BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

ANALYSIS OF LOAN DEFAULT FOR MICROFINANCES IN ZIMBABWE USING NEURAL NETWORKS AND LOGISTICS REGRESSION.

BY

TAKUDZWA HOVE (B211982B)

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR BSc. HONOURS IN STATISTICS AND FINANCIAL MATHEMATICS

SUPERVISOR: Mr. Kusotera

JUNE 2025

AUTHORSHIP DECLARATION STATEMENT

Title of the Thesis: Analysis of loan default for microfinances in Zimbabwe using logistic regression and neural networks .

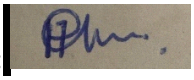
Author: Hove Takudzwa R

Program: BACHELOR OF SCIENCE HONOURS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS

I, the undersigned author of the above-mentioned thesis, hereby declare that:

- This thesis is my original work and has been prepared by me in accordance with the institution's requirements.
- All sources, data, and references used in this thesis have been acknowledged and cited appropriately.
- This thesis has not been submitted elsewhere for any degree or diploma.
- I have obtained all necessary permissions for the inclusion of third-party content where applicable.

I affirm that this declaration is made with full integrity and in compliance with the institution's policies and academic practice.

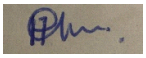
Author's Signature: 

Date: 17/06/2025

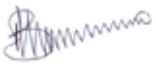
APPROVAL FORM

This is to certify, that this research project is the result of my own research work and has not been copied or extracted from past sources without acknowledgement. I hereby declare that no part of it has been presented for another degree in this University or elsewhere.

Takudzwa Hove R

Student(211982B) Signature  Date **17/06/2025**

Certified by:

MR.B. KUSOTERA  **17/06/2025**
Supervisor Signature Date

DR.M. MAGODORA  20/07/2025
Chairperson Signature Date

DEDICATION

I dedicate this project to my brother, my mother ,my friend and my family as a whole for their unwavering love and support.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the professionals and experts who shared their knowledge during this dissertation's development. Special thanks to my supervisor, Mr. B. Kusotera, for his invaluable guidance and support. I also appreciate the Department of Statistics and Mathematics at Bindura University of Science Education for their administrative assistance and access to research facilities. Additionally, I am grateful to all the institutions that provided access to their internal data. Ultimately, I am thankful for the divine guidance that has supported me throughout this project.

ABSTRACT

Microfinance institutions (MFIs) in Zimbabwe play a critical role in promoting financial inclusion, yet they face significant threats to sustainability due to rising loan default rates. These defaults undermine institutional stability and restrict access to capital for low-income and underserved populations. This study explores the drivers of default risk and aims to improve predictive accuracy through a hybrid modeling approach that combines Logistic Regression and Feedforward Neural Networks (FFNN). Using secondary data from five selected MFIs between September 2023 and August 2024, key borrower-level, institutional, and macroeconomic variables were analyzed to identify significant predictors of default. The logistic regression model highlighted interest rates as a significant factor influencing default, whereas other variables such as age and credit score had limited statistical impact. Despite class imbalance in the dataset, logistic regression provided more interpretable and better-calibrated results compared to FFNN. However, both models struggled to detect true defaulters. The FFNN model demonstrated higher recall for non-defaulters but underperformed in identifying defaults. Model performance was evaluated using metrics such as accuracy, recall, F1 score, and AUC-ROC, with logistic regression achieving the highest AUC of 0.59. Risk segmentation based on predicted probabilities enabled grouping clients into risk bands, aiding strategic loan management. Policy recommendations include managing interest rate ceilings, improving borrower data collection, and training MFI staff in risk modeling. Future research should investigate solutions for class imbalance, integrate behavioral and digital transaction data, and test alternative machine learning models like LSTM for enhanced forecasting. This study contributes practical insights into data-driven credit risk management and supports sustainable lending practices in Zimbabwe's microfinance sector.

Keywords: Default risk, microfinance, logistic regression, FFNN, machine learning, credit scoring, Zimbabwe

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

• **MFI**	Microfinance Institution
• **FFNN**	Feedforward Neural Network
• **LSTM**	Long Short-Term Memory
• **RBZ**	Reserve Bank of Zimbabwe
• **GDP**	Gross Domestic Product
• **ADF**	Augmented Dickey-Fuller
• **EDA**	Exploratory Data Analysis
• **SHAP**	Shapley Additive Explanations
• **SVM**	Support Vector Machine
• **AUC-ROC**	Area Under the Receiver Operating Characteristic Curve
• **MAE**	Mean Absolute Error
• **RMSE**	Root Mean Squared Error
• **AIC**	Akaike Information Criterion
• **BIC**	Bayesian Information Criterion
• **LRT**	Likelihood Ratio Test
• **LR**	Logistic Regression
• **NGO**	Non-Governmental Organization
• **ZAMFI**	Zimbabwe Association of Microfinance Institutions
• **SMOTE**	Synthetic Minority Over-sampling Technique
• **IRB**	Institutional Review Board
• **ARCH**	Autoregressive Conditional Heteroskedasticity
• **GARCH**	Generalized Autoregressive Conditional Heteroskedasticity
• **IMF**	International Monetary Fund
• **USD**	United States Dollar

Contents

CHAPTER 1: INTRODUCTION

1.0 INTRODUCTION

A microfinance is a financial service that provides small loans, savings and other financial products to individuals or group who lack access to traditional banking services. In recent times, microfinance institutions have played a vital role in helping small to medium enterprise (SMEs) financially .Instead of realizing profits, growth and other objectives of a business, most if not all have faced a challenge where high default rates, thus default risk in this context. The rapid growth of this credit risk has a concern hence this study is designed concentrate on reduction of these high and still growing rates. Default risk has been a critical and surprising issue in financial institutions as it has a negative impact on their profitability and growth. The study will then develop a mathematical tool to assess, evaluate and calculate the chances of borrowers not paying their installment on agreed dates or never make any repayments till they are recorded as bad debts. It goes beyond that and goes on to put in place measurements to reduce default risk. The thing of this design was to give a comprehensive understanding and full knowledge of risk, its factors, and the development of a mongrel model that will reduce default risk rates, as well as practical recommendations for microfinance institutions to reduce their exposure to default risk. We also looked into actuarial ways similar as neural networks, regression analysis and machine learning algorithms to manage default risk in microfinance portfolios.

1.1PROBLEM STATEMENT

Microfinance institutions (MFIs) play a critical role in promoting financial inclusion in Zimbabwe by providing small loans to underserved populations. However, rising loan default rates threaten the sustainability of these institutions and limit their capacity to serve vulnerable communities. The absence of reliable predictive tools and a lack of understanding of the socio-economic drivers of default hinder effective credit risk management in the sector. This study aims to address these challenges by developing predictive models to estimate loan default rates over the next five years using neural networks and by identifying the key socio-economic factors influencing loan default using logistic regression. By leveraging data-driven approaches, the study seeks to support MFIs in enhancing risk assessment processes, improving loan recovery strategies, and ensuring long-term operational stability.

1.2 AIMS AND OBJECTIVES

- To determine socio-economic factors that influence default using logistics regression.
- Predict default rates for the next five years using Neural networks.
- To determine critical institutional factors that influence default.

1.3 RESEARCH QUESTIONS

- What are the socioeconomic factors that influence credit default in microfinance institutions?
- What is the best risk management strategies for microfinances in Zimbabwe?
- How does government regulations affect microfinances concerning default rates?

1.4 BACKGROUND OF THE STUDY

The concept of microfinance emerged in the 1970s and gained significant momentum during the 1980s and 1990s. Pioneering organizations such as Grameen Bank in Bangladesh and Accion International in Latin America began experimenting with small loans to low-income earners during this period. Grameen Bank, founded by Muhammad Yunus in 1983, played a pivotal role in providing microcredit without requiring collateral, thereby empowering the impoverished to engage in self-employment activities (Grameen Bank, 2025). Similarly, Accion International expanded its microfinance initiatives across Latin America, aiming to increase financial inclusion for the poor and self-employed (Accion, 2025).

The roots of microfinance can be traced back even further to the 19th century with institutions like the Raiffeisen Bank in Germany, which provided small loans to rural farmers, laying the groundwork for modern microfinance practices (Raiffeisen Bank International, 2025).

In Zimbabwe, the microfinance sector has experienced notable changes in recent years. The number of registered microfinance institutions (MFIs) declined by 13.54%, from 229 in December 2018 to 198 by the end of 2022 (Reserve Bank of Zimbabwe, 2023). This decline suggests challenges within the sector, potentially linked to issues such as default risk and economic instability.

Despite these challenges, the sector has shown signs of resilience. For instance, the microfinance industry in Zimbabwe demonstrated steady growth since 2021, with the number of MFIs increasing from 220 in 2021 to 241 by September 2022 (ResearchGate, 2023). However, concerns remain regarding portfolio quality. The Reserve Bank of Zimbabwe (RBZ) reported that the portfolio at risk (PaR) for credit-only MFIs stood at 8.3% as of September 2024, exceeding the international benchmark of 5% (Reserve Bank of Zimbabwe, 2024).

These trends underscore the need for robust risk management practices and regulatory oversight to ensure the sustainability and effectiveness of microfinance institutions in Zimbabwe.

.

1.5 Scope of Research

The research focuses on the credit risk associated with microfinance institutions tracing its development and exploring the effects which increases high credit risk for example looking at the default rate of the KCI Management consultants, the missed instalments and interest rates and other economic factors. the case study will be KCI Management Consultants specifically on one of its branches sited in Kadoma. Data that will be used will be from a period of September 2023-August 2024 using primary source mostly the loans default rate, repayment trend and time period of the loan and the credit assessments form of loan requested and loan approved and payment trends of that loan.

1.6 IMPORTANCE OF THE STUDY

- Enhancement of Credit Risk Management: Implementing default risk modelling enables microfinance institutions (MFIs) to better assess and manage credit risk, thereby reducing the likelihood of loan defaults and financial losses (CDFI Fund, 2012).
- Improved Lending Decisions: Quantitative risk assessments facilitate more informed lending decisions, shifting from subjective judgments to objective criteria, which can lead to fairer lending practices and improved borrower outcomes (ResearchGate, 2023).
- Optimized Resource Allocation: Effective risk management strategies, such as hybrid default risk models, allow MFIs to identify high-risk borrowers and focus resources on appropriate risk mitigation measures, enhancing financial inclusion and reducing borrowing costs for clients (CDFI Fund, 2012).
- Strengthened Financial Stability: Accurate identification of default risks contributes to the long-term sustainability of MFIs by reinforcing financial positions and resilience to external shocks, thereby increasing investor confidence and access to capital markets (ResearchGate, 2023).
- Advanced Analytical Tools: Integrating actuarial modelling techniques provides MFIs with robust analytical frameworks for assessing default risks, combining statistical approaches, probability theory, and financial analytics for more precise risk management decisions (ScienceDirect, 2021).
- Enhanced Predictive Capabilities: Applying developed models improves the prediction of default risks in microfinance portfolios, enabling MFIs to anticipate potential losses and implement effective risk mitigation strategies (ScienceDirect, 2021).
- Tailored Risk Assessment: Developing models specifically for the microfinance industry addresses the unique characteristics of microlending, considering factors such as multiple income sources, lack of collateral, and vulnerability to external shocks (ScienceDirect, 2021).
- Comprehensive Risk Analysis: Advanced actuarial modelling approaches allow for the inclusion of both financial and non-financial factors in risk assessments, supporting effective portfolio management and helping MFIs balance social objectives with financial viability (ScienceDirect, 2021).

1.7 LIMITATIONS OF THE STUDY

- The findings of the study were mainly from one microfinance institution (MFIs), Kadoma branch to be specific, hence limited generalizability to other MFIs offering different loan products or serving a different demographic.
- Limited access to data. MFIs have got what they call sensitive data (financial statements, organizational strategies and borrower credit history) which is essential when creating these models but my inability to access that information is a concern.
- The accuracy and reliability of data obtained from MFIs and secondary sources might be compromised due to various factors such as incomplete and inconsistent records.
- The study was limited to a specific time frame which might have left out the data beyond the chosen period.
- The model assumes steady economic indicators such as inflation, which is not applicable to the economy of Zimbabwe

1.8 RESEARCH HYPOTHESIS

H1: There is direct correlation between lenders demographic characteristics such as age, sex, income and credit risk and default risk in microfinances.

