

LOAN DEFAULT PREDICTION USING MACHINE  
LEARNING: AN EMPIRICAL STUDY OF  
MICROFINANCE BANKS IN PAKISTAN

BY

ANAM SOOMRO

A thesis submitted in fulfilment of the requirement for the  
degree of Doctor of Philosophy in Islamic Banking and  
Finance

IIUM Institute of Islamic Banking and Finance  
International Islamic University Malaysia

SEPTEMBER 2025

## ABSTRACT

Loan default poses a significant challenge for the financial well-being of microfinance banks of Pakistan, affecting their profitability and overall financial stability. Traditional credit evaluation techniques such as the 5Cs model (Character, Capital, Capacity, Collateral, and Conditions) are human-centric and inherently inefficient, at risk of inaccuracies. Consequently, numerous borrowers at risk of default continue to obtain loans, thus increasing financial risk exposure. To address these challenges, this study applies machine-learning techniques to develop a data-driven loan default prediction model aimed at enhancing risk assessment in microfinance banks. Methodology: This research is a mixed-mode study based on quantitative and qualitative methods. Quantitatively four machine-learning algorithms, such as Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB) are trained and validated on a large real-world dataset gathered from two microfinance banks of Pakistan. The dataset was split into training and testing sets, with 70% of the data used for training and 30% for testing. Feature selection and preprocessing strategies like handling missing values, outlier detection, and normalization of data were employed to improve predictive performance. Whereas the performance metrics accuracy, precision, recall, and F1 score were used to assess the predictive performance of the models. The qualitative aspect of the study involved surveys of loan officers. This was to verify that what was produced by Machine learning was crosschecked against real lending behaviour. Findings: The research determined the most influential predictors of loan defaults to be debt-to-income ratio, instalment amount, and annual income, total number of accounts, loan amount, interest rate, open account, public records, loan term and purpose of debt consolidation. The results indicate that Gradient Boosting performed better than the other models with the best accuracy (77.75%) and precision (55.33%), although recall (11.33%) and F1 score of (18.81%) limited it. Comparatively Random Forest showed slightly less accuracy (77.58%), precision (53.85%), recall (10.19%) and F1 score of (17.14%) whereas Logistic Regression shown reasonable accuracy (77.39%), precision (51.74%) but had the lowest recall (8.74%) and F1 score of (14.95%). The Decision Tree model, while interpretable, had the least overall performance, achieving an accuracy of (68.9%) precision (31.15%), recall (30.32%) and an F1 score of (30.73%). In addition, to crosscheck the machine learning results, practitioner perceptions were included, and it was reaffirmed that Debt to income ratio, instalment amount, and income level were some of the most dominant drivers in human credit evaluation. The research also points out differences between machine learning models and practitioner perceptions, including the use of credit history, climate risk, and borrower reputation drivers usually undervalued by automated models but significant in real-world lending situations. Implications: The findings of this study have strong implications for microfinance banks, policymakers, and regulators are advised to adopt machine learning based credit risk models for enhancing loan screening, minimizing Non-Performing Loans, and maintaining financial sustainability. This study also adds to the literature on financial technology (FinTech) by illustrating how machine learning can enhance credit risk analysis in emerging economies. Recommendations: Future research should focus on hybrid models that combine quantitative machine Learning techniques with qualitative expert judgment through structured interviews to give better complete credit risk assessment. Regulatory frameworks need to be developed to ensure fairness, transparency, and moral

application of AI in credit risk management. Moreover, the incorporation of behavioural and macroeconomic factors in the dataset could improve loan default forecasting, as well.



## خلاصة البحث

يشكل التخلف عن سداد القروض تحديًا كبيرًا للرفاهية المالية لبنوك التمويل الأصغر في باكستان، مما يؤثر على ربحيتها واستقرارها المالي العام. إن تقنيات تقييم الائتمان التقليدية مثل نموذج Cs5 (الشخصية ورأس المال والقدرة والضمانات والشروط) تتمحور حول الإنسان وغير فعالة بطبيعتها، ومعرضة لخطر عدم الدقة. ونتيجة لذلك، يواصل العديد من المقترضين المعرضين لخطر التخلف عن السداد الحصول على القروض، مما يزيد من التعرض للمخاطر المالية. ولمعالجة هذه التحديات، تطبق هذه الدراسة تقنيات التعلم الآلي لتطوير نموذج تنبؤ بالتخلف عن سداد القروض قائم على البيانات يهدف إلى تحسين تقييم المخاطر في بنوك التمويل الأصغر. المنهجية: هذا البحث هو دراسة مختلطة تعتمد على الأساليب الكمية والنوعية. من الناحية الكمية، يتم تدريب أربع خوارزميات للتعلم الآلي، مثل شجرة القرار (DT) والانحدار اللوجستي (LR) والغابة العشوائية (RF) وتعزيز التدرج (GB) والتحقق من صحتها على مجموعة بيانات كبيرة من العالم الحقيقي تم جمعها من بنكين للتمويل الأصغر في باكستان. تم تقسيم مجموعة البيانات إلى مجموعات تدريب واختبار، مع استخدام 70% من البيانات للتدريب و 30% للاختبار. تم استخدام استراتيجيات اختيار الميزات والمعالجة المسبقة مثل التعامل مع القيم المفقودة واكتشاف القيم المتطرفة وتطبيع البيانات لتحسين الأداء التنبؤي. في حين تم استخدام دقة مقاييس الأداء والدقة والاسترجاع ودرجة F1 لتقييم الأداء التنبؤي للنماذج. تضمن الجانب النوعي للدراسة استطلاعات لموظفي القروض. كان هذا للتحقق من أن ما تم إنتاجه بواسطة التعلم الآلي قد تم التحقق منه مقابل سلوك الإقراض الحقيقي. النتائج: حدد البحث أن أكثر العوامل المؤثرة في التخلف عن سداد القروض هي نسبة الدين إلى الدخل ومبلغ القسط والدخل السنوي وإجمالي عدد الحسابات ومبلغ القرض ومعدل الفائدة والحساب المفتوح والسجلات العامة ومدة القرض والغرض من توحيد الديون. تشير النتائج إلى أن تعزيز التدرج كان أفضل من النماذج الأخرى بأفضل دقة (77.75%) ودقة (55.33%)، على الرغم من أن التذكر (11.33%) ودرجة F1 (18.81%) حدثا منه.

بالمقارنة، أظهرت الغابة العشوائية دقة أقل قليلاً (77.58٪) ودقة (53.85٪) وتذكر (10.19٪) ودرجة F1 (17.14٪) بينما أظهر الانحدار اللوجستي دقة معقولة (77.39٪) ودقة (51.74٪) ولكنه كان لديه أدنى تذكر (8.74٪) ودرجة F1 (14.95٪). كان نموذج شجرة القرار، على الرغم من قابليته للتفسير، أقل أداءً إجماليًا، محققًا دقة (68.9٪) ودقة (31.15٪) وتذكر (30.32٪) ودرجة F1 (30.73٪). بالإضافة إلى ذلك، للتحقق من نتائج التعلم الآلي، تم تضمين تصورات الممارسين، وتم التأكيد على أن نسبة الدين إلى الدخل، ومبلغ القسط، ومستوى الدخل كانت من أكثر العوامل المهيمنة في تقييم الائتمان البشري. يشير البحث أيضًا إلى الاختلافات بين نماذج التعلم الآلي وتصورات الممارسين، بما في ذلك استخدام التاريخ الائتماني، ومخاطر المناخ، ومحركات سمعة المقترض التي عادةً ما يتم التقليل من قيمتها بواسطة النماذج الآلية ولكنها مهمة في حالات الإقراض في العالم الحقيقي. الآثار المترتبة: لنتائج هذه الدراسة آثار قوية على بنوك التمويل الأصغر، وصانعي السياسات، والهيئات التنظيمية، ويُصحح بتبني نماذج مخاطر الائتمان القائمة على التعلم الآلي لتحسين فحص القروض، وتقليل القروض المتعثرة، والحفاظ على الاستدامة المالية. تضيف هذه الدراسة أيضًا إلى الأدبيات المتعلقة بالتكنولوجيا المالية (FinTech) من خلال توضيح كيف يمكن للتعلم الآلي أن يعزز تحليل مخاطر الائتمان في الاقتصادات الناشئة. التوصيات: يجب أن تركز الأبحاث المستقبلية على النماذج الهجينة التي تجمع بين تقنيات التعلم الآلي الكمية وحكم الخبراء النوعي من خلال المقابلات المنظمة لإعطاء تقييم أفضل لمخاطر الائتمان. يجب تطوير أطر تنظيمية لضمان العدالة والشفافية والتطبيق الأخلاقي للذكاء الاصطناعي في إدارة مخاطر الائتمان. علاوة على ذلك، فإن دمج العوامل السلوكية والاقتصادية الكلية في مجموعة البيانات من شأنه أن يُحسِّن التنبؤ بالتخلف عن سداد القروض.

## APPROVAL PAGE

The thesis of Anam Soomro has been approved by the following:

---

Habeebullah Zakariyah  
Main Supervisor

---

Asadullah Shah  
Co-Supervisor

---

Co-Supervisor

---

Internal Examiner

---

External Examiner

---

Chairman

## DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Anam Soomro

Signature .....

Date 25/03/2025



**INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA  
DECLARATION OF COPYRIGHT AND AFFIRMATION OF  
FAIR USE OF UNPUBLISHED RESEARCH**

**LOAN DEFAULT PREDICTION USING MACHINE  
LEARNING: AN EMPIRICAL STUDY OF MICROFINANCE  
BANKS IN PAKISTAN**

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

Copyright © 2025 Anam Soomro and International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below.

1. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.
3. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Anam Soomro

25/03/2025

.....  
Signature

.....  
Date

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my main supervisor Associate Professor Dr Habeebullah Zakariyah and Co- supervisor Professor Dr Asadullah Shah for their continuous support, invaluable guidance, and patience throughout my PhD journey. Their insights, encouragement, and constructive feedback have played a crucial role in shaping this research and my growth as a researcher.

I extend my heartfelt appreciation to my parents, whose unconditional love, sacrifices, and unwavering belief in me have been my greatest source of strength. Their support has been the foundation of my perseverance and dedication. Without their encouragement, this journey would not have been possible.

I am also deeply grateful to my friends, who have been my pillars of support, offering motivation, companionship, and moments of respite during challenging times. Their encouragement and belief in me kept me going, even in the most difficult moments.

Additionally, I sincerely appreciate the financial and institutional support provided by CPS which made this research possible.

Finally, I dedicate this thesis to all those who have stood by my side, inspiring and motivating me every step of the way.

## TABLE OF CONTENTS

Abstract .....	ii
Abstract in Arabic .....	iv
Approval Page.....	vi
Declaration .....	vii
Copyright .....	viii
Acknowledgements.....	ix
List of Tables .....	xiv
List of Figures .....	xv
List of Abbreviation.....	xvi
<b>CHAPTER ONE: INTRODUCTION .....</b>	<b>1</b>
1.1 Background of the Study .....	1
1.2 Financial System of Pakistan.....	2
1.3 Microfinance Industry in Pakistan.....	4
1.3.1 Microfinance Bank.....	5
1.3.2 Non-Bank Microfinance Company .....	5
1.4 Problem Statement.....	7
1.5 Research Questions.....	8
1.6 Research Objectives.....	8
1.7 Significance of the Study.....	10
1.8 Summary.....	10
<b>CHAPTER TWO: LITERATURE REVIEW.....</b>	<b>13</b>
2.1 Overview of Microfinance in Pakistan .....	13
2.2 Issues And Challenges of Microfinance in Pakistan .....	14
2.2.1 Limited Outreach .....	15
2.2.2 High Operating Costs.....	15
2.2.3 Limited Access to Funding .....	15
2.2.4 Lack of Innovation in Products .....	15
2.2.5 Regulatory Challenges .....	16
2.2.6 Limited Financial Literacy.....	17
2.2.7 Gender Discrimination .....	17
2.2.8 Political Instability .....	17
2.3 Major Issues of the Microfinance Sector .....	17
2.3.1 Low Outreach.....	18
2.3.2 Excess Liquidity.....	19
2.3.3 Loan Default.....	19
2.3.4 Size of the Loan .....	20
2.3.5 Limited Industry Expertise among MFI Owners .....	20
2.4 Challenges in Islamic Microfinance Sector.....	20
2.4.1 Overreliance on a Single Mode of Islamic Financing.....	21
2.4.2 High Transaction Cost .....	22
2.4.3 Market Penetration.....	23
2.4.4 High Interest Rates.....	24
2.5 Loan Risk in Microfinance Institutions .....	25
2.6 Credit Scoring.....	26

2.7 Methods For Evaluating Default Risk .....	27
2.8 Worldwide Perspective On Machine Learning Methods For Credit Scoring: A Review of Research.....	28
2.9 Research Gap.....	46
2.10 Summary.....	53
<b>CHAPTER THREE: RESEARCH METHODOLOGY .....</b>	<b>56</b>
3.1 Introduction.....	56
3.2 Research Design .....	56
3.3 Quantitative Research Method .....	57
3.3.1 Data Collection.....	57
3.3.2 Sample Size.....	58
3.3.3 Unit of Analysis .....	58
3.3.4 Data Analysis .....	59
3.3.4.1 Data Preprocessing .....	59
3.3.4.2 Handling Class Imbalance .....	60
3.3.4.3 Model Training and Evaluation.....	60
3.3.4.4 Interpretation of Results .....	60
3.4 Qualitative Research Method .....	61
3.4.1 Data Collection.....	62
3.4.2 Sampling Techniques .....	62
3.4.3 Respondents and Research Instruments.....	63
3.4.4 Data Analysis .....	63
3.5 Machine Learning.....	63
3.5.1 Supervised Machine Learning.....	64
3.5.1.1 Logistic Regression .....	65
3.5.1.2 Decision Trees .....	66
3.5.1.3 Random Forest.....	67
3.5.1.4 Extreme Gradient Boosting .....	68
3.6 Evaluation Metrics For Classification Model.....	69
3.6.1 Performance Metrics .....	69
3.6.2 Confusion Matrix .....	70
3.6.3 Precision.....	71
3.6.4 Recall.....	72
3.6.5 F1 Score .....	72
3.7 Summary.....	72
<b>CHAPTER FOUR: RESULTS AND FINDINGS.....</b>	<b>74</b>
4.1 Introduction.....	74
4.2 Objective-01:To Identify The Key Borrower And Loan Related Factors That Influence Loan Default in Microfinance Banks in Pakistan .....	75
4.2.1 Data Collection.....	78
4.2.2 Description of Considered Factors.....	78
4.2.3 Loan Specific Variable.....	79
4.2.4 Credit History and Financial Profile .....	80
4.2.5 Rationale for Factor Selection.....	82
4.2.5.1 Factors Typically Used in Microfinance .....	83
4.2.5.2 Implications for Default Predictions .....	83
4.2.5.3 Data Availability .....	83
4.2.6 Analysis Approach.....	84

4.2.7 Summary of Findings.....	85
4.3 Objective-02: To Evaluate And Compare the Performance of Various Machine Learning Algorithm In Predicting Loan Default .....	86
4.3.1 Accuracy .....	86
4.3.2 Precision.....	86
4.3.3 Recall.....	87
4.3.4 F1 Score .....	87
4.3.5 Confusion Matrix .....	87
4.3.6 Justification for Machine Learning Model Performance .....	88
4.3.7 Practical Implications for Microfinance Institution Pakistan.....	103
4.3.7.1 Strengthen Borrower Screening and Monitoring .....	103
4.3.7.2 Diversify Loan Portfolio.....	103
4.3.7.3 Enhance Borrower Financial Literacy .....	103
4.3.7.4 Strengthen Borrower Relationships.....	104
4.3.7.5 Utilize Innovative Lending Models.....	104
4.4 Objective-03: To Determine the Most Influential Features Contributing to Loan Default Identified by Machine Learning Models .....	105
4.4.1 DTI (Debt-To-Income Ratio).....	106
4.4.2 Installment Amount.....	106
4.4.3 Annual Income .....	106
4.4.4 Total Number of Accounts.....	107
4.4.5 Loan Amount .....	107
4.4.6 Interest-Rate .....	107
4.4.7 Open Account.....	108
4.4.8 Public Records .....	108
4.4.9 Loan Terms .....	108
4.4.10 Purpose of Debt Consolidation .....	108
4.5 Objective 04: To Validate the Practical Relevance and Interpretability of Top MI Identified Predictors Through Expert Insights From Bank Loan Officers .....	109
4.5.1 Key Insights from Practitioners .....	110
4.5.2 Alignment with Machine Learning Finding.....	110
4.5.3 New Insights from Practitioners.....	111
4.5.4 Practical Implications.....	112
4.5.5 Conclusion .....	112
4.6 Summary.....	115

<b>CHAPTER FIVE: DISCUSSION, CONCLUSION AND FUTURE DIRECTIONS .....</b>	<b>117</b>
5.1 Introduction.....	117
5.1.1 Algorithm Comparison .....	118
5.2 Practical Implication for Microfinance banks .....	121
5.2.1 Strategies to Reduce Loan Defaults .....	121
5.2.2 Benefits of Adopting ML .....	121
5.3 Key borrower-level risk factors .....	122
5.4 Machine learning model performance across studies.....	125
5.5 Top predictive factors and their weight .....	126
5.6 Comparison of objective 04: To Validate the top Factors Influencing Loan Default Through Bank Loan Officers.....	129
5.7 Critical Review of the Work.....	130

5.8 Theoretical Contribution.....	130
5.9 Practical Contribution.....	131
5.10 Methodological Contribution and Applied Contribution .....	131
5.11 Conclusion.....	132
5.11.1 Model Performance Summary .....	132
5.11.2 Research Contributions .....	132
5.11.3 Future Directions.....	133
5.11.3.1 Use of Advanced Ensemble Techniques .....	133
5.11.3.2 Incorporating Alternative Data Sources .....	133
5.11.3.3 Addressing Data Quality and Privacy Concerns .....	133
5.11.3.4 Longitudinal Studies.....	134
5.11.3.5 Real-time Predictive Models .....	134
5.11.3.6 Impact of Policy Changes.....	134
5.11.3.7 Training for Loan Officers .....	134
5.11.3.8 Sector-Specific Models.....	134
5.11.3.9 Macroeconomic Factors .....	135
5.11.4 Policy and Practical Implications.....	136
5.11.5 Ethical Considerations and Alignment with Islamic Finance....	137
5.12 Summary.....	138
<b>REFERENCES.....</b>	<b>139</b>
<b>APPENDIX A .....</b>	<b>162</b>

## LIST OF TABLES

Table 1.1	Non-Bank Microfinance Company	6
Table 1.2	Research Questions and Objectives	11
Table 2.1	Comparative Analysis and Findings	56
Table 3.1	Loan default prediction model attributes	70
Table 4.1	Factors Influencing Loan Default Prediction	98
Table 4.2	Loan-Specific Variables	103
Table 4.3	Credit History and Financial Profile	105
Table 4.4	Borrowers Related Factors	106
Table 4.5	Performance Metric for DT (Decision Tree) Classifier	117
Table 4.6	Classification Report of Decision Tree	120
Table 4.7	Performance Metric for LR (Logistic Regression) Classifier	121
Table 4.8	Classification Report of LR	124
Table 4.9	Performance Metric for RF (Random Forest) Classifier	125
Table 4.10	Classification report of RF	128
Table 4.11	Performance Metric for GB (Gradient Boosting) Classifier	129
Table 4.12	Classification Report of Gradient Boosting	132
Table 4.13	Comparison of Performance Metrics	133
Table 4.14	Comparison of the objectives	149
Table 5.1	Comparative Analysis of ML Models	155
Table 5.2	Loan Default Prediction within MFBs in Pakistan	159
Table 5.3	Comparative Analysis of ML Performance	162
Table 5.4	Comparative analysis synthesizes findings from various studies	165

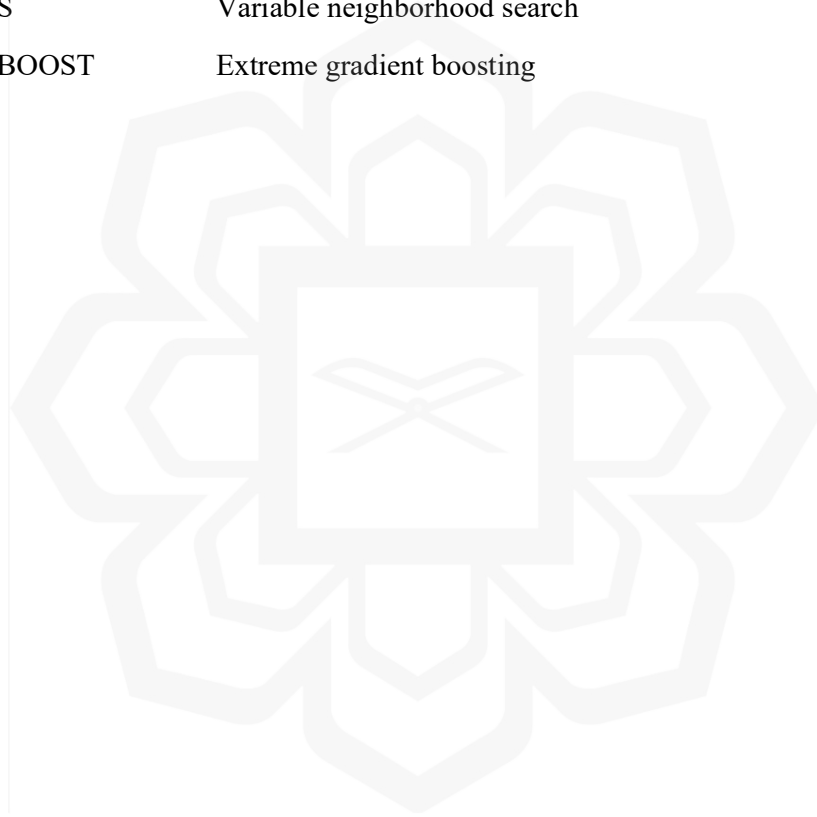
## LIST OF FIGURES

Figure 1.1	Microfinance outreach in Pakistan	4
Figure 3.1	Proposed model	80
Figure 3.2	Classification Tree (Bell, 2014)	88
Figure 3.3	Random Forest (Chakure, 2019)	90
Figure 3.4	Confusion matrix (Granström and Abrahamsson, 2019)	93
Figure 4.1	Pictorial Representation of DT	117
Figure 4.2	Confusion Matrix for Decision Tree	119
Figure 4.3	Pictorial Representation of LR	121
Figure 4.4	Confusion Matrix for LR	123
Figure 4.5	Graphical Representation of RF	125
Figure 4.6	Confusion Matrix for Random Regression	127
Figure 4.7	Graphical Representation of GB	129
Figure 4.8	Confusion Matrix for Gradient Boosting	131
Figure 4.9	Graphical Representation of Comparative Performance Metrics	134
Figure 4.10	Top Factors Contribution to Charge Off Used to Predict Loan Default	138

## LIST OF ABBREVIATION

AI	Artificial intelligence
ANN	Artificial neural network
AUC	Area Under the Curve
DT	Decision tree
DTI	Debt to income ratio
DFI	Development finance institution
FINTECH	Financial technology
FICO	Fair Isaac cooperation
GB	Gradient boosting
GDP	Gross domestic product
GLP	Gross loan portfolio
IMFI	Islamic microfinance institution
LDA	Linear discriminant analysis
LR	Logistic regression
LS	Local search
MAR	Missing at random
MCAR	Missing completely at random
MARS	Multivariate adaptive regression splines
MNAR	Missing not at random
MFI	Microfinance institution
ML	Machine learning
MFB	Microfinance bank
NAP	New approach
NBFC	Non-banking microfinance company
NPL	Nonperforming loan
PLS	Partial least square
P2P	Per-to-peer lending

RF	Random forest
ROC	Receiver operating characteristics
RBF	Radial basis function
SBP	State bank of Pakistan
SCF	Survey of consumer finances
SECP	Securities and exchange commission of Pakistan
SMOTE	Synthetic minority over sampling technique
SVM	Support vector machine
VNS	Variable neighborhood search
XGBOOST	Extreme gradient boosting



# CHAPTER ONE

## INTRODUCTION

### 1.1 BACKGROUND OF THE STUDY

Bank loans have emerged as a very important source of external finance to businesses and individuals due to financial constraints. Although loans are a significant source of revenue to the banks, lending is a risky business in itself. Credit risk is one of the most important ones, meaning that it is a risk that a borrower will default on the contractual obligations of the debt (Hull,2018). This risk occurs when the borrowers fail to pay the loan as per the agreed terms or time. When repayment is successful, the lender will receive returns in terms of interest income. Nonetheless, when there is default the creditor suffers losses on the principal and anticipated interest. The Credit Risk Theory states that lenders have to determine the likelihood of default in order to ensure that their risk exposure is well managed (Crouhy, Galai, & Mark, 2000).

Calcagnini et al. (2018) stressed that credit risk is a major cause of financial instability especially in the event of economic crises, and its impact is felt throughout the financial system. Banks recognize that lending is not solely profit-generating but also carries substantial risks. To mitigate the risks associated with loan defaults, banks employ strategies such as evaluating the borrower's repayment capacity and requiring collateral before extending loans. Employing skilled experts within banks who thoroughly assess the suitability of loan applicants, considering various factors, and generating a numerical score to determine their eligibility for loan approval or rejection perform this task. With the progress of technology, machine learning algorithms and neural networks can automatically predict an individual's credit score using historical data, enabling the identification of potential loan defaulters well in advance of loan disbursement (Marques et al., 2012 & Adieus et al., 2016).

The prediction of loan defaults is a crucial and significant challenge faced by lending institutions, including banks and other financial organizations, as it dramatically affects their profitability and growth. Over the years, machine-learning techniques have been employed to assess and forecast default risk by analyzing individual borrowers'

historical data. Research findings have shown that ML models can help lower non-performing loan (NPL) ratios and enhance early identification of at-risk borrowers (Prasad & Venkatesh, 2023; Wang et al., 2022). This research proves valuable for financial institutions in evaluating customers' creditworthiness and making informed decisions by considering multiple variables to anticipate borrower risk. Obtaining loans for diverse purposes, such as home, student, vehicle, and business loans, has become integral to our daily lives through various financial entities like credit unions and banks. However, most individuals struggle to accurately assess their ability to repay the total amount of credit they acquire.

Financial institutions process many loan requests from a diverse range of customers. Nevertheless, banks or financial institutions do not approve all loan customers. Many banks analyze loan applications using credit scoring models and risk assessment techniques to decide whether to grant the application. Machine learning techniques can analyze customers' previous credit data and understand their repayment data, which will help anticipate the chance of default and allow Pakistan microfinance banks to enhance their decision-making process for future loan applicants.

Machine learning assigns numerical credit scores to individuals, enabling the assessment of their default risk and creditworthiness. It also predicts and evaluates an individual's credit risk by analyzing historical data. This project examines existing literature on diverse approaches and studies on forecasting the risk assessment of loan default using supervised machine learning algorithms. Its objective is to determine the likelihood of firms or individuals being unable to meet their debt obligations or repayments.

## **1.2 FINANCIAL SYSTEM OF PAKISTAN**

Financial sustainability is the condition in which the financial system, financial intermediaries, financial markets, and financial market infrastructure assist in the smooth cash movement among shareholders and savers in an organized and trustworthy manner. Given the importance of economic stability in ensuring a robust economy, preserving the financial system's integrity has become a paramount concern for central banks and regulatory bodies worldwide. To safeguard the financial system's solidity and

flexibility, central banks must detect flaws in the system and adopt mitigation actions to prevent the system's spillage to other sectors of the economy and maintain public trust in the system. SBP plays a crucial role in upholding financial sector integrity. According to the State Bank of Pakistan Act of 1956, the bank was entrusted with regulating the monetary and credit system of Pakistan and fostering its growth in the best interest of the nation.

