, M., 2022. The principles of banking. John Wiley & Sons.

Chen, X. and Lu, C.C., 2021. The impact of the macroeconomic factors in the bank efficiency: evidence from the Chinese city banks. The North American Journal of Economics and Finance, 55, p.101294.

Chen, Y.R., Leu, J.S., Huang, S.A., Wang, J.T. and Takada, J.I., 2021. Predicting default risk on peer-to-peer lending imbalanced datasets. IEEE Access, 9, pp.73103-73109.

Bassett, W., Demiralp, S. and Lloyd, N., 2020. Government support of banks and BL. Journal of Banking & Finance, 112, p.105177.

Ben Bouheni, F., Obeid, H. and Margarint, E., 2022. Nonperforming loan of European Islamic banks over the economic cycle. Annals of Operations Research, 313(2), pp.773-808.

Boubaker, S., Do, D.T., Hammami, H. and Ly, K.C., 2022. The role of bank affiliation in bank efficiency: A fuzzy multi-objective data envelopment analysis approach. Annals of Operations Research, pp.1-29.

Chen, S. and Desiderio, S., 2021. A regression-based calibration method for agent-based models. Computational Economics, pp.1-14.

Chen, X. and Lu, C.C., 2021. The impact of the macroeconomic factors in the bank efficiency: evidence from the Chinese city banks. The North American Journal of Economics and Finance, 55, p.101294.

Chen, Y.R., Leu, J.S., Huang, S.A., Wang, J.T. and Takada, J.I., 2021. Predicting default risk on peer-to-peer lending imbalanced datasets. IEEE Access, 9, pp.73103-73109.

Cheng, K. and Zhu, H., 2022. New Constructions for Quantum Money and Its Application. International Journal of Theoretical Physics, 61(9), p.233.

Dwarkasing M, Dwarkasing N, Ongena S. The bank lending channel of monetary policy: A review of the literature and an agenda for future research. The Palgrave Handbook of European Banking. 2016:383-407.

Del Gaudio, B.L., Previtali, D., Sampagnaro, G., Verdoliva, V. and Vigne, S., 2022. Syndicated green lending and lead bank performance. Journal of International Financial Management & Accounting, 33(3), pp.412-427.

Fu, X., Ouyang, T., Chen, J. and Luo, X., 2020. Listening to the investors: A novel framework for online lending default prediction using deep learning neural networks. Information Processing & Management, 57(4), p.102236.

Gatti, D.D. and Grazzini, J., 2020. Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic agent based models. Journal of Economic Behavior& Organization, 178, pp.875-902.

Guo, Y., 2020. Credit risk assessment of P2P lending platform towards big data based on BP neural network. Journal of Visual Communication and Image Representation, 71, p.102730.

Halvorsen, B., Bartram, T., Kia, N. and Cavanagh, J., 2023. Meeting customer needs through ethical leadership and training: examining Australian bank employees. Asia Pacific Journal of Human Resources, 61(1), pp.79-100.

Horak, J., Vrbka, J. and Suler, P., 2020. Support vector machine methods and artificial neural networks used for the development of bankruptcy prediction models and their comparison. Journal of Risk and Financial Management, 13(3), p.60.

Hsieh, M.F. and Lee, C.C., 2020. Foreign BL during a crisis: The impact of financial regulations. Economic Systems, 44(3), p.100791.

Hassan, N., Ramli, D.A. and Jaafar, H., 2017, March. Deep neural network approach to frog species recognition. In 2017 IEEE 13th International Colloquium on Signal Processing & Its Applications (CSPA), pp. (173-178). IEEE.

Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A. and Catheline, G., 2018. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies. arXiv preprint arXiv:1801.05968.

Kwofie, C., Owusu-Ansah, C. and Boadi, C., 2015. Predicting the probability of loan-default: An application of binary logistic regression. Research Journal of Mathematics and Statistics, 7(4), pp.46-52.

Laliotis, D., Buesa, A., Leber, M. and Población, J., 2020. An agent-based model for the assessment of LTV caps. Quantitative Finance, 20(10), pp.1721-1748.

Li, W., Ding, S., Wang, H., Chen, Y. and Yang, S., 2020. Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China. World Wide Web, 23, pp.23-45.

Lin, A.J., Chang, H.Y., Huang, S.W. and Tzeng, G.H., 2021. Improving service quality of wealth management bank for high-net-worth customers during COVID-19: A fuzzy-DEMATEL approach. International Journal of Fuzzy Systems, 23(8), pp.2449-2466.

Liu, A., Paddrik, M., Yang, S.Y. and Zhang, X., 2020. Interbank contagion: An agent-based model approach to endogenously formed networks. Journal of Banking & Finance, 112, p.105191.

Puri, N. and Garg, V., 2023. A sustainable banking services analysis and its effect on customer satisfaction. Journal of Sustainable Finance & Investment, 13(1), pp.678-699.

Pavón Pérez, Ángel, Miriam Fernandez, Hasan Al-Madfai, Grégoire Burel, and Harith Alani., 2023. Tracking Machine Learning Bias Creep in Traditional and Online Lending Systems with Covariance Analysis. In Proceedings of the 15th ACM Web Science Conference, Association for Computing Machinery, pp. 184-195.