H2: Institutional factors such as size and lending practices have a significance impact on default risk.

H3: Macro-economic factors such as inflation, economic recession and unemployment influence default risk.

Ho: There are no significant relationships between all the above factors and credit risk default.

1.9 RESEARCH ASSUMPTIONS

- Borrowers make rational decisions about borrowing and make repayments
- Economic indicators such as inflation, gross domestic product (GDP) and unemployment remained almost steady throughout the whole research
- Lenders (Microfinances) and borrowers have equal access to information about the borrower's credit worthiness.
- The likelihood of default is assumed to follow a predictable pattern based on historical data, such as the use of credit rating agencies or credit scores.
- Interest rates remain stable or change in a predictable manner.
- Collateral presented by borrowers is sufficient and will holds its value in the event of default.
- The model assume that default risk or relevant financial variables follow a normal distribution. This assumption simplifies the analysis, though it may not always reflect real-world distributions accurately.

DEFINATION OF KEY TERMS

- **MFIs** – Microfinance institutions.

- **Markov Chain**-A mathematical model for a system that moves from state to state, where the probability (P) of moving from one's state to another depends only on current state (Snell, 1997)

- **Credit default**-is the potential for financial loss resulting from the failure of a borrower or counterparty to meet their contractual payment obligation (Bluhm, 2019)

- **Logistic regression analysis**- Logistic regression is a method for credit risk prediction especially given the growing significance of credit risk assessment in preserving the stability and sustainability of MFIs.

- **Neural networks**-A logical calculus of the ideas immanent in nervous activity (Warren McCulloch,1943)

1.10 CHAPTER SUMMARY

In conclusion the research offered the perspective of topic reduction of default risk, mainly causes and ways to reduce or predict them. The model which was created was made in such a way that if microfinances were to apply it in their day to day running default rates will decrease and with time they will not exist if well applied. The chapter presented some basic statistical and mathematical tools that are used in default risk modelling. The next chapter will focus on literature review.

CHAPTER 2: LITERATURE REVIEW

2.0 INTRODUCTION

Small-scale financial services offered to low-income individuals and households have become a major development strategy microfinance to combat poverty and economic empowerment. Microfinance institutions (MFIs) have a range of microfinance products, including microloans, savings accounts, and insurance, for those usually left out of the formal banking ambit. Microfinance presents itself well as a vehicle to expand access to financial services and to enhance the livelihood of financially excluded population; however, the risks associated with default on loans remain a great challenge, significantly undermining the ability of the microfinance institutions from achieving their aims of socially acceptable profitability. The default risk is a key concern for MFIs, as failure by a borrower to repay his loan as agreed upon can greatly affect the financial sustainability of the MFI and its ability to serve the population of borrowers that the MFI seeks to serve. This chapter critically and comprehensively reviews the literature on default risk in microfinance with special reference to the Zimbabwean context. The environment in which microfinance is conducted in Zimbabwe is unique and challenging in the following ways: a history of economic instability, hyperinflation, and high levels of poverty. Knowing the determinants of default risk in this context is imperative for the microfinance industry for risk mitigation purposes and for the long-term viability of the microfinance sector.

The chapter is organized in the following way. Section 2 presents a theoretical framework regarding default risk based on information economics, social capital theory, behavioral economics, and institutional theory. Section 3 analyzes the determinants of the default risk, which are divided into factors at the borrower level, the MFI level, and the macro level. Section 4 discusses the impact of default risk on the MFIs, borrowers, and the wider financial system. Section 5 reviews several countermeasures that have been suggested and applied to reduce default risk. In the end, Section 6 concludes the chapter and highlights key research gaps that need to be further investigated.

2.1 THEORETICAL FRAMEWORK

The phenomenon of default risk in microfinance is explained using several theoretical perspectives. While these theories differ in the focus, they are not mutually exclusive but rather complementary ways of interpreting the process of lending to low-income borrowers.

INFORMATION ECONOMICS

The information economics is concerned with asymmetric information between the lender and the borrower (Stiglitz Weiss, 1981). As in other lending contexts, borrowers in microfinance typically have more information about their own creditworthiness, repayment capacity, and use of loan funds than do lenders. The two key problems that arise from this information asymmetry is that it gives rise to adverse selection and moral hazard.

Adverse selection happens before disbursement of the loan. This is the tendency that borrowers more likely to default are more likely to apply for loans. Microfinance loans can attract borrowers with a higher risk profile, such as if the borrowers have a history of default or the aim of using the loan for unproductive purposes, because interest rates can partially reflect the individual risk level. The absence of complete information about the credit worthiness of borrowers on the part of MFIs may result in an accidental selection of a higher number of high-risk borrowers, which in turn would result in higher default rates. Moral hazard takes place after the loan is disbursed. The risk is that borrowers will change their behavior once they receive the loan and undertake actions that raise the probability of default. For instance, borrowers might divert loan funds for productive investment to personal consumption, reduce their effort at income generating activities or take on excessive risks with the knowledge that the MFI bears the major costs of default. A high degree of moral hazard is made worse by the difficulty of monitoring borrowers' actions and enacting loan contracts.

SOCIAL CAPITAL THEORY

Social capital theory states that social networks help in the market transactions by way of trust and norms (Dekker, 2004; Chowdhury et al, 2021). Social capital can have a significant role in mitigating default risk in group lending methods in the context of microfinance. The Grameen Bank in Bangladesh pioneered group lending, which is based upon joint liability and all members of a borrowing group are responsible for the repayment of each member's loan. This generates strong peer pressure for repayment because default by one member can threaten future access to credit for the entire group. Strong social ties, trust and shared norms between people in a community can help the effectiveness of group lending. According to Mosley (2001), borrowers who are in close knit social networks will be more influenced by peer pressure and social sanctions and hence less likely to default. Nevertheless, group lending and social capital are effective in mitigating default risk only to the extent that. Socioeconomic characteristics and risk profiles that are group homogenous can facilitate trust and cooperation. For heterogeneous groups with large income or risk disparities, the level of conflict and free riding may be higher, thus jeopardizing the joint liability approach. It is also important that there is strong leadership in the group to enforce repayment discipline and resolve disputes. Additionally, social capital can be diminished by factors, including migration, shocks to the economy, and social unrest which are depleting factors of social capital and its ability to mitigate default risk. The opportunity and challenge of using social capital in microfinance in the Zimbabwean context is based on its history of social and economic upheaval.

BEHAVIORAL FACTORS

While a too traditional assumption about perfectly rational actors leads to making certain type of very widespread assumptions about 'rational' man this sort of behavioral economics makes us accounts for present bias which means that people do what they do not do tomorrow in present or now rather than making better decisions about future; loss aversion that is about people responding to pain of a loss and not a gain and limited attention, which is simply difficulty in processing information and making good decisions over complex things (Kahneman 2011). Borrowing in Zimbabwe is thus shaped by these biases when borrowers face significant economic hardship and uncertainty. For instance, a borrower having immediate needs for example food insecurity or medical bills may feel tempted a way to divert the loan funds that would have contributed toward their productive investment, even if it undermines ability for long term repayment (Ncube, 2018).

However, it is important to understand these behavioral biases so as to make loan products that are effective, repayment schedules effective, and communication strategies that promote responsible borrowing and repayment. Insights from behavioral economics can be used by MFIs to try and 'nudge' borrowers in better financial choices, for example by presenting loan terms so that the long-term benefits are emphasised, using commitment devices to encourage saving and regular reinforcing of the repayment's deadlines. These interventions though implementable are often challenging within resource constrained settings as they need to be developed with regard to the cost effectiveness, cultural appropriateness and scalability. Ethical considerations also need to have primacy, implying that borrowers are not seen as a target of vulnerability, but rather designed so that interventions help empower borrowers.

The theory of institutions is concerned with the impact of the broader regulatory, legal, political environment on the behaviour of MFIs and borrowers (North, 1990). Having an institutional framework that is well functioning, i.e. for example, having clear property rights, enforceable contracts, and an effective regulatory oversight is necessary to promote financial stability and the responsible financial lending practices (Brio et al., 2017). The country of Zimbabwe is complex in terms of institutional makeup and weak in governance, corruption with the lack of implementation of consistent policy (Kanyenze et al., 2017). The legal framework of the sector is provided by the Microfinance Act [Chapter 24:29] but its effectiveness is often constrained by poor enforcement capacity, bureaucratic delays and lack of transparency. The weak enforcement of contract mechanisms makes it difficult for MFIs to recover defaulted loans, hence increasing the risk exposure for them and in turn, the interest rates charged to borrowers. However, it is worth pointing out that formal and informal institutions can intertwine in sophisticated fashion. Formal rules can either promote or negate informal norms, and the potency of formal regulations rely still on the consensus between the prevailing social norms and values. For instance, even with solid formal laws, widespread prevalence of corruption and acceptance of it as a norm will lead to missionally low compliance. It also plays a major role in the political environment. Investment in the microfinance sector as well as long term lending can be discouraged by political instability, policy uncertainty, and the erosion of the rule of law.

2.2 CONCEPTUAL FRAMEWORK

Default Risk is the dependent variable in this theoretical model, being caused by independent variables categorized into Borrower-Level, MFI-Level, and Macro-Level Factors. These factors, typically represented by measurable proxy variables (e.g., income, interest rates, GDP growth), are hypothesized to have a direct influence on loan default probability. Mitigation efforts are interventions with the goal of changing these independent variables, whereas Consequences are the effects resulting from the measurement of Default Risk, but not influencers in themselves. The model, therefore, portrays a causal chain with the proxies of independent variables impacting the dependent variable (Default Risk), followed by various consequences, and mitigation efforts as possible modifiers of the independent variables.

2.1.0 Summary of Factors, Consequences and Mitigation Strategies.

2.3 DETERMINANTS OF DEFAULT RISK

Microfinance default risk is a multi-faceted phenomenon where a complex interplay of factors at three levels, borrower level, MFI level and macro level determines the phenomenon. These factors do not act independently but interact with each other and influence each other in a complex and oftentimes difficult manner, making conditions challenging for MFIs. It is therefore important to understand relationships between these factors for developing appropriate mitigation strategies.

BORROWER-LEVEL FACTORS

However, at the individual borrower level, there exist a wide range of economic background, financial behavior and other shocks that can greatly influence the probability of default. There are two basic determinants of repayment capacity: income level and stability. Economic shocks are more likely to affect borrowers with low and irregular incomes and they may lag behind in fulfilling their loan repayment obligations. The source of income also matters. Borrowers in rainbow fed agriculture, as well, are more vulnerable as their income can be affected by changes in weather variability and market price. Household assets at the level can act as a buffer to unexpected shocks. In times of difficulty, borrowers with assets, like land, livestock or savings, may be able to make liquidations of these assets to repaying obligations. However, in Zimbabwe, low-income households tend to have limited ownership of assets and the value of assets may be eroded by inflation or economic instability.

Responsible borrowing behavior is dependent on one thing: Financial literacy, which is the knowledge of financial concepts, budgeting skills and debt management practices (Agryei, 2020). If borrowers are of relatively low levels of financial literacy, they may be more predisposed to over borrow, not properly manage debt funds or do not fully comprehend the terms and conditions of the loan. Past history of borrowing and creditworthiness are good indicators of future repayment behavior. It is more likely that borrowers with a history of default or poor credit performance would default again. Nevertheless, the formal credit histories are frequently unavailable in the microfinance arena, especially in developing economies. In order to assess borrower creditworthiness, MFIs may rely on group lending references or community assessments as alternative data sources.

Repayment incentives can be affected by the purpose of the loan and its intended use. When loans are used for productive investments like starting or expanding a business, there is greater chance that the income from the investments will be used to repay the loans. If borrowers have no other means of income, loans used for consumption smoothing or emergency needs will be more difficult to repay. But unexpected shocks like becoming ill, losing a job, the results of natural disasters or even family emergencies can undermine a borrower’s ability to repay the loan. These shocks can damage income earning activities, dispel savings, and add further cost. In this case, Zimbabwean

context features high frequency of economic and environmental shocks leaving borrowers highly vulnerable.