The financial system in Pakistan encompasses a variety of institutions, including banks, both conventional and Islamic, Development Finance Institutions (DFIs), Microfinance Banks (MFBs), Non-banking Finance Companies (NBFCs), insurance companies, Modarabas, and other financial intermediaries. Microfinance banks operate under the jurisdiction of the Microfinance Institutions Ordinance 2001. Banks dominate Pakistan's financial industry, constituting a significant portion of financial assets with GDP. The State Bank of Pakistan (SBP) regulates banks, DFIs, Exchange Companies, and MFBs, while the Securities and Exchange Commission of Pakistan (SECP) regulates NBFCs, insurance companies, and Modaraba Companies. (Source: sbp.org.pk).

The Pakistan Microfinance Network (PMN), Established in 1997, emerged as a national entity representing retail microfinance participants. Initially known as the Microfinance Group-Pakistan (MFG-P), it started as an informal association of industry professionals. The organization was incorporated with the Securities and Exchange Commission in April 2001 under the Companies Act as the nature and intensity of activity and membership increased over time. Since then, it has been recognized as the Pakistan Microfinance Network. Currently, the network comprises 44 Microfinance Providers, consisting of Microfinance Banks (licensed by the State Bank of Pakistan, SBP) and Non-Bank Microfinance Companies (regulated by the Securities and Exchange Commission of Pakistan, SECP).

## 1.3 MICROFINANCE INDUSTRY IN PAKISTAN

Powered by ANSAS

	Quarter		Change	
	Q4 - 2023	Q3 - 2023	Units	%
Number of Branches/Units	4,153	4,073	80	2.0
Number of Districts Covered	135	136	-1	-0.7
Penetration Rate (%)	23.5	22.6	1	3.6
Active Borrowers	9,389,408	9,061,178	328,230	3.6
Gross Loan Portfolio (PKR Millions)	546,132	529,830	16,302	3.1
Number of Loans Disbursed	10,629,259	8,427,324	2,201,935	26.1
Disbursements (PKR Millions)	154,045	149,353	4,693	3.1
Average Loan Size (PKR)	14,493	17,722	-3,230	-18.2
Number of Savers	108,687,815	104,963,882	3,723,933	3.6
Value of Savings (PKR Millions)	596,619	535,266	61,353	11.5
Average Saving Balance (PKR)	5,489	5,100	390	7.6
Number of Policy Holders	8,572,031	9,258,205	-686,174	-7.4
Sum Insured (PKR Millions)	325,523	347,022	-21,499	-6.2

Source: PMN, issue 70: Q4 (OCT-DEC) 2023

Figure 1.1 Microfinance outreach in Pakistan

In the final quarter of 2023, the microfinance sector substantially increased active borrowers and the total loan portfolio. In Q4, the industry reached a notable milestone by acquiring 23.5% of the potential clientele, amounting to around 9.4 million active borrowers.

This signifies a 3.6% increase from the previous quarter, resulting in nearly 330,000 additional borrowers within a single quarter. Among these 9.4 million active debtors, Microfinance Banks (MFBs) contributed 67%, an increase from 65% in the preceding quarter. Telenor Microfinance Bank exhibited the most substantial impact among the MFBs, achieving a 40% rise in clientele and a 13% expansion in its gross loan portfolio (GLP).

Conversely, whereas MFBs exhibited a positive trend, Non-Banking Microfinance Companies (NBMFCs) encountered a marginal decrease in market share during the prior quarter. The share of NBMFCs in the total active clientele decreased from 34.5% in Q3 to 33% in Q4 of 2023. Significantly, 35% of the overall active clientele consists of nano loan consumers.

The gross loan portfolio experienced a 3.1% growth, resulting in 546 billion. The market share allocation between MFBs and NBMFCs remained stable from the prior quarter, at 78% and 22%, respectively. Nano loans accounted for 2% of the total GLP. The most notable alteration was evident in the volume of loans disbursed, which surged by 26%, predominantly propelled by LOLC Pakistan with a 187% quarter-on-quarter escalation and SRSO with a 98% quarter-on-quarter rise. Microfinance Banks offer Nano loans, including MMFB, TMFB, and HBL MFB.

Source: (Pakistan Microfinance network, Micro watch-issue 70, Quarter 4, (OCT-DEC).

### 1.3.1 Microfinance Bank

The SBP has prudentially authorized a bank to serve the microfinance industry. The first MFB was founded by presidential order in 2000. The Microfinance Institutions Ordinance 2001 has since licensed 11 MFBs. MFBs are legally permitted to receive and process public deposits. MFBs hold 31% of the share.

### 1.3.2 Non-Bank Microfinance Company

Following the establishment of the Securities and Exchange Commission of Pakistan's (SECP) regulatory framework for non-bank microfinance firms in 2015, it is now mandatory for institutions offering microfinance services to obtain a license from SECP as Non-Bank Microfinance Companies (NBMFCs). Currently, 25 PMN member organizations hold the NBMFC authorization. NBMFC grasps 69% of the total. The details of non-bank microfinance companies are in Table 1.1.

Table 1.1 Non-Bank Microfinance Company

<b>Microfinance Banks</b>	<b>Microfinance institutions</b>	<b>Rural supports program</b>
Advance microfinance bank	Aghae Pakistan	National Rural Support Programme

<b>Microfinance Banks</b>	<b>Microfinance institutions</b>	<b>Rural supports program</b>
Apna micro finance bank	Akhuwat Islamic Microfinance Company	Punjab Rural support Programme
Finca microfinance bank	AMARDO Foundation	Sarhad Rural Support Programme
Moblink microfinance bank	ASA Pakistan	Sindh rural support organization
Khushali microfinance bank	Badhban community development Microfinance Company	Ghazi barotha taraqiyyati idara
Telenor microfinance bbank	CSC empowerment and inclusion program	
Sindh microfinance bank	Damen support program	
NRSP microfinance bank	FFO support program	
HBL microfinance bank	Islamic relief Pakistan	
PAK OMAN microfinance bank	Kashf foundation	
U microfinance bank	Islamic relief Pakistan	

The existence of 11 SBP-licensed microfinance banks and the wide variety of microfinance institutions and rural support programs depict the institutional maturity and reach of the microfinance sector in Pakistan. The size and variety of this scale explains why sophisticated credit risk management models, like machine learning (ML), are necessary to guarantee efficiency, consistency, and scalability in the evaluation of borrowers..

## 1.4 PROBLEM STATEMENT

In Pakistan, people are relying more on financial institutions to obtain credit to meet their basic needs and entrepreneurial activities (Madaan et al., 2021). In the current unstable economy and the increasing competitiveness, taking out a loan has become a necessity to many layers of the population (Aslam., et al., 2019).

Meanwhile, financial institutions such as microfinance companies to commercial banks depend largely on loan disbursement as a major source of revenue. Nevertheless, loans do not only bring in profits but also subject institutions to large credit risks. This is because borrowers cannot settle their debts in time and this causes losses both in the principal and the interest (Abbas. et al. 2014).

To reduce this risk, the conventional lending process usually uses the 5Cs of credit-Character, Capital, Capacity, Collateral, and Conditions (Li et al., 2019). Although useful, these criteria are highly subjective and depend on personal experience, thus they are subject to inconsistencies and are not effective in determining high-risk borrowers, particularly in the microfinance sector.

Although machine learning (ML) approaches promise to automate and improve credit risk assessment, the existing literature does not agree on the best models to use in the microfinance banks of Pakistan (Lin et al., 2022). Most of the current models like Logistic Regression, Decision Trees, and Support Vector Machines are not flexible to local lending patterns or their predictive performance is erratic.

Moreover, the urgency of the sector is supported by the recent statistics: as of Q4 2023, the level of borrower penetration in the microfinance banks of Pakistan reached 23.5%, and the Non-Performing Loans have been increasing noticeably (Pakistan Microfinance Network, 2023). This brings to the fore a dire necessity of more precise, data-based credit screening tools that will guarantee financial sustainability.

Thus, the study proposes a machine learning-based loan default prediction model of microfinance banks in Pakistan based on the most important characteristics of borrowers based on historical loan data. The objective is to enhance default prediction, credit risk assessment and informed lending decisions.

## 1.5 RESEARCH QUESTIONS

1. What key borrower- or loan-related factors significantly influence loan default prediction in microfinance banks in Pakistan?
2. Which machine learning algorithm demonstrates the highest predictive accuracy and reliability in identifying loan defaults in Pakistani microfinance banks?
3. Which borrower and financial attributes are identified as most influential by machine learning models in predicting loan default?
4. How do bank loan officers evaluate the practical relevance and interpretability of the top predictors identified by machine learning models in assessing loan default risk?

## 1.6 RESEARCH OBJECTIVES

1. To identify the key borrower- and loan-related factors that influence loan default in microfinance banks in Pakistan.
2. To evaluate and compare the performance of various machine learning algorithms in predicting loan defaults.
3. To determine the most influential features contributing to loan default, as identified by machine learning models.
4. To validate the practical relevance and interpretability of top ML-identified predictors through expert insights from bank loan officers.

Table 1.2 Research Questions and Objectives

Research questions	Research objectives	Research method	Data collection	Data Analysis
1. What key borrower- or loan-related factors significantly influence loan default prediction in microfinance	To identify the key borrower- and loan-related factors that influence loan default in microfinance banks in Pakistan.	Quantitative research method.	Data was collected from microfinance banks in Pakistan on their loan portfolios and the characteristics of their borrowers, such as age, gender, income level,	Statistical tools such as descriptive statistics, correlation analysis, and regression analysis were used to analyze the data and to identify the key factors contributing to loan defaults.

Research questions	Research objectives	Research method	Data collection	Data Analysis
banks in Pakistan?			education level, and occupation.	
2. Which machine learning algorithm demonstrates the highest predictive accuracy and reliability in identifying loan defaults in Pakistani microfinance banks?	To evaluate and compare the performance of various machine learning algorithms in predicting loan defaults.	Quantitative and descriptive research method.	Data on loan portfolios, borrower characteristics, and repayment histories was collected from microfinance banks. Data preprocessing will include handling missing values, encoding categorical variables, and scaling numerical features.	Random Forest, Decision Trees, XGBoost, and Logistic Regression) The choice of these models was based on considerations for interpretability, robustness in modeling imbalanced datasets and proven success in previously solving credit risk-related problems. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics.
3. Which borrower and financial attributes are identified as most influential by machine learning models in predicting loan default?	To determine the most influential features contributing to loan default, as identified by machine learning models.	Quantitative research method using machine-learning techniques.	Data on loan portfolios, borrower characteristics, and repayment histories was collected from microfinance banks.	Feature selection techniques (e.g., SHAP thus gives us interpretability; Recursive Feature Elimination ensures we have optimal predictors. Variable importance is estimated using Information Gain) were used to rank factors based on their importance in predicting loan defaults.
4. To validate the practical relevance and interpretability of top ML-identified predictors through expert insights from bank loan officers.	How do bank loan officers evaluate the practical relevance and interpretability of the top predictors identified by machine learning models	Mixed-method approach (quantitative and descriptive analysis.	Data was collected from microfinance practitioners through structured surveys, interviews, and direct data sharing.	Practitioner feedback was compared with machine learning findings using descriptive statistics and validation matrices.

Research questions	Research objectives	Research method	Data collection	Data Analysis
	in assessing loan default risk?			

## 1.7 SIGNIFICANCE OF THE STUDY

The relevance of this study is that it can potentially transform the credit risk analysis process of microfinance banks (MFBs) in Pakistan through the application of machine learning (ML) algorithms. The traditional credit scoring frameworks, such as the 5Cs model (Character, Capital, Capacity, Collateral, and Conditions), are highly driven by individual judgments, hence subject to inefficiencies and personal biases. Machine learning algorithms are applied in this research to facilitate a data-driven solution for identifying high-risk borrowers, consequently lowering loan default rates and financial instability in the microfinance business sector. The shift to predictive analytics and automation will allow MFBs to make better-informed lending decisions and, ultimately, higher profitability and sustainability.

This study significantly contributes to the banking industry because microfinance banks' findings are expected to reflect improved default risk management and risk mitigation processes. Furthermore, this study will add to the current literature and fill up the gap on loan default of microfinance customers and the effect of default risk on the performance of Microfinance banks. Forecasting loan defaults is a difficult task. Microfinancing banks in Pakistan currently employ a different way to detect loan defaulters. Manual credit scoring methods and risk measurement models have been used to evaluate loan defaulters.

## 1.8 SUMMARY

The text discusses the importance of bank loans as a source of external funding for enterprises and individuals due to financial concerns. However, this increased lending

is linked with a range of hazards, like default risk, which pertains to the borrower's default in repaying the loan within the stipulated timeframe and according to the established terms.

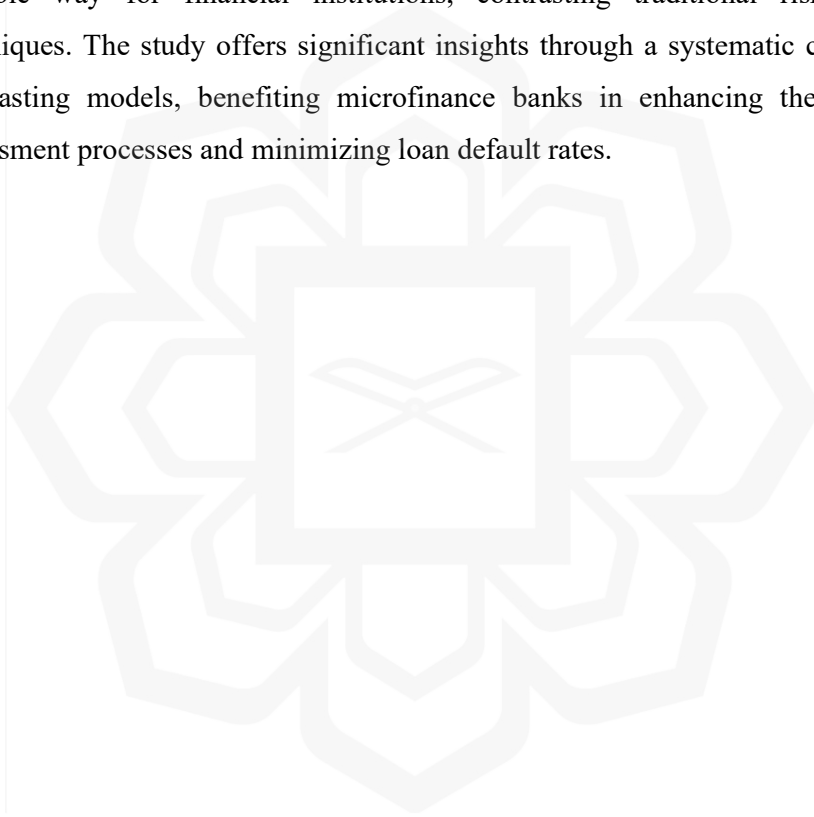
Banks employ highly skilled professionals to evaluate an applicant's eligibility for loan acceptance or rejection depending on several parameters, reducing the chance of default. The machine learning methods can automatically forecast an individual's credit score and identify borrowers who default before loan approval. Loan default prediction is among lending institutions' most vital and significant issues. Machine learning assigns numerical credit scores to individuals to assess their default risk and creditworthiness. The financial system of Pakistan, including banks, conventional and Islamic banks, microfinance banks, non-banking finance companies, insurance firms, modarabas, and other financial intermediaries, is also discussed, with microfinance banks coming under the microfinance institutions ordinance 2001.

The chapter discusses the state of microfinance in Pakistan and the risks associated with banking. Despite the political and economic turmoil, the microfinance sector in Pakistan has shown stability, with metrics indicating growth in outreach and active borrowers. Microfinance Banks (MFB) hold a 31% share, while Non-Bank Microfinance Companies (NBMFC) have a 69% share. The chapter continues by discussing the credit risk. Credit risk is considered one of the most significant risks that banks face and is defined as the likelihood of a half or complete loss of an unsettled loan due to late payment or default. The text also presents conflicting research findings regarding the effect of credit risk on bank profitability.

The chapter discusses the importance of loans in the financial industry and the risks associated with granting loans, known as credit risk. The traditional method of evaluating a borrower's creditworthiness using the '5C concept' has significant drawbacks, leading to the adoption of credit scoring based on statistical algorithms and, more recently, machine learning and deep learning algorithms to forecast a borrower's creditworthiness. The study aims to predict loan defaults of microfinance institutions in Pakistan using machine learning methodologies. The research aims are to examine historical loan data, ascertain the principal factors contributing to loan defaults, and evaluate the effectiveness of various machine-learning models in forecasting loan defaults.

The research questions investigate the factors influencing loan defaults, the efficacy of several machine-learning algorithms, and the most effective algorithm for predicting defaults in microfinance institutions. The study seeks to enhance credit risk management methods in microfinance banks by data-driven predictions of loan defaults.

This research addresses a significant deficiency in the literature by employing machine-learning approaches to enhance the accuracy of loan failure predictions, an area that has been neglected in the context of microfinance institutions in Pakistan. This research empirically assesses various machine-learning models to determine the most suitable way for financial institutions, contrasting traditional risk assessment techniques. The study offers significant insights through a systematic comparison of forecasting models, benefiting microfinance banks in enhancing their credit risk assessment processes and minimizing loan default rates.



## **CHAPTER TWO**

### **LITERATURE REVIEW**

This chapter offers a general discussion of loan risk and assessment in microfinance, as well as a thorough examination of the literature review in academic journals on using machine learning techniques in loan risk prediction.

#### **2.1 OVERVIEW OF MICROFINANCE IN PAKISTAN**

The Pakistan Microfinance Network (PMN,2021) states that the active borrowers in the microfinance sector in Pakistan have grown by 10.5 percent to 9.9 million between June 2020 and June 2021. During the same time, gross loan portfolio increased by 15.8 percent to PKR 510 billion (USD 3.2 billion). Although this growth is an indication of the growing financial inclusion, it also creates concerns of the growing credit exposure. With increasing volumes of borrowers, institutions might feel the pressure to keep the quality of underwriting, and this can lead to poorer risk evaluations and, eventually, to the higher default rates. This is the reason why automated and precise default prediction systems are of high demand, like machine learning (ML) to make sure that credit quality is expanding as well.

Another impressive demographic fact is that 69 percent of all microfinance borrowers are women (PMN, 2021) . Although this is in line with most microfinance programs around the world, there is a concern that in Pakistan, gender-disaggregated credit risk data is not available to determine whether existing products and scoring mechanisms adequately capture women repayment behavior, financial literacy, or access to collateral. The given demographic dynamic should be taken into account in predictive models to prevent systemic bias and enhance equity in access to credit.

The State Bank of Pakistan (SBP, 2020) has been a proponent of financial inclusion through the introduction of simplified regulatory systems of microfinance banks, such as improved reporting standards. These reforms are meant to enhance institutional supervision and credit transparency, which are the key facilitators of data-

driven credit modeling. Nevertheless, regulatory change might not be sufficient to reduce the increasing credit risk without a parallel development of analytical tools.

The vulnerability of the microfinance sector was further confirmed by the effects of the COVID-19 pandemic. The ability to repay loans was also badly affected because of the prevalence of income shocks, especially on informal workers. To counter this, institutions came up with loan restructuring policies. Although the sector was resilient, resilience was not universal; most MFIs reported an increase in delinquency ratios. (PMN, 2021) data shows a post-COVID increase in portfolio-at-risk (PAR > 30 days) of 4.2 percent in 2019 to 6.8 percent in 2021. These shocks show how inadequate the conventional risk scoring can be and why it is necessary to have powerful predictive systems that can be flexible in times of crisis.

Although structural improvement, such as the introduction of the National Financial Inclusion Strategy and the encouragement of digital financial ecosystems, has been achieved, there are still considerable gaps in credit access. Females, rural communities and first-time borrowers continue to have limited access to formal credit because of the lack of financial history (SBP, 2021). This underscores the promise of ML-based scoring models that have the capability of utilizing non-traditional data (e.g., transaction history, behavioral data) to increase safe lending without necessarily being limited to traditional metrics.

## **2.2 ISSUES AND CHALLENGES OF MICROFINANCE IN PAKISTAN**

Microfinance entails offering financial services to individuals and small businesses with limited access to traditional banking facilities, primarily those with low incomes. In Pakistan, microfinance institutions (MFIs), including microfinance banks, have been instrumental in combating poverty and fostering economic development. Nevertheless, despite notable advancements in the sector in recent years, microfinance banks encounter diverse obstacles that impede their growth and long-term viability. Here are some key challenges confronted by microfinance banks in Pakistan:

### **2.2.1 Limited Outreach**

Despite the growing demand for microfinance services, microfinance banks have limited outreach in Pakistan, particularly in rural areas. According to a report by the State Bank of Pakistan (SBP), only 2.7% of the total population had access to microfinance services in 2020 (Ghalib, 2013; Rauf & Mahmood, 2009).

### **2.2.2 High Operating Costs**

According to Muhammad (2010), microfinance banks in Pakistan face high operating costs due to several factors, including the lack of infrastructure in rural areas, high staff turnover, and the need for technology investments. These costs can make it challenging for microfinance banks to provide affordable financial services to their clients.

### **2.2.3 Limited Access to Funding**

Microfinance banks in Pakistan frequently encounter difficulties getting funds due to insufficient collateral, restricted credit histories, and the higher risk of lending to low-income clients, making it difficult for microfinance banks to expand their operations and reach more clients (Razzaq et al. 2019).

### **2.2.4 Lack of Innovation in Products**

Microfinance aims at serving the financially excluded groups of people, especially the low-income groups with varied and usually complicated financial requirements. This involves constant product design and delivery innovation to make sure that it is in line with the changing needs of the underserved communities (Muhammad, 2010). Despite the fact that one of the main objectives of the microfinance sector is to empower women, a lot of microfinance institutions (MFIs) still provide gender-neutral loan products that fail to take into consideration the peculiarities of financial behavior and needs of female borrowers. As an example, in Pakistan, only a small number of MFIs offer micro-loans that are customized to suit the needs of women entrepreneurs and have characteristics

like a flexible repayment schedule or childcare-related financial planning (PMN, 2022). Such absence of specialized products can restrict the effective involvement of women in the microfinance programs, although they constitute the majority of borrowers.

Besides the limitation of products, MFIs have difficulties in spreading their wings because of poor segmentation of clients and low financial literacy among the target markets. The limited scope of services, which is usually restricted to microcredit, limits their capacity to attract and maintain a wider client base. Despite the fact that microfinance involves services like savings, insurance, and remittances, most institutions in Pakistan are still stuck on offering credit services only, thus not fully exploiting the opportunities of financial inclusion (Muhammad, 2010).

Moreover, the limited education and awareness of potential clients inhibits the successful market research. Marketing to these groups may need special marketing strategies and staff. Nevertheless, the cost of hiring such employees is an added burden to the operations of the institutions particularly the rural support programs (RSPs) and community-based MFIs that are already working on thin margins. Such cost pressures can hinder the expansion of the sector in terms of breadth and depth without scalable data-driven models or other delivery channels.

### **2.2.5 Regulatory Challenges**

Microfinance banks in Pakistan are subjected to stringent regulations by the SBP, which can create challenges for them regarding compliance and reporting. The rules can also limit their capacity to develop and provide new products and services.

According to the research of Muhammad (2010), over the past decade, the sector in question has remained largely unregulated, and there is still much work to be done in this regard. The current regulations often do not fit the bill and can hinder growth in the sector. To foster growth, regulations and restrictions must be reduced. Some rules put in place at the outset may need to be reconsidered, as they can pose significant obstacles as the sector begins to take off. Additionally, it is essential to streamline the complex regulatory norms that have driven up costs, such as those governing funds transfers, insurance, and savings deposits. These areas require urgent attention and reform.

### **2.2.6 Limited Financial Literacy**

Following Razzaq et al. (2019), limited financial literacy among the low-income population in Pakistan is another challenge for microfinance banks, as it limits the uptake of their products and services, leading to low customer retention rates and high delinquency rates.

### **2.2.7 Gender Discrimination**

As per the findings of Hassan and Jebran (2016), it appears that funds are primarily offered to male customers, leaving female customers with limited access to available funds and discouraging females from seeking financial support. It is essential to provide equal opportunities for male and female customers and distribute funds equally. Encouraging and supporting female customers can lead to a more balanced and fair distribution of financial resources.

### **2.2.8 Political Instability**

Pakistan has a history of political instability, which can challenge microfinance banks in attracting investors and maintaining operations (Muhammad 2010). The country has also faced security issues in some regions, which can create additional challenges for microfinance banks operating in those areas.

## **2.3 MAJOR ISSUES OF THE MICROFINANCE SECTOR**

The issues and challenges the microfinance sector faces around the globe in providing financial access to people experiencing poverty can be divided into several categories. This section discusses the microfinance sector's problems and difficulties in delivering financing services.

### 2.3.1 Low Outreach

Four dimensions of studying microfinance banks and institutions' outreach are breadth, depth, scope, and worth. The breathe of outreach: "It conceals the number of low-income individuals served by MFBs and MFIs. The outreach's breadth can be calculated using an estimated total number of active borrowers (Schreiner, 2002). The depth of outreach demonstrates that those not served by MFB or MFI are excluded from the financial industry. The proportion of active female borrowers to all borrowers calculates depth (Schreiner, 2002).

The scope of outreach: The variety of services offered by microfinance banks (MFBs) or microfinance institutions (MFIs) (Tchakoute Tchuigoua and Soumare, 2019).

The worth of outreach: The terms and conditions of microloans impact both borrowers and the sustainability of a microfinance institution (MFI). The value of microloans increases when the terms and conditions align with the borrowers' requirements regarding loan size, repayment duration, installment payments, guarantees, interest rates, fees, and other factors. The average loan size measures the extent of outreach (Tchakoute-Tchuigoua and Soumare, 2019; Hermes and Lensink, 2011).

As Guays (2011) and Nasir (2013) witnessed, many MF programs pay much attention to women. It is believed that women are more likely to save and borrow than men, and they are also more dependable and have a better repayment record. These characteristics support the idea that certain financial institutions are more likely to satisfy women's needs. However, they cannot ignore men as clients to increase their outreach. As per Quayes (2012), low-disclosure MFIs' financial sustainability declines when the breadth of their outreach rises and vice versa. The negative association between depth of outreach and financial sustainability shows the trade-off between these two factors in the case of MFIs with limited disclosure.

### **2.3.2 Excess Liquidity**

According to the González (2010), more liquidity offers additional security, but excessive amounts can result in riskier lending. An MFI with abundant liquidity may be inclined to promote rapid expansion, mainly according to the overall number of clients for each branch. A significant increase in microloans, backed by surplus liquidity, might reduce lending rules. This relaxation of regulations will raise the probability of default (Cermeno et al., 2011). As a result, increased liquidity is projected to be significantly linked to the likelihood of an MFI crisis.

### **2.3.3 Loan Default**

Loan default is a serious problem that has an impact on the sustainability and scalability of microfinance institutions (MFIs) in Pakistan. Although certain case studies conducted in the early days have reported high rates of delinquency in certain areas or during informal lending circumstances, the rates do not reflect the average rates of the sector. Based on the Pakistan Microfinance Network (PMN, 2023), the total Non-Performing Loan (NPL) ratio in the microfinance industry in Pakistan was about 5.3 percent, which is not a high level of credit risk but a consistent one.