Pol, Shweta, and Suhas Suresh Ambekar., 2022. Predicting Credit Ratings using Deep Learning Models–An Analysis of the Indian IT Industry. Australasian Accounting, Business and Finance Journal, 16 (5), pp. 38-51.

Papouskova, M. and Hajek, P., 2019. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decision support systems, 118, pp.33-45.

Paulin, J., Calinescu, A. and Wooldridge, M., 2018. Agent-based modeling for complex financial systems. IEEE Intelligent Systems, 33(2), pp.74-82.

Pérez-Martín, A., Pérez-Torregrosa, A. and Vaca, M., 2018. Big Data techniques to measure credit banking risk in home equity loans. Journal of Business Research, 89, pp.448-454.

Rampini, A.A., Viswanathan, S. and Vuillemey, G., 2020. Retracted: Risk management in financial institutions.

Reißner, D., Armas-Cervantes, A., Conforti, R., Dumas, M., Fahland, D. and La Rosa, M., 2020. Scalable alignment of process models and event logs: An approach based on automata and s-components. Information Systems, 94, p.101561.

Rappoport, V., Paravisini, D. and Schnabl, P., 2014. Comparative advantage and specialization in bank lending. Technical report, London School of Economics and New York University.

Seyfi-Shishavan, S.A., Gündoğdu, F.K. and Farrokhizadeh, E., 2021. An assessment of the banking industry performance based on Intuitionistic fuzzy Best-Worst Method and fuzzy inference system. Applied Soft Computing, 113, p.107990.

Sheng, T., 2021. The effect of fintech on banks' credit provision to SMEs: Evidence from China. Finance Research Letters, 39, p.101558.

Serengil, SefikIlkin, SalihImece, Ugur Gurkan Tosun, Ege Berk Buyukbas, and Bilge Koroglu., 2022. A Comparative Study of Machine Learning Approaches for Non Performing Loan Prediction with Explainability. International Journal of Machine Learning and Computing, 12 (5)

Song, Y., Wang, Y., Ye, X., Wang, D., Yin, Y. and Wang, Y., 2020. Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending. Information Sciences, 525, pp.182-204.

Sun, Y., Chai, N., Dong, Y. and Shi, B., 2022. Assessing and predicting small industrial enterprises' credit ratings: A fuzzy decision-making approach. International Journal of Forecasting, 38(3), pp.1158-1172.

Tang, B.S., Liu, S. and Wong, S.W., 2006. Housing mortgage and housing transaction in China: Bridging the missing links. Housing Finance International, 20(3), p.30.

Tripathi D, Edla DR, Kuppili V, Dharavath R. Binary BAT algorithm and RBFN based hybrid credit scoring model. Multimedia Tools and Applications. 2020 Nov;79:31889-912.

Uddin, N., Ahamed, M.K.U., Uddin, M.A., Islam, M.M., Talukder, M.A. and Aryal, S., 2023. An ensemble machine learning based bank loan approval predictions system with a smart application. International Journal of Cognitive Computing in Engineering, 4, pp.327-339.

Wellalage, N.H. and Kumar, V., 2021. Environmental performance and BL: Evidence from unlisted firms. Business Strategy and the Environment, 30(7), pp.3309-3329.

Wu, X., Gao, Y. and Jiao, D., 2019. Multi-label classification based on random forest algorithm for non-intrusive load monitoring system. Processes, 7(6), p.337.

Xia, Y., Zhao, J., He, L., Li, Y. and Niu, M., 2020. A novel tree-based dynamic heterogeneous ensemble method for credit scoring. Expert Systems with Applications, 159, p.113615.

Yan, C., Siddik, A.B., Yong, L., Dong, Q., Zheng, G.W. and Rahman, M.N., 2022. A two-staged SEM-artificial neural network approach to analyze the impact of FinTech adoption on the sustainability performance of banking firms: The mediating effect of green finance and innovation. Systems, 10(5), p.148.

Yin, W., Kirkulak-Uludag, B., Zhu, D. and Zhou, Z., 2023. Stacking ensemble method for personal credit risk assessment in Peer-to-Peer lending. Applied Soft Computing, 142, p.110302.

Yu, M.M., Lin, C.I., Chen, K.C. and Chen, L.H., 2021. Measuring Taiwanese bank performance: A two-system dynamic network data envelopment analysis approach. Omega, 98, p.102145.

Zhang, J., 2020. Investment risk model based on intelligent fuzzy neural network and VaR. Journal of Computational and Applied Mathematics, 371, p.112707.

Zhang, X. and Li, J., 2018. Credit and market risks measurement in carbon financing for Chinese banks. Energy Economics, 76, pp.549-557.

Zhang, Y., Ye, S., Liu, J. and Du, L., 2023. Impact of the development of FinTech by commercial banks on bank credit risk. Finance Research Letters, p.103857.

Zhang, Z., 2018, June. Improved adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th international symposium on quality of service, IWQoS), pp. 1-2, IEEE.