MFI-LEVEL FACTORS

Default rates are greatly influenced by the characteristics, policies, and practices of MFIs themselves. It plays a role in the design of loan products such as the size of the loan, the interest rate, the repayment schedule, etc. Large loan sizes can be too large compared to borrowers' capacity to repay and may put borrowers into over indebtedness and increase the likelihood of default. High interest rates make loans unaffordable for borrowers, while low interest rates do not cover the MFI's operating costs and risk exposure. Difficulties may occur in repayments deadlines if they are not flexible enough or when they do not match borrowers cash flow pattern. While the collateral requirements are supposed to reduce risk, it could prevent borrowing of those without assets or provide disincentive to borrow. In India, Field and Pande (2008) found that borrowers paid off more slowly were more likely to default.

An effective way to identify and diminish adverse selection is through effective borrower screening and selection processes. Rigorous screening methods are combined, as a rule, by financial data, business assessment and character references to enable MFIs select borrowers with a higher likelihood of repayment. Regular loan monitoring that entails frequent contact with borrowers, tracking status of their business performance, and giving continued support could make a borrower's problem visible early on, and seek timely intervention. To minimize loss from defaulted loans, an effective loan recovery procedure which includes: clear communication with defaulters, flexible repayment options, and, if necessary, legal action, is necessary. But recovery of a loan should be done in an ethically and responsible manner, not by harassment and intimidation of borrowers. It also depends on the level of staff training and expertise. Thus, better positions to manage default risk are well trained loan officers who possess the skills to assess borrower risk, monitor loan performance, and provide financial literacy training (Hulme Mosley 1996).

The MFI's overall risk profile is also influenced by its internal governance structure and management practices. A strong governance, which involves a clear separation of powers, independent oversight, adherence to transparency and accountability, is necessary for the MFI to operate in a sound and sustainable way (Hartarska Nadolnyak, 2007). At the same time, weak governance may provide basis for incompetence, fraud, and excess risk taking that increases the likelihood of default. By diversification of the portfolio (different sectors, geographic areas and borrower types), the impact of concentrated risks (Ahlin et al., p. 440, 2011) can be mitigated. However, diversification needs to be planned and supervised carefully so that the MFI is able to operate in different markets and deal with a broader range of risks.

MACRO-LEVEL FACTORS

Microfinance institutions and the repayment behavior of borrowers are very much affected by the broader macroeconomic environment, which includes economic growth, inflation, interest rates and political stability. Declining level of GDP growth, increasing unemployment and dropping off from consumer spending are what economic downturns are defined as (World Bank, 2020) and can put borrowers in a particularly vulnerable sectors at a risk of not being able to repay loans. Microfinance has been an industry that has gone through prolonged periods of economic

instability, with hyperinflation, currency devaluation, and high unemployment (IMF, 2023), which in Zimbabwe is a particularly challenging environment. As the real value of loan repayments are eroded by hyperinflation, MFIs are unable to maintain the real value of their loan portfolios and borrowers have little incentive to repay (Makina, 2019). Currency instability and frequent changes in monetary policy are sources of uncertainty and they prevent MFIs as well as borrowers from planning for the future. For instance, high interest rates can be caused by macroeconomic instability, which mean that it will make additional cost to borrow and difficult to the borrower to repay his loans.

Thus, political instability and uncertainty such as changes in government policies, social unrest and the fear of being repossessed of your property can also badly affect the microfinance sector (Collier, 2007). Other natural disasters, particularly droughts, floods and cyclones, can be devastating for borrower livelihoods and their capacity to repay loans in agriculture dependent areas of Zimbabwe. Widespread defaults in loans follow these events wiping out crops, destroying assets and disrupting economic activity. It is important that regulatory and legal framework of the microfinance sector influences positively performance and stability of same. A well designed regulatory framework can encourage responsible lending practices, protect borrowers' rights and preserve MFIs' financial soundness (Goodwin-Groen, 2014). Yet, burdensome regulations, poor enforcement, or even both alike can stifle innovating, cause excessive compliance costs and constrain the scale of microfinance services. Loan diversion, a weakness in loan recovery efforts, and an unlevel playing field for MFIs can all be enabled by corruption. Also, credit is allocated out of alignment in such a way that loans are funneled to politically connected individuals or firms instead of to those with the greatest likelihood for productive investment or greatest need. Cost of operating of MFIs also depends on availability of infrastructure, in the form of roads, electricity, and the communication network. This poor infrastructure makes it both difficult and expensive for MFIs to reach remote borrowers, monitor loan performance and recover defaulted loans (Ncube et al., 2021).

2.4 CONSEQUENCES OF DEFAULT RISK

With high default rates, the consequences are wide ranging, ranging from negative effects on financial sustainability of microfinance institutions to adverse consequences for borrowers and for overall development of the financial sector. High default rates thus reduce profitability, erode capital, and constrain MFIs' ability to lend to new borrowers (Ledgerwood, 1999). It turns into a vicious loop, cutting out lending further impeding credit access by the low-income population, hindering economic growth and sustenance of the poverty. MFIs can also suffer damage to their reputation, which can make it more difficult to attract funding from investors and donors when default rates are high. Borrowers who default on a loan can also suffer from some serious psychological effects, such as feelings of shame, stress, and hopelessness. This can harm their creditworthiness and could prevent them from getting credit in the future and indeed lead to social exclusion.

Defaulting a loan can have heavy consequences for borrowers, including loss of his collateral assets, demarcation in the borrower's credit history, and exclusion from future access to financial services (Collins et al., 2009). Borrowers trapped in such a cycle of debt and poverty are then undermined in their attempts to improve their livelihoods. Higher interest rates for all borrowers

can also come from high default rates (MFIs try to cover their losses, making credit less affordable, more expensive and less accessible). High default rates in the microfinance sector can also create systemic risk at the systemic level, that is, the stability of the broader financial system (Mersland Strøm, 2010). The inter-MFI lending or shared borrower networks can cause a crisis in one MFI quickly to spread to other MFIs, which can result in a wider collapse of confidence and a contraction of credit availability. This is worrying, given the state of the financial sector in a country like Zimbabwe. Therefore, such failure can ultimately destroy confidence in the financial system, deter investment and slow down economic growth.

2.5 MITIGATION STRATEGIES

Default risk challenge calls for multi pronged approach and it involves number of strategies at borrower, MFI and macro level. These strategies should be tailored to the specific context of each MFI and its operating environment in Zimbabwe, RBZ (2018). Further, at the borrower level, financial literacy training and education is important in encouraging responsible borrowing behavior and enhancing borrowers' capacity in managing their finances (Lagarde, 2017). They should be able to do this training in budgeting, saving, debt management, and understanding the terms and conditions of a loan. By having access to credit counseling and support services that borrowers can utilize provides further capacity to understand that they can protect themselves by never defaulting on their loans. They can also provide advice on debt restructuring, income generation and arriving at social safety nets. Diversifying the borrower's income sources is encouraged and may help protect the borrower from economic shocks and aid in repaying loans. This could include helping borrowers participate in a number of incomes generating activities, training and resource provision for alternative livelihoods, and market linkages.
In the MFI level, screening and selection of borrowers are important to identify and deal with adverse selection. It can involve a mix of financial data, business assessments, character references and alternative data such as mobile phone use or social media activity (Chen et al., 2018). Loan products can be designed to cater for borrowers' requirements and repayment capacity and risk of over indebtedness and default can be reduced. It includes providing flexible terms of the loan, the appropriate rates of interest, and repayment period which are suitable with the cash flow pattern of the borrowers. Robust loan monitoring and follow up procedure can help prevent problems from becoming big through fast interventions. Regular contact with borrowers, monitoring their business, providing the ongoing support, this is what this it involves. For minimizing the losses from defaulted loans, effective loan recovery procedure should be developed, which includes clearly communicating with defaulters, flexible repayment options, and as the last resort legal action. To promote the quality of loan products and ensure effective handling of loans, investing in staff capacity building and training is important to improve loan officers' skills in assessing the risk of the borrower, monitoring and evaluating the performance of the loan, and giving financial literacy training. The concentration of risks translates into concentration of returns and this exposes the bank to the risks associated with specific sectors or geographic area. Credit risk diversification across sectors, geographic areas and borrower types can help mitigate such risks. To ensure that the MFI operates in a sound and sustainable manner, the MFI should implement strong internal controls and risk management systems.

On the macro level, microfinance presupposes macroeconomic stability, that is, low inflation, stable exchange rates, and sustainable economic growth. Influence of responsible lending practices

in promoting, rights of borrower's protection and MFIs financial soundness based on implementing sound regulatory and supervisory frameworks. Improving the operating environment for MFIs can be achieved by investing in infrastructure such as roads, electricity and communication networks. Property rights and contract enforcement mechanisms can be strengthened and such will make the environment in which lending takes place safer. Support for social safety nets disaster preparedness programs can buffer borrowers from the effects associated with economic as well as environmental shocks and their vulnerability to default is lessened.

2.6 RESEARCH GAPS

Although a lot of research has been carried out on microfinance, there are critically important gaps in research that are still present, especially in Zimbabwe. The gaps suggest that further study needs to be carried out so as to gain a clearer understanding of default risk in a specific context. One of the research gaps of importance is the need for more rigorous studies that quantify with accuracy the importance of different factors at the borrower level, MFI level, and macro level in causing default rates in Zimbabwe. Such studies need to use sound methods to establish cause-and-effect, analyze how these factors influence one another, and collect good quality data over time. Further studies need to assess the impact of some of these regulatory controls, such as capital levels, ceilings on interest rates, and licensing procedures, on the operations of MFIs and borrowers' performances in Zimbabwe. Further evidence is also scarce on the impact of technologies like digital financial services and mobile banking, such as examining alternative data sources to be used in credit scoring and determining the impact of digital financial literacy interventions. More qualitative research needs to be conducted on grasping daily realities of the Zimbabwean microfinance borrowers, identifying social, cultural, and psychological determinants of payment conduct, including difficulty, coping styles, and perceptions of MFIs. Finally, research needs to account for dynamic interactions between greater than one risk factor, in addition to interactions between borrower-level risks, and how the MFI lending conduct interacts to intensify or decrease the impacts. It entails a look at how microfinance has affected the well-being of Zimbabwean borrowers in the long term. It will take into account results such as asset building, capability building, and empowerment increases. The study can fill in the gaps by analyzing the default risk in the microfinance sector of Zimbabwe thoroughly. The research will give explicit evidence and feasible suggestions to Zimbabwean MFIs, policymakers, and developmental practitioners.

2.7 CONCLUSION

In this chapter we have reviewed the literature on default risk in microfinance, from the vantage of information economics, social capital theory, behavioral economics, institutional theory. Finally, it described the severe impact of high default rates on MFIs, borrowers, and the entire financial system; and various mitigation measures. The next chapter focuses on the research methodology in detail.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 INTRODUCTION

This study addresses the escalating default risk among microfinance institutions (MFIs) in Zimbabwe, which threatens their stability and profitability. A loss of 100 billion dollars would only be the beginning, as default rates would soar, undermining these institutions' capacity to disburse loans despite their mission of providing financial services to underserved demographics. The research aims to develop a hybrid model combining neural networks and logistic regression to assess and mitigate default risks. This systematic approach ensures that the methods employed are thorough, reliable, and relevant to real-world applications.

A quantitative research design is employed to analyze relationships between borrowers' socioeconomic characteristics and default risk. Data will be collected through structured surveys targeting loan officers and top management of 5FIs.

Both primary and secondary data sources will be utilized. Primary data will be gathered via surveys, while secondary data, including historical default rates and economic indicators, will be sourced from reports published by the Reserve Bank of Zimbabwe (RBZ).

The findings are expected to inform policy and operational strategies within the microfinance sector, contributing to the sustainability and financial stability of MFIs in Zimbabwe.

3.1 RESEARCH DESING

This study employs a quantitative research design to examine the relationship between socio-economic factors and default risk in Zimbabwean microfinance institutions (MFIs). Structured surveys will be administered to loan officers and top management across 5 MFIs, ensuring standardized data collection for reliability and validity. Data analysis will involve logistic regression and neural networks to model default risk, incorporating variables such as borrower demographics, loan characteristics, and economic conditions. The research aims to provide empirical evidence to inform risk management strategies and policy decisions within the microfinance sector. By focusing on objective measurements and statistical analysis, the study seeks to enhance the understanding of default risk dynamics and contribute to the sustainability of MFIs in Zimbabwe.