The main factors that lead to default are lack of financial literacy among borrowers, lack of monitoring and misuse of loan funds on non-productive activities. Moreover, other common causes have been cited in various studies including delayed loan approvals, poor credit appraisals, lack of knowledge in business management, bad weather conditions and unwillingness to repay by the borrower (Awan, Nadeem and Malghani, 2015). These difficulties imply the necessity of the better borrower evaluation methods and emphasize the opportunities of machine learning models in detecting the risky borrowers earlier in the credit cycle, thus enhancing the portfolio quality.

### **2.3.4 Size of the Loan**

An organization's average loan size reflects its borrowers' average income. The lower the income level of its customers, the lower the loan. This variable's association with the likelihood of a crisis in an MFI might be positive or detrimental. On the one side, a smaller average loan amount is projected to increase risk. In other words, making microloans for small amounts can encourage customers who will have more difficulty paying the loans (Roslan and Karim, 2009). A more considerable average loan amount will result in a higher risk because borrowers with higher monthly installments seem more prone to default (Guttman, 2008; Eze, Ibekwe, 2007).

### **2.3.5 Limited Industry Expertise among MFI Owners**

Essentially, microfinance offers low-interest loans to the poor to lift them out of poverty or catastrophic circumstances (Obaidullah and Khan 2008). As a result, this profession requires more specialized financial expertise and a unique blend of talents, including an understanding of social science, regional language, and customs. On the contrary, many owners, particularly the Chief Executive Officers (CEOs), run the organization as a business and often establish their own after receiving minimal training in other MFIs. Most of these entrepreneurs displayed little concern for the consequences of their actions.

Young financial professionals' feedback to the company was neglected because they had little choice; they let the bosses have their way of keeping their employment (Antwi, 2016; Owusu-Nuamah, 2014). Most proprietors run their businesses as they consider fit, with little or no supervisory oversight. Indeed, the MFI community is responsible for the severe setbacks experienced by the incompetence of governance and administration (Khan 2008).

## **2.4 CHALLENGES IN ISLAMIC MICROFINANCE SECTOR**

The Pakistani Islamic microfinance is characterized by complicated operational and structural issues. Abdul Rahman and Dean (2013) list the main limitations as the

scarcity of Shariah-compliant funds, lack of regulatory clarity, operational expenses, and governance concerns, especially in mudarabah-based models, where the borrowers can underreport profits, which is detrimental to the share of the institution. These problems limit scalability, and undermine institutional sustainability.

Segrado (2005) also stressed that non-participation in the Islamic financial services can be caused by high service charges, poor product knowledge and self-exclusion based on previous dissatisfaction. Akhtar (2007) also cited the barriers to access in rural areas, inadequate law enforcement, and absence of formal documentation among target populations.

Among the more contentious issues is the fact that certain MFIs impose fixed service charges which can be similar to interest (riba) and this poses ethical and perception problems. Nevertheless, the scholars of Islamic finance explain that these fees (ujrah) are acceptable when they are imposed on certain services and are not associated with time or loan value. This practice is supported by the Handbook of Islamic Banking Products & Services (2017) published by the State Bank of Pakistan, which necessitates the approval of the Shariah board and complete disclosure of prices.

#### **2.4.1 Overreliance on a Single Mode of Islamic Financing**

In Pakistan, Qard al-Hasan an interest-free benevolent loan is usually used as a basic product offered by Islamic microfinance institutions (IMFIs). Although theoretically Shariah-compliant, in actuality the majority of IMFIs charge a fixed percentage called ujarah (management fee) as a way to absorb administration expenses. The validity of the service fees within Shariah has been questioned by various scholars on the basis that a fee that is not necessarily tied directly with the cost of the service or that is variable in relation to the magnitude of the loan or the term of the loan can be unlawful as they amount to prohibited riba (Saad, 2012; Hassan et al., 2013). In others these charges can distort the boundary between acceptable ujarah and unacceptable interest, which might be a deterrent to pious Muslim clients. Besides, fee structuring may give rise to gharar (excessive uncertainty) because of a lack of transparency, thereby posing a further problem to conformance according to the Islamic commercial jurisprudence.

Despite the using of alternatives like Murabahah and Bay Muajjal, which have a more asset base and transparent structure, these options are not used widely by MFIs. Murabahah contract deals with a situation in which the MFI buys goods and sells to the borrower at an agreed markup so that a transaction is determined by a real asset thus, avoiding riba. On the same note, Bay Muajjal also includes deferred payment, where goods are supplied by MFIs immediately and awaits payment at the end of the period, in lump sum or instalment (Obaidullah, 2008; Alam, 2008). The contracts are more Shariah compliant since they do not include a debt-for-debt exchange, but instead the trade (Bai 2018).

Although they are beneficial, these tools are not heavily utilized as they are exposed to practical and operational limitation. As an example, Murabahah has asks the MFIs to take the ownership and inventory risks in the form of temporary. Many institutions do not have the background infrastructure to handle these risks effectively. Moreover, the philanthropic nature of delivery of such products as well as the paperwork involved and absence of trained Shariah compliance officers constrains the scalability of such models (Abdullah et al., 2017; Parveen, 2009). Consequently, most MFIs revert to the easier cash model creditor references, even though compliances are concerned..

#### **2.4.2 High Transaction Cost**

The Islamic debt and lease based products and instruments in the form of Murabahah, Ijara, and Bay muajjal have been favored by Islamic MFIs because of the sophistication and uncertainties in profit and loss sharing based instruments Mudarabah and Musharakah (Chowdhury, 2006). Nevertheless, like the traditional players, Islamic MFIs encounter significant transaction costs due to the monitoring, enforcement, and information costs particularly, in the rural financial markets.

The main distinction is the impact of information asymmetry to each of the models. Moral hazard and adverse selection are usually the problems resulting in information gaps in the usual MFIs. Islamic MFIs on the other hand deal with increased asymmetry as a result of their non-collateralized lending model based on trust which requires increased level of scrutiny to ascertain their Shariah-compliance (Karim et al.,

2008). Also, Islamic financing based on asset-driven finance deals are common thus needing a lot more paperwork, the confirmation of ownership of asset, and approval by Shariah board more costly operations alongside that of interest-based lending.

There are higher layers of religious and law compliances involved in transaction costs with Islamic MFIs. As an illustration, every financing contract is supposed to spell out the asset, conditions of sale or lease and certification of compliance thus multiplying time and administrative expenditure. Standard MFIs, in contrast, work with standardized and interest-based contracts that because of their standardization can be processed more efficiently, particularly in large volume.

Islamic MFIs are also usually located in infrastructure inadequate locations where they may lack access to good market information which increases the cost of due diligence (Akhtar & Pearce, 2010). Morocco implicates the similar client base and the absence of centralized information in a situation where several borrowings and over-indebtedness of clients took place as a result of information asymmetry, in its classic CIF form (Akhtar & Pearce, 2010). In order to address these challenges, Islamic MFIs need to invest in digital infrastructure and credit information systems that will reduce the overall costs and make it transparent (Karim et al., 2008).

### **2.4.3 Market Penetration**

Islamic microfinance has a small market share, as indicated by (J Brandsma and L Hart 2002); the absence of a branch network presents the most expensive and time-consuming hurdle when reaching many microfinance customers in Egypt. Other challenges have included inadequate internal standards for performance, protocol, and manager and employee capacity. The goal of Islamic MFIs to empower women is rarely met (Ahmed, 2002). The ease with which women can obtain loans creates tension and increases marital violence. Male members kept female members under pressure to secure a loan. Females are accountable for payback despite not using the funds, putting pressure on their male colleagues to repay the debt.

According to a study by Shehu et al. (2022), the challenges encompass inadequate infrastructure, societal misconceptions, an unfavorable legal and regulatory

framework, intense competition from other financial institutions, neglect of essential microfinance functions, and a shortage of skilled personnel. Despite the numerous obstacles, the investigation identified where these institutions may profit. These opportunities include growing government involvement, considerable financially excluded remote areas, and a high proportion of underprivileged individuals. Challenges and prospects may be resolved by following the research, with appropriate regulatory reforms, and committing other stakeholders to the core goal of microfinance banking.

#### **2.4.4 High Interest Rates**

The issue of high interest rates that high-income clients can afford more disproportionately, presents a dilemma to the micro-finance institutions (MFIs). It is fuelled by high operating costs especially in servicing the small and rural loans. According to Gonzalez (2010), transaction and administrative cost was the major cause of high interest rates in the micro finance sector. The logistics of geographic dispersion, screening of borrowers and follow up after the loan further increases those costs. Lewis (2008) lamented about some traditional MFIs as to their high interest rates, and thus referred to them as microloan sharks because of exploitive tendencies. In the same light, Cull et al. (2007) noted that though individual lending contributed to increased profitability of MFIs, group-model based lending was financially inconsistent.

Contrary to the traditional MFIs, the Islamic MFIs warning systems rely on strict ethical and religious codes of conduct specifically the aspect of eliminating the use of riba (interest) that poses a special pricing structure challenge. Islamic MFIs will commonly use service fee rather than charge interest (ujrah) or implement profit-driven instruments such as Murabahah (cost-plus financing) and Ijara (leasing). These mechanisms are meant to recover cost, and yet escape the tag of interest. Yet the critics suggest that some Islamic MFIs have levied ujrah charges which are similar to interest rates in magnitudes- questioning the actual compliance to the Shariah (Karim et al., 2008).

In resolving the conflict between Operational sustainability and Islamic ethics, the scholars recommend the pricing mechanism that is transparent, cost-based, and

associated with real services or assets. To take one example, Murabahah profit margin must be taken to be based on real administrative and risk costs, rather than than interest standards in the market. On the same note, agency models through Wakalah may provide fee-based arrangement in which the client charges the MFI to represent their interest without infringing the riba lies. In order to be credible, such pricing structures must be verified by non-affiliated Shariah boards and audited regularly on its compliances.

In conclusion, the Islamic MFIs should be innovative in a restricted ethical environment to balance between reaching out to the poor and financial sustainability. This will increase their validity, it will reduce reputational risks and boost trust of pious Muslim customers.

## **2.5 LOAN RISK IN MICROFINANCE INSTITUTIONS**

Kou et al. (2014) note that the complexity of the task of default risk assessment is the reason why intensive studies on this issue have been conducted in recent decades. This risk is pivotal in maintaining profitability of institutions and it is in this capacity that the loans still rank first among credit risk exposures of financial institutions. As pointed out by Ruziqa (2013), bad debts are good indicators of the worsening credit quality and this has severe consequences on the stability of the financial market and the macroeconomic projection.

In microfinance the issue of default presents enhanced difficulties especially in the application of Islamic Microfinance Institutions (IMFIs). The financial institutions have little safety cushions and this increases their susceptibility to loan losses. According to Yoseph (2017), even a minor increase in defaults may translate into serious capital erosion, which poses a risk to the sustainability of the institution. It is therefore necessary to make provisions on expected credit losses and a rise in loan loss provision ratio is usually an indication of increased difficulty in asset recovery (Asrat, 2018).

Whereas the traditional MFIs are mostly concerned about interest income and collateralization as a way of mitigating risk, the Islamic MFIs uses Islamic precepts

(including: risk-sharing (musharakah, mudarabah), ethical lending (qard al-hasan)). Such models transfer the financial risk of lender and the borrower that encourages the two to be responsible to each other. Furthermore, debt forgiveness is strongly encouraged in Islamic jurisprudence when the borrower faces genuine hardship, as stated in the Qur'an (2:280): "And if the debtor is in a hard time, then grant him time until it is easy [for him to repay]. But if you remit it by way of charity, that is better for you..."

To mitigate default risk, Islamic MFIs can implement Shariah-compliant strategies such as:

- By effective documentation and following of the installments using lending like Murabahah which is asset backed.
- Using social collateral by utilizing group lending models built on society trust.
- Promoting financial responsibility among borrowers in regard to Islamic ethics.
- Forming Zakat funds/Sadaqah pools to help unlucky yet honest borrowers.

Practice shows that instances of defaults are usually less in the IMFIs as compared to the conventional MFIs because they have ethical lending structures in place and operate on community-based plans (Obaidullah & Khan, 2008). Nevertheless, insufficient official credit bureaus and excessive indebtedness are continuously being an issue in a number of Muslim-dominant nations.

## **2.6 CREDIT SCORING**

Credit scoring is the key strategy for lowering the credit risk associated with a loan request (Siddiqi 2017). This method involves assessing borrowers' financial stability, predicting default probability, and mitigating non-repayment risk to an acceptable threshold. Credit scoring is commonly used to determine credit risk (Anderson, 2007). All financial institutions must develop relevant, effective, and efficient credit risk assessment tools. Credit scoring has emerged as an essential instrument for formulating

management strategies and assessing credit risk (Thomas et al., 2002). The banking industry has placed much emphasis on credit scoring because this subject holds significant importance within financial risk management. Mainly because of the 2007–2008 financial crisis (Leo et al. 2019).

Whereas in traditional credit scoring numerical indicators are important, IMFI grant credit with consideration of ethical and social aspects in credit assessment. Take, Islamic organizations such as the Akhuwat in Pakistan have guaranteed enviably more modest default rates (0.150) relative to an average of 2.290 of other MFIs--which, in great part, can be credited to their group lending framework and mutual liability contracts, which, effectively, counter moral hazard and adverse selection (M,khan and AS Bedi, 2022). Drawing on community consensus and social status of an applicant, Akhuwat does not base the approval of applicants on numeric scores, which differs significantly with the informal but effective peer-based credit qualification system.

This assessment of trust is based on Islamic ethics: social responsibility, and personal integrity have the same significance as official financial measures. The group members take collective responsibility of repayment, and the unabridged members of the group usually warrant the applicants. These mechanisms echo deeper values of Islamic jurisprudence (amanah (trustworthiness), ta'āwun (mutual cooperation) and risk-sharing) in that they make financial risk management compatible with faith-based ethics (Obaidullah & Khan, 2008) The application shows that Islamic creditworthiness is assessed qualitatively, rather than numerically, incorporating social trustworthiness as well as moral integrity into the realm of financial decision-making.

## **2.7 METHODS FOR EVALUATING DEFAULT RISK**

Financial credit loss is the possibility of nonpayment by a borrower or other obligations in a creditor-debtor relationship with the risk of defaulting or illiquidity. The use of the Linear Discriminant Analysis (LDA), Logistic Regression (LRA), Multivariate Adaptive Regression Splines (MARS), and other traditional models to conduct credit scoring has been in use since long (Thomas, 2000).

More recently, Artificial Neural Networks (ANN) and Support Vector Machines or SVM especially were found to be better accurate measures in predicting credit risks (Bekhet & Eletter, 2014; Liu & Chen, 2006). Special attention can be paid to the fact that hybrid methods that suggest the combination of Genetic Algorithms with SVM or neural networks have been also shown to improve the level of prediction results even more (Min & Lee, 2005; Huang et al., 2004).

Nevertheless, there are quite a number of ethical and operation constraints that present in the application of the methodologies to Islamic Microfinance Institutions (IMFIs). In particular, interpretability (it can be against Islamic values of avoiding excess uncertainty (gharar)) and the lack of ethical screening of purposes and character of borrower can be seen as major issues (Iqbal et al., 2025).

Hybrid credit scoring models are applicable in resolving these limitations. Those models, which integrate quantitative evaluation of investment with an ethical filtering and a local validation process, are in line with ta Society. They can incorporate mutualized assurances (takaful) or peer-endorsement schemes, and improve fairness and trust without undermining projects of understanding future probabilities.

## **2.8 WORLDWIDE PERSPECTIVE ON MACHINE LEARNING METHODS FOR CREDIT SCORING: A REVIEW OF RESEARCH**

A study by Yuan et al. (2022) evaluated the default risk of borrowers in peer-to-peer (P2P) online lending networks. The researchers introduced an innovative approach for classifying default risk, incorporating feature extraction and data cleaning to improve the accurateness of risk evaluation. The authors created a risk assessment model utilizing low-rank representation (LRR) and discriminant analysis, employing borrower data from Lending Club. Before implementing a discriminant analysis method to minimize the dimensionality of the cleaned data, the authors utilized three LRR models to remove outliers and noise from the high-dimensional borrower dataset. Machine learning classifiers, including k-nearest neighbor, support vector machine, and artificial neural network, were employed to assess and classify default hazards within the reduced feature space. The results indicated that the borrower data had considerable noise and redundancy. Specifically, the LRR models with local manifold regularization

effectively cleaned such data. The local Fisher-supervised discriminant analysis model effectively extracted low-dimensional and discriminative features, thereby improving the accuracy of the final risk assessment models. Such methods as LRR, GBDT, and SMOTE applied to diverse studies assist enhance performance by managing such problems as noise, redundancy, and class imbalance (Yuan et al., 2022; Wenxue et al., 2019; Zhu et al., 2019).

Traditional methods, including credit scoring, Logistic Regression (LR), and Bayesian Decision Rules, have been employed to identify and predict default risk proactively. A study by Sharma et al. (2023) utilized sequential feature selection alongside machine learning models, specifically Random Forest and XGBoosting Classifier, to predict default risk on a peer-to-peer lending platform. The analysis used data sourced from the Lending Club. It is necessary to note that Lending Club information does not necessarily extend to local lending activity in an informal and community-based context of developing nations like MFIs (Sharma et al., 2023; Yuan et al., 2022). The experimental results demonstrated that the Random Forest and XGBoost Classifier attained an accuracy of 0.97. In the context of the ROC (Receiver Operating Characteristic), the Random Forest model demonstrated superior performance to the XGBoosting Classifier.

Machado and Karray (2022) employed hybrid machine learning algorithms that combined supervised and unsupervised ML techniques to predict credit scores for business clients. They implemented various strategies and compared the effectiveness of these hybrid models with standalone supervised ML models. The findings revealed that the hybrid models outperformed the individual supervised models in accurately predicting credit scores for commercial clients. Although ANN and ensemble solutions such as Random Forest tend to be more predictive, some use common models such as Logistic Regression or Decision Trees as their openness and interpretability are primary in regulatory or low-tech contexts (Kumar et al., 2021). Additionally, the researchers discovered that incorporating historical credit ratings further enhanced the predictive capabilities of the hybrid models, which had been overlooked in previous research. Consequently, the authors successfully assessed credit risk for commercial customers using hybrid machine learning algorithms.

In their research, Condori-Alejo et al. (2021) proposed a model for assessing microcredits in rural areas based on a thorough analysis of rural variables and relevant literature. This model is a decision-support tool for rural business advisors to mitigate default risk for rural microfinance institutions. The data underwent pre-processing, and machine learning models were selected, trained, validated, and evaluated using various metrics. The study assessed the most significant variables in the financial and microfinance sectors. Among the evaluated models for granting rural microcredit, the ANN achieved the highest level of accuracy at 93.72%. ANNs are especially appropriate when it comes to rural MFIs because noise and unstructured data are common aspects of informal lending contexts (Condori-Alejo et al., 2021). Although ANN is always highly accurate in predicting (e.g., Condori-Alejo et al., 2021), its overfitting and inability to provide interpretations are the drawbacks to taking it into practice, particularly in an environment with regulations or low technology. Logistic Regression demonstrated an accuracy of 86.07%, Random Forest achieved 66.35%, SVM scored 84.44%, Decision Tree yielded 88.80%, and k-Nearest Neighbor achieved 65.98%. These models were effective in the evaluation process for granting rural microcredit, depending on the variables and data used for the specific institution. Moreover, the ANN model's 93.72% assertiveness rating is higher than the institution's previous 76.81% technique, indicating a 16.91% enhancement in the index of default clients. Nevertheless ANNs can also overfit particularly when the training on smaller datasets (possibly imbalanced) lack sufficient regularization and validation measures (Yeh and Lien, 2009).

In their study, Kumar et al. (2021) aim to identify the limitations of various AI-ML-based credit scoring techniques currently employed by banking and non-banking institutions. The research adopts a systematic approach to the literature review, analyzing existing empirical studies to compare and contrast the most appropriate AI-ML-based models used by different financial institutions worldwide. The primary objective of this research is to highlight several ML algorithms that the researchers have previously identified. The ethical implications of AI-based credit models also bring into line such issues as unjustness, openness, and prejudice (Kumar et al., 2021), which are more related to the instructions of ethical screening of models. Furthermore, the study raises intriguing questions about financial institutions' successful integration of

traditional and digital approaches, emphasizing the need to address potential ethical issues.

Ampountolas et al. (2021) evaluated the efficacy of various ML algorithms in classifying borrowers using actual micro-lending data. Utilizing easily accessible customer data, the study demonstrated that commercial multi-class classifiers, including random forest algorithms, can effectively accomplish this task (e.g., age, occupation, and location). It provides cost-effective and reliable methods for micro-lending organizations in developing countries to assess borrower creditworthiness without access to credit histories or centralized credit databases. The analysis results demonstrate that machine learning algorithms are applicable for modeling credit risk in microcredit settings, even without a central credit database or prior credit history. Empirical analysis indicated that tree-based machine learning algorithms outperform alternative methods, with ensemble classifiers identified as the most effective models.

In the study conducted by Ozgur et al. (2021), it was highlighted that the determinants of bank lending exhibit non-linear and non-parametric relationships with outstanding bank loans. The researchers examined the impact of 19 macroeconomic, local, and international variables on bank loans in Turkey from 2002Q4 to 2019Q2. To assess the influence of these factors, they compared the performance of a regression model with ML-based approaches. The authors observed that ML-based algorithms offer flexibility in handling the complexity of variables, which is particularly valuable when dealing with large-dimensional datasets. In contrast, conventional linear regression approaches struggle to manage such datasets effectively. Therefore, it is always advised to have a reliable strategy in place for predicting debt repayment before extending any credit to the debtors.

Assef and Steiner (2020) aimed to identify ML techniques suitable for categorizing borrowers. They utilized the characteristics of 5,432 customers from a Brazilian financial institution to classify each client into one of three categories: non-defaulters, temporary defaulters, or defaulters. The findings revealed that the radial basis functions method of artificial neural networks achieved the highest level of accuracy in the risk assessment process.

Aniceto et al. (2020) assessed the creditworthiness of borrowers using ML techniques, specifically random forest (RF) and AdaBoost. The researchers utilized a loan database from a Brazilian bank to investigate the effectiveness of these ML approaches in evaluating borrowers. The dataset consisted of low-income borrowers from Brazilian financial institutions, with a default rate of approximately 48%. By developing an ML model based on real-world data, the authors demonstrated that RF and AdaBoost outperformed other models in accurately assessing borrower adequacy.

Provenzano et al. (2020) introduced ML algorithms for estimating credit ratings and predicting defaults. They utilized financial indicators such as historical balance sheets, bankruptcy regulations, and macroeconomic factors sourced from Moody's dataset. Through their analysis, the authors discovered that employing machine learning models yielded impressive out-of-sample performance results in mitigating bankruptcy risk and improving credit scores.

Shen et al. (2020) proposed an innovative credit-scoring model based on cost-sensitive logistic regression, utilizing a multi-objective optimization framework to accomplish the objectives. A multi-objective particle swarm optimization (MOPSO) method was used to optimize the parameters of cost-sensitive logistic regression. The established model was applied to assess the credit scoring of a leading P2P organization in China. The findings indicated that the suggested model surpassed existing prevalent credit scoring models by decreasing type II error rates and overall classification error costs and enhancing measures such as AUC, F1 scores (a harmonic mean of Recall and Precision), and G-means. Moreover, the established model was evaluated against other multi-objective optimization techniques, validating the superiority of MOPSO in the realm of cost-sensitive logistic regression credit scoring models.

Malhotra et al. (2020) aimed to explore the effectiveness of decision trees, AdaBoost, and SVMs in identifying potential bad loans. Their findings indicated that AdaBoost performed superior to decision trees and SVM models in accurately predicting excellent and bad credit clients.

Teply and Polena's (2020) key contribution was the robust evaluation of 10 distinct categorization techniques based on a set of actual P2P loan data. Their data collection, which includes 212,252 records and 23 different variables, was obtained

from Lending Club and spans 2009 through 2013. Writers use a data sample of every loan's ultimate loan settlement. Six different categorization performance evaluations and a 5-fold cross-validation procedure were used. According to their findings, the three most effective algorithms using the lending club data are linear discriminant analysis, ANN, and logistic regression. The two lowest classification techniques, on the other hand, are the classification and regression tree and k-nearest neighbors.

According to the research conducted by Zhu (2019), the study discusses the challenges faced by P2P online lending platforms concerning user loan defaults despite the opportunities offered by e-commerce and big data technology. These platforms' sustainable and healthy development relies on managing such risks effectively. The research focuses on developing a loan default prediction model using real-world customer loan data from lending clubs. Nevertheless, the sample reported in the Lending Club, the most frequently researched body of data on this topic, belongs to a P2P service hosted in the United States and thus is not representative of informal finance or unbanked individuals stereotypical of microfinance environments. The RF algorithm is employed as the primary approach while addressing the issue of imbalanced class in the dataset by utilizing the SMOTE technique. The research encompasses various steps, including data cleaning and dimension reduction. The experimental results demonstrate that the RF approach outperforms logistic regression, decision trees, and machine learning algorithms in accurately forecasting default samples.

Wang and Li et al. (2019) compared the efficacy of five leading classifiers employed in credit scoring. The classifiers comprise the Naive Bayesian Model, Logistic Regression Analysis, Random Forest, Decision Tree, and K-Nearest Neighbor classifier. Every classifier has distinct advantages and disadvantages, complicating the identification of the most effective option. The research findings demonstrate that the Random Forest classifier outperforms other classifiers in precision, accuracy, AUC (area under the curve), and precision.

Wenxue et al. (2019) utilized a hybrid methodology incorporating GBDT (Gradient Boosting Decision Trees) and logistic regression models to assess personal credit risk. Applying the GBDT model successfully extracted significant combinatorial aspects from the original data, preventing the oversight of essential information inside the bank credit data. The integration of GBDT and LR in the personal credit risk

assessment model significantly improved prediction accuracy, given the multiple indicators and complicated noise characteristics inherent in bank credit data.

Munkhdalai et al. (2019) intended to build a novel benchmark utilizing authentic consumer data and to present machine-learning methodologies as a foundation for this benchmark. They extensively compared machine learning methodologies and the FICO credit scoring system, which relies on human experience, utilizing data from the Survey of Consumer Finances (SCF). Given that the SCF data comprises several variables, they used two strategies to choose the most representative characteristics for effective modeling. The initial method utilized hypothesis testing, correlation analysis, and feature value assessments based on random forests, but the second method presented an innovative random forest-based technique termed NAP (New Approach).

The researchers developed regression models utilizing diverse machine learning algorithms, such as ensemble gradient boosted trees, support vector machines, logistic regression, and deep neural networks. The findings suggested that if financial institutions had created their risk assessment models utilizing the examined machine learning approaches, the expected credit losses in the 2000s could have been reduced and made more sustainable. Significantly, deep neural networks attained the highest AUC, although XGBoost methods exhibited the highest accuracy when trained on the subset identified by NAP.

Moradi et al. (2019) have discussed the socioeconomic impact on determining a client's creditworthiness; earlier research studies mainly concentrated on a customer's demographic and other statistical factors. The writers emphasized that political abnormalities have an impact on economic aspects as well. To estimate credit risk, they also took politico-economic issues into account. They created an adaptive network-based fuzzy inference system to anticipate whether a specific loan is performing. Due to COVID-19, the socioeconomic impact is essential for most lending institutions. Researchers evaluated their customers' credit scores using data from an Iranian bank, particularly during unique political and economic circumstances. Using Iranian bank clients' behavioral characteristics, they evaluated credit scoring using the fuzzy inference method, outperforming other standard models, especially in financial uncertainty.