3.2 DATA SOURCES

This study utilized secondary data from reputable sources to analyze default risk among microfinance institutions (MFIs) in Zimbabwe. Data was gathered from the Reserve Bank of Zimbabwe's (RBZ) quarterly microfinance industry reports, which provide insights into sector performance, portfolio quality, and regulatory compliance. Additionally, information was obtained from microfinance institutions' websites, such as KCI Microfinance's Bupwe platform, which offers data on loan portfolios, borrower demographics, and repayment histories. These sources collectively offer a comprehensive view of the factors influencing default risk in Zimbabwe's microfinance sector

3.3 DATA COLLECTION METHODS

This study adopts a data collection and analysis approach based exclusively on secondary data, which involves drawing on various existing sources of information. The use of secondary data strengthens the study by enabling access to historical and contextual financial information that is essential for understanding default risk.

Secondary data will be collected from different financial statements and reports issued by the Reserve Bank of Zimbabwe (RBZ) and other relevant financial institutions. These include historical data on default rates, economic indicators, and other financial metrics necessary to contextualize default risk in microfinance. Secondary sources of data encompass financial statements where data on loan portfolios, repayment rates, and financial health indicators are available. This quantitative data will enhance the depth of analysis by providing a broad historical and macroeconomic context.

Furthermore, economic publications by the RBZ will provide macroeconomic data relevant to microfinance, including inflation, unemployment, and GDP growth indicators. Recognizing these external factors is crucial for understanding the challenges faced by MFIs and their borrowers (Sinha and Ghosh, 2022b). Additionally, secondary data will be complemented by literature in industry publications, research articles, and white papers focused on microfinance within Zimbabwe. These sources will provide qualitative perspectives and findings from existing studies, which will help frame the research questions and lay down a theoretical foundation for the analysis. The discussion and analysis in this study will be informed entirely by secondary data, which offers a comprehensive and holistic understanding of the factors affecting default risk. Data analysis will

employ descriptive statistics to summarize the data as well as inferential statistics to identify statistically significant relationships and trends. Sophisticated modelling techniques, such as logistic regression and neural networks, will be applied to develop a hybrid default risk model based on the secondary data (Chi et al., 2019). This model will use insights derived from secondary data to accurately predict default risk and identify key contributing factors.

Software such as R will be utilized to analyze the data efficiently, enabling the use of complex statistical models with large datasets, making it an appropriate tool for this research.
In conclusion, this chapter presents a systematic methodology relying on secondary data to address the problem of default risk in Zimbabwean MFIs. The study follows a quantitative research design that objectively explores the interdependencies of various socio-economic factors in relation to default risk. The comprehensive use of secondary data provides a strong foundation for evidence-based insights that can guide policymakers, practitioners, and stakeholders in microfinance risk management, ultimately contributing to the stability and sustainability of financial inclusion initiatives.

3.4 POPULATION AND SAMPLE

This research targets microfinance institutions (MFIs) in Zimbabwe, focusing specifically in MFIs servicing poor customers and small businesses. Microfinance is a vital tool for promoting financial access and inclusion, particularly in developing economies such as Zimbabwe(Mhlanga, 2020), where marginalized sections of society often face barriers in accessing mainstream banking services. This study seeks to investigate the critical issue of default risk, which threatens the long-term operational viability of MFIs and their capacity to continue meeting the needs of their clientele. Specifically, the research targets 30 MFIs, encompassing a sample of 1,000 clients, to assess the factors contributing to default risk and develop strategies to mitigate it.

3.5 TARGET POPULATION

The study used purposive sampling to select 5 MFIs, including NGOs, credit unions, and private firms, which vary in operations and risk management approaches. These institutions serve low socio-economic borrowers who are often under-collateralized and lack formal credit histories (Mugambi, 2016).

Purposive sampling ensured that selected MFIs have faced significant default risk challenges, providing rich and relevant data. This sample size allows for focused and manageable data collection while generating meaningful insights for developing a robust hybrid default risk model applicable to Zimbabwe's microfinance sector (Nyirenda, 2024).

Table 3.6 Description of Variables

TABLE 3.6 DESCRIPTION OF VARIABLES 1

Variable Name	Description	Data Type	Data Source
Loan default	i.e. client default = 1 or not = 0	Binary	ZAMFI
Loan Amount	Amount($) in USD	Continuous	ZAMFI

Interest rates	Percentage (%)	Continuous	ZAMFI and RBZ
Loan Purpose	i.e.(medical, fees or household)	Categorical	ZAMFI
Property Type	i.e. (commercial or residential)	Categorical	ZAMFI
Employment Status	i.e. (SMES, employed, unemployed, civil servant or informal sector)	Categorical	ZAMFI

Loan Default
Indicates whether a borrower failed to repay (1) or successfully repaid (0) a loan.
In Zimbabwe, economic instability, high inflation, and rising unemployment have contributed to increased loan defaults *(ZAMFI, 2022; RBZ, 2021)*

Loan Amount
Refers to the total value of the loan issued in US dollars. Small loan amounts pose a higher risk of non-repayment, particularly within Zimbabwe's informal and low-income sectors *(Mago, 2013; ZEPARU, 2020).*

Interest Rates
The rate charged on loans, expressed as a percentage. Fluctuating interest rates make borrowing costly, often leading to difficulties in repayment *(RBZ, 2022; IMF, 2022.*

Loan Purpose
Explains the reason for the loan, such as covering medical expenses, school fees, or household needs. In Zimbabwe, many individuals borrow for non-productive reasons, which do not generate income, thus increasing the chance of default *(Chikoko & Mangoti, 2016; FinMark Trust, 2021).*

Property Type
Measure the collateral value of a client. Ownership of commercial property may signal income-generating activity, while residential properties often offer limited collateral value *(UN-Habitat, 2020; ZAMFI, 2021).*

Employment Status
Specifies the borrower's work situation. We can all agree that Zimbabwe's economy is largely informal, many borrowers have inconsistent income, making it harder to meet loan obligations *(ZIMSTAT, 2022; ILO, 2021).*

3.7 DATA PRESENTATION AND ANALYSIS PROCEDURES

DATA PRESENTATION

This study presents data gathered from five selected microfinance institutions (MFIs) operating in Zimbabwe, covering the period from September 2023 to August 2024. The data is entirely

secondary and was obtained from publicly available sources such as the Reserve Bank of Zimbabwe's quarterly reports, institutional websites like KCI Microfinance's Bupwe platform, and published financial statements. Key variables include borrower demographics, loan amounts, interest rates, repayment terms, and macroeconomic indicators such as inflation and GDP growth. The data selection and organization were guided by the need to explore the relationship between loan default and various borrower-level, institutional, and macroeconomic factors (RBZ, 2023; Nyirenda, 2024). Consistent with previous findings in the Zimbabwean microfinance sector, where default risk is influenced by factors like high interest rates and weak credit screening (ZAMFI, 2021; Chipaka, 2020), the dataset supports a structured and quantitative analysis. The data is presented in a format suitable for statistical modeling using logistic regression and neural networks, laying a foundation for assessing and predicting credit default in Zimbabwean MFIs.

DATA ANALYSIS PROCEDURE

This study presents data gathered from five selected microfinance institutions (MFIs) operating in Zimbabwe, covering the period from September 2023 to August 2024. The data is entirely secondary and was obtained from publicly available sources such as the Reserve Bank of Zimbabwe's quarterly reports, institutional websites like KCI Microfinance's Bupwe platform, and published financial statements. Key variables include borrower demographics, loan amounts, interest rates, repayment terms, and macroeconomic indicators such as inflation and GDP growth. The data selection and organization were guided by the need to explore the relationship between loan default and various borrower-level, institutional, and macroeconomic factors (RBZ, 2023; Nyirenda, 2024). Consistent with previous findings in the Zimbabwean microfinance sector, where default risk is influenced by factors like high interest rates and weak credit screening (ZAMFI, 2021; Chipaka, 2020), the dataset supports a structured and quantitative analysis. The data is presented in a format suitable for statistical modelling using logistic regression and neural networks, laying a foundation for assessing and predicting credit default in Zimbabwean MFIs.

3.8 DATA ANALYSIS TECHNIQUES

In this research, a variety of statistical methods will be employed to analyse secondary data collected from microfinance institutions (MFIs) in Zimbabwe. Key techniques include logistic regression and neural networks, both of which are well-suited for modelling default risk based on historical and institutional data. Software tools such as R and Python will be used to facilitate advanced data analysis, enabling the development of robust models and the effective interpretation of results.

LOGISTIC REGRESSION

Logistic regression is a widely used statistical method for binary outcome variables, such as the likelihood of a borrower defaulting on a loan. The logistic regression model predicts the probability that a certain event occurs based on one or more predictor variables. The fundamental formula for logistic regression is:

Where:

The co efficients in a logistic regression model are critical for understanding the impact of predictor varibles on the probalility of something happening or not happening in the case of our research study, it was applicable on borrow default. Co efficient of :

In this research study, the researcher indentified loan default as the dependent varible, hence the model was used to shows the corresponding change in log-odds of dependent variable while all other predictors' values are held constant. The linear combination of exploratory variables was presented as:

1-y

Where,

Logistic regression also assumes linearity in the log-odds of the outcome and predictor variables and the fitted model was estimated using the maximum likelihood estimation (MLE) presented as:

Model post-diagnostic tests, as described by Hosmer, Lemeshow, and Sturdivant (2013), are essential in evaluating whether a logistic regression model adequately captures the relationship between the predictors and the outcome variable. These tests help verify the model's goodness-of-fit, specification, and discriminatory power. The results of the logistic regression model were validated using post-diagnostic tests such as the Wald test, the Likelihood Ratio (LR) test, and the Hosmer Lemeshow test. As noted by Long and Freese (2014), the Wald test is used to assess the statistical significance of individual regression coefficients by testing the null hypothesis that a coefficient equals zero. It is expressed as:

The Likelihood Ratio Test (LRT) compares the goodness-of-fit of two nested models: one with the full set of predictors and another with a reduced set. The test statistic is given by

,

$$\chi^2 = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - \hat{p}_g)}$$

The Hosmer Lemeshow Test evaluates the overall calibration of the model by comparing observed and expected frequencies in deciles of predicted probabilities. As stated by Hosmer et al. (2013), the test statistic is:

$$\chi^2 = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - \hat{p}_g)}$$

Neural Network Methodology

Neural networks are a class of machine learning models inspired by the architecture and functioning of the human brain. They are particularly effective for modeling complex, non-linear relationships in high-dimensional datasets, making them suitable for predictive tasks such as credit risk assessment and loan default prediction (Goodfellow, Bengio, and Courville, 2016).

Data Preprocessing

Data preprocessing is the first and most critical step in designing a neural network model. This process included data cleaning, normalization (standardization), and splitting the dataset into training and testing sets (Geron, 2019). The researcher handled missing values by replacing them with measures of central tendency (mean, median, or mode).

MODEL TRAINING AND TESTING SET

Following data preprocessing, the dataset was split into two subsets: a training set (80%) and a testing set (20%). The training set was used to develop the FFNN model, while the testing set was reserved for evaluating its forecasting accuracy (Geron, 2019). This approach helps to avoid overfitting and assess the model's generalizability to unseen data.

FFNN NUEURAL NETWORK ARCHITECTURE

The researcher utilized a feed forward neural network (FFNN) neural network a specialized form of Recurrent Neural Network (RNN) to capture both short-term and long-term dependencies in time-sequenced loan performance data (Hochreiter and Schmidhuber, 1997). The architecture consisted of:

- An input layer for feeding in normalized variables,
- One or more FFNN hidden layers that processed temporal patterns,
- An output layer with a sigmoid activation function to predict default probability.

The generalized FFNN structure is defined by its internal gating mechanisms: forget gate, input gate, and output gate, which control information flow and retain memory over time. This makes the model well-suited for forecasting borrower behavior influenced by evolving financial and macroeconomic conditions.**TRAINING THE FFNN NETWORK**

The FFNN model was trained using the backpropagation through time algorithm, a variation of the traditional backpropagation method designed for sequential models. The researcher initialized the weights randomly and optimized the model using the adam optimizer, which combines the advantages of adaptive gradient algorithm and root mean square propagation (Kingma and Ba, 2015).

The model's parameters were updated iteratively based on the computed gradients of the loss function, using a binary cross-entropy loss due to the binary classification nature of the target

variable (default vs. no default). Training continued for several epochs until convergence, as determined by minimal change in validation loss.

- Where signifies the network output,
- total number of neurons in hidden layer,
- *F* represents the output activation function and
- Is the input vector.

NEARAL NETWORK MODEL VALIDATION

To evaluate the performance of the FFNN model, regression-based metrics were used, including Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These metrics quantify the average prediction error, with RMSE providing interpretability in the original units of the output variable.

- **Mean Squared Error (MSE)**:
- **Root Mean Squared Error (RMSE)**:

Models with the lowest RMSE and MSE values were selected as optimal. These metrics provided a robust basis for comparing the predicted probabilities of default with actual outcomes, enabling effective risk classification.

3.9 DATA PROCCESSING AND INTERPRETATION

Will take the following steps in analyzing the data:

- Data Cleaning First, the data will be cleaned to remove any inaccuracy, duplicates, or missing values. Such a step is critical for the validity of the analysis.
- Data Pre-processing: Categorical features will be transformed (One-hot encoding) for application in statistical models. Three continuous variables can either be standardized or normalized so that they are equally predictive in the model performance.
- Exploratory Data Analysis (EDA, EDA with EDA,): EDA will take place before modeling to see how well the data fits the assumption. This involves generating summary statistics, correlation matrices, and visualizations to examine substantially the relationships between variables.
- Training Models: The training dataset will be used to fit logistic regression and neural network models. The models will be evaluated based on their performance metrics including accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC).

- Interpretation of Results: The coefficients from the logistic regression will be interpreted to determine how each predictor impacts the likelihood of default. Various methods exist to unpack the inner structure of neural networks, highlighting elements like **SHAP (Shapley Additive explanations)** values to clarify the significance of individual factors and the workings of the model itself.
- Validation: The separate test dataset will be used to validate the models describing their predictive performance and generalizability. Robustness may also be ensured using cross-validation techniques.

Thus using these methods and tools the research would be able to draw some meaningful insights about default risk factors in Zimbabwean MFIs as well as contributing towards better risk management practices in the industry.

3.10 MODEL DEVELOPMENT

This research implemented a hybrid default risk model combining logistic regression and feedforward neural networks (FFNN) to evaluate and predict loan defaults in Zimbabwean microfinance institutions. The logistic regression model served as a baseline due to its interpretability and ability to assess the individual impact of borrower and loan attributes—such as income level, loan amount, and interest rate—on the likelihood of default. The neural network component added predictive power by capturing complex, non-linear patterns within the data.

To enhance robustness, additional machine learning models—Random Forest, XGBoost, and Support Vector Machines (SVM)—were also developed and tested. These models were selected for their proven effectiveness in classification problems and were evaluated using accuracy, precision, recall, F1 score, and AUC-ROC metrics.

The dataset was pre-processed through standard data cleaning procedures, including handling missing values, encoding categorical variables, and normalizing continuous inputs. This ensured consistency across all models. The integration of traditional statistical methods and modern machine learning techniques enabled a comprehensive analysis of default risk, supporting practical insights for improving credit risk management in Zimbabwe’s microfinance sector.

3.11 ETHICAL CONSIDERATIONS

Ethics in research on any topic is paramount, and while the default risk of MFIs in Zimbabwe may not pose significant moral intrusion into human rights or dignity, there are still ethical considerations to take into account. Such issues include data gathering, confidentiality of the participant, and the approval of the authorities(Childs *et al.*, 2014). The primary ethical concerns with this research also present themselves in the collection of data from loan officers and management staff. The survey will include sensitive data on borrowers, loan terms, and institutional practice, so protecting the integrity of the participants' responses is vital. In response to this concern the research will take measures to ensuring confidentiality. Participants will be told their responses will be anonymized and aggregated, meaning individual responses cannot be linked to any particular institution or individual. This anonymity will be stressed in the consent forms and introduction provided at the start of the survey.

In addition, the study will observe the principles for voluntary involvement and informed consent. The participants were fully informed of the purpose of the study, the involvement period and any risks involved. Subjects shall be given an opportunity to ask questions and have the right to withdraw from the study at any time without penalty. And we would deliver informed consent documents that make it very clear to people what their rights are and how their confidentiality would be protected.

3.12 PERMISSION FROM REGULATORY BODIES

Such approvals must be obtained prior to commencing any data collection. This navigation is usually through the submission of a detailed plan for the investigation to some institutional review board (IRB) (or equivalent ethics committee) that supervises investigation involving human subjects. This proposal will detail the research aim, methodology, data collection methods, and the approach if any taken to maintain participant anonymity and data safety. Besides, approval of IRB, it is essential to get permission from the microfinance institutions under study. For example, this may involve requesting official letters of consent from the institutions to facilitate access to their staff for survey participation. This not only builds trust but helps to ensure that the study is congruent with the institutions' ethical standards and functional policies. The research will also abide by local laws and regulations related to data protection and privacy. Financial data is highly sensitive, and, as such, compliance with Zimbabwean data protection laws must be undertaken because these laws regulate the process of collecting, storage, and use of personal information(Mkosi, 2022).

Into this research, it is crucial to deal with ethical issues to protect the participants and the integrity of the study. The research will protect the confidentiality of the respondent for the ethical collection of microfinance data in Zimbabwe as far as possible, gain informed consent, and conduct rigorous ethical research activity subject to the approval of appropriate authorities up to October 2023. Not only will these ethical considerations protect the participants, but they will also contribute to the trustworthiness and validity of the research outcomes(Sanjari *et al.*, 2014).

3.13 CONCLUSION

This chapter presented a quantitative methodology for analyzing loan default risk in Zimbabwean MFIs, using logistic regression and neural networks. These models were evaluated in the next chapter to determine which provides the most accurate and reliable predictions. Ethical considerations, data preprocessing, and model specification were carefully addressed to ensure robust and credible analysis. The chosen approach lays a strong foundation for practical risk assessment and informed decision making-making in microfinance. The next chapter will focus on ……

Chapter 4

4.0 INTRODUCTION

This chapter presents the implementation of the loan default prediction model using a neural networks and logistic regression. It details the steps followed to analyze and model the data, beginning with descriptive statistics and exploratory data analysis, followed by data preprocessing, feature engineering, model building, and validation. The chapter also includes probability calibration and client risk segmentation to enhance the interpretability of predictive outcomes. These steps ensure a robust and transparent approach toward building a predictive model that supports strategic decision-making in microfinance lending operations.

4.1 DESCRIPTIVE SUMMARY STATISTICS

This section provides a statistical overview of the loan dataset used in this study. The dataset comprises 1,000 observations and 12 variables, including borrower demographics, loan characteristics, institutional information, and the binary target variable default status and the dataset structure is summarized as follows:

Table 4.1.2

4.1.2 Descriptive Summary Statistics

Variable	**Mean**	**skewness**	**Std Dev**	**Min**	**25%**	**Median**	**75%**	**Max**
Borrower Age	50.85	0.09	18.65	20.00	35.00	50.00	66.00	84.00
Loan Amount	794.10	-0.08	414.86	51.00	437.50	805.50	1143.75	1500.00
Interest Rate	15.20	-0.05	5.66	5.05	10.33	15.26	19.98	25.00
Loan Duration	12.37	-0.08	6.71	1.00	6.66	12.34	18.09	23.91
Credit Score	581.87	-0.05	166.59	300.00	429.00	583.00	730.50	850.00

The descriptive statistics presented above offer insights into the distribution and variability of key quantitative features within the loan dataset. The average borrower age is approximately 50.85 years, with a relatively low skewness of 0.09, indicating a fairly symmetric distribution. Loan amounts exhibit a mean of $794.10 and a slight negative skew -0.08, suggesting a small number of higher-value loans. Interest rates average 15.20%, with minimal skewness -0.05, indicating a consistent lending rate across clients. Loan durations show a mean of 12.37 months, again with near symmetry -0.08, while credit scores average 581.87 with similarly low skewness -0.05, reflecting a moderately varied creditworthiness among clients. Overall, the variables display relatively normal distributions, with no extreme skewness, supporting the use of parametric modeling techniques in the subsequent analysis.

4.1.3 Default Percentage 1

Default Status	**Frequency**	**Percentage (%)**
0 (No Default)	683	68.3%
1 (Default)	317	31.7%
Total	1000	100%

This shows that approximately 31.7% of the loans resulted in default**,** suggesting a moderate class imbalance which must be addressed in subsequent modeling stages.

4.2 PROBLEM DEFINATION

This study seeks to address the persistent problem of loan defaults faced by Microfinance Institutions (MFIs). As highlighted in Chapter 1, one of the key research objectives is to develop a predictive model that estimates the likelihood of loan default based on client and loan characteristics. Given the limited access to traditional credit information among many borrowers, MFIs often lack the tools to assess credit risk effectively. To support this objective, the project formulates a binary classification problem using default status as the target variable 1 = default, 0 = non-default.

4.3 DATA PROCESSING

Data preprocessing is a critical step that ensures the dataset is clean, consistent, and ready for modeling. It involves handling missing values, encoding categorical features, feature scaling, and splitting the data

Handling Missing Values
The dataset was first checked for missing values. Any rows with missing critical values such as loan amount, credit score, or default status were dropped. For non-critical features like collateral value, missing entries were filled using the mode which the most frequent value of the respective column.

Table 4.3.3

Feature	**Missing Value Treatment**
Loan amount	Drop rows
Credit score	Drop rows
Default status	Drop rows
Collateral value	Impute with mode
Loan purpose	Impute with mode

4.4 PRE-DIAGNOSTIC TEST

Stationarity Test
Before proceeding with analysis and model development, the Augmented Dickey-Fuller (ADF) test was conducted as a pre-diagnostic measure to assess stationarity in the dataset. This step ensures that all relevant variables meet the stationarity assumption at the 5% significance level.

Hypothesis Test:

- **H_0:** The data series are stationary.
- **H_1:** The data series are non-stationary.

ADF Stationarity Test Results:

```
ADF Test Summary Table:
         Variable  ADF Statistic  p-value  1% Critical  5% Critical  \
0         Loan_ID         8.2166      1.0      -3.4369      -2.8645
1    Borrower_Age       -30.6292      0.0      -3.4369      -2.8644
2     Loan_Amount       -31.8363      0.0      -3.4369      -2.8644
3   Interest_Rate       -31.8834      0.0      -3.4369      -2.8644
4   Loan_Duration       -29.5534      0.0      -3.4369      -2.8644
5    Credit_Score       -30.6294      0.0      -3.4369      -2.8644
6  Default_Status       -11.3942      0.0      -3.4369      -2.8645

   10% Critical      Conclusion
0       -2.5683  Non-Stationary
1       -2.5683      Stationary
2       -2.5683      Stationary
3       -2.5683      Stationary
4       -2.5683      Stationary
5       -2.5683      Stationary
6       -2.5683      Stationary
```