Linh et al. (2019) focused on the socioeconomic factors present in Vietnam's rural credit markets, where farmers' access to credit is mainly based on their output level and household income. The formal, semi-formal, and informal channels are the primary sources of financing for farmers. However, the author does not mention the factors that semi-formal lenders consider. Additionally, it exclusively applies to Vietnam's rural loan markets and does not apply to the marketplaces of other developing nations. Since the 1990s, credit scoring has gained popularity as a method to assess the creditworthiness of borrowers. Instead of simply accepting or rejecting a loan application, credit scoring predicts the future behavior of borrowers, distinguishing them as either defaulters or non-defaulters according to lenders' classifications.

Wang and Li (2019) utilized decision-support methodologies to develop a credit risk analysis model. They employed support vector machine models and linear regression to examine the data and assess credit risk. The empirical findings demonstrated that decision support can improve borrowers' credit risk evaluation accuracy.

Maher et al. (2018) presented an efficient and reliable way of measuring credit risk in the context of Jordanian banks. The study aimed to explore and evaluate the implementation of credit scoring models. Logistic regression, ANN, and SVM approaches were employed to develop these scoring models. The findings indicated that the LR model outperformed both ANN and SVM regarding various performance measures.

Kurapati and Bhansali (2018) showcased the effectiveness of loan default prediction models. They concluded that the Random Forest method outperformed other models like Decision Tree and Gradient Boosting in accurately identifying credit defaulters. Addo et al. (2018) proposed that selecting factors aligned with business objectives and the algorithms employed in decision-making are crucial components in loan-granting activities within organizations.

Nalica and Svraka (2018) highlighted the traditional approach of lenders and financial institutions, where qualified professionals were responsible for assessing the creditworthiness of borrowers before making loan decisions. However, in recent years, these institutions have adopted alternative loan analysis models to determine loan

approval or rejection based on credit scores and repayment ability. The authors utilized ML techniques and ANN to predict potential defaulters among borrowers. These models leverage historical data to assign a numerical score (1 or 0) indicating the creditworthiness of borrowers.

There has been significant interest in predicting loan defaults, leading to the testing of various algorithms. Ensemble algorithms have been recommended for better performance. To enhance the classification model's effectiveness, research has emphasized the importance of incorporating "concrete information" such as income, age, FICO score, debt details, and borrowers' "soft information." Social network information has gained attention in this regard. While some scholars have explored the relationship between online social network information and loan default, limited research has focused on the role of social network information beyond online platforms. Although Cnudde (2019) collected Facebook social network data to improve prediction accuracy, this approach is not viable in many countries where Facebook is inaccessible.

Furthermore, many individuals are reluctant to share their online social network data or do not possess social media profiles. Additionally, some users' data may be inadequate, making obtaining their social network information impossible. These factors limit the effectiveness of this technique and impede the improvement of its risk prediction capabilities.

Based on Zhu, L et al. (2019) findings, this study focuses on developing a loan default prediction model using actual user loan data from lending clubs. The SMOTE approach addresses the imbalanced classes in the dataset, followed by various procedures, including data cleaning and preprocessing. The experimental results indicate that the Random Forest algorithm outperforms logistic regression, decision trees, and machine learning algorithms in predicting default samples. Another study by Moradi and Mokhatab Rafiei (2019) utilizes monthly data from a customer profile dataset to train an adaptive network-based fuzzy inference system (ANFIS), and a subsequent tranche evaluation in a fuzzy inference system is then conducted and uses the newly established variables and their underlying rules. To analyze the credit risk of the consumers, they compiled a database of poor clients regularly and built an evolving model by conducting an empirical investigation.

Khemakhem and Boujelbene (2018) utilized decision trees and artificial neural networks to forecast credit risk. They observed that imbalances highly influenced both methods in class distribution. Nevertheless, decision trees outperformed artificial neural networks when the data was balanced in accurately predicting credit risk.

The study conducted by Ala'raj et al. (2018) aimed to explore the application of credit scoring models in Jordanian banks and assess their effectiveness in measuring credit risk. The research methodology consisted of two main steps. Firstly, the models were developed using LR, ANN, and SVM approaches. The best-performing model was selected for further analysis. The second phase involved a profitability test to evaluate whether the chosen model aligned with the strategic and business goals of the banks.

Additionally, credit expert information was incorporated to enhance the results of the scoring model. The findings indicated that the LR model outperformed ANN and SVM across various performance measures. Moreover, the LR model proved to be most effective in achieving the strategic and commercial objectives of the banks.

Khemakhem and Boujelbene (2018) aimed to explore the importance and effectiveness of integrating sampling models with statistical prediction and artificial intelligence techniques to predict and quantify default likelihood using real-world credit data. To address the issue of unbalanced data, a database from a Tunisian commercial bank was utilized, and random over-sampling (ROS) and synthetic minority over-sampling techniques (SMOTE) were employed. Performance assessment was conducted using the confusion matrix and receiver operating characteristic curve. The findings showed that amalgamating advanced statistical techniques and resampling methods is beneficial for managing default rates and provides precise credit prediction.

Addo et al. (2018) constructed binary classifiers using deep learning and machine learning models with actual data to predict the probability of loan default. The performance of these models was compared on various datasets, and the top 10 significant features were selected from these models to assess the reliability of the binary classifiers. The findings indicated that tree-based models were more robust than multilayer artificial neural networks.

Wijewardhana et al. (2018) utilized historical data from a collection agency based in the US to predict the repayment patterns of clients regarding their debts. They

assessed highly accurately a customer's debt payback behavior using mathematics, data mining, and statistical models. However, the sample selection issue is the most prevalent problem for most research authors in the literature on consumer credit. It is crucial to remember that the data sets used for the research study directly affect the forecast's accuracy. For this reason, proper, acceptable, and relevant data is the foundation for most prediction studies when evaluating credit scores. Furthermore, banking institutions face significant challenges in developing advanced machine learning algorithms to improve key banking processes such as fraud detection, credit assessment, customer churn prediction, and more, primarily due to limited available data sources.

Most researchers have emphasized that a categorization issue exists with credit scoring and have compared the credit data sets from Germany and Australia with well-known classifier benchmarks (Boughaci and Alkhawaldeh 2018). To determine the credit score, researchers combined the support vector machine model with local search methods (LS), stochastic local search techniques (SLS), and variable neighborhood search (VNS) methods.

Khemakhem and Boujelbene (2018) identified significant financial and non-financial factors to develop a credit score model and assess a company's creditworthiness. The synthetic minority oversampling technique addressed the class imbalance and improved the classifier's performance, such as decision trees and ANN, to forecast default risk. The results indicated that essential predictors of default included profitability ratios, repayment capacity, solvency, credit report length, guarantees, company size and number of loans, ownership structure, and duration of the corporate banking relationship. Furthermore, it was observed that both algorithms were sensitive to class disparity, but decision trees outperformed artificial neural networks in terms of predicted accuracy with balanced data.

Zhang et al. (2018) utilized multiple instance learning (MIL) to develop a novel credit rating model. Papouskova and Hajek (2019) constructed a two-stage consumer credit risk model using heterogeneous ensemble learning, which employed class-imbalanced ensemble learning for credit scoring prediction and regression ensemble for exposure at default prediction.

Boz et al. (2018) investigated the application of machine learning to reevaluate applicants and suggest an additional screening step to an already-in-use score system. They use actual data from one of Turkey's well-known loan businesses. The variables in their study were created by the information the applicants gave. According to their continued payments, the clients have previously been classified as excellent and flawed by the company's experts. Using this labeled data set, the authors employed several techniques to categorize the undesirable applicants and the key factors in this classification. Most clients were rated good because the data set only includes candidates who passed the initial grading procedure. They used several ways to enhance the performance of identifying the applicants who were more likely to default to tackle the issue of the uneven nature of the situation. The second part of this study aims to forecast client payment behavior using static (demographic and financial data) and dynamic (payment) data. The authors also investigate the effects of the length of the payment history and the robustness of the proposed prediction models. Overall, they found that the methods for customer monitoring, SVM, and random forest perform well and are effective.

DM et al. (2018) conducted a study to evaluate the predictive accuracy of logistic regression and support vector machine models for individual loan defaults in Kenya. Equity Bank provided the data for the years 2006 through 2016. One thousand loan applicants with approved loans were utilized as a sample. The factors considered in the analysis encompassed various aspects such as credit history, loan purpose, loan size, savings account status, employment status, gender, age, security, and location. The logistic regression and support vector machine models were trained using the train data. The study utilized logistic regression and SVM models to examine loan default prediction. The results indicated that the support vector machine outperformed the logistic regression model's performance. Based on these findings, the study recommended that financial institutions consider employing support vector machines to anticipate loan defaults.

Chen et al., 2019; Zhou et al. (2018) have employed a variety of borrower list information, fundamental information for assessing peer-to-peer default risk, and financial data to determine risk. Several studies have indicated a significant association between borrower list information and financial information with default risk. These

factors have been found to play a crucial role in assessing the likelihood of default by borrowers. The features of borrowers' repayment capacity, willingness to repay, and online browsing activity were also considered. To enhance the accuracy of risk assessment, various factors play a role in determining the likelihood of borrowers defaulting on their debts. These factors include borrowers' financial data, such as income and credit rating; historical information, such as borrowing frequency and overdue frequency; borrowing details, such as loan amount and term; and basic personal information, including age, education, and marital status (Chen et al., 2019). The risk assessment process can be improved by gathering comprehensive information about potential borrowers, leading to more accurate predictions.

Attigeri et al. (2017) carried out risk analysis using diagnostic and cross-sectional studies using a framework based on research methodology. Classification models for credit risk assessment were developed using supervised machine learning algorithms and an empirical approach to build models. The chi-square statistical test creates and evaluates the logistic regression and neural network classification models. This statistical test played a crucial role in assessing the variables' significance and impact on the classification performance. By incorporating this methodology, the models aimed to enhance the accuracy and effectiveness of credit risk assessment. This study infers the significance of employing machine learning algorithms to anticipate unsatisfactory clients. Logistic regression performed better for this work's data set and parameters.

Behr and Weinblat (2017) compared the effectiveness of three distinct data mining approaches. Using the LOGIT model, decision tree, and random forest methods, this analysis used balance sheet data from 446,464 UK, France, Germany, Italy, Portugal, and Spain enterprises. Classification trees and random forests account for the extreme imbalance in the solvency state. Resampling approaches thoroughly optimize all three model specifications, solely utilizing the training sample. Out-of-sample forecasts are the only criteria used to evaluate model performance. Random forest significantly outperforms the classification tree and LOGIT model in all investigated years and nations.

According to Wójcicka (2017), there is an ongoing effort to develop various techniques and models to address a significant challenge banks face: accurately

identifying borrowers and calculating credit risk. This challenge emphasizes the need for effective methods and tools to assess borrowers' creditworthiness and mitigate potential lending risks. By continuously exploring and refining these techniques and models, banks can improve their ability to make informed credit assessment and risk management decisions. The numerous techniques used aim to enhance the reliability of creditworthiness projections, perhaps leading to a more significant and successful loan portfolio. The study's findings indicate that neural networks play a crucial role in the credit risk evaluation process and have strategic importance. The model has predicted firm instability and effectively enhanced credit risk assessment accuracy. Incorporating neural networks in the credit risk judgment process has proven valuable in providing reliable and insightful insights for assessing and managing credit risks.

Aledanjro (2016) discovered that classical logistic regression did not perform well when the numerous characteristic variables had extensive nonlinear correlations. While the machine learning model outperforms the LR model in prediction accuracy, the LR model still holds an advantage regarding variable applicability and stability. Recognizing this, researchers have improved and applied the LR model to predict borrower default behavior. By leveraging the strengths of LR, they have aimed to enhance its performance and effectiveness in assessing credit risk and predicting default probabilities.

Butaru et al. (2016) utilized data from six U.S. banks to analyze consumer delinquency. They employed decision trees, LR, and random forest models for prediction. The findings revealed that both the random forest and decision tree models exhibited similar levels of prediction accuracy, outperforming the LR model consistently.

Malekipirbazari and Aksakalli (2015) introduced a credit risk assessment model with a random forest machine-learning algorithm. The model incorporated 15 borrower-provided feature data points and non-standard financial data variables to enhance the accuracy of the forecasts. The study's results indicated that the credit risk assessment model based on the random forest algorithm exhibited greater precision in predicting borrower defaults than the commonly used method relying on FICO credit scores.

Emekter et al. (2015) developed an LR model to predict default risk using historical data obtained from lending clubs. According to the empirical findings, the debtor's credit score significantly impacts their default likelihood. The efficiency of credit score calculation using a multilayer perceptron (MLP) neural network by Zhao et al. (2015) employed a dataset of German credit to train the model and determine its accuracy. According to their findings, an MLP model with nine hidden units had a classification accuracy of 87%, which was greater than that of prior trials of a similar nature. By increasing the number of hidden units, their study's findings demonstrated a trend in the accuracy of MLP models' scores.

Bekhet and Eletter (2014) assist Jordanian commercial banks in making loan choices; this study suggested two credit scoring models. Assessment of the loan application would reduce analysis time and expense while increasing the efficiency of credit decisions and managing loan office responsibilities. The credit rating models were created using loan applications from various Jordanian commercial banks, both accepted and denied. Based on the findings, the logistic regression model demonstrated a slightly higher average accuracy rate than the radial basis function model, as reported by the authors (Zhao et al., 2015). The radial basis function, however, was more effective at detecting clients who could default.

Harris (2013) used credit rating models based on SVM to assess credit risk. The author compared models using broad criteria (less than 90 days past due) and narrow criteria (more than 90 days past due) for defining defaults. The results indicated that the models based on the broad definition of default outperformed those based on the narrow definition.

The study also aimed to develop accurate credit-scoring models for a credit union in Barbados. Empirical testing demonstrated that these quantitative credit risk models could enhance credit risk evaluation at the Barbados-based institution, surpassing the judgment-based approach. The performance of five popular machine learning classifiers, namely Naive Bayesian Model, Logistic Regression Analysis, Random Forest, Decision Tree, and K-Nearest Neighbor classifier, was compared for credit scoring purposes. It is important to note that each classifier has its strengths and weaknesses, making it challenging to assert the superiority of one over another. However, the experimental findings indicated that the random forest classifier

outperformed the other models regarding precision, recall, area under the curve, and accuracy. These performance measures suggest that this study's random forest classifier exhibited better predictive credit scoring capabilities.

According to Cubiles-De-La-Vega and Blanco-Oliver (2013), supervised classification algorithms have been successfully utilized for credit scoring in various non-microfinance contexts. However, applying credit scoring using these algorithms in the microfinance sector is relatively new, and current research primarily relies on conventional statistical techniques. This gap is surprising, considering that adopting credit scoring with supervised classification algorithms can enhance the efficiency of microfinance institutions and enable them to compete more effectively in a regulated environment.

This study addresses this gap by empirically examining various statistical learning methods for microfinance credit assessment. The study utilizes a dataset of microcredits from a Peruvian microfinance institution. It applies multiple models such as classification trees, LR, MLP, SVM, linear and quadratic discriminant analysis, and ensemble approaches based on bagging and boosting algorithms. The objective is to differentiate between default and non-default credits and evaluate the performance of these models in a microfinance context.

The outcomes point to employing a multilayer perceptron tuned with a second-order algorithm in the R statistical system. Furthermore, their results demonstrate that the MFI misclassification costs decreased to 13.7% with the deployment of our MLP-based model compared to the application of other traditional models. Wang et al. (2012) focused on the use of decision trees (DT) in credit scoring models. They found that DT-based models often underperform compared to other methods due to two main factors: noisy data and redundant features. To overcome these challenges and improve classification accuracy, the authors proposed two ensemble strategies: RS-Bagging DT and BaggingRS DT, which combine bagging and random subspace techniques.

The study employed two real-world credit datasets from Australia and Germany to evaluate the effectiveness of the proposed solutions. Experimental results revealed that the single DT classifier had the lowest average accuracy among the five classifiers tested (Logistic Regression Analysis, Linear Discriminant Analysis, MLP, Radial Basis

Function Network). On the other hand, RS-Bagging DT and Bagging-RS DT outperformed not only the single classifiers but also four well-known ensemble classifiers (Bagging DT, Random Subspace DT, Random Forest, and Rotation Forest) in terms of performance. These findings indicate that RS-Bagging DT and Bagging-RS DT are promising alternatives for credit-scoring tasks. By leveraging ensemble techniques and addressing the challenges of noisy and redundant data, these approaches offer improved classification accuracy and potentially enhance credit scoring models.

According to the research findings by Brown and Mues (2012), the random forest and gradient-boosting classifiers exhibit excellent performance in credit scoring, even when faced with significant class imbalances within the datasets. The study also revealed that in scenarios involving a substantial class imbalance, the C4.5 decision tree technique, quadratic discriminant analysis, and k-nearest neighbors classifiers outperformed those initially considered the best performers. These results indicate that these alternative classifiers have demonstrated their capability to effectively handle imbalanced data and yield superior performance in credit scoring tasks.

In a study by Kumar et al. (2021), a decision tree approach classifies borrowers into performing or non-performing loan risks. The researchers suggested the adoption of the C5.0 algorithm, a decision tree model, specifically for rural Indonesian banks (Bank Perkreditan Rakyat), as it can significantly reduce the risk of non-performing loans (Mandala et al., 2012). The study aimed to identify key factors rural banks should consider when evaluating credit applications to mitigate non-performing loans. A decision tree model for credit evaluation using data mining techniques. The study identified eight continuous variables, including age, monthly income, credit amount, monthly expenses, current payment amount, savings, collateral values, and loan duration, as well as five discrete variables, including gender, type of collateral, type of business activities, source of funding, credit status, and loan usage. Among these variables, the study emphasized the importance of collateral value as a significant factor for rural banks to consider when assessing the creditworthiness of borrowers, aiming to reduce the occurrence of non-performing loans.

According to the research conducted by Brown et al. (2012), random forest and gradient-boosting classifiers are highly effective in credit scoring, mainly when dealing

with imbalanced class distributions in the datasets. Utilizing the applicant's transaction history data and sociodemographic and loan application data.

Conferring to the findings of Yap et al. (2011), applying data mining techniques is essential in developing credit scoring models. These models utilize credit scoring algorithms to identify relevant demographic characteristics linked to credit risk. By incorporating historical payment data, demographic variables, and statistical methodologies, these algorithms generate a credit score for each client. This study discussed enhancing credit scoring models' evaluation of credit worthiness. Due to privacy issues and the lack of financial data from banks, this study uses payment history information from club members to apply credit scoring methodologies. The club has already been struggling with an increase in the number of people who aren't paying their monthly dues. To identify potential defaulters, administrators seek a model that compares the classification performance of the decision tree, logistic regression, and credit scorecard models. The credit scorecard model, logistic regression, and decision tree exhibited classification error rates of 27.9%, 28.8%, and 28.1%, respectively. Scorecards are generally much easier to use in actual applications, even though no model beats the others.

As stated by Wu et al. (2010), credit risk assessment forms the basis of credit risk management in financial institutions. Credit scoring has been highlighted in this context as a crucial instrument to evaluate the risk, enhance cash flow, lower the potential dangers, and take management decisions. Khandani et al. (2010) present a machine learning technique based on genetic algorithms. When performed using actual lending data, decision trees, and SVM, up to 25% cost reductions are possible.

Yeh and Lien (2009) aimed to analyze customers' default payments in Taiwan and evaluate the predictive accuracy of six data mining methods in determining the probability of default. The research highlights the importance of estimating the actual probability of default for effective risk management rather than simply classifying clients as credible or not credible. To address this concern, a novel approach called the "sorting smoothing method" was introduced to estimate the actual probability of default. A simple linear regression model was applied, with the actual probability of default as the response variable (Y) and the predictive probability of default as the independent variable (X). The findings indicated that the ANN method demonstrated the highest

coefficient of determination and exhibited regression intercept (A) closest to zero and regression coefficient (B) closest to one. Thus, among the six data mining techniques examined, the artificial neural network method proved to be the most accurate in estimating the actual probability of default.

Nanni and Lumini (2009) observed that machine learning techniques like ensemble methods outperformed standalone techniques when applied to financial datasets from Australia, Germany, and Japan. Though several research have used modern computational technology to investigate corporate solvency,

Galindo and Tamayo (2000) used mortgage-loan data to test decision tree algorithms for default detection and assessed their outcomes to the K-nearest neighbor and probate models. They discovered that neural networks generated the most accurate outcomes with the lowest error, closely by the K-Nearest Neighbor algorithm, and that probate models underperformed as they projected with the highest inaccuracy.

## **2.9 RESEARCH GAP**

As per the above-mentioned studies, there is a lack of studies on applying machine learning algorithms that were explicitly done to predict loan defaults in microfinance banks in Pakistan. At the same time, there are studies on loan default prediction in other financial institutions and countries and different data sets. Limited research on the impact of loan default prediction models on the lending practices of microfinance banks in Pakistan. Even if a loan default prediction model is accurate, it is essential to understand how it will be used in practice and its potential impact on borrowers and lenders, which could also influence loan defaults in microfinance banks in Pakistan.

Table 2.1 Comparative Analysis and Findings

Title	Authors name	Research methodology	Findings	Research gap
1. Default Risk Prediction Using Random Forest and XGBoosting Classifier	AK Sharma, et al. (2022)	Sequential feature selection with machine learning models (Random Forest and XGBoosting Classifier)	The Random Forest and XGBoosting Classifier have an accuracy of 0.97; the Random Forest model performs better when considering ROC.	The research gap that can be identified from this study is that the research was conducted on the Lending Club data, which is limited to the US market. There is a need for more studies that investigate the use of machine learning models for predicting default risk in peer-to-peer lending platforms and financial institutions in other countries, particularly in developing countries where microfinance institutions play a vital role in providing financial services to underserved communities. Additionally, there is a need for more studies that explore the interpretability of machine learning models for lending decisions to ensure fair and ethical lending practices.
2. Rural micro-credit assessment using machine learning in a Peruvian microfinance institution	HI Condori-Alejo et al. (2021)	Determination analysis of rural variables, pre-processing data, machine learning models (Artificial Neural Network, Logistic Regression, Random Forest, Support Vector Machine, Decision Tree, k-Nearest Neighbor)	ANN model has a 93.72% assertiveness rating, Logistic Regression (86.07%), Decision Tree (88.80%), and SVM (84.44%) also performed well. Random Forest (66.35%) and k-Nearest Neighbor (65.98%) were less assertive.	The research gap identified from this study is that it was conducted on rural microfinance institutions, and it is unclear if the findings can be generalized to other microfinance institutions. Additionally, the study only used a limited number of machine learning models, and it would be beneficial to explore different models that may perform better in predicting default risk for microcredit clients. Finally, the study did not address the issue of interpretability of machine learning models, which is essential

Title	Authors name	Research methodology	Findings	Research gap
				to ensure that lending decisions are fair and ethical.
3. A machine learning approach for micro-credit scoring.	A Ampountolas et al. (2021)	Machine learning algorithms, including random forest	Using actual micro-lending data, tree-based machine learning algorithms outperformed others in sorting borrowers into different credit categories.	The study did not explore the potential challenges and limitations of using machine learning algorithms to evaluate borrower creditworthiness in developing countries with no credit histories or centralized credit databases.
4. Machine learning approach to drivers of bank lending: evidence from an emerging economy.	O Ozgur et al. (2021)	The authors contrasted a regression model with machine learning-based approaches to determine the effects of 19 macroeconomic, local, and international variables on bank loans in Turkey from 2002Q4 to 2019Q2	The authors found that ML-based algorithms have the flexibility to accept the complex nature of variables, but typical linear regression approaches could not manage the enormous dimensional datasets. Therefore, it is essential to have a reliable strategy in place for predicting debt repayment before extending credit to debtors.	While the study highlights the importance of machine learning-based approaches for predicting debt repayment, it does not provide a specific methodology or model for achieving this goal. Future research could focus on developing and testing such models in various contexts.
5. Machine Learning Techniques in Bank Credit Analysis.	FM Assef and MTA Steiner (2020).	The study utilized machine learning techniques to categorize borrowers based on the characteristics of 5432 customers of a Brazilian financial institution.	The artificial neural network's radial basis functions (ANN RBF) method offered the highest level of accuracy in the risk assessment process.	The study was limited to only one financial institution; further research is needed to validate the findings.
6. Best classification algorithms in peer-to-peer lending	P Teply and M Polena (2020)	Data sample analysis using six categorization performance evaluations and a 5-fold cross-validation procedure.	The authors evaluated 10 categorization techniques for P2P loan data obtained from Lending Club from 2009 to 2013, including 212,252 records and 23 variables. Their findings showed that	The research could benefit from a broader range of P2P loan data sources and the evaluation of more sophisticated machine learning techniques.

Title	Authors name	Research methodology	Findings	Research gap
			linear discriminant analysis, artificial neural networks, and logistic regression were the most effective algorithms, while classification, regression trees, and k-nearest neighbors were the least effective.	
7. A study on predicting loan default based on the random forest algorithm.	L. Zhu (2019)	The research methodology included using the SMOTE approach to address the issue of imbalance class in the dataset, followed by data cleaning and dimension reduction.	The experimental findings revealed that the Random Forest approach outperformed logistic regression, decision trees, and other machine learning algorithms in predicting default samples.	The research could benefit from evaluating additional machine learning techniques and using datasets from multiple P2P lending platforms.
1. An empirical comparison of machine-learning methods on bank client credit assessments.	Munkhdalai et al. (2019)	The researchers used the Survey of Consumer Finances (SCF) data and two methods for selecting the best representative features for modeling. They also used a variety of machine-learning algorithms, including an ensemble of gradient-boosted trees, support vector machines, logistic regression, and deep neural networks.	The study found that machine learning techniques outperformed the FICO credit rating system. The deep neural networks and XGBoost algorithms showed the highest AUC and accuracy.	Despite their potential benefits, the study did not investigate why lending institutions did not adopt machine-learning techniques earlier. Additionally, the study did not evaluate the performance of the machine-learning models on data outside of the SCF dataset.

Title	Authors name	Research methodology	Findings	Research gap
2. Predicting credit risk on the basis of financial and non-financial variables and data mining.	Khemakhem and Boujelbene (2018)	A case study using an actual database from a Tunisian commercial bank and employing the random over-sampling (ROS) and synthetic minority over-sampling techniques (SMOTE), confusion matrix, and receiver operating characteristic (ROC) curve used to assess performance.	The combination of advanced statistical techniques and resampling methods is beneficial for managing default rates and predicting credit outcomes.	There is a need for further research to improve the results of data mining methods used for credit risk assessment. Specifically, the study highlights the limitation of the unavailability of certain information that could have enriched the research and the results. The authors suggest increasing the sample size and including qualitative variables in the model would improve the results. Additionally, the authors propose the need for other non-parametric prediction models and feature selection techniques to enhance credit risk assessment. Thus, the research gap in this study is the need for further research to improve the effectiveness of credit risk assessment methods by including qualitative variables and exploring non-parametric prediction models and feature selection techniques.
3. Comparison of Accuracy of Support Vector Machine Model and Logistic Regression Model in Predicting Individual Loan Defaults.	Obare DM and Muraya MM (2018).	The study used a sample of 1000 loan applicants whose loans had been approved by Equity Bank in Kenya between 2006 and 2016. The variables considered were credit history, purpose of the loan, loan amount, saving account status, employment status, gender, age, security, and area of residence. The data was split	The SVM (linear kernel) model outperformed the LR model regarding accuracy and precision. The LR model showed an accuracy of 0.7727 with train data and 0.7333 with test data, while the SVM (linear kernel) model showed an accuracy of 0.8829 with train data and 0.8612 with test data. The SVM (linear kernel) model also showed a precision of 0.8785 with train data and 0.7831 with test data, while the LR model	The study did not mention any research gap. However, it can be noted that the study only compared two models and many other models for loan default prediction. Additionally, the study only used data from one bank in Kenya, and it is unclear whether the results can be generalized to other banks or countries.