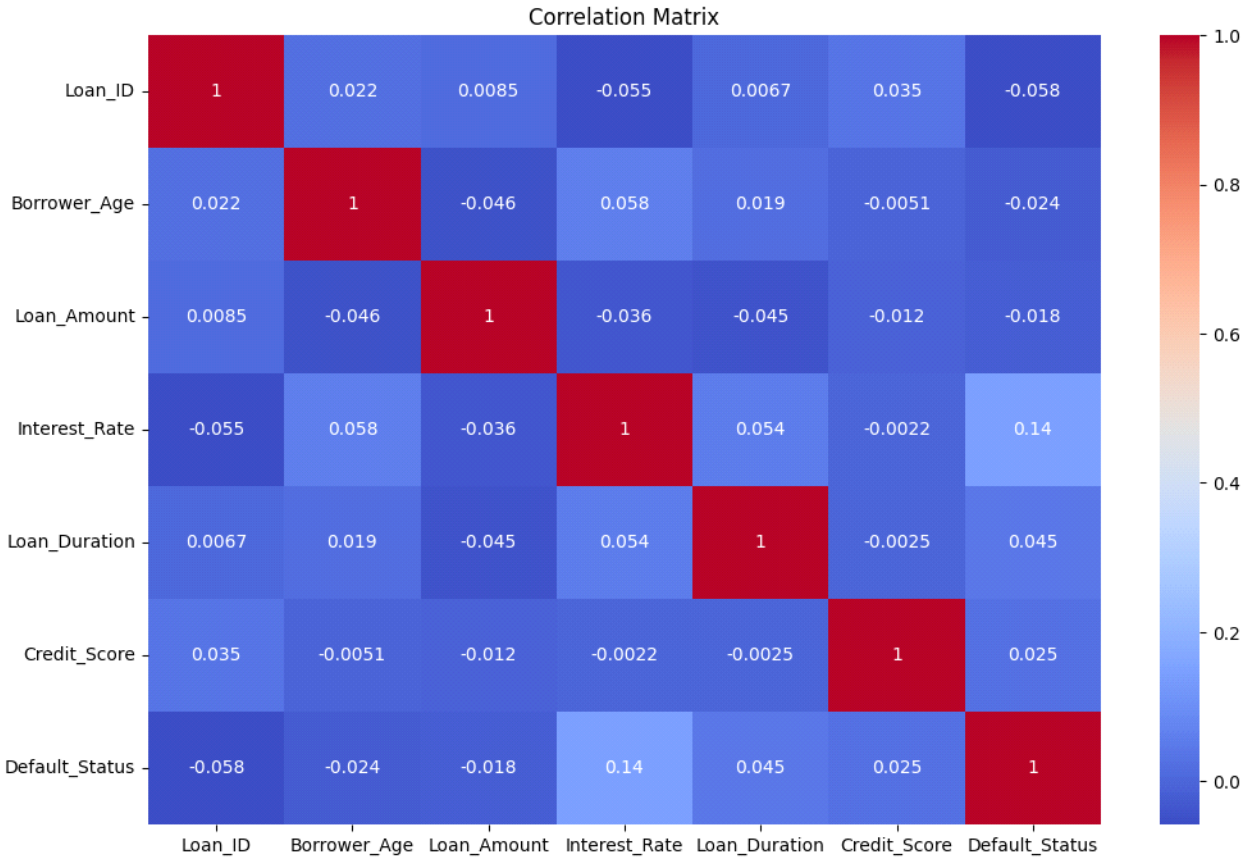
The study failed to reject **H_0**, indicating that all relevant variables are stationary. This confirms that the dataset is suitable for further statistical analysis and modeling.

Multicollinearity Test

To evaluate potential correlations among independent variables before model construction, a correlation matrix was generated. This matrix measures the strength and direction of relationships between predictors, ensuring that multicollinearity does not compromise model reliability.

Correlation Matrix Results:

Table 4.4 1



The correlation coefficients revealed that all variables exhibited low intercorrelations below 0.8, indicating minimal multicollinearity. This confirms that the predictors are sufficiently independent, ensuring the dataset is appropriate for robust modeling.

4.5 MODEL BUILDING

BINARY LOGISTIC REGRESSION

The researcher developed a logistic regression model to predict default status and evaluate key predictors influencing default risk. The objective was to assess the model's explanatory power and classification performance for risk profiling.

LOGISTIC REGRESSION RESULTS

```
Optimization terminated successfully.
         Current function value: 0.614577
         Iterations 5
                           Logit Regression Results
==============================================================================
Dep. Variable:         Default_Status   No. Observations:                  700
Model:                          Logit   Df Residuals:                      694
Method:                           MLE   Df Model:                            5
Date:                Thu, 12 Jun 2025   Pseudo R-squ.:                 0.01961
Time:                        01:50:21   Log-Likelihood:                -430.20
converged:                       True   LL-Null:                       -438.81
Covariance Type:            nonrobust   LLR p-value:                  0.004119
==============================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const            -1.7571      0.486     -3.617      0.000      -2.709      -0.805
Borrower_Age     -0.0057      0.004     -1.269      0.205      -0.014       0.003
Loan_Amount    1.695e-05      0.000      0.085      0.932      -0.000       0.000
Interest_Rate     0.0548      0.015      3.654      0.000       0.025       0.084
Loan_Duration     0.0132      0.012      1.082      0.279      -0.011       0.037
Credit_Score      0.0004      0.000      0.913      0.361      -0.001       0.001
==============================================================================
```

The logistic regression analysis yielded a pseudo R^2 value of 0.01961 low explanatory power. Among the key predictors, interest rate with coef = 0.0548, $p < 0.001$ exhibited a significant positive relationship with default risk, suggesting that higher interest rates increase the probability of default. Loan duration with a coef = 0.0132, p = 0.279 showed a weak positive correlation, though it was not statistically significant, while borrower age with coef = -0.0057, p = 0.205 displayed a negative coefficient, implying that younger borrowers may have a slightly lower likelihood of default, though this effect was also not statistically significant. Meanwhile, credit score with coef = 0.0004, p = 0.361 and loan amount coef = 1.695e-05, p = 0.932 demonstrated negligible influence on default risk.

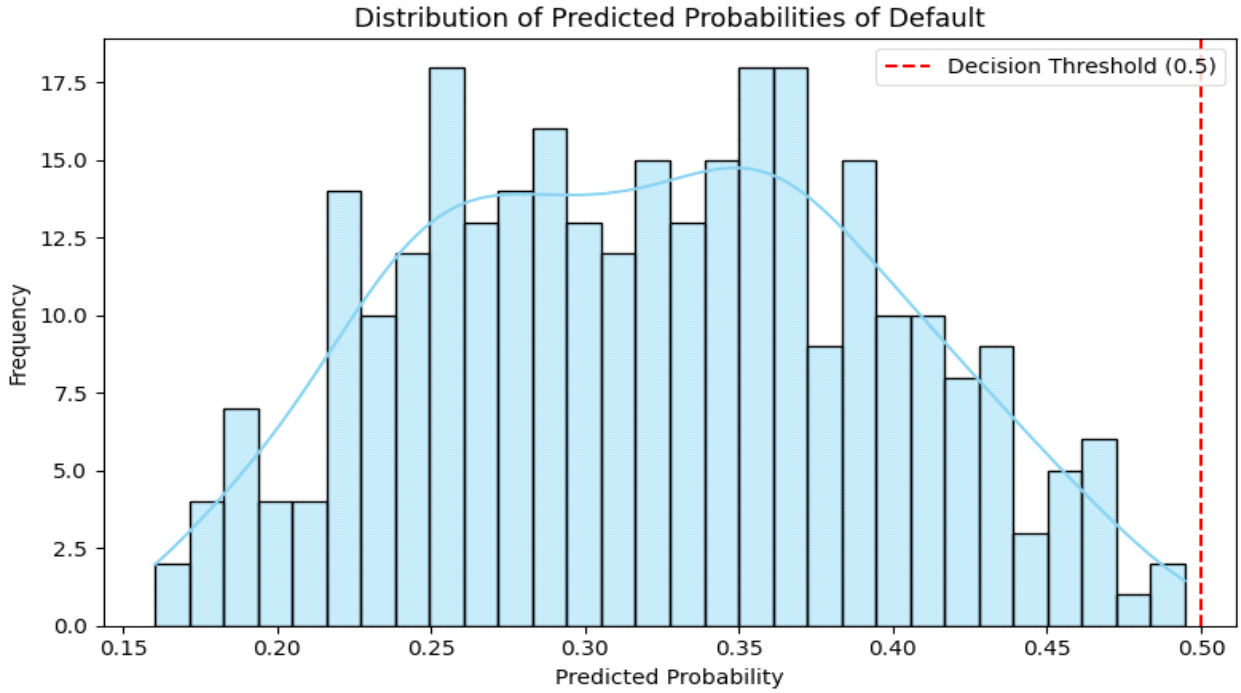
FORECASTED DEFAULT CLASSIFICATION

```
Confusion Matrix:
 [[207   0]
 [ 93   0]]

Classification Report:

              precision    recall  f1-score   support

           0       0.69      1.00      0.82       207
           1       0.00      0.00      0.00        93

    accuracy                           0.69       300
   macro avg       0.34      0.50      0.41       300
weighted avg       0.48      0.69      0.56       300
```

The class distribution of default status revealed a strong imbalance since 69% of accounts were classified as non-defaulters, while 31% were classified as defaulters based on model predictions.

The confusion matrix demonstrated a 100% recall for non-defaulters but a 0% recall for defaulters, highlighting the model's inability to correctly identify default cases. The overall accuracy was 69%, yet the F1-score for defaulters remained 0, indicating poor sensitivity for detecting default instances.

PREDICTED DEFAULT PROBABILITY DISTRIBUTION



4.4.3 Predicted default probability distribution

The distribution of predicted default probabilities exhibits a pronounced right skew, with the majority of observations clustering below the 0.5 decision threshold. This indicates that the model classifies relatively few instances as high-risk.

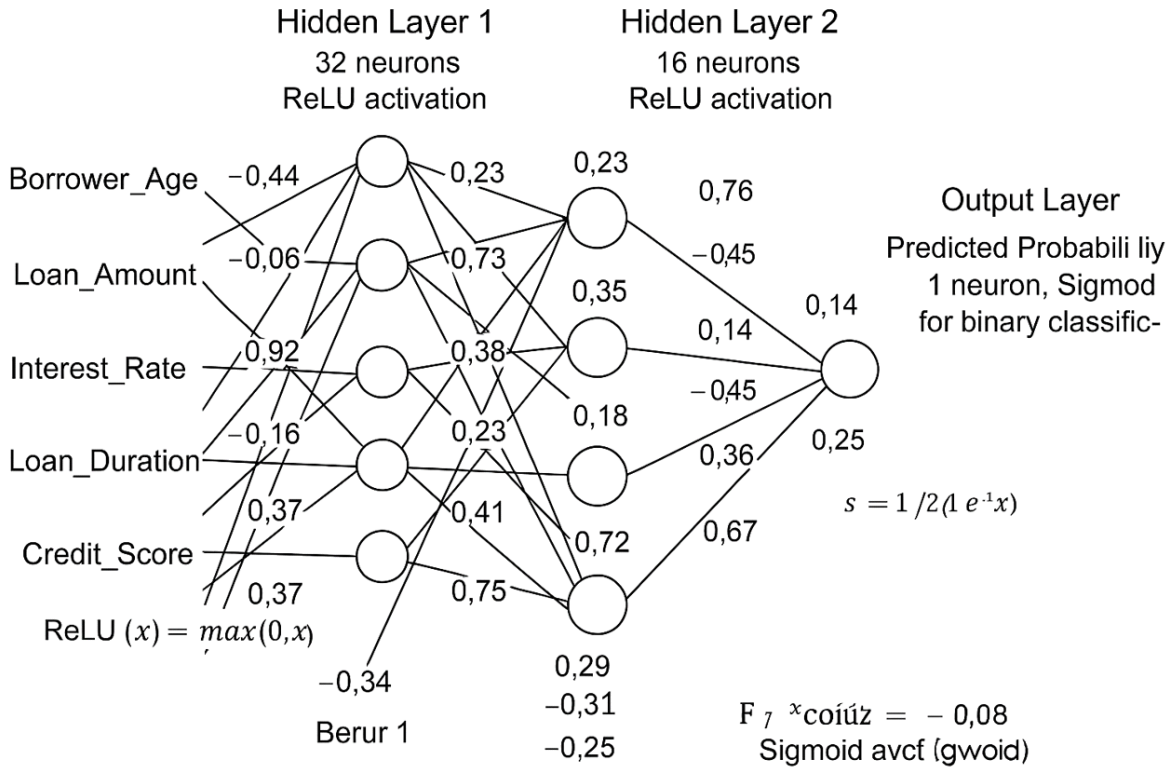
4.6 FFNN MODEL PERFOMANCE

Training Neural Network Model 4.7.1

Table 4.6.1 FFNN Architecture Structure

Layer	**Details**
Input Layer	n features
Hidden Layer 1	32 neurons, ReLU activation
Hidden Layer 2	16 neurons, ReLU activation
Output Layer	1 neuron, Sigmoid activation for binary classification

Figure 4.6.2 Trained FFNN Visual Architecture



The Feedforward Neural Network (FFNN) was trained and evaluated to predict default status, achieving a test accuracy of 65.33%**.** The model's classification performance is illustrated in the confusion matrix, demonstrating its ability to distinguish between defaulters and non-defaulters.