Title	Authors name	Research methodology	Findings	Research gap
		into training and test data, and the train data was used to train the SVM and LR models. The performance of the models was evaluated using accuracy and precision with both train and test data.	showed a precision of 0.8440 with train data and 0.8244 with test data.	
4. A machine learning approach for micro-credit scoring.	A Ampountolas et al. (2021)	Machine learning algorithms, including random forest	Tree-based machine learning algorithms outperformed others in sorting borrowers into different credit categories using actual micro-lending data in 2013, including 212,252 records and 23 variables. Their findings showed that linear discriminant analysis, artificial neural networks, and logistic regression were the most effective algorithms, while classification, regression tree, and k nearest neighbors were the least effective.	The study did not explore the potential challenges and limitations of using machine learning algorithms for evaluating borrower creditworthiness in developing countries lacking credit histories or centralized sophisticated machine learning techniques.
5. A study on predicting loan default based on the random forest algorithm.	L. Zhu (2019)	The SMOTE approach was employed in the research methodology to address the imbalance class problem in the dataset, which was executed by data cleaning and dimension reduction.	The experimental results showed that the Random Forest approach better-predicted default samples than logistic regression, decision trees, and other machine learning algorithms.	The research could benefit from evaluating additional machine learning techniques and using datasets from multiple P2P lending platforms.
6. Machine learning approach to drivers of bank lending:	O Ozgur et al. (2021)	The authors compared a regression model with machine learning-based approaches to	The scientists discovered that whereas traditional linear regression techniques could not handle the enormous	While the study highlights the importance of machine learning-based approaches for

Title	Authors name	Research methodology	Findings	Research gap
evidence from an emerging economy.		ascertain the effects of 19 macroeconomic, regional, and global variables on bank loans in Turkey from 2002Q4 to 2019Q2.	dimensional datasets, ML-based algorithms have the flexibility to embrace the complicated nature of variables. Therefore, it is essential to have a reliable strategy in place for predicting debt repayment before extending credit to debtors.	predicting debt repayment, it does not provide a specific Methodology or model for achieving this goal. Future research could focus on developing and testing such models in various contexts.
7. Best classification algorithms in peer-to-peer lending	P Teply and M Polena (2020)	Data sample analysis using six categorization performance evaluations and a 5-fold cross-validation procedure.	The authors evaluated 10 categorization techniques for P2P loan data obtained from Lending Club from 2009 to	The research could benefit from a broader range of P2P loan data sources and the evaluation of more credit databases.

## 2.10 SUMMARY

The literature review chapter provides an overview of microfinance in Pakistan, an essential financial service that aims to assist low-income individuals and small enterprises in obtaining financial services they otherwise would not have access to. Over 60 microfinance institutions operate in Pakistan and offer financial products like microcredit, micro insurance, and micro-savings. Although the government and the State Bank of Pakistan have made considerable efforts to encourage the growth of the microfinance sector in the country, there are still challenges facing the industry, including limited outreach, high operating costs, and limited access to funding, lack of innovation in products, regulatory challenges, and limited financial literacy among the low-income population in Pakistan. Furthermore, it discusses the research on applying machine learning techniques to predict loan risk.

Moreover, it discusses the challenges microfinance institutions (MFIs) face, particularly regarding MFI owners' inadequate knowledge of the industry and challenges in Islamic microfinance. Regarding the lack of sufficient knowledge, it is observed that many business owners operate their businesses without proper supervision or oversight, leading to significant setbacks in the MFI community, primarily due to ineffective governance and administration. In terms of Islamic microfinance, the text highlights the lack of supply of funds, legal barriers, mounting operating costs, and grass-root level awareness as significant challenges, along with the inability of MFIs to fully monitor clients in a mudaraba microfinancing, which makes them an easy target of clients to manipulate documents and under-report profits. The text also notes the challenges Islamic MFIs face in relying on a single Islamic mode of financing, high transaction costs due to information asymmetry issues, and inadequate infrastructure that makes collecting data on potential consumers challenging.

In addition, it discusses loan risk in MFIs and strategies for managing it. Loan risk refers to the possibility that loans will lose value and become worthless, which could cause an MFI to fail. To mitigate this risk, MFIs must allocate reserves for potential loan losses in their financial records, with higher provisions reflecting higher levels of risk. Credit scoring is a widely used method for assessing credit risk, which involves evaluating a borrower's financial strength and the likelihood of default. Loan officers at MFIs commonly employ the "five Cs" (character, capacity, capital, collateral,

and conditions) as criteria to assess the eligibility of borrowers for loans. However, this method is limited by the inability to process many applications, prompting the development of scorecards that create a score to evaluate the risk associated with lending. The passage also discusses financial credit risk assessment methods, which aim to estimate the likelihood of a borrower going insolvent or conforming to a risk category. Some methods encompass statistical and optimization models like linear discriminant analysis, logistic regression analysis, and artificial intelligence approaches such as artificial neural networks and support vector machines.

Additionally, it explores various empirical research conducted in financial institutions utilizing machine learning techniques to assess credit risk. For instance, Yuan et al. (2022) introduced a novel approach employing low-rank representation and discriminant analysis to evaluate default risk among borrowers in peer-to-peer online lending networks. Similarly, Sharma et al. (2023) employed sequential feature selection and machine learning models to predict default risk in a peer-to-peer lending platform. Machado and Karray (2022) used hybrid machine learning algorithms to predict the credit scores of commercial clients. Condori-Alejo et al. (2021) suggested a model for assessing rural microcredits using various machine learning models, and Kumar et al. (2021) analyzed the limitations of AI-ML-based credit scoring techniques. Ampountolas et al. (2021) evaluated the performance of numerous machine learning algorithms in classifying borrowers into specific credit categories utilizing actual micro-lending data.

Additionally, It includes multiple studies undertaken to construct credit risk assessment models for borrowers. The studies utilize various decision support techniques and algorithms, such as logistic regression, artificial neural networks, support vector machines, and ensemble algorithms. These models consider multiple factors, including socio-economic factors, behavioral characteristics of clients, and politico-economic issues. The authors of the studies also highlight the importance of selecting appropriate factors and algorithms to ensure accurate credit risk assessment.

Furthermore, it summarizes several studies that examine the application of machine learning to credit risk assessment. The research studies utilize various methods such as support vector machines, decision trees, artificial neural networks, and ensemble learning. The studies demonstrate that machine learning algorithms

effectively predict credit risk and can help financial institutions anticipate unsatisfactory clients. Key indicators of potential default include profitability ratios, repayment capability, solvency, credit history duration, collateral, company size, loan quantity, ownership composition, and the length of the relationship with the corporate bank. However, the studies also note that algorithms are sensitive to class disparity, and data imbalance can affect their performance. The chapter suggests that as the availability of high-dimensional data expands, machine learning and data mining-based risk assessment algorithms will become increasingly important in this field.

Lastly, it discusses various studies and models developed to tackle the banking sector's challenge of default risk assessment. These studies used multiple techniques and models, such as neural networks, logistic regression, decision trees, random forests, and support vector machines. The analyses revealed that the random forest algorithm was more accurate in forecasting defaulter defaults than the widely used FICO credit scores method. Other studies suggested using multilayer perceptron (MLP) models tuned with a second-order algorithm in the R statistical system, to reduce the misclassification costs. Such strategies have the purpose to increase the precision of creditworthiness forecasts and lead towards a healthier and more profitable loan portfolio.

## **CHAPTER THREE**

### **RESEARCH METHODOLOGY**

#### **3.1 INTRODUCTION**

This chapter presents an overview of the research methodology employed to accomplish the objectives. The methodology is divided into five main sections: research design, data collection process, data analysis, machine learning models, and evaluation metrics.

#### **3.2 RESEARCH DESIGN**

Research design incorporates the model upon which the arrangement of the main components of the study such as the research problem, literature review, data collection, data analysis, and interpretation has to be made. In the current research, explanatory sequential mixed-methods research with a quantitative-dominant design is embraced. In this way more nuanced and practically oriented insight into loan default prediction can be achieved as the combination of statistical modeling and expert interpretation is attained.

The research starts with a quantitative part, in which the dataset of the attributes related to borrowers and loans in Pakistani microfinance banks is utilized to train supervised machine learning algorithms. These models should single out the main predictors related to default and measure their performance by such parameters as accuracy, precision, recall, F1-score.

After this, the structured interviews did not occur at the quantitative stage of study design, because research design focused on the investigation of the same data analysis in the area of loan procedures in participating microfinance banks. In order to ensure that the predictions features identified by the machine learning models are relevant in the context and interpretable, clarification was made through mails to loan officers. This strategy enabled identification of specific feedbacks regarding the data-derived results and curtailed interference in working schedules. Whereas not as interactive as formal interviews, such an approach was able to offer adequate domain

insight within the realm and boundaries of the research. This step promotes the credibility, contextual relevance, and interpretability of machine learning.

This design is guided by Creswell and Plano Clark (2018), who point out such parameters as explanatory sequential mixed-methods being highly relevant when quantitative findings must be elaborated by qualitative research. This combination of the two approaches guarantees a better comprehensive insight into the dynamics of loan default, which is the key to developing data-driven, yet operationally viable, credit risk solutions in microfinance.

### **3.3 QUANTITATIVE RESEARCH METHOD**

The study is conducted using a quantitative method because this is the most emphatic part of an explanatory, sequential mixed-methods approach. The quantitative stage implies model construction with the help of supervised machine learning as it employs a detailed anonymized collection of 171,844 loan data of Pakistani microfinance Banks (MFBs).

Instead of the experiment to test the particular hypotheses, the aim is to find out the patterns and associations within the data to define the major factors linked with the loan defaults. Logistic regression, decision trees, random forest, and gradient boosting algorithms are used to mine information that can inform data-driven credit risk evaluation as applied to microfinances.

#### **3.3.1 Data Collection**

The data set was available over 2015-2023 and contains demographics of borrowers, loan traits, repayment patterns. It was supplied by partner MFBs on a formal data-sharing agreement, and all data records were de-anonymized.

Census sampling strategy was adopted, whereby the whole data was not picked but all data was employed. The original rate of completeness was 89 percent, which increased to 92.6 percent after a routine processing procedure including imputation, encoding, and duplicate resolution. Some issues like the difference in formatting and

incomplete behavioral data on previous years were addressed by normalization plus field alignment.

In order to confirm the ML results, opinions of loan officers were enlisted once model was trained, with regard to the prioritization of foreseeable variables. This schedule eliminated the confirmation bias and offered practical and professional-level verification of the information-oriented results.

### **3.3.2 Sample Size**

This work had utilized census sampling strategy whereby the complete dataset containing 171,844 anonymized loan data (obtained by microfinance banks in Pakistan) was utilized. No sample was needed to be taken, no sample size estimation or method of selection was necessary since the complete dataset was accessible in partner institutions. The abundance of the data allows making the machine learning models introduced in this paper more generalizable and robust. Determining the sample size involves calculating the minimum number of participants necessary to obtain statistically significant results with satisfactory confidence. The chosen sample size is favored to attain the intended level of accuracy or significance for the collected data. It is crucial to determine the optimal sample size before collecting the data to prevent the sample from being too small after processing it (Thietart, 2014).

The sample size for a loan default prediction model would depend on the number of loans made by the microfinance banks in Pakistan and the number of loans that have defaulted. A larger sample size would generally provide more accurate predictions, but a relatively small sample size can help develop a predictive model. It is important to note that the specific variables and sample size used in any particular loan default prediction model can vary depending on the study and the data available.

### **3.3.3 Unit of Analysis**

The individual loan account is the unit of analysis of the research. The dataset has multiple examples of such loans by microfinance banks and each example has many

characteristics of the loan like borrower demographics, loan features and other repayment data. This unit of analysis allows the formulation of the predicting models, which evaluation of default risk is provided at the loan level, which gives glimpse into the factors which can be evaluated at a detailed level.

### **3.3.4 Data Analysis**

The quantitative analysis was organized in a number of main stages which were the preprocessing of the data, feature engineering, algorithm selection, training, evaluation, and validation of the model. We aimed at building a powerful machine learning model of predicting loans defaults based on past loan data in the microfinance institutions in Pakistan.

#### ***3.3.4.1 Data Preprocessing***

It included 171,844 loan records with its initial percentage of completeness of 89 and grew to 92.6 after cleaning. The following pre-treatments were used:

- **Missing Values Imputation:** We randomly assigned numbers fields with the mean or mode, and the mode-mode imputation to categorical variables, which was proposed by Mueller & Massaron (2016).
- **Outlier Detection and Treatment:** Outliers were detected with focus on the domain-specific thresholds and statistical tools as recommended in the work by Kwak & Kim (2017).

**Encoding:** These required categorical variables were encoded as one-hot and label encoding (Cornelissen, 2018).

- **Normalization:** All the numerical features were put on a scale (Z score) using normalization (Zhang & Qi, 2002).

To select features three approaches were used:

- Such techniques as filter methods (e.g., correlation analysis),
- Wrapper techniques (e.g. Recursive Feature Elimination),

i.e., embedded techniques (i.e., tree-based feature importance scores).

Despite being considered (Koutanaei et al., 2015), PCA was not used because of the decreased level of interpretability that is vital to making decisions in predicting the level of credit risks.

#### ***3.3.4.2 Handling Class Imbalance***

There were about 18 percent defaults in the data. In order to solve this imbalance:

- SYN (Synthetic Minority Oversampling Technique) has been applied according to Chawla et al. (2004) and Brownlee (2019): it was only used on the training set.
- Parameters: `sampling_strategy='minority'`, `[cls] k_neighbors=5`.
- No Tomek links or ENN was applied, which proved to be of little use during the first tests.
- The imbalanced class proportion was estimated to around 1:1 even after resampling thus enhancing model sensitivity to defaulters.

#### ***3.3.4.3 Model Training and Evaluation***

- Train/Test Split: A stratified split ratio 70/30 was carried out so that the distribution of classes is not violated.
- Cross-Validation: stratified 5-fold cross-validation was used to give stability across splits (Dahiya et al., 2017).
- Hyperparameter Tuning: This is done through grid search on hyper parameters as `max_depth`, `learning_rate` and `n_estimators`.
- Metrics of Performance: The overall performance was compared using accuracy, precision, recall, F1-score, and AUC-ROC.

#### ***3.3.4.4 Interpretation of Results***

SHAP values and feature importance scores were utilized to make inferences on model outputs. The validation of the best features found by the ML models was done among

the loan officers during the training of the models, as our method of minimizing the confirmation bias and ensuring the actual applicability of the results to the real world (Laborda & Ryoo, 2021).

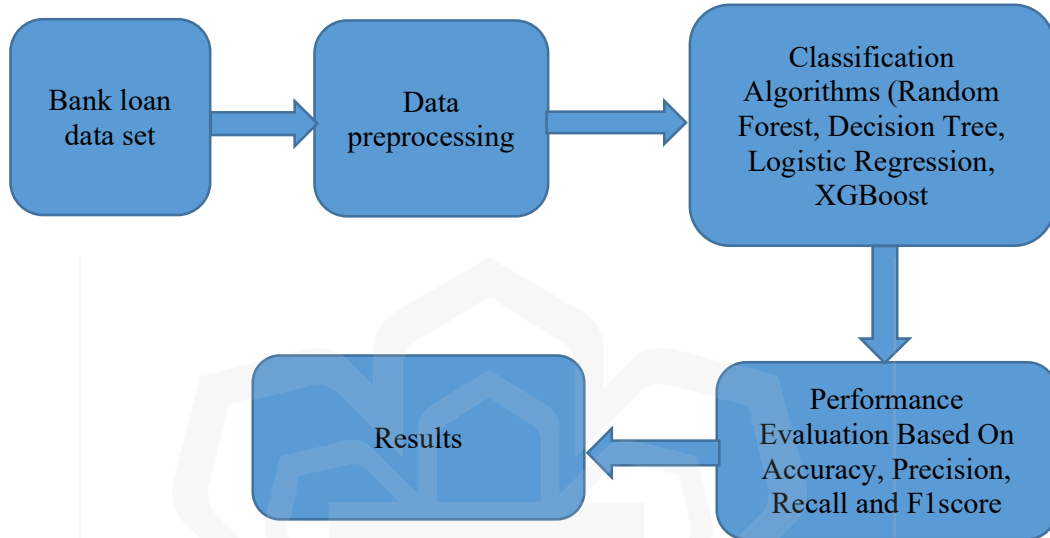


Figure 3.1 Proposed Model

### 3.4 QUALITATIVE RESEARCH METHOD

The experimental sequential mixed-methods design embraced by this research presupposed that the qualitative analysis was held after the quantitative analysis based on machine learning and focused on confirming the main predictive factors. Instead of undertaking the common interviews, practitioner feedback was gathered through structured emails to the loan officers in the Pakistani microfinance Banks. This method made it possible to include the realistic view of operative scenarios in the model validation cycle and directly covered Objective 4 of the study.

The strategy is in line with the validation approaches in empirical studies, where expert elicitation is applied to evaluate the practical essence and congruency of the statistically found outcomes (Creswell & Plano Clark, 2018).

### 3.4.1 Data Collection

The microfinance practitioners were selected and a systematic email was sent to them. The message held:

1. An introduction of the research briefly.
2. An introduction to the list of the most important predictive features that were found based on the machine learning models.
3. Four specific questions, which are open-ended to request a response of the expert on:
  - The practicality and significance of such factors.
  - Any neglected or lost variables.
  - Comparing these factors to practical lending.
  - Some recommendations concerning the enhancement of the predictive model.

It was this email that was the qualitative data collection research tool. The respondents were taken through a survey which called upon them to express their opinion in writing, this was then analyzed to determine where the opinion differed or converged with that as offered by the model.

### 3.4.2 Sampling Techniques

The loan officers, credit analysts, and risk managers in Pakistan were the sampling units focused on using a purposive expert sampling. The participants were chosen by professional competence and the familiarity with the operation of the underwriting of loans and evaluation of the default risk. Three (03) responses were obtained including practitioners who had different years of experience in microfinance lending.

### **3.4.3 Respondents and Research Instruments**

The content of the email was drafted to ensure it corresponded with the research purpose as well as to solicit responses of targeted practitioners. Prior to its such wider circulation it was pilot-tested informally with three microfinance officer to see its clarity and relevancy. The four open-ended questions were content-coded according to the themes of relevance, completeness, practical alignment and model improvement, and in such a way, that the feedback response could directly be used in validating the predictive variables.

### **3.4.4 Data Analysis**

Thematic content analysis was applied to analyse the responses. Main phrases and concepts were clustered on such themes as:

1. Concurrence of model-considered variables.
2. Ideas of lacking or context-specific characteristics.
3. Operational issues of automated model implementation.
4. Discrepancy of statistical output and life on the ground by the practitioner.

This thematic analysis facilitated the process of triangulating the quantitative findings of the study with insights at a field-level, which reinforces the validity of the findings with respect to their situational context, and makes sure that the eventual policy recommendations are evidence-based not only on the data science but lending practice, as well.

## **3.5 MACHINE LEARNING**

In consonance with Sandhu (2018), machine learning is the discipline that empowers computers to learn and make predictions without explicit programming. It falls under the umbrella of artificial intelligence and aims to address problems based on historical or past instances. Unlike other artificial intelligence applications, machine learning

involves uncovering hidden patterns in data through data mining techniques. Models utilize these patterns for categorization or forecasting related to the problem ( Libbrecht, Noble, 2015).

The goal of machine learning is to enable algorithms to learn and make predictions based on feedback and data. Machine learning aims to create methods and tools that can accurately and timely automate the time-consuming human workforce. Machine learning-based techniques accomplish this objective by identifying recurring patterns in a training data sample, allowing hypotheses about the data domain. As an artificial intelligence discipline, the effectiveness of a machine learning-based system must increase as it gains experience or data (Michiel, 2018).

The two main categories of machine learning techniques are unsupervised and supervised machine learning. The degree to which the examples presented in the learning process are labeled or not makes up the fundamental difference between these two learning methods. Both supervised and unsupervised machine learning have been heavily utilized in default risk analysis (Bao et al., 2019).

### **3.5.1 Supervised Machine Learning**

Supervised learning predicts an unknown mapping between inputs and outputs using available samples where the outputs are labeled, such as in classification and regression tasks (Naqa, Murphy, 2015). Supervised machine learning algorithms are utilized within credit scoring models to establish the relationship between customer characteristics and credit default risk. These algorithms further facilitate the prediction of default categorization, typically represented in binary form. Extensive research has demonstrated the high predictive accuracy of supervised machine learning algorithms in credit-scoring models (Ben-David & Frank, 2009; Sohn et al., 2016; Twala, 2010).

According to Kotsiantis et al. (2007), supervised machine learning involves using input variables and employing specified output variables. These algorithms aim to predict and categorize a predefined attribute and their accuracy, misclassification, and other performance indicators on the successful prediction or categorization of the predefined attribute. It is important to note that once the algorithm achieves a certain

performance standard, it stops learning. In practical terms, supervised algorithms initially analyze training data to create functions that map new attribute instances, as Libbrecht and Noble (2015) outlined. The algorithms require predefined maximum values for the desired outcome and performance levels. Based on the machine learning approach, it has been observed that utilizing a training subset of approximately 66% is reasonable and yields the expected results without requiring additional computational resources.

Machine learning uses learning algorithms to solve problems. The algorithms adjust the outputs in response to the inputs to resolve the situation in light of the information they have learned. As Mueller & Massaron (2016) pointed out, the algorithms change how the computer interprets the data. The aim of this part is a detailed description of the research algorithms. The four learning algorithms are XGBoost, classification trees, random forests, and logistic regression.

### ***3.5.1.1 Logistic Regression***

Classification or regression issues are two different problems that might arise with supervised or predictive learning. Supervised learning aims to acquire knowledge from a supervisor by effectively learning the mapping between input and output values (Alpaydin, 2014; Murphy, 2012). In binary classification problems, logistic regression is a suitable method (Kim et al., 2018; Siddique et al., 2017). Logistic regression, as highlighted by Hosmer, Lemeshow, and Sturdivant (2013), is widely recognized as the preferred approach for establishing the relationship between the dependent (outcome or response) variable and one or more independent (predictor or explanatory) variables. Loans can be categorized as defaulted or not defaulted, which corresponds to the logistic model's distribution.

Based on Hosmer et al. (2013) work, the logistic regression model differs from linear regression because it is specifically designed for binary outcome variables. Consequently, logistic regression is not used as a regression technique, as its name may suggest, but rather as a classification method. In binary problems, the outcome variable can take two distinct values: true or false, typically encoded as 1 and 0. Examples of

binary outcome variables in datasets include the ability to repay a debt or default, win or lose, or be ill or well.

### ***3.5.1.2 Decision Trees***

The decision tree classifier is highly popular due to its inherent logic-based structure. Kobacoff (2015) indicates that decision trees are commonly used to predict binary outcome variables. They are constructed by making binary splits based on predictor variables, forming a tree-like structure. This tree is then used to divide incoming observations into two distinct groups based on the applied splits. Hence, decision trees are applied and named for their representation as a tree structure.

However, classification trees forecast a qualitative response, and regression trees are quantitative. The classification tree's prediction is based on the class that happens the most frequently. Regression trees define the projected class as the average number of observations that belong to an identical terminal node.

In the study conducted by Chizi et al. (2009), a detailed explanation of decision tree classifiers was presented. A decision tree comprises interconnected nodes that collectively form a rooted tree structure, with a single root node serving as the starting point. Nodes with outgoing edges are internal nodes, and all subsequent nodes have a single incoming edge. Each internal node divides the dataset based on a specific logic, often related to the importance of a particular attribute. On the other hand, leaves are nodes that receive incoming edges but do not have outgoing edges.

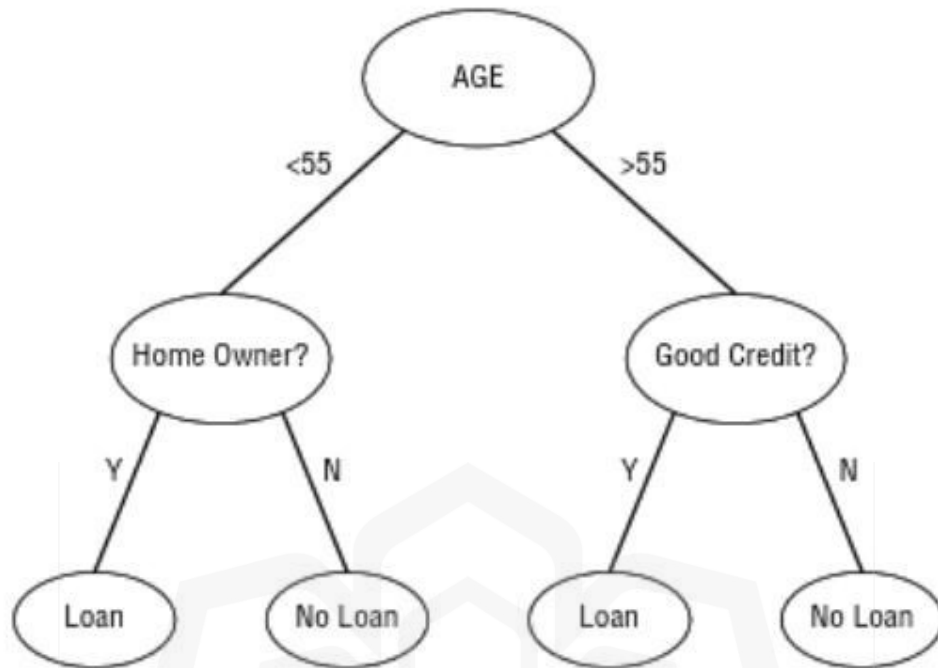


Figure 3.2 Classification Tree (Bell, 2014)

The decision nodes in the tree mentioned in Figure 3.2 are "Home owner?" and "Good credit?" and the root node is "Age." The leaves that hold the outcome value are located at the bottom. The tree first determines whether the borrower is older than or less than 55. If the applicant is younger than 55, the tree determines whether or not they are a home owner. If not, a loan is not issued; otherwise, a loan is. Those over 55 make the same choice based on good or bad credit.

### 3.5.1.3 Random Forest

The random forest algorithm is a supervised learning technique for classification and regression tasks. It operates by generating numerous decision trees during the training phase and subsequently predicting the class representing the individual trees' mode. The core concept of this classifier involves training multiple decision trees and utilizing them collectively to classify data, as explained by Breiman (2001) and Kamath & Kamat (2016).

During the training of each decision tree in a random forest, a subset of the training data is randomly selected with replacement. This process is similar to training a regular decision tree, except for one difference. At each split in the tree, a random sample of features gets selected, and the split is on one of those selected features. The number of features to be drawn at each split is usually determined by taking the square root of the total number of available features, as Hastie et al. (2009) described. Random feature selection has to lessen the association among the individual trees.

According to research conducted by Yaman et al. in 2018, the random forest algorithm is a learning technique that can solve classification and regression tasks. Additionally, random forests can separate crucial classification characteristics, avoid overfitting, and are appropriate for large amounts of data with significant noise (Deng et al., 2018).

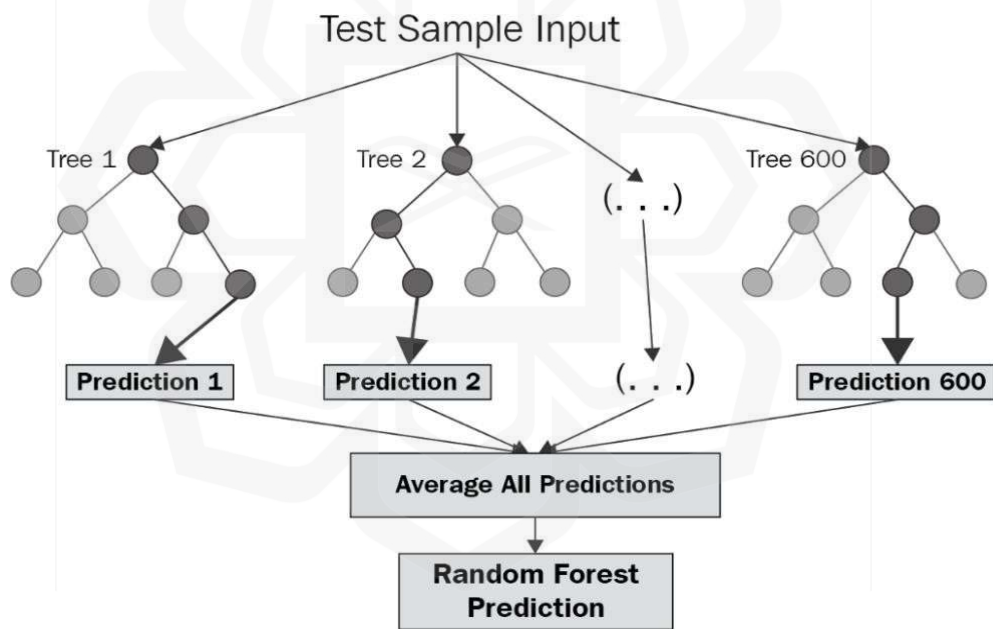


Figure 3.3 Random Forest (Chakure, 2019)

#### 3.5.1.4 Extreme Gradient Boosting

The tree-based method, Extreme Gradient Boosting (XGBoost), has also been popular in recent years. As Brownlee (2016) indicated, it supports regression and classification issues and are incredibly effective with various machine-learning challenges. For

instance, XGBoost-based techniques have dominated Kaggle competitions. XGBoost scales to billions of examples and is more than ten times faster than other systems (Chen and Guestrin, 2016).

In contrast to Kaggle competitions, many studies on loan default forecasts use XGBoost. Odegua (2020) and Xia et al. (2017) used XGBoost in their research. R Odegua (2020) used Python and only XGBoost to research loan default predictions and suggests using grid search to identify the ideal parameters and 5-fold cross-validation to lessen bias.

According to Chen & Guestrin (2016), the method used by XGBoost, tree boosting, is highly effective. As per Gandhi (2018), boosting is a strategy for transforming weak learners into strong ones. A poor learner might have a 0.5 error rate, where the predictive power would equal a coin toss. An excellent learner makes no mistakes, having an error rate of 0.0. When weak learners are joined and decided upon, the result is a group of strong learners.

## **3.6 EVALUATION METRICS FOR CLASSIFICATION MODEL**

### **3.6.1 Performance Metrics**

The data used in the study is a loan data which has a high degree of class imbalance. In this section, model evaluation statistics are appropriate in such imbalanced dataset such as the ROC curve, confusion matrix, accuracy, recall, precision, and F1-score. Sensitivity and specificity are also introduced which help in the digestion of ROC curve and confusion matrix.

Such evaluation metrics are used to gauge not to optimize the performance of machine learning models. They can be used to compare two or more algorithms, and they do so by giving quantitative measures of predictive quality. In some works these metrics have been used not to directly tune the model but to guide model selection and tuning. In their study, Anand et al. (2022) applied the presence of Confusion Matrix, Accuracy, F1-score, Recall, Precision, ROC-AUC, and Feature Importance as the main performance indicators of their classification research.

Since the dependent variable used in this research paper is binary (i.e. default or non-default) these measures of evaluation are ideally suited to classification algorithms. Although accuracy is used frequently it is not always the most important metric, precision, the recall, and F1-score of the accuracy are particularly significant when there is a class imbalance. Such measures enable problems with model performance (false positives, false negatives, etc.) to be found and facilitate the preference of models which perform better in practice.

### 3.6.2 Confusion Matrix

A confusion matrix is a widely utilized table that summarizes the efficiency of a classification model or classifier on a particular set of test data where the true values are known. It is an effective measure for assessing model performance in binary response conditions. The confusion matrix distinguishes between positives, representing observed default cases, and negatives, which correspond to non-default cases. True positives (TP) are instances where the model correctly predicts defaulting customers as defaulters, while true negatives (TN) indicate accurate predictions of non-defaulting customers. On the other hand, false positives (FP) occur when non-defaulting customers are incorrectly predicted as defaulters and false negatives (FN) arise when defaulted customers are wrongly predicted as non-defaulters.

We can determine various metrics using a confusion matrix, with accuracy being one of the more prevalent indicators. The researchers can calculate accuracy by the sum of true positive and true negative values divided by the number of predictions and classifications. It represents the ratio of correctly classified observations out of the total number of observations, and numerous studies have used it, including those conducted by (Erie, 2019) and Wang et al. 2020).

It is described mathematically as (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

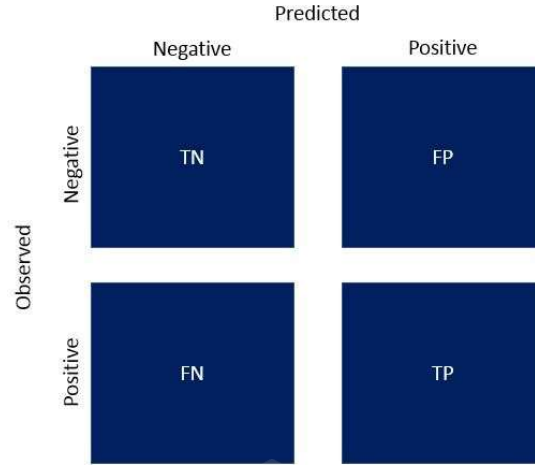


Figure 3.4 Confusion matrix (Granström and Abrahamsson, 2019)

Using accuracy as a metric to evaluate imbalanced data can be problematic. When the majority class dominates the dataset, and all predictions are assigned to that class, it is possible to achieve a high accuracy, such as 99%, if the dataset consists of 99% of that class. However, this can be misleading and not comprehensively evaluate the model's performance. In such cases, specificity becomes a more relevant metric and can be computed by (2). Specificity measures the proportion of true negatives (TN) correctly identified by the model relative to the sum of false positives (FP) and true negatives. It focuses on the model's ability to accurately classify the negative class, which is particularly important in this context.

$$Specificity = \frac{TN}{FP + TN} \quad (2)$$

### 3.6.3 Precision

Accuracy of positive predictions is measured by precision, it denotes the fraction of accurately predicted defaulters among all others predicted to be defaulters."Zia Ereiz (2019). Sensitivity measures the number of defaulting clients that the model captures. The mathematical representation of sensitivity and precision is (3) and (4):

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

### 3.6.4 Recall

Recall is the ratio of true positives over total positives as (5).

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

### 3.6.5 F1 Score

The balanced average of Precision and recall is called the F1 score, sometimes called a balanced F Score, and can be represented by (6).

$$F = 2 * \frac{precision.Recall}{precision+ Recall} \quad (6)$$

## 3.7 SUMMARY

The study aims to analyze the relationship between loan default and microfinance bank data in Pakistan using a quantitative research design that employs a predictive modeling approach. The data used in the study includes loan applications, borrower information, loan amounts, loan repayment history, and other relevant financial information collected over five years from 2015 to 2023. Data preparation, feature engineering, model selection, model training, and model evaluation are all steps in the data analysis process. Ensuring the privacy and confidentiality of the data used is one of the ethical considerations. The various steps involved in data preprocessing include data cleaning, feature selection, and handling unbalanced data. The section describes logistic regression, classification trees, random forests, and XGBoost as four learning algorithms. The evaluation metrics employed include the ROC curve, confusion matrix, accuracy, recall, precision, and F1 score, with the AUC when comparing the models. The importance of understanding the data when addressing missing values and outliers is emphasized, and the genetic algorithm, gain ratio, and principal component analysis

techniques have outperformed others in feature selection strategies for credit score prediction.



# CHAPTER FOUR

## RESULTS AND FINDINGS

### 4.1 INTRODUCTION

This chapter illustrates the performance of different machine learning algorithms to predict loan default in microfinance industry of Pakistan. The research organization is based on four major objectives, i.e. (1) to determine the variables of borrower and loan which affect the defaults; (2) to examine and compare the effectiveness of the chosen machine-learning models, (3) to identify which are the most important variables impacting the defaults; and (4) to confirm the practical usefulness of such variables using the experiences of bank loan officers. This knowledge could help in reducing credit risks especially considering the importance of loan repayment to financial well-being of the microfinance institutions (Ahmed, Tahir, & Ahmad, 2023).

To tackle these goals, the evaluation metrics that have been widely accepted, namely, accuracy, precision, recall and F1-score were used to evaluate the work of four predictive models, namely, Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Gradient Boosting (GB). There was also the use of the confusion matrices in analyzing model performance by giving a fine-grained look at how the models predict the actual observations including true and false positive rates (Odegua, 2020).

Data on two Pakistan microfinance banks was used in order to answer the first objective. This was a set of varying borrower demographics and loan characteristics directly taken out of the institutional credit evaluation modifying systems. This robustness of these features gives practical understanding of indicators that should not be neglected; hence it gives a good base as far as training and interpreting a model is concerned.

The second goal was comparisons of the performance of the chosen machine learning algorithms to locate the most effective one to predict loan defaults. Such comparison was also necessary to the make-sure that the selected approach will be not only predictively accurate but also practically useful.

The third goal was aimed at mining the most significant attributes that influence charge-offs on the loan. The identification of these key predictors makes decision-making in the lending institutions to be more focused and model interpretability be easier to understand.

Finally, to ascertain that features ranked on top were practical, the researchers wanted to establish validation of domain experts. Although there was no formal interview conducted, clarification and feedback were acquired through email contact with loan officers in the institutions in which they participated. In this approach, contextual validation was achieved without intermittent operational interruptions.

The results of this section provide a basis of future studies in the field of credit risk analytics and can provide practical guidance to financial institutions that want to improve the way they deal with their loan portfolio (Ting, Mia, Hossain, & Wah, 2024). Machine learning techniques presented a high potential in default predictions and minimization of financial losses, etc., and will make the microfinance industry more inclusive and resilient.

#### **4.2 OBJECTIVE-01: TO IDENTIFY THE KEY BORROWER AND LOAN RELATED FACTORS THAT INFLUENCE LOAN DEFAULT IN MICROFINANCE BANKS IN PAKISTAN**

These borrower characteristics may include key demographic factors such as age, income level, credit history, employment status, and education level, collectively providing a comprehensive profile of the debtor's financial behavior and reliability. The loan characteristics examined include crucial elements such as loan amount, interest rate, repayment terms, and purpose of the loan and repayment schedule, all of which significantly determine the risk associated with each loan. These factors were based on these institutions' internal credit evaluation criteria, which may encompass aspects such as debt-to-income ratio, previous repayment behavior, collateral offered, and overall financial stability of the borrower. This approach not only aids in identifying potential risk but also offers real insights into the aspects of loan default that are considered high risk, thereby enabling more informed decision-making for lenders.

This research aims to identify the critical elements influencing loan default within MFBs in Pakistan. The study intends to enhance loan evaluation processes and develop more accurate predictive models by leveraging real-world data from two distinct institutions. Ultimately, the goal is to mitigate default risks, fostering financial stability within the sector.

Table 4.2 Factors Influencing Loan Default Prediction

<b>Factor</b>	<b>Characteristics</b>
1. Loan Amount (loan_amnt)	The total sum borrowed typically ranges from PKR 60000 to 100000
2. Loan Term (term)	The duration of repayment is commonly set at 12 to 36 months.
3. Interest Rate (int_rate)	The cost of the loan is expressed as a percentage, generally between 18% and 35%.
4. Installment Size	Monthly payment amount, which can range from pkr 3000 to 5000
5. Grade	Classification of the loan based on risk assessment, often represented by letters (e.g., A, B, C).
6. Sub-Grade	More detailed classification within each grade, providing further granularity in risk assessment.
7. Employment Title (emp_title)	Job title of the borrower, reflecting their professional background.
8. Employment Length (emp_length)	Duration of employment in years, indicating job stability (e.g., less than 1 year to over 10 years).
9. Home Ownership	Status of home ownership, categorized as own, mortgage, or rent.
10. Annual Income (annual_inc)	The total yearly income of the borrower influences repayment capacity.
11. Verification Status	Indicates whether the borrower's income and employment have been verified (e.g., verified, not verified).

Factor		Characteristics
12. Issue Date (issue_d)		The date when the loan was issued is relevant for tracking loan performance over time.
13. Loan Status		Current status of the loan (e.g., Fully Paid, Charged Off), indicating repayment outcome.
14. Purpose of the Loan (purpose)		Intended use of the loan funds (e.g., car, business, education, wedding).
15. Debt-to-Income Ratio (dti)		Ratio comparing monthly debt payments to income; lower values indicate better financial health.
16. Earliest Credit Line (earliest_cr_line)		The year when the borrower opened their first credit account reflects the length of the credit history.
17. Number of Open Accounts (open_acc)		Total number of active credit accounts held by the borrower.
18. Total Number of Public Records (pub_rec)		The number of adverse public records (e.g., bankruptcies), typically 0 or 1, indicates lower risk.
19. Total Credit Accounts (total_acc)		The borrower's overall number of credit accounts ranges from 3 to 58.

Table 4.1 deals with the factors influencing loan default prediction in Pakistani microfinance banks and provides a comprehensive overview of the various elements that affect borrowers' ability to repay loans. Each factor, ranging from loan amount and terms to borrowers' characteristics, such as income and credit history, is crucial in assessing default risk. For example, higher loan amounts and longer repayment terms can increase the financial burden on borrowers, while a lower debt-to-income ratio typically indicates better financial health. Additionally, factors like employment stability and the purpose of the loan further inform lenders about potential repayment challenges.

The significance of this objective lies in its potential to enhance the lending practices of microfinance institutions in Pakistan. By understanding and analyzing these factors, institutions can develop more effective risk assessment models that improve their loan evaluation processes and reduce the likelihood of defaults, which is particularly significant in microfinance, where maintaining economic stability is essential for lenders and borrowers. Eventually, this research aims to foster a more sustainable microfinance atmosphere, ensuring that loans are granted responsibly and that borrowers are supported in their repayment efforts, thereby paying for the region's overall economic growth.

#### **4.2.1 Data Collection**

The loan officers of the two microfinance banks meticulously maintained the data from the loan records. Rather than conducting formal interviews, insights were derived from the practical knowledge and established criteria employed by these loan officers during the loan evaluation process. The dataset encompasses a diverse array of variables, including loan characteristics, borrowers demographic and credit history, making it essential to understand the dynamics surrounding loan default and repayments comprehensively. The internal policies of the banks, combined with the expertise of the loan officers, significantly influenced the selection of factors that constituted this dataset. Notably, it included both 'Fully paid' and 'Charge Off' loans, facilitating a detailed analysis of successful repayment and defaults. This comparative analysis has proven invaluable in identifying the key factors that differentiate between repaid loans and those resulting in default.

#### **4.2.2 Description of Considered Factors**

The factors collected from microfinance banks are grouped into three categories: loan-specific factors, borrowers-related factors, and credit history and financial profile. Each of the categories captures different dimension that may influence a borrower's ability to meet loan repayment obligations:

### 4.2.3 Loan Specific Variable

- **Loan Amount:** Large sums may increase the like hood of default due to heightened repayment pressure.
- **Loan Terms:** Extended repayment duration introduced uncertainties that can elevate risk.
- **Interest Rate:** Higher rates increase the cost of loans, adding to borrowers' financial strains.
- **Installment Size:** Large installments can lead to financial stress, complicating timely repayments.
- **Loan Purpose:** Loans intended for income generation tend to be less risky than those meant for consumption.
- **Borrower Characteristic:** These reflect borrowers' financial and personal backgrounds, significantly influencing their repayment capacity.

Following this, Table 4. 2 represents the loan-specific variables of the debtors:

Table 4.3 Loan-Specific Variables

Factor	Range/Values
Loan Amount	Pkr 60000 - 100000
Loan Term	12 to 36 months
Interest Rate	18% - 35%
Installment Size	Pkr 3000 - 5000
Purpose of the Loan	Car, Business, Education, Wedding

Table 4.2 represents the loan-specific factors that reveal critical elements that significantly impact the like hood of loan default in microfinance institutions. The loan amount ranges from pkr 60000 to 100000, indicating that large sums can increase the repayment burden, especially for borrowers with limited financial resources, thereby heightening default risk. Loan terms of 12 to 36 months introduced uncertainties over time, which can further elevate risk. The interest rate from 18% to 35% reflects the cost of borrowing; higher rates increase the expense of servicing loans, contributing to

financial strains and potential defaults. Monthly installments between pkr 3000 to 5000 can create additional financial stress for borrowers, making timely repayment more challenging. Finally, the purpose of the loan for cars, small businesses, and large businesses, as well as education and weddings, plays a crucial role in risk assessment as loans for non-income-generating purposes may carry higher risk. Collectively, these factors shape the borrower's risk profile and are essential for improving lending practices and minimizing default in microfinance.

#### 4.2.4 Credit History and Financial Profile

These elements are taken from the data given by microfinance banks, how the borrowers behaved in the past with credit and what obligation they have,

- **Debt-to-Income Ratio:** Elevated DTI Ratios suggest a more significant financial burden than income.
- **Employment History and Income Level:** Stable income sources and job security reduce the probability of defaults.
- **Credit History Indicators:** These Provide insights into the borrower's financial behavior.
- **Length of Credit History:** A longer credit history is often associated with greater financial maturity.
- **Negative Public Records:** Instances such as bankruptcies or liens indicate potential repayment challenges.
- **Number of Open Accounts:** Excessive reliance on credit may signal financial instability.

Similar to loan-specific variables, Table 4.3 represents credit history and financial profile.

Table 4.4 Credit History and Financial Profile

Factor	Range/Values
Debt-to-Income Ratio (DTI)	0% to 40%

Length of Credit History	1 to 10 years
Number of Open Accounts	1 to 3
Total Number of Public Records	0, 1
Total Credit Accounts	1 to 5

Table 4.3 presents key factors to a borrower’s credit history and financial profile, which are critical for assessing their repayment capacity and default risk. The DTI ranges from 0% to 40%, indicating that lower ratios are generally favorable, as they suggest borrowers are not overly burdened by debt relative to their income. The length of credit history spans from 1 to 10 years , with longer histories typically associated with more responsible credit behavior and greater economic maturity. The number of open accounts varies between 1 to 3, providing insight into the debtor’s credit management; a moderate number of accounts can indicate responsible use, while extreme accounts may signal potential financial instability. Additionally, the total number of public records shows either 0 or 1, with the absence of negative records indicating lower risk for lenders. Lastly, the total credit accounts range from 1 to 5, reflecting the borrower's overall credit utilization; a higher number of accounts can suggest more extensive credit experience but may also raise concerns about over reliance on credit. These factors offer a complete view of a debtor’s financial health and creditworthiness. Table 4.4 presents borrower-related factors.

Table 4.5 Borrowers Related Factors

<b>Factor</b>	<b>Description</b>	<b>Potential Influence on Default</b>	<b>Rationale for Inclusion</b>
<b>Debt-to-Income Ratio (DTI)</b>	A ratio indicates the borrower's total debt proportion to their income.	Higher DTI ratios may indicate financial stress and a greater risk of default, as a more significant portion of income is already committed to debt repayment.	Directly reflects financial stress and the borrower's ability to manage existing debt obligations.

<b>Factor</b>	<b>Description</b>	<b>Potential Influence on Default</b>	<b>Rationale for Inclusion</b>
<b>Length of Credit History (earliest_cr_line)</b>	The date when the borrower opened their first credit account.	A more extended credit history may indicate more experience managing credit and a lower risk of default. However, the quality of that history is also essential.	Indicates the borrower's experience with credit and their long-term financial behavior.
<b>Number of Open Accounts (open_acc)</b>	The number of credit accounts the borrower currently has open.	Having too many open accounts may suggest overextension and a higher risk of default. It could indicate a reliance on credit to manage finances.	Provides insight into the borrower's current credit usage and potential overextension.
<b>Total Number of Public Records (pub_rec)</b>	The number of adverse public records, such as bankruptcies or liens, on file for the borrower.	A higher number of public records significantly increases the likelihood of default, indicating past financial difficulties.	Reflects past financial problems and a higher risk profile.
<b>Total Credit Accounts (total_acc)</b>	The borrower holds the total number of credit accounts throughout their credit history (both open and closed).	A high number of total accounts can provide an overview of the borrower's credit experience, but it needs to be considered in conjunction with other factors like DTI and payment history.	Provides a broad overview of the borrower's credit experience and history of managing credit.

#### 4.2.5 Rationale for Factor Selection

The factors identified align closely with practical metrologies employed by microfinance institutions in assessing loan applications. These factors reflect practical considerations and align with established practices in evaluating creditworthiness. By incorporating a variety of factors, loan officers can obtain a holistic view of the borrower's finances and the specific characteristics of the loan itself. This comprehensive approach is essential for making informed lending decisions that

minimize risk and enhance repayment rates. Several key criteria were employed to determine which factors to include in the analysis.

#### ***4.2.5.1 Factors Typically Used in Microfinance***

The chosen variables are widely recognized within the microfinance sector as standard indicators of default risk. Institutions often rely on these factors to create robust models that predict debtor's behavior and assess potential financial risks. By adhering to these industry norms, the analysis benefits from a foundation built on proven methodologies that have validation through extensive use of real-world lending scenarios (Maenuddin, Hamid, Fahlevi, Nassir, & Hashim, 2023).

#### ***4.2.5.2 Implications for Default Predictions***

Each factor selected carries significant implications for understanding a borrower's capacity to service their loan. For instance, the DTI ratio is a crucial metric that reflects the financial stress a debtor may experience. A high DTI ratio indicates that a substantial portion of the debtor's income is allocated to debt repayment, leading to improved financial strain and a higher likelihood of default. Similarly, the purpose of the loan is to serve as an essential indicator of potential economic return or improvement. Loans taken out for productive purposes, such as business expansion or education, may be associated with better repayment outcomes than loans for consumption or non-essential expenditures (Guerrero, Vargas, Sánchez, & Fariás, 2025).

#### ***4.2.5.3 Data Availability***

The viability of analyzing these factors is greatly enhanced by their obtainability within loan records. The data collected by loan officers is inclusive and readily available, allowing for efficient analysis and application in predictive modeling. This convenience ensures that the study can be conducted promptly, providing valuable insights into debtor behavior without widespread additional data collection efforts (Alagic et al., 2024).

The dataset containing these selected factors is pivotal in identifying patterns and correlations associated with loan defaults. By examining the characteristics of loans and the correlation related to loan defaults compared to those that were fully repaid, the analysis aims to uncover critical insights into which elements significantly influence default risk within the microfinance sectors (Uddin, Akter, Mollah, & Al Mahi, 2022). This structured approach not only aids in understanding borrowers' behavior but also enhances risk assessment capabilities, ultimately informing better leading practices and decision-making processes.

The rationale for factor selection in this study is rooted in established microfinance practices, partial implications for predicting defaults, and relevant data availability. By focusing on these critical aspects, the lender can develop more effective strategies for managing credit risk and improving overall financial performance within their institutions.

#### **4.2.6 Analysis Approach**

In this objective, analysis was done using real-life loan data taken under the microfinance banks in Pakistan. The data provided both the loans in the category of repayment termed as Fully Paid and also loans which were Charged Off enabling a comparative scenario of the re-payment experience. Instead of using statistical tests or hypothesis testing the approach was to find operationally relevant variables that were being used by the loan officers in the assessment of the loan applicants.

The factors mentioned above, key borrowers, loans, and aspects related to credit, were chosen because of the regularity of their appearance in the institutional data and because they coincided with the internal credit assessment standards. These factors were not statistically significant as identified with the help of appropriate statistical correlation or regression analysis but based on expert knowledge that was part of the dataset. The practical relevance was of interest as opposed to statistical validation which would indicate the practical decision-making processes adopted by microfinance practitioners.

This method gives a case-based investigation of the issues of what are risk in actual loan portfolios. It preconditions the further analysis based on machine learning in order to measure and confirm the predictive ability of these features.

#### **4.2.7 Summary of Findings**

The discovery of the core issues which affect loan default in the Pakistani microfinance banks in the following section was based on the institutional information provided by the two partner microfinance institutions. These variables included aspects of the demographics and credit history of the borrower as well as features of the loans themselves; they were chosen because of their use in the internal credit assessment schemes adopted by loan officers in carrying out both the underwriting and the monitoring of the loans.

Under this objective, no formal statistical tests (i.e., correlation, chi-square or regression analysis) or descriptive statistics (i.e., means, standard deviations) were performed. Rather, it was precisely about properly structuring and displaying the variables considered to be relevant towards the prediction of defaults solely by means of the structured data taken out of the institutional records. The fact that loans may be categorized into such categories as Fully Paid and Charged Off allowed an observational distinction at the high-level into performing and non-performing loans but with no modeling and hypothesis tests.

This goal is not providing training and analysis of the models, performance and validation of predictive accuracy, which must be discussed in separate sections of the thesis. Besides, no qualitative feedback by practitioners was requested at this point when expert validation under Objective 4 was already done separately.

Having recorded the factors found in the real world data in regards to lending microfinance banks, this section gives background knowledge of the indicators used in the real world as factors to be taken into account with regards to loan defaults. This empirical foundation forms the basis to subsequent (quantitative) modeling and validation efforts and forms part of a practical, contextually and contextually aware approach to credit risk modeling in the microfinance industry.

### **4.3 OBJECTIVE-02: TO EVALUATE AND COMPARE THE PERFORMANCE OF VARIOUS MACHINE LEARNING ALGORITHM IN PREDICTING LOAN DEFAULT**

Using machine learning algorithms to forecast loan default is crucial for microfinance institutions in Pakistan, where many borrowers lack formal credit histories (Alam et al., 2020). Machine learning techniques can analyze large datasets to identify patterns and predict borrower behavior, enhancing the ability of MFIs to manage risk effectively (Omowole, Urefe, Mokogwu, & Ewim, 2024). The algorithms can assess credit risk using diverse data sources, including transaction histories and alternative data. This capability allows for more accurate predictions of borrower creditworthiness, which is essential for MFIs serving underserved populations.

F1-Measures, probability of detection, value of precision, and exactness are some of the standard metrics used for evaluating the performance of organization models like logistic regression, decision tree random forest, and gradient boosting (Khan & Gupta, 2024).

#### **4.3.1 Accuracy**

Exactness is an essential metric for evaluating performance metrics on machine learning models. Accuracy is mainly simple and intuitive, and it is an overall performance indicator, baseline comparison, and applicability across algorithms for improving performance tracking.

#### **4.3.2 Precision**

The critical metric in Precision is evaluating the execution of classification models focused on positive class, handling imbalanced datasets, complementing other metrics, model comparison, and optimization in training.

### 4.3.3 Recall

Recall is an essential metric for evaluating performance by identifying the positive instance, handling an imbalanced dataset, complementing precision, identifying the improvement area, and comparing the model's performance.

### 4.3.4 F1 Score

The primary reason for using the F1 Score in metrics is to balance between Recall and Precision and also handle class imbalance, robustness for false negative and positive, optimization, and tuning. It compares the model's effectiveness in metrics to attain better performance (Anand, Velu, & Whig, 2022).

### 4.3.5 Confusion Matrix

The confusion matrix and its derived matrix are crucial for evaluating the execution of classification models in machine learning. Based on the actual yes and no predictions, the table is classified. The confusion matrix represents the count from predicted and actual values. The significance of Confusion matrices are:

- The confusion matrix delivers a wide-ranging opinion of model performance by comparing the predicted class to the true class.
- It also allows the calculation of various performance metrics like F1 scores, recall, precision, and accuracy (Gao, Yang, & Zhao, 2023).
- It is likewise applicable for both binary and multiclass classification problems.
- Comparing the confusion matrix of the production model to the reference period can help monitor model quality over time.
- It assesses the execution of classification models when the predictions are on test data and states how good the classification model's performance is (Zhu, Chu, Song, Hu, & Peng, 2023).