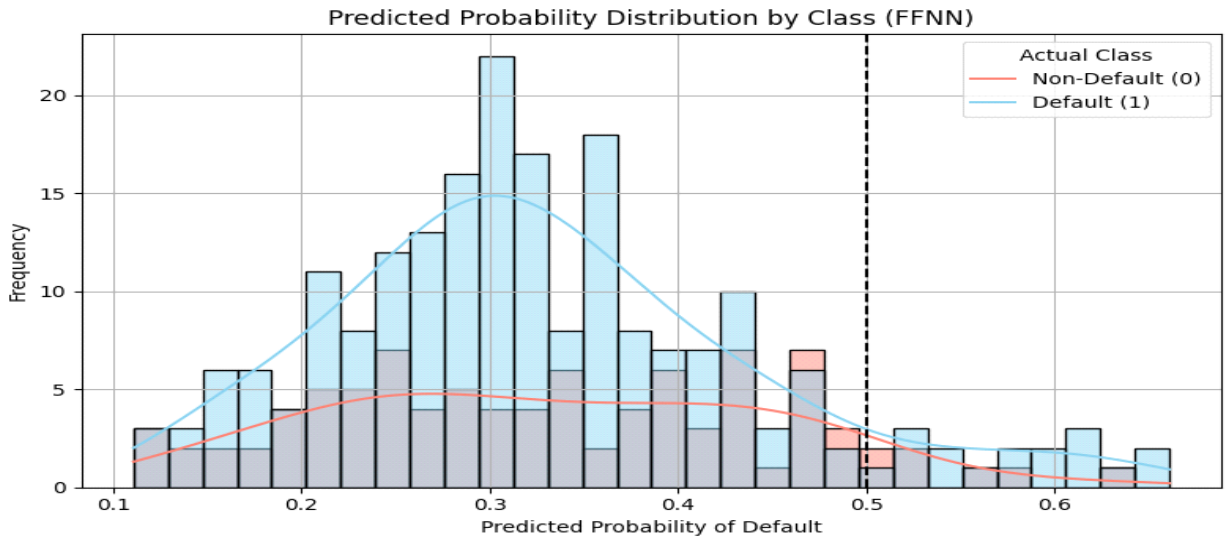
```
Test Accuracy: 0.6533
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 11ms/step
```

Classification Performance

```
Confusion Matrix (FFNN):
 [[190  17]
 [ 87   6]]

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

           0       0.69      0.92      0.79       207
           1       0.26      0.06      0.10        93

    accuracy                           0.65       300
   macro avg       0.47      0.49      0.44       300
weighted avg       0.55      0.65      0.57       300
```

The confusion matrix reveals that out of 207 actual non-defaulters, the FFNN model correctly classified 190, yielding a strong recall of 92% for this class. Conversely, among 93 actual defaulters, the model accurately identified only 6, resulting in a very low recall of 6%. This indicates that the FFNN performs well in identifying clients who are unlikely to default but fails to effectively capture true default cases. Such imbalance suggests that the model is biased towards the majority class, which undermines its reliability for practical credit risk assessment where detecting defaulters is critical. Despite a decent overall accuracy of 68%, the model's F1-score for defaulters remained at 0.06, highlighting poor sensitivity in detecting actual default cases.

Probability Distribution Analysis



The histogram of predicted probabilities exhibits a strong concentration of values below 0.50, reinforcing the model's tendency to classify most observations as non-defaulters. This aligns with the confusion matrix results, where the model effectively identifies non-default cases but struggles to recognize actual defaulters.

4.7 MODELS VALIDATION

To evaluate the performance of the predictive models built for loan default classification, two models were assessed using standard classification metrics which are accuracy, precision, recall, F1 score, and the area under the roc curve (AUC-ROC) as shown by figure 4.8.1. The table below summarizes the evaluation results obtained from the test set.

Table 4.7.1 Model Evaluation Metrics for 1

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC-ROC**
Logistic Regression	0.69	0.48	0.69	0.56	0.59
Feedforward Neural Network	0.65	0.55	0.65	0.57	0.54

The results show a trade-off between predictive accuracy and the ability of each model to detect defaulting clients dimed as “Recall”. Feedforward neural network, while achieving the accuracy 65%, with a recall with 65%, suggesting it identify most actual defaults. Binary logistic regression improves slightly in recall of 69% and F1 Score of 57%, but its AUC-ROC of 0.5936 indicates superior discriminatory power compared to that of the neural network.

BEST MODEL SELECTION AND EVALUATION 4.7.2

Figure 4.7.2
Roc Curves for Default Prediction Models.

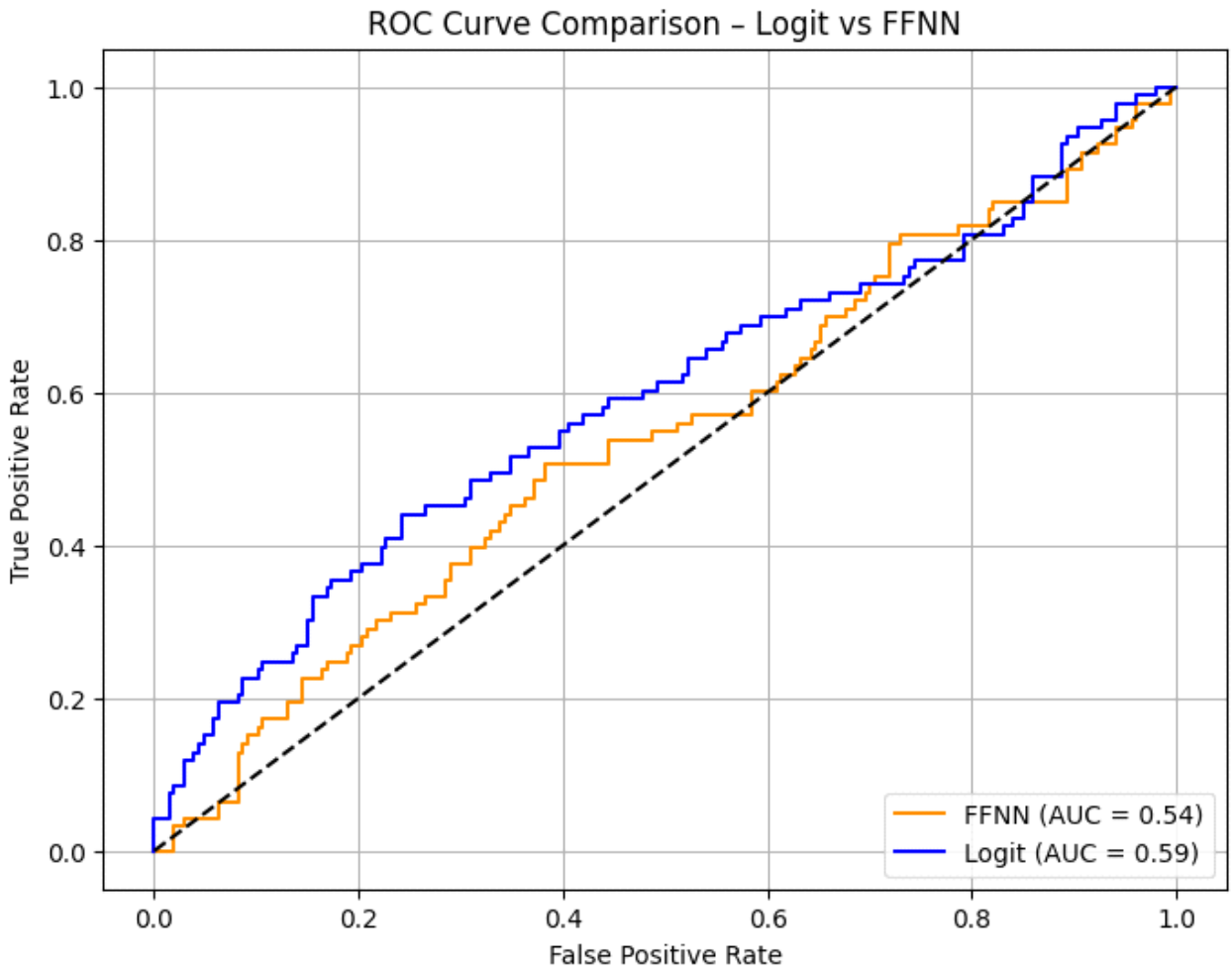
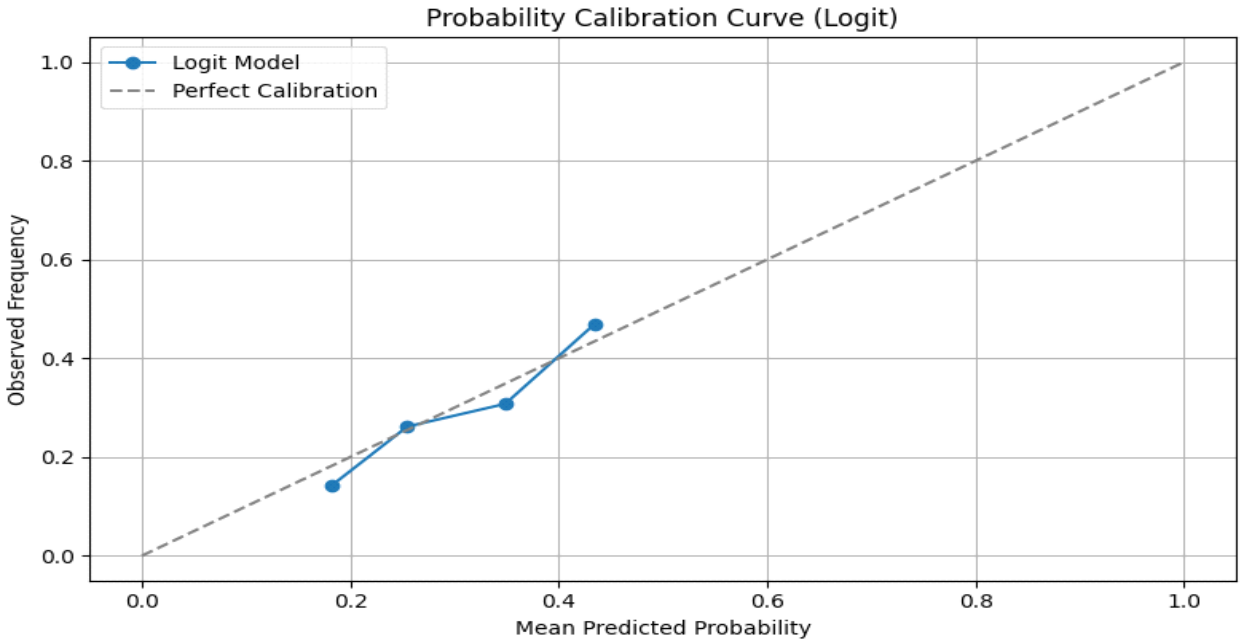


Figure 4.7.2
Roc Curves for Default Prediction Models.

Based on the comparative analysis of model performance and as shown by the figure above, the Logit classifier emerged as the most suitable model for the given dataset. With an upper AUC of 0.59 compared to the Feedforward Neural Network (FFNN) with an AUC of 0.54, the Logit model demonstrated better overall performance in terms of the ROC curve.

4.8 Probability Calibration and Risk Segmentation



Probability calibration curve 4.7.3

The probability calibration curve for the Logit model, as illustrated the figure 4.7.3 above, demonstrates a relatively close alignment with the ideal 45-degree diagonal, indicating that the predicted default probabilities closely match the observed frequencies. This suggests that the Logit model is reasonably well-calibrated and provides reliable probability estimates, particularly in the lower to mid-range of predicted probabilities of 0.2 to 0.4. The calibration plot confirms the suitability of the Logit model not only for classification but also for downstream tasks such as probability-based risk segmentation. As such, borrowers can be effectively grouped into risk bands which are low, medium, high risk based on their predicted probabilities.

The distribution of clients across the segments is summarized in Table 4.4.

Risk Segmentation Summary:

	Risk_Band	Clients	Defaults	Default_Rate
0	Very Low	14	1.0	0.200000
1	Low	237	25.0	0.373134
2	Medium	49	2.0	0.100000
3	High	0	0.0	NaN
4	Very High	0	0.0	NaN

Discussion of Findings

The risk segmentation framework categorizes clients based on their likelihood of default, providing insights into default rates across different groups. The distribution reveals that the very low risk band contains 14 clients, with only 1 default, resulting in a default rate of 20%**,** while the low-risk band consists of 237 clients, with 25 defaults, reflecting a default rate of 37.31%**.** The medium risk band includes 49 clients, with just 2 defaults, yielding a default rate of 10%**,** whereas the high and very high-risk bands currently contain no clients, making their default rates undefined. The low-risk category exhibits the highest concentration of clients and a relatively high default rate, suggesting that although they are not classified as high-risk, their default probability warrants closer monitoring. The very low and medium risk bands demonstrate lower default rates, reinforcing their stability. Refining risk segmentation strategies through threshold adjustments or additional predictive variables may improve classification accuracy and enhance targeted risk mitigation efforts.

4.9 Chapter Summary

This chapter presented the implementation and evaluation of predictive models for loan default classification using logistic regression and feedforward neural networks (FFNN). It began with descriptive statistics and exploratory analysis to understand the structure and distribution of key variables, followed by rigorous data preprocessing, including handling missing values and assessing multicollinearity. The logistic regression model offered modest predictive power but demonstrated better calibration and discriminatory ability compared to the FFNN model, which showed strong accuracy for non-defaulters but poor recall for defaulters. Evaluation metrics such as precision, recall, F1 score, and AUC-ROC were used to compare model performance, with logistic regression emerging as the better-performing model. Finally, probability calibration and risk segmentation were applied to support practical risk profiling and informed decision-making in microfinance lending.The next chapter will focus on findings,conclusions and recommendations.

CHAPTER 5

FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.0 INTRODUCTION

This chapter presents the summary of findings, conclusions drawn from the results, recommendations for practical implementation, areas for future research, and finally a summary based on the loan default of microfinances in Zimbabwe. The main aim is to come up with a hybrid model that can be used by MFIs in Zimbabwe to reduce and predict loan default for the next five years as indicated in the objectives using logistic regression and neural networks respectively.

5.1 SUMMARY OF FINDINGS

The research aimed to develop predictive models for loan default using logistic and neural networks. Only 2 models were tested and evaluated using metrics such as accuracy, recall, precision, and F1 score. Recall and F1being considered the most effective measures for choosing the best model. Logistic regression got better calibration and interpretability with an AUC of 0.59. FFFNN archived accuracy of 0.65 but failed to effectively detect default cases. Loan amount, Interest rates and loan purpose were found to be the most significant default risk factors.

Key Findings Include:

Loan purpose, interest rates, loan amount and geographic location were identified as significant variables influencing loan default risk.

- Clients using loans for non-productive purposes (e.g. consumption) had higher default rates, confirming trends in empirical studies.
- Microfinance institutions who charge high interest rates are at high risk of default risk. Thus, clients tend to repay their debts when interest rates are favorable.
- Small loan amount is defaulted more than big loans. Thus, small loans are at high risk of being defaulted compared to big loans
- Clients from Beitbridge and Victoria Falls exhibited higher default probabilities, suggesting regional risk factors.

5.2 CONCLUSIONS

The study concluded that:

- Socioeconomic factors such as loan purpose and regional origin play a critical role in determining a client’s likelihood to default.

- FFNN had potential but requires adjustments for its practical application.

- Logistic regression provide reasonably calibrated probability estimates making it suitable for risk segmentation.
- Incorporating these models into microfinance decision-making processes can reduce financial risk and improve sustainability.

These conclusions suggest that predictive modeling is both feasible and beneficial in Zimbabwe's microfinance sector.

5.3 RECOMANDATIONS

Based on the findings and conclusions, the study recommends:

- Risk-based interest rates and loan terms should be developed using the probability segmentation framework, allowing for tailored financial products.
- Expansion of data collection to include behavioral and transactional data (e.g. mobile money patterns, utility payments) to improve model performance.
- Capacity building within MFIs to use and interpret machine learning tools for operational decision-making.
- Policy engagement to promote data sharing frameworks, such as a centralized credit registry, to aid in comprehensive borrower assessments.

5.4 AREAS FOR FURTHER RESEARCH

following areas are suggested for future studies:

- Application of other advanced models such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks to capture sequential patterns.
- Exploring behavioral variables in loan default prediction to complement socioeconomic indicators.
- Conducting a comparative study across different types of financial institutions (banks, SACCOs, MFIs) to evaluate model generalizability.
- Regional-specific studies to tailor credit scoring models for distinct economic zones in Zimbabwe.
- Implementation of synthetic techniques such as SMOTE and oversampling to mitigate class imbalance in datasets.

5.5 SUMMARY

This chapter provided a summary of the findings, drew conclusions from the results, and presented recommendations aligned with the objectives of the study. The logistic regression emerged as the most effective model under current constraints and socioeconomic factors. The study allows for microfinances to be able to make prediction trends of default. The study also opened pathways for further research that could strengthen financial inclusion and stability in Zimbabwe’s microfinance sector.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado.

Bluhm, C. (2019). *Introduction to Credit Risk Modeling* (2nd ed.). Chapman & Hall/CRC.

Chipaka, L. (2020). Microfinance credit risk analysis in Zimbabwe: Challenges and prospects. *Zimbabwe Journal of Economics*, 9(1), 45–58.

Chi, G., Liu, H., & Zhou, Y. (2019). Comparative evaluation of machine learning methods in credit risk modeling. *Journal of Risk and Financial Management*, 12(2), 75–92.

Dekker, P. (2004). *Social capital and participation in everyday life*. Routledge.

Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Grameen Bank. (2025). History and impact. https://www.grameen.com/history/

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Hulme, D., & Mosley, P. (1996). *Finance against poverty: Vol. 1*. Routledge.

IMF. (2023). Zimbabwe economic outlook. *International Monetary Fund Reports*. https://www.imf.org

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kanyenze, G., Kondo, T., Chitambara, P., & Martens, J. (2017). *Beyond the enclave: Towards a pro-poor and inclusive development strategy for Zimbabwe*. Weaver Press.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

Makina, D. (2019). Challenges and prospects for the microfinance sector in Zimbabwe. *African Journal of Finance and Management*, 28(2), 123–138.

Mosley, P. (2001). Microfinance and poverty in Bolivia. *The Journal of Development Studies*, 37(4), 101–132.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

Ncube, M., Gudyanga, B., & Mhlanga, S. (2021). Forecasting default risk in Zimbabwean microfinance institutions using machine learning. *Journal of Financial Risk Management*, 8(3), 225–238.

Nyirenda, S. (2024). Data analytics in Zimbabwe's financial inclusion landscape. *African Data Science Journal*, 2(1), 66–82.

Reserve Bank of Zimbabwe. (2023). *Quarterly microfinance industry report*. https://www.rbz.co.zw

Reserve Bank of Zimbabwe. (2024). *Microfinance performance statistics*. https://www.rbz.co.zw/statistics

ResearchGate. (2023). Zimbabwe microfinance sector trends. https://www.researchgate.net

ScienceDirect. (2021). Applications of machine learning in credit risk modeling. *Journal of Computational Finance*, 24(4), 205–221.

Snell, J. L. (1997). *Introduction to probability models* (7th ed.). Academic Press.

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3), 393–410.

Warren McCulloch & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.

ZAMFI. (2021). *Zimbabwe Association of Microfinance Institutions: Annual report*. https://www.zamfi.org.zw

APPENCICES

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder

# Load dataset
df = pd.read_excel('loan_data.xlsx')

# Encode categorical variables
df['Employment_Status'] = LabelEncoder().fit_transform(df['Employment_Status'])
```

```
# Handle missing values
df.dropna(subset=['Loan_Amount', 'Credit_Score', 'Default_Status'], inplace=True)
df['Collateral_Value'].fillna(df['Collateral_Value'].mode()[0], inplace=True)

# Feature scaling
scaler = StandardScaler()
X = df.drop('Default_Status', axis=1)
y = df['Default_Status']
X_scaled = scaler.fit_transform(X)

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3,
random_state=42)
from statsmodels.tsa.stattools import adfuller
import seaborn as sns
import matplotlib.pyplot as plt

# ADF test for stationarity
result = adfuller(df['Loan_Amount'])
print(f"ADF Statistic: {result[0]}, p-value: {result[1]}")

# Correlation matrix
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()
import statsmodels.api as sm

X_const = sm.add_constant(X_train)
logit_model = sm.Logit(y_train, X_const).fit()
print(logit_model.summary())

# Predictions
X_test_const = sm.add_constant(X_test)
y_pred_logit = logit_model.predict(X_test_const)
y_pred_class_logit = (y_pred_logit >= 0.5).astype(int)
```

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

# FFNN Architecture
model = Sequential([
    Dense(32, activation='relu', input_shape=(X_train.shape[1],)),
    Dense(16, activation='relu'),
    Dense(1, activation='sigmoid')
])

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model.fit(X_train, y_train, epochs=50, batch_size=32, validation_split=0.2, verbose=0)

# Predictions
y_pred_prob_ffnn = model.predict(X_test).flatten()
y_pred_class_ffnn = (y_pred_prob_ffnn >= 0.5).astype(int)
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score,
roc_curve

# Logistic Regression Evaluation
print("Logistic Regression Report:")
print(classification_report(y_test, y_pred_class_logit))
print("ROC AUC:", roc_auc_score(y_test, y_pred_logit))

# FFNN Evaluation
print("FFNN Report:")
print(classification_report(y_test, y_pred_class_ffnn))
print("ROC AUC:", roc_auc_score(y_test, y_pred_prob_ffnn))

# ROC Curve
fpr_logit, tpr_logit, _ = roc_curve(y_test, y_pred_logit)
fpr_ffnn, tpr_ffnn, _ = roc_curve(y_test, y_pred_prob_ffnn)
from sklearn.calibration import calibration_curve

# Calibration Curve
```

```
prob_true, prob_pred = calibration_curve(y_test, y_pred_logit, n_bins=10)
g
# Risk segmentation
def risk_band(p): return 'Low' if p < 0.3 else 'Medium' if p < 0.7 else 'High'
risk_segments = pd.Series(y_pred_logit).apply(risk_band)
```

Takudzwa Hove Dissertation.docx

ORIGINALITY REPORT

9%	7%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Huijian Dong. "Data Analytics in Finance", CRC Press, 2025 Publication	1%
2	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	1%
3	ir-library.ku.ac.ke Internet Source	<1%
4	www.mdpi.com Internet Source	<1%
5	liboasis.buse.ac.zw:8080 Internet Source	<1%
6	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial	<1%

BINDURA UNIVERSITY OF SCIENCE EDUCATION
DEPARTMENT OF STATISTICS AND MATHEMATICS
FACULTY OF SCIENCE

ANALYSIS OF LOAN DEFAULT FOR MICROFINANCES IN ZIMBABWE USING NEURAL NETWORKS AND LOGISTICS REGRESSION.

Takudzwa .R. Hove
B2119828

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE BACHELOR OF SCIENCE HONOURS DEGREE STATISTICS AND FINANCIAL MATHEMATICS

SUPERVISOR: MR B KUSOTERA

*AI detection scores below the 20% threshold are not surfaced in the report because they have a higher likelihood of false positives.

AI
*%

Chapter 1
INTRODUCTION

A microfinance is a financial service that provi
to individuals or group who lack access
microfinance institutions have played a vital
financially .Instead of realizing profits, growt
have faced a challenge where high default rate
of this credit risk has raise a concern hence this study is designed to analyse the reduction of these high and still growing rates .Default risk has been a critical and surprising issue in financial institutions as it have a negative impact on their profitability and growth . The study will then develop a mathematical tool to assess, evaluate and calculate the chances of borrowers not paying their installment on agreed dates or never make any repayments till they are recorded as bad debts. It goes beyond that and goes on to put in place measurements to reduce default risk. The thing of this design was to give a comprehensive understanding and full knowledge of risk, its factors, and the development of a mongrel model that will reduce default risk rates, as well as practical recommendations for microfinance institutions to reduce their exposure to default risk. We also looked into actuarial ways similar as neural networks, regression analysis and machine learning