Various representations, like Random Forests, decision trees, logistic regression, and gradient boosting, have effectively predicted loan defaults. Research indicates that random forests often outperform other models in outcome by providing robust predictions based on complex datasets.

#### 4.3.6 Justification for Machine Learning Model Performance

The outcomes of the machine learning models exhibit differing levels of efficacy in forecasting loan default. Presented above is a concise justification for the performance of each model:

- **Gradient Boosting (GB) (Optimal Model):** GB surpassed other models because of its iterative learning methodology, which reduces errors by progressively enhancing weak learners (Odegua, 2020). This model is adept at managing intricate, non-linear correlations in financial data. Nonetheless, it entails extensive adjustment to prevent overfitting (Khan & Gupta, 2024).
- **Random Forest (RF) (High Accuracy but Lower Recall):** RF succeeded because of its ensemble learning methodology, which mitigates overfitting relative to individual decision trees (Fan et al., 2024). It offers enhanced feature importance estimation but has challenges with recall, resulting in the omission of certain defaulters (Al-Khazaleh et al., 2024).
- **Decision Tree (DT) (Interpretability with High Overfitting):** DT is highly interpretable yet subject to overfitting, as it tends to memorize training data instead of generalizing patterns (Ahmed et al., 2023). DT ultimately ends up with lower accuracy in forecasting new loan defaults (Bader et al., 2024)
- **Logistic Regression (LR) (lowest Recall, Poor Non-Linear Data Performance):** LR was poor because it assumed linearity between dependent and independent variables (Brygała & Korol, 2024). Since the financial risk indicators have complex associations, LR failed to quantify non-linear associations properly, yielding the worst recall (i.e., inappropriate classification of numerous default cases) (Li et al., 2023).

Table 4.6 Performance Metric for DT (Decision Tree) Classifier

Performance Analysis of Decision Tree	
Metrics	Values
Accuracy	0.689
Precision	0.3115
Recall	0.3032
F1 Score	0.3073

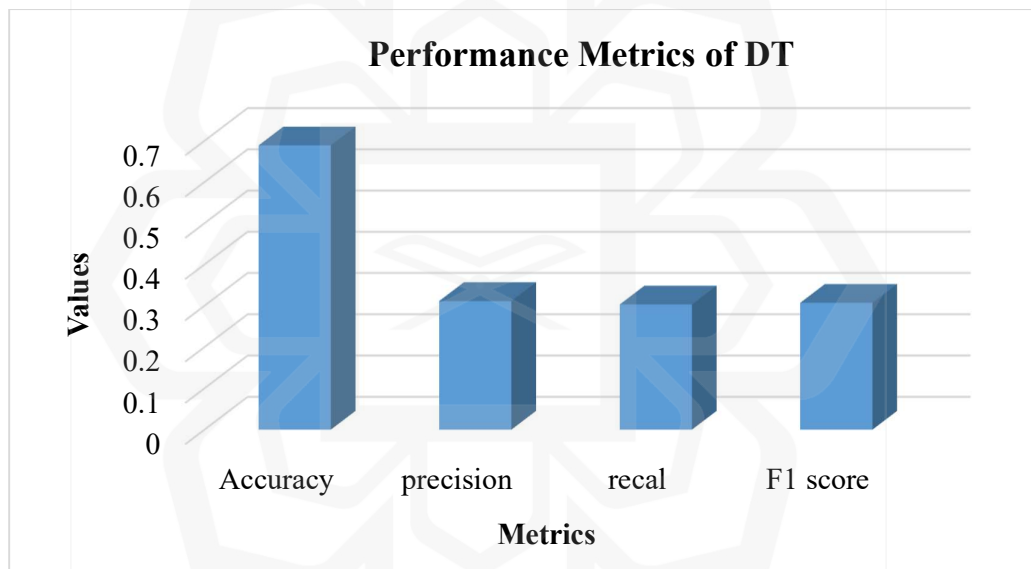


Figure 4.1 Pictorial Representation of DT

Table 4.5 and Figure 4.1 summarize the decision tree's performance analysis, highlighting four critical indicators: accuracy, precision value, recall, and F1 score. The model attains an exactness of 0.689, representing that it appropriately predicts results in almost 68.9% of the cases. The value of precision is notably low at 0.3115, meaning that only about 31.15% of the instances predicted by way of positive are actual values. Similarly, recall stands at 0.3032, revealing that the model identifies only 30.32% of the true positive cases in which the value is shown in the confusion matrix. The F1 measure, which balances precision and recall, is 0.3073, reflecting moderate overall performance.

DTs are easy to interpret and visualize, making them helpful in understanding decision-making processes. At the same time, they can easily over-fit the training data, which results in poor generalization of unseen data to improve the performance and avoid over-fitting, employing techniques like pruning, setting minimum samples for splits, using ensemble methods, conducting features selection, applying cross-validation, and tuning hyperparameters. Feature selection helps improve the generalization accuracy of unseen data and maintain interpretability. These outcomes are proposed while demonstrating reasonable accuracy. In addition, the pictorial representation of Table 4.5 is illustrated in Figure 4.1.

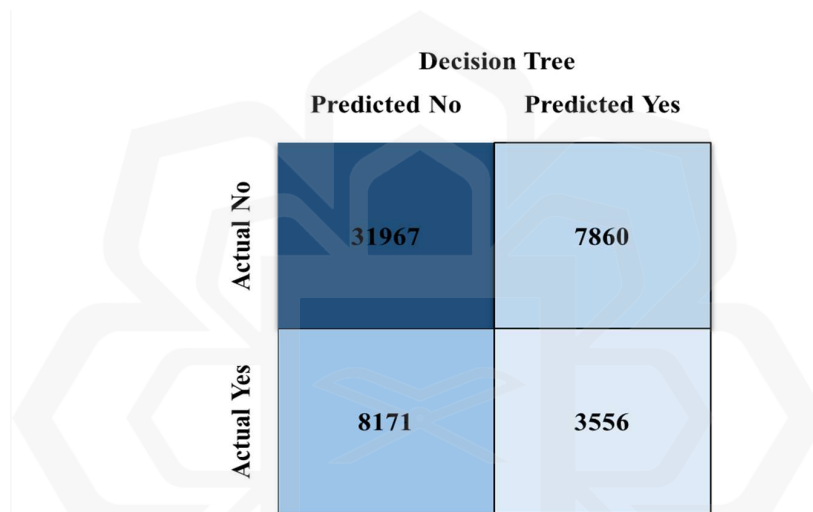


Figure 4.2 Confusion Matrix for Decision Tree

Figure 4.2 shows the decision tree model's classification matrix based on the outcome's value. The matrix visualizes the comparison between the actual and predicted labels, with the numbers corresponding to:

- True negative (38871): correct prediction for Actual no.
- False positive (7860): Instances where class 0 was predicted as Actual yes.
- False Negatives (8171): Instances where Actual yes is predicted as Actual no.
- True Positives (3556): Correct predictions for Actual yes.

Table 4.7 Classification Report of Decision Tree

Class	Precision	Recall	F1-Score	Support
0	0.8	0.8	0.8	39,827
1	0.31	0.3	0.31	11,727
Exactness (Accuracy)	---	---	0.69	51,554
Macro Average	0.55	0.55	0.55	51,554
Weighted Average	0.69	0.69	0.69	51,554

Table 4.6 shows the value obtained in the classification report for Decision Tree models and provides a comprehensive overview of their performance across two actual no and yes metrics based on the value of precision, probability of detection, and F1-measures, as well as support.

**Actual no:** The decision tree models also perform well on the majority actual no with a precision value of 0.80 and a probability of detection of 0.80, with an effect in F1-measures of 0.80. The value that indicates a balanced performance in classifying prediction 1.

**Actual yes:** Like the logistic regression model, the decision tree exhibits low performance in minority prediction yes with a precision of 0.31 and a recall of 0.30, leading to F1 measures of 0.31, highlighting the model's difficulty in accurately predicting true positives for prediction yes.

**Overall Accuracy:** The complete exactness of the decision tree prototype is slightly lowest at 0.69, meaning it correctly predicts 69% of all instances.

**Average:** The average value of precision, probability of detection, and F1 measures are 0.55, representing a balanced though reasonable performance across both classes. A weighted average, reflecting the Actual yes and no supply, is 0.69 for the precision and recall, suggesting a more favorable performance when counting for the highest number.

Table 4.8 Performance Metric for LR (Logistic Regression) Classifier

Performance Analysis for Logistic Regression	
Metrics	Values
Accuracy	0.7739
Precision	0.5174
Recall	0.0874
F1 Score	0.1495

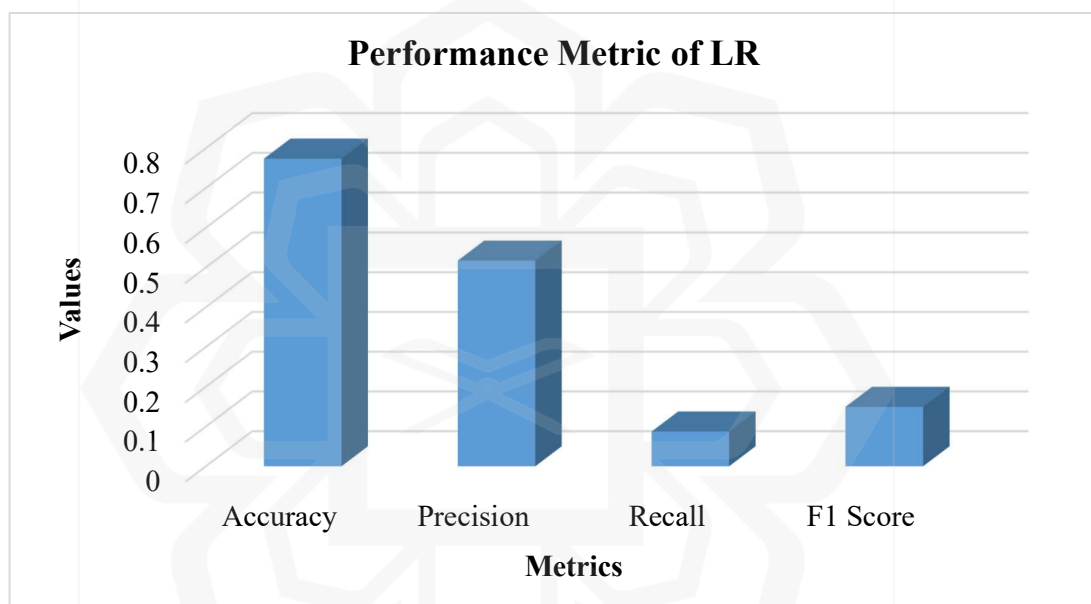


Figure 4.3 Pictorial Representation of LR

Table 4.7 exhibits the performance evaluation of an LR model across four key metrics. The model achieves an accuracy of 0.7739, correctly predicting outcomes in 77.39% of cases. The precision is 0.5174, indicating that 51.74% of examples forecast as positive are actually correct. However, recall is notably low at 0.0874, suggesting the model successfully identifies only 8.74% of the true positive instances. The F measures, corresponding precision, and probability of detection value is 0.1495, replicating a poor overall performance. These results indicate that while the logistic regression model exhibits substantial accuracy, it struggles significantly with recall, correctly identifying only a tiny fraction of positive cases.

LR has good speed, simplicity, and computational efficiency, making it a good baseline model. The model shows very low recall, which indicates that it fails to capture many true positives. To enhance the performance of LR, focus on feature engineering to create more information predictors and apply techniques like regularization to prevent over fitting. Cross-validation can also help select the best model parameters and assess performance on unseen data. Further optimization is necessary to improve the model's ability to detect true positives while maintaining its precision. Table 4.7 graphical illustration is demonstrated in Figure 4.3.

		<b>Logistic Regression</b>	
		<b>Predicted No</b>	<b>Predicted Yes</b>
<b>Actual No</b>	<b>Actual No</b>	<b>38871</b>	<b>956</b>
	<b>Actual Yes</b>	<b>10702</b>	<b>1025</b>

Figure 4.4 Confusion Matrix for LR

Figure 4.4 shows the classification matrix illustration of the logistic regression model based on the value you provided. The metrics visualization is the comparison between the actual and predicted labels, with the numbers corresponding to:

- True Negatives (38871): Correct predictions for actual no.
- False Positives (956): Instances where actual No was predicted as predicted Yes.
- False Negatives (10702): Instances where actual Yes was predicted as predicted No.
- True Positives (1025): Correct predictions for predicted Yes.

Table 4.9 Classification Report of LR

Class	Precision	Recall	F1-Score	Support
0	0.78	0.98	0.87	39,827
1	0.52	0.09	0.15	11,727
Accuracy	---	---	0.77	51,554
Macro Average	0.65	0.53	0.51	51,554
Weighted Average	0.72	0.77	0.71	51,554

The classification report for the Logistic Regression prototype delivers the complete presentation overview across two classifiers, as shown in Table 4.8.

**Actual no:** The model illustrates the high execution with a precision value of 0.78 and a high probability of detection of 0.98, resulting in F measures of 0.87. Results specify that models are highly efficient in correctly identifying the majority of Actual no and correct prediction of 78% in positive predictions.

**Actual yes:** Opposite to Actual no, the model struggles with lower Actual yes or prediction yes by attaining a precision of only 0.52 and a probability of detection of 0.09, leading to F measures for 0.15. Models suggest that it frequently fails to identify true positives for prediction. Yes, the outcome is a high number of false negatives.

**Overall Accuracy:** The exactness of logistic regression is 0.77, representing the correct predictions in 77% of all cases.

**Averages:** The average precision and recall are 0.65 and 0.53, reflecting the model's overall performance across both Prediction yes and no. The weighted averages, which consider class imbalance, are 0.72 for precision and 0.77 for recall, indicating better performance when considering the supple of Prediction yes.

Table 4.10 Performance Metric for RF (Random Forest) Classifier

Performance Analysis for RF	
Metrics	Values
Accuracy	0.7758
Precision	0.5385
Recall	0.1019
F1 Score	0.1714

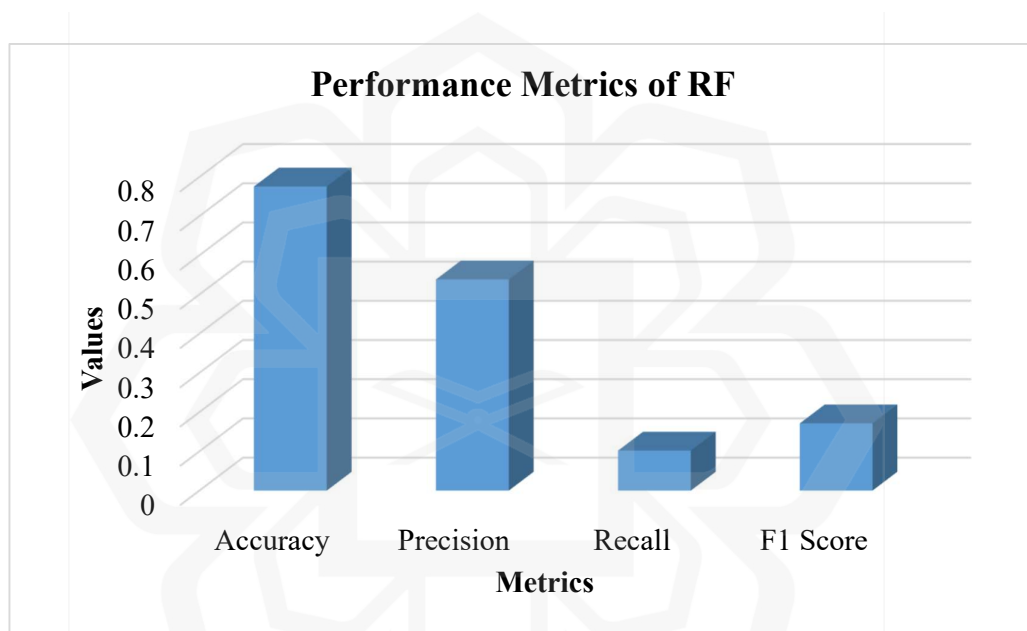


Figure 4.5 Graphical Representation of RF

Table 4.9 provides a performance analysis of Random Forest models, showcasing four critical metrics followed by before metrics: exactness, value of Precision, probability of detection, and F1 measures. A model attains exactness in 0.7758, demonstrating that it appropriately predicts outcomes in 77.58%. The precision measures at 0.5385, meaning that 53.85% of the instance prediction has positive accuracy. However, the recall is notably low at 0.1019, revealing that the model identifies only 10.19% of actual positive cases. The F-measures score, which combines precision and recall into a single metric, stands at 0.1714, reflecting a poor overall performance. RF improved performance over decision trees by combining multiple DTs to enhance forecast exactness and reduce overfitting risk compared to single decision trees. It performs better than DT in terms of precision probability of detection but still

misses many TP values. To improve the performance, leverage the ensemble learning technique by combining multiple DT-trained bootstrapped subsets of the data, which enhances accuracy and reduces over-fitting. These results propose that the random forest prototype demonstrates responsible accuracy and precision. Figure 4.5 is the pictorial illustration of Table 4.9.

		Random Forest	
		Predicted No	Predicted Yes
Actual	No	38803	1024
	Yes	10532	1195

Figure 4.6 Confusion Matrix for Random Regression

Figure 4.6 is the classification of the matrix illustration for the Random Forest prototype based on the values. The matrix visualizes the comparison between the actual and predicted labels, with the numbers corresponding to:

- True Negatives (38803): Correct predictions for Actual no.
- False Positives (1024): Instances where Actual no was predicted as Predicted yes.
- False Negatives (10532): Instances where Actual yes was predicted as Prediction no.
- True Positives (1195): Correct predictions for Prediction yes.

Table 4.11 Classification report of RF

Class	Precision	Recall	F1-Score	Support
0	0.79	0.97	0.87	39,827
1	0.54	0.1	0.17	11,727
Accuracy	---	---	0.78	51,554
Macro Average	0.66	0.54	0.52	51,554
Weighted Average	0.73	0.78	0.71	51,554

The classification report for the Random Forest is provided in Table 4.10 as a detailed assessment of their performance across two Actual yes and no using values of precision, probability of detection, F-measures, and provision.

**Actual no:** The Random Forest prototype demonstrates strong presentation in the majority for actual no and actual yes, with an exactness of 0.79 and a probability of detection of 0.97, causing F1 scores of 0.87. The indication of the model is highly efficient in correctly classifying instances of prediction no, with a high rate of true positives and a comparatively low rate of false negatives.

**Actual yes:** Opposite of the prototype struggles in the sectional prediction, yes, attaining a precision value of 0.54 and probability of detection is 0.10, leading to F measures of 0.17. Results suggest the model frequently fails to identify true positives for Actual no, and the outcome is often false negatives.

**Overall Accuracy:** The complete exactness of random forest models of 0.78 correctly predicts 78% of all instances.

**Average:** The instruction average value of precision and probability in detection are 0.66 and 0.54, respectively, indicating the model through all presentations across the Actual class. The averages, which account for the class distribution, are 0.73 for precision, and for recall, they are 0.78, suggesting better performance when considering the high number of instances in class 0.

Table 4.12 Performance Metric for GB (Gradient Boosting) Classifier

Performance Analysis for GB	
Metrics	Values
Accuracy	0.7775
Precision	0.5533
Recall	0.1133
F1 Score	0.1881

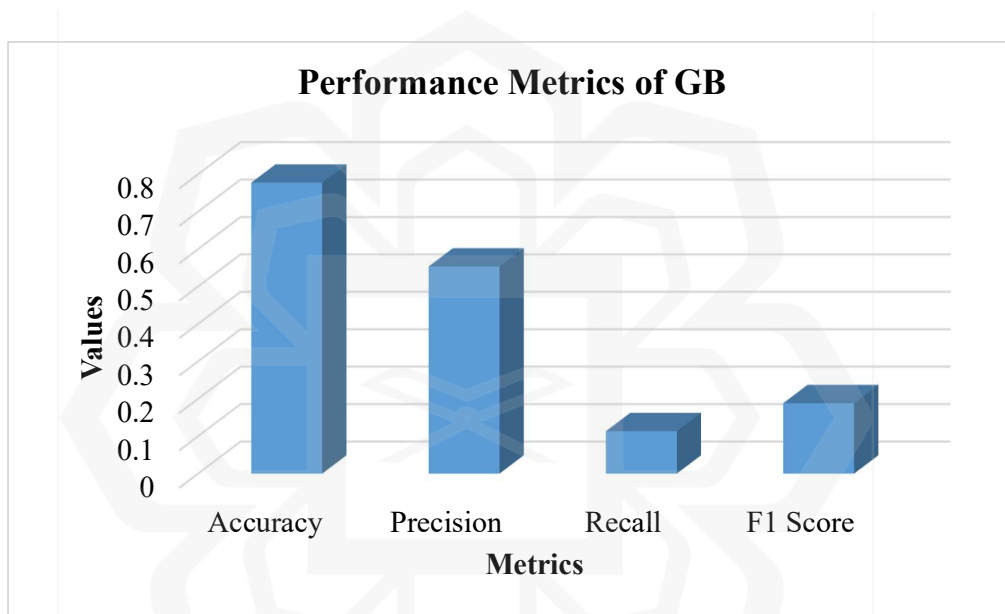


Figure 4.7 Graphical Representation of GB

Table 4. 11 outlines the execution analysis of a GB model, highlighting four metrics. The prototype attains an exactness of 0.7775, representing that it correctly predicts outcomes in 77.75% of the cases. The precision is 0.5533, meaning that 55.33% of the predicted positive instances are accurate. However, the recall is an instance. The F measures, combining the value of precision and probability of detection, stand at 0.1881, replicating the overall performance. These results indicate that while the Gradient Boosting model demonstrates reasonable accuracy and precision, it significantly struggles with recall and needs additional optimization to enhance its effectiveness in detecting true positives. Also, GB attains high exactness among the models evaluated, indicating its overall effectiveness in correctly forecasting outcomes.

It also employs techniques like shrinkage and regularization to mitigate over-fitting risk, making it suitable for complex datasets. At the same time, it has some drawbacks despite high exactness, such as the recall value being relatively low, meaning it struggles to identify all relevant instances. This model may prioritize precision over recall, leading to a conservative approach in classifying positive instances. Table 4. 11 pictorial representation is shown in Figure 4. 8, where the values are measured in graphs for better understanding.

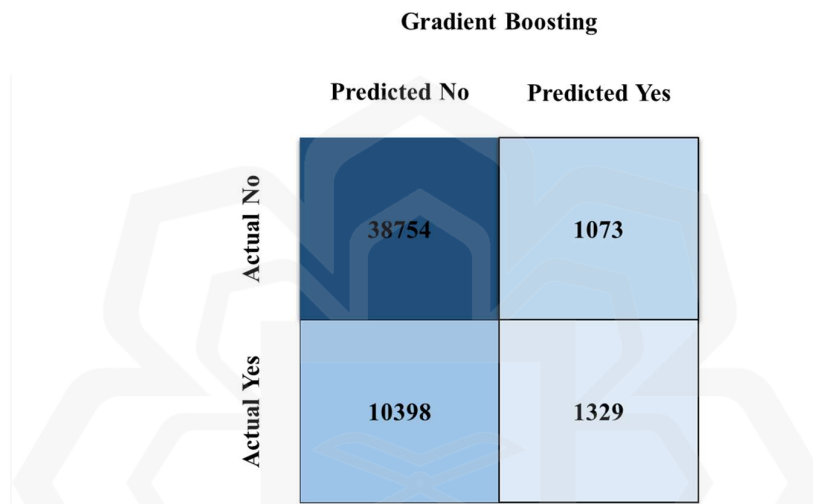


Figure 4.8 Confusion Matrix for Gradient Boosting

Figure 4.8 is the confusion matrix diagram for the gradient boosting on the basis of the values' results. The matrix visualizes the comparison between the actual and predicted labels, with the numbers corresponding to:

- True Negatives (38754): Correct predictions for Actual no.
- False Positives (1073): Instances where Actual no was predicted as Prediction yes.
- False Negatives (10398): Instances where Actual yes was predicted as Prediction no.
- True Positives (1329): Correct predictions for Prediction yes.

Table 4.13 Classification Report of Gradient Boosting

Class	Precision	Recall	F1-Score	Support
0	0.79	0.97	0.87	39,827
1	0.54	0.1	0.17	11,727
Accuracy	---	---	0.78	51,554
Macro Average	0.66	0.54	0.52	51,554
Weighted Average	0.73	0.78	0.71	51,554

Similarly, the gradient boosting classification report presented in Table 4.12 as performance across two actual and prediction classes using recall, precision, and F1 measures is also supported.

**Actual no:** The gradient boosting model reflects the random forest model on the higher actual no class. By attaining the precision value of 0.79 and a probability in detection of 0.97, the F1 measures' outcome is 0.87. The indication of consistent efficiency in the identification of prediction is no instances.

**Actual yes:** Similar to the previous model, the gradient boosting model also fights with the lower actual classes, with a precision value of 0.54 and a probability of detection of 0.10, resulting in F1- measures of 0.17. The highpoints of the prototype difficulties in accurately forecasting true positives for the prediction class yes.

**Overall accuracy:** The complete exactness of the gradient boosting prototype is also 0.78, indicating it correctly predicts 78% of all instances.

**Average:** The major average precision value, probability of detection, and the F1 scores are 0.66, 0.54, and 0.52, respectively. The reflection of the balance is moderate performance across both classes. The averages, which consider the class distribution, are 0.73 for precision and 0.78 for recall. The favorable performance when considering the distribution of classes.

The DT, LR, RF, and GB models prove similar metrics, particularly their ability to identify the highest Classes in Actual no and Prediction no. The model shows

significant challenges in accurately predicting the lower classes, which are called actual yes and prediction yes. Despite attaining complete performance accuracy, 0.78 highlights the result needed for optimization and improving the detection of true positives in class 1, such as adjusting class weights or employing techniques.

Table 4.14 Comparison of Performance Metrics

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.689	0.3115	0.3032	0.3073
Logistic Regression	0.7739	0.5174	0.0874	0.1495
Random Forest	0.7758	0.5385	0.1019	0.1714
Gradient Boosting	0.7775	0.5533	0.1133	0.1881

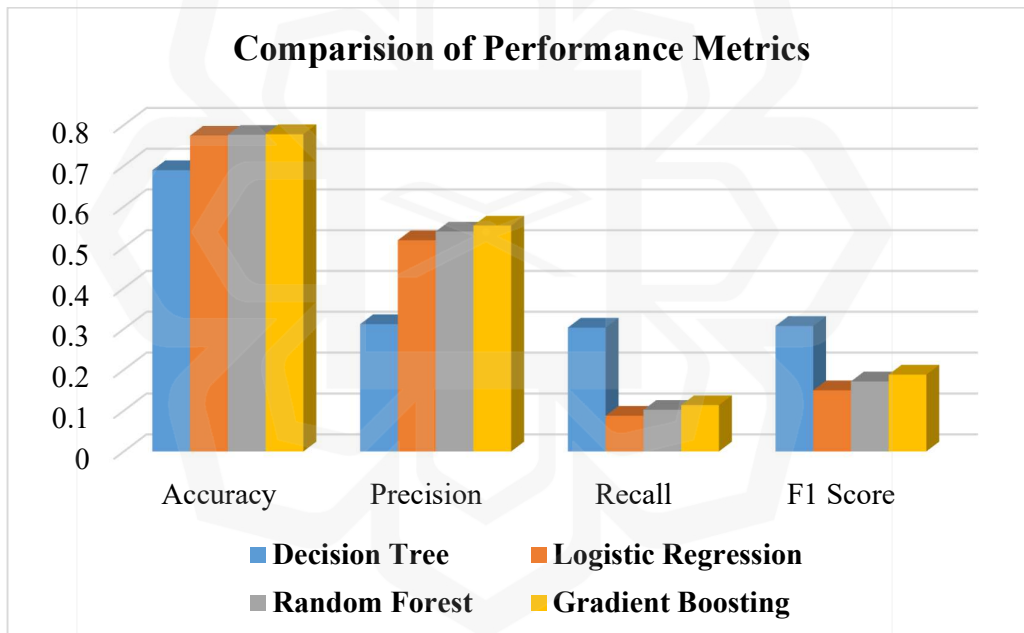


Figure 4.9 Graphical Representation of Comparative Performance Metrics

Figure 4.9 demonstrates the graphical representation in Table 4. 13, which summarizes the performance metrics of comparing four different machine algorithm models: DT, LR, RF, and GB. All prototypes are estimated based on four key metrics: exactness or accuracy, value of precision, probability of detection, and F1 measures. GB outperforms other models like DT, LR, and RF in exactness, value of precision, and F measures score by effectively combining multiple weak learners and focusing on

errors to advance predictions. Its ability to adapt over hyperparameter tuning and handling imbalanced datasets allows it to achieve a better balance between precision and recall, making it suitable for complex applications.

While GB excels in exactness due to its ensemble learning approach that combines multiple weak learnings into a strong model, it tends to sacrifice recall for precision. This behavior underscores the trade-offs inherent in model selection, where optimizing for one metric can adversely affect another. Understanding these dynamics is crucial when choosing a model based on the specific application needs, mainly when dealing with imbalanced datasets or when the cost of false negatives is high. So, GB gives better performance when compared to other metrics.

**Accuracy:** Accuracy is the ratio of appropriately projected instances to the entire occurrence. It delivers a general measure of the overall performance of models. In some cases, the gradient boosting model has the highest accuracy at 77.75%, indicating it correctly predicts the majority of instances.

**Precision:** The value of precision deals with the quantity of TP (true positive) results in every positive prediction the model produces. It is vital when the cost for false positives is higher, and the gradient boosting model again leads with a precision of 0.5533, suggesting that over half of its progressive forecasts are correct.

**Recall:** Recall, a delicacy, indicates the proportions of actual yes appropriately known by the model. The performance metrics are essential when the missing positive (false negative) is costly. Here, all models exhibit relatively low recall values, with Gradient boosting achieving 0.1133, indicating that it identifies only a tiny fraction of actual positives.

**F1 scores:** The vocal mean of precision value and probability of detection, provided that a balance among two metrics is called F measures. It is specifically used when handling an imbalanced dataset. The gradient boosting prototype again performs best with an F measure of 0.1881, imitating a compromise between its precision and recall.

In summary, while GB outclasses in exactness due to its ensemble learning approach that combines multiple weak learners into a strong model, it tends to sacrifice

recall for precision. This behavior underscores the trade-offs inherent in model selection, where optimizing for one metric can adversely affect another (Odegua, 2020). Understanding these dynamics is crucial when choosing a model based on specific application needs, mainly when dealing with imbalanced datasets or when the cost of false negatives is high.

#### **4.3.7 Practical Implications for Microfinance Institution Pakistan**

The findings from the proposed model on community influence and microfinance loan default during crises have several practical implications for Pakistan's microfinance institutions.

##### ***4.3.7.1 Strengthen Borrower Screening and Monitoring***

MFIs should thoroughly assess the borrower's DTI ratio in risk evaluation and help to ensure the borrower has sufficient income to repay the loan, even during economic downturns. Ongoing monitoring of the borrower's financial situation is also critical to identifying any early emerging repayment issues.

##### ***4.3.7.2 Diversify Loan Portfolio***

Concentrating loans in a single geographic area or industry makes MFIs threats to localized shocks. Expanding the loan portfolio across various regions, sectors, and individual borrowers helps mitigate systemic risk that reduces the contagion effect where defaults by one borrower trigger defaults by their peers.

##### ***4.3.7.3 Enhance Borrower Financial Literacy***

Educating borrowers on sound financial management practices can improve their resilience to economic crises. MFIs should provide training on budgeting, debt

management, and building emergency savings, as well as empower borrowers to withstand income shocks better and maintain loan repayments.

#### ***4.3.7.4 Strengthen Borrower Relationships***

Maintaining strong relationships with borrowers built on trust and open communication is key, and it allows MFIs to understand the borrower's situation better and provide support if needed. It also facilitates renegotiating loan terms or providing emergency relief in crises to prevent defaults.

#### ***4.3.7.5 Utilize Innovative Lending Models***

Exploring models like individual liability loans or dynamic incentives can help MFIs manage default risk more effectively. These approaches reduce community influence on repayment, making borrowers more accountable for their obligations. These are real-world implication strategies (Malik et al., 2020).

#### 4.4 OBJECTIVE-03: TO DETERMINE THE MOST INFLUENTIAL FEATURES CONTRIBUTING TO LOAN DEFAULT IDENTIFIED BY MACHINE LEARNING MODELS

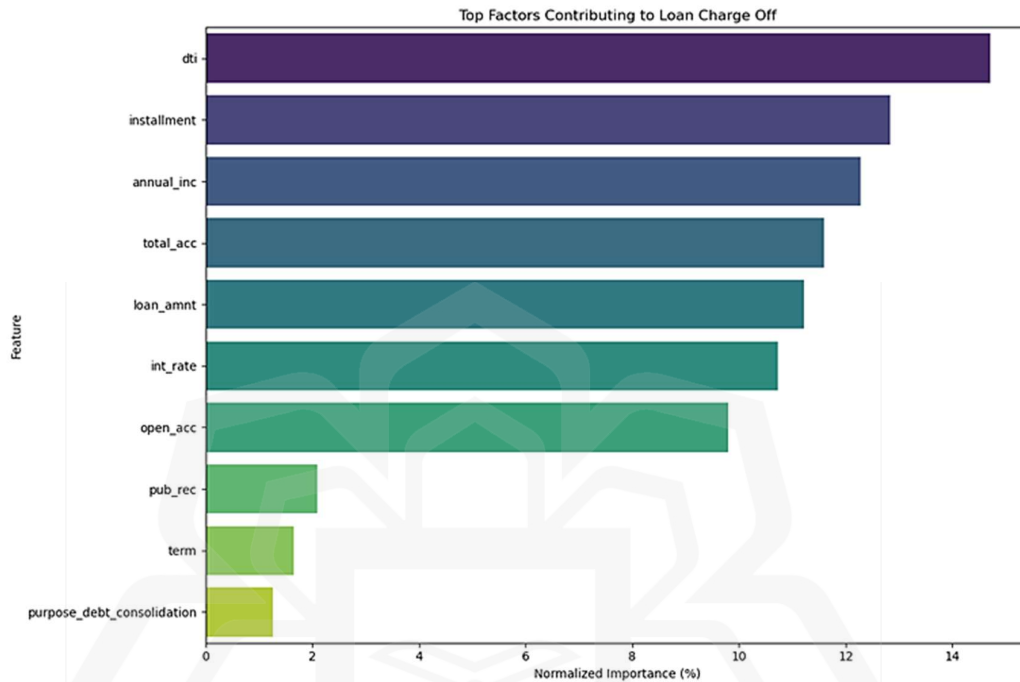


Figure 4.10 Top Factors Contribution to Charge Off Used to Predict Loan Default

The pictorial bar graph shows in Figure 4.10 the top factors contributing to loan charges off, ranked by their normalized importance of percentage (%). A loan charge-off is when a borrower defaults and the lender writes off the loan as uncollectible (Li, Li, Bellotti, & Yao, 2023). This chart aims to highlight the most influential factors leading to such a scenario.

The normalized importance (%) is represented in the horizontal axis, which shows each factor's relative influence on loan charge-offs. The vertical axis lists the specific features or variables influencing the charge-off. The top contribution factors are DTI, installment, annual income, total number of accounts, loan amount, and interest rate. In comparison, the Low contributing factors are open accounts, public reports, loan terms, and the purpose of debt consolidation.

#### **4.4.1 DTI (Debt-To-Income Ratio)**

The DTI proportion estimates a borrower's monthly debt payment relative to gross regular income, serving as a critical indicator of financial health (Akinjole, Shobayo, Popoola, Okoyeigbo, & Ogunleye, 2024). With normalized importance close to 14%, an advanced DTI proportion proposes that a larger ratio of a borrower's income has determined the repayment of debt, which may cause economic strain and increased effort on keeping up with loan payments. Consequently, this elevated risk of default may signal a higher likelihood of charge-offs, emphasizing the importance of managing debt relative to income.

#### **4.4.2 Installment Amount**

The second most important factor, with just over 12% importance, represents the amount the borrower must pay each period, where a large amount might increase default risk, which refers to the fixed monthly payment that a borrower must make to repay their loan. A higher installment amount can impose a significant financial burden, especially if the borrower's income is inadequate to cover these payments alongside their other financial obligations. When borrowers face large installments, they may find it challenging to manage their finances effectively, leading to increased financial strain and a higher likelihood of default (Akinjole et al., 2024).

#### **4.4.3 Annual Income**

With roughly 11% important, a borrower's income also significantly affects charge-offs, indicating that lower-income might correlate with high default rates. The borrower's income directly influences the capability to meet loan repayment obligations, as those with lower incomes often face significant challenges in managing their financial responsibilities, mainly when unexpected expenses arise. This correlation between income and default risk is essential; lower annual incomes are frequently associated with higher loan default rates (Owusu, Quainoo, Mensah, & Appati, 2023).

#### **4.4.4 Total Number of Accounts**

Indicating that the borrower's credit statement and the total number of credit accounts play a significant role in predicting a default of about 10%. This variable is essential because many accounts may suggest that a borrower has multiple existing debt obligations, which can complicate financial management. Juggling numerous accounts can be challenging, as it increases the likelihood of missed payments, leading to a heightened risk of defaults (Uddin et al., 2023).

#### **4.4.5 Loan Amount**

The total sum of the borrowed money is known as the loan amount. Around 9% is important; this suggests that the size of the loan itself is a significant factor. Large loans could pose a higher risk. As a result, the loan size can significantly impact a borrower's financial stability and repayment capacity, thereby increasing the risk of default. Understanding the dynamics of loan amounts is essential for both borrowers and lenders to ensure sustainable borrowing and lending practices (Krasovytskyi & Stavvytskyy, 2024).

#### **4.4.6 Interest-Rate**

The interest rate signifies the borrowing cost, the loan percentage in amount. Just below 9%, higher interest rates might increase the financial burden, leading to more defaults. High interest rates may significantly elevate the overall loan cost, increasing the financial burden on borrowers. The added cost could be particularly problematic for those already financially stretched, as it reduces their disposable income and increases the likelihood of missed payments, ultimately raising the risk of default (Kouser, TP, Sruthi, & Jadhav, 2024).

#### **4.4.7 Open Account**

With roughly 7%, the number of active credit accounts is less influential than the factors above. While this factor is less influential than others, it still plays an essential role in measuring a borrower's economic well-being. A higher number of open accounts can indicate greater credit exposure, suggesting that the borrower may have multiple outstanding debts. This situation can strain their capability to manage supplementary debt effectively, thereby increasing the default risk (Dansana et al., 2024).

#### **4.4.8 Public Records**

Public records refer to negative financial events, including derogatory items such as bankruptcies, tax liens, or judgments recorded against the borrower (Stevenson, Mues, & Bravo, 2021). Contributing around 4%, these records can indicate past financial difficulties, suggesting a higher risk of default on current loans. Such derogatory public records highlight the borrower's previous struggles with managing debt, which may impact their ability to meet future financial obligations.

#### **4.4.9 Loan Terms**

At about 2%, this factor represents the length of time for loan repayment, for example, 36 to 60 months. Its lower importance suggests that while the loan term affects charge-offs, it's not a primary driver.

#### **4.4.10 Purpose of Debt Consolidation**

Finally, this factor refers to loans taken out specifically to consolidate debt. While debt consolidation can simplify a borrower's financial obligations by combining multiple debts into one, it may also suggest that the borrower struggled to manage their previous debts (Alonso Robisco & Carbo Martinez, 2022). If the underlying financial issues are not resolved, this can lead to a higher risk of default on the new, consolidated loan. However, with the lowest importance slightly above 1%, a loan taken out specifically

for debt consolidation might slightly predict a higher risk of charge-offs. However, its influence is minimal compared to the other factors, indicating that while debt consolidation can be a sign of financial difficulty, it is not a strong predictor of loan defaults.

In this context of charge-off prediction, features such as DTI, Installment amount, Annual income, total number of accounts, total number of accounts, loan amount, interest rate, open account, public records, loan terms, and purpose of Debt consolidation are some significant contributors. The specific ranking of features can vary based on the dataset and model parameters.

Lenders can use these insights to assess the risk of loan defaults better. Focusing on a borrower's Debt-To-Income ratio and ensuring that installment amounts are manageable might reduce the risk of charge-offs. Additionally, paying attention to the borrower's income and loan terms could provide a more holistic view of the borrower's ability to repay.

#### **4.5 OBJECTIVE 04: TO VALIDATE THE PRACTICAL RELEVANCE AND INTRPRETABILITY OF TOP ML IDENTIFIED PREDICTORS THROUGH EXPERT INSIGHTS FROM BANK LOAN OFFICERS**

Objective 4 is centrally focused on validating the top predictors identified by machine learning models for predicting loan defaults with bank loan officers' invaluable practical experiences and nuanced insights. This validation is not simply a perfunctory exercise but an essential step in guaranteeing that predictive models are not operating in isolation, disconnected from the complexities of the real world. As an alternative, this process ensures that these models are for machine learning grounded in the on-the-ground realities of lending practices and incorporate vital qualitative considerations that purely data-driven approaches might easily overlook. By actively integrating this wealth of practitioner knowledge, the objective is to enhance the robustness, relevance significantly, and overall practical applicability of the loan default prediction framework within MFIs in Pakistan. The critical goal is to cultivate a symbiotic relationship between the quantitative power and precision of machine learning and the qualitative understanding of experienced lending professionals. This synergy will, in

turn, forge a more comprehensive, contextually sensitive, and ultimately more effective approach to credit risk management within the microfinance sector.

#### **4.5.1 Key Insights from Practitioners**

Insights from experienced loan officers underscore key factors influencing loan defaults, enriching the understanding gained from machine learning models. Practitioners consistently highlighted the importance of a borrower's material status and income levels, noting that higher income generally correlates with lower default rates, aligning with quantitative analyses. Material status, though challenging to quantify directly, significantly impacts financial stability. Loan officers emphasized the need to monitor loan utilization meticulously to prevent fund diversion, a significant cause of defaults, and highlighted the importance of robust loan tracking and proactive borrower engagement to ensure funds for approved purposes.

Economic factors, such as prevailing interest rates and market volatility, also significantly impact loan repayment capabilities. Economic downturns or interest rate spikes can strain borrowers, increasing default likelihood. Incorporating macroeconomic indicators into predictive models is crucial to account for external factors affecting borrower solvency. These qualitative insights from practitioners complement quantitative data, providing a holistic view of factors influencing loan defaults. The emphasis on material status, loan utilization monitoring, and macroeconomic considerations enhances the ability of Microfinance Institutions (MFIs) to refine lending practices and risk management strategies. This comprehensive approach supports more informed decision-making and fosters financial stability for MFIs and their borrowers.

#### **4.5.2 Alignment with Machine Learning Finding**

We observed several striking areas of significant alignment when comparing these invaluable practitioner insights with the core findings derived from our sophisticated machine-learning models. For instance, both the meticulously crafted machine learning models and the experienced practitioners independently recognized and emphasized the

significance of the debt-to-income ratio (DTI), the size of installment payments, and the borrower's annual income as key predictors of potential loan defaults. This demonstrable convergence of perspectives across quantitative and qualitative assessment methods powerfully reinforces these factors' undeniable importance in any comprehensive credit risk assessment. Furthermore, it strongly suggests that these factors should have significant weight when making informed loan approval decisions.

However, despite these areas of compelling agreement, certain interesting discrepancies also emerged, notably in the inclusion of climate-related risks and qualitative factors such as the borrower's overall community reputation. While the machine learning models primarily relied on quantitative data analysis, the practitioners made a strong case for considering a range of external factors, such as the potential impacts of climate change. They rightly pointed out that climate change, particularly in regions heavily dependent on agriculture, can significantly threaten borrowers' ability to repay their loans. The practitioners also articulated the importance of qualitative information, such as community feedback concerning the borrower's reliability and overall character, arguing that these factors are often overlooked in purely data-driven models but can provide critical context and insight.

#### **4.5.3 New Insights from Practitioners**

The discussions with experienced practitioners also yielded a wealth of new and valuable insights that our sophisticated machine-learning models did not initially capture. The practitioners suggested a range of additional factors that could lead to more accurate default predictions, including a more thorough and granular cash flow analysis and the implementation of borrower education programs. They strongly emphasized that meticulously analyzing a borrower's daily or monthly cash flow is critical for understanding their true repayment capacity. They also argued that implementing robust borrower education and financial literacy programs can significantly improve borrowers' understanding of their loan obligations and promote more responsible and sustainable financial behavior.

Furthermore, the practitioners emphasized the valuable role of qualitative data, such as the borrower's community reputation and any available informal credit metrics.

They noted that in many rural areas, feedback from the local community is often considered when assessing credit risk and that this feedback can provide invaluable insights into the borrower's reliability and overall character. Practitioners also highlighted the importance of carefully considering a borrower's household income and expenses, noting that in many cases, a borrower's ability to repay their loan depends on the broader financial context of their entire household.

#### **4.5.4 Practical Implications**

The practical implications of these invaluable practitioner insights are both wide-ranging and highly significant. By incorporating real-world perspectives and those all-important qualitative considerations that traditional quantitative data may not fully capture, we can significantly refine and improve the accuracy of our predictive models, making them more relevant and effective in practice. Moreover, by actively integrating often overlooked factors, such as the potential risks associated with climate change, into the machine learning frameworks, we can dramatically improve these models' robustness and overall applicability, especially in economic sectors particularly vulnerable to environmental changes.

Similarly, the judicious incorporation of qualitative data, for example, the borrower's standing in their local community and any available informal credit metrics, can enhance the accuracy and reliability of our risk assessments, allowing us to build a more complete and nuanced understanding of each borrower's profile. The central point is that by combining the power of machine learning insights with the ground-level knowledge of experienced practitioners, MFIs can develop and implement lending practices that are more effective, more context-aware, and better suited to the diverse needs of their target communities.

#### **4.5.5 Conclusion**

In conclusion, validating sophisticated machine learning findings with bank loan officers' extensive insights and practical experiences is essential to enhancing the overall effectiveness of predictive models and ensuring their firm alignment with real-

world lending practices. This integration goes far beyond mere theoretical consideration; it is crucial to ensure that these models are robust, contextually relevant, and ultimately highly effective in achieving their intended goals. By diligently integrating practitioner knowledge, Microfinance Institutions (MFIs) can machine learning incorporate a wide range of qualitative factors, thoroughly account for sector-specific dynamics, and skillfully manage the external risks that may often be overlooked by relying solely on data-driven approaches.

This comprehensive validation process ensures that loan default prediction frameworks are highly accurate and thoughtfully adapted to the specific circumstances in which they will be applied. This, in turn, will lead to more informed and strategically sound lending decisions, an appreciable reduction in loan default rates, and a tangible improvement in overall financial inclusion within the microfinance sector. In the final analysis, the thoughtful synthesis of quantitative precision with qualitative insights represents a uniquely powerful approach to managing credit risk, one that has the potential to empower MFIs to serve their communities better while simultaneously minimizing the financial risks associated with lending. By embracing this symbiotic path, MFIs can unlock a new era of sustainable and inclusive growth, fostering lasting positive change within their communities. Table 4.14 is provided to compare Objective III and Objective IV.

Table 4.15 Comparison of the objectives

<b>Factor</b>	<b>Machine Learning Insights (Objective 3)</b>	<b>Practitioner Validation (Objective 4)</b>	<b>Alignment/Discrepancy</b>
Debt-to-Income Ratio (DTI)	Ranked as the most critical factor for default prediction.	Confirmed as essential for assessing repayment capacity and financial health.	Strong alignment.

Installment Size	Highlighted as a significant predictor of financial strain.	Emphasized for its role in affordability and reducing default risk.	Strong alignment.
Annual Income	I identified as a top contributor, which impacts repayment ability.	Confirmed as vital for understanding earning patterns and loan eligibility.	Strong alignment.
Loan Amount	More significant loan amounts are linked to higher default risks.	Validated for its influence on overall debt burden and sustainability.	Strong alignment.
Interest Rate	Higher rates are associated with increased default risk.	Highlighted for its impact on affordability and repayment behavior.	Strong alignment.
Purpose of Loan	Loans for consumption are deemed riskier than income-generating ones.	Supported for its role in risk assessment and borrower motivation.	Strong alignment.
Credit History (Accounts)	It was found relevant but less critical for predicting defaults.	Emphasized for providing deeper insights into borrower credit behavior.	Moderate alignment; practitioners value more.
Public Records	Minimal contribution to prediction models.	Acknowledged but noted as outdated or incomplete for practical use.	Partial alignment; practitioners rely on less.

Loan Terms	Longer terms add uncertainty but have a smaller impact overall.	Noted as secondary but relevant in specific cases.	Strong alignment on limited influence.
Climate Risks	Not considered in machine learning analysis.	Highlighted as increasingly crucial in agriculture-dependent regions.	Discrepancy: practitioners suggest inclusion.

An alignment/ discrepancy column will be added to highlight areas of agreement or gaps between the two perspectives to enhance the comparative analysis between machine learning findings and practitioners' insights regarding loan default factors. Practitioners emphasized climate risks, cash flow analysis, and community reputation, which are crucial in real-world lending but often not captured in traditional machine learning models. This column will offer actionable insights, guiding the refinement of predictive models to incorporate quantitative data, like DTI ratio and installment size, and qualitative insights, such as assessing climate-related risk. This comprehensive approach aims to improve the accuracy and relevance of the risk assessments, leading to more informed lending decisions and better financial inclusion outcomes.

#### 4.6 SUMMARY

The chapter presents the study's core findings and detailed analysis, focusing on the effectiveness and performance of the proposed model in achieving the research objectives. It begins with an overview of the model performance based on collected data and evaluation metrics, emphasizing both preprocessing steps and methodologies, assessing model performance using metrics such as exactness, value of precision, recall, and F1 measure, which objectively evaluated its performance outside controlled environments, comparison it to existing models and highlighted advantages and area for improvement. Statistical analysis is presented through tables and charts, demonstrating the model's strengths in precision and recall while identifying optimization opportunities. The discussion contextualizes findings relative to original research

questions, interpreting results, connecting them to existing literature, and addressing implementation challenges that affect effectiveness. The chapter concludes by summarizing key results and discussing their implications for the field, suggesting future directions for model development. Overall, the chapter effectively communicates the study's outcomes with a comprehensive analysis of the model's performance, providing insights into both practical applications and theoretical significance while outlining pathways for future research.



## CHAPTER FIVE

### DISCUSSION, CONCLUSION AND FUTURE DIRECTIONS

#### 5.1 INTRODUCTION

The prediction of loan default in Microfinance banks (MFBs) represents a significant challenge, particularly within Pakistan's dynamic and economically diverse context (Maeenuddin et al., 2024). This study critically evaluates how the ML model can be leveraged to mitigate loan default risks, improve lending decisions, and enhance financial sustainability for MFBs using actual world data from Pakistani MFBs. This study identifies key borrower and loan-specific factors contributing to loan default while determining the most effective ML algorithm for accurate prediction.

In addressing these objectives, the study first seeks to uncover the specific factors that lead to loan default in microfinance banks in Pakistan. Understanding these factors is essential for developing targeted risk management strategies that can enhance lenders' decision-making process (Rehman, Muhammad, Sarwar, & Raz, 2019). Additionally, assessing various ML algorithms within the context of Pakistani Microfinance banks will provide insights into which models offer the best predictive performance. This evaluation is crucial as it informs MFBs about the most suitable tools for forecasting defaults, enabling them to allocate resources more effectively and reduce potential losses.

Moreover, the study recognizes the importance of macroeconomic factors influencing loan default predictions. By identifying these external variables, MFBs can better understand the broader economic landscape in which they operate and adjust their lending practices accordingly; this holistic approach not only aids in improving credit risk assessment but also contributes to the overall stability and resilience of the microfinance sector in Pakistan (Sarfo, Zhang, O'Kane, & O'Kane, 2024).

Tackling credit risk is paramount for enhancing lending practices in developed economies like Pakistan. Improved risk assessment methodologies can lead to more sustainable financial systems, fostering increased access to credit for underserved populations (Ali & Sajid, 2020). Integrating advanced ML techniques into traditional

leading frameworks holds great promise in this regard. As MFBs adopt these innovative approaches, they can enhance their capacity to predict default accurately, ultimately leading to more informed lending decisions and improved financial outcomes.

Moreover, the given discussion is based on the two theoretical prisms Credit Risk Theory and Adverse Selection Theory. The Credit Risk Theory focuses on the consideration of financial and behavioral profiles of the borrowers that allows estimating the probability of the loan default. On the other hand, Adverse Selection Theory proposes that the borrowers who are the keenest to obtain the loans usually have the highest probability of defaulting, and it becomes very prudent that the lenders should come up with a strict screening system. The idea behind the interpretation of the predictive performance of the machine learning algorithms in detecting the default-prone borrowers can be found in these theories.

### 5.1.1 Algorithm Comparison

Comparing some different machine learning algorithms of loan default prediction training Gradient Boosting (GB), Random Forest (RF), Logistic Regression (LR) and Decision Tree (DT), a few points can be noted:

- **RFs Superior Performance:** RF algorithms excel in handling complex datasets due to their ensemble nature, which combines multiple decision trees to improve prediction accuracy. This robustness effectively allows them to manage various input features and interactions (Lu, Li, & Wu, 2024).
- **Limitations of Decision Trees and LR:** While DTs are easy to interpret, they are prone to overfitting, mainly when dealing with noisy data. LR, on the other hand, may struggle with capturing complex relationships among variables, leading to limitations in recall metrics. Both algorithms may not perform well when identifying true defaulters is critical (Ahmed, Nizam, Sajid, & Qamar, 2024).
- **GB Balance between Value of Precision and Probability of Detection:** GB algorithms provide a favorable balance between precision and recall by

building models iteratively. This iterative approach allows them to correct errors made by previous models, improving overall performance. GB is particularly effective in scenarios where false positives and negatives carry significant consequences (Máté, Raza, & Ahmad, 2023).

The marginal learning performance of GB against RF can be explained by sequential learning mechanism of GB where one tree learns to correct the mistakes of an earlier tree. This supercharging technique promotes GB through much better management of bias-variance trade-offs, as well as offering the opportunity to fine-tune predictions in unbalanced or noisy data. Thus, GB is better able to strike the right balance between the precision and recall at loan default prediction tasks (Mát

In contrast to the individuals who found competitive results using Logistic Regression (Rehman et al., 2019), in this study, LR has shown poor results in recall and F1-score. Such difference could be attributed to the fact that LR cannot record non-linear relations that are found in MFB data. On the same note, although Zhang et al. (2023) focused on credit score and employment status, our models focused on DTI and income // indicates that the nature of the borrowers tends to be different across regions.

Table 5.1 Comparative Analysis of ML Models

Model	Accuracy	Precision	Recall	F1-Score
LR	77.39%	51.74%	8.74%	14.95%
DT	68.9%	31.15%	30.32%	30.73%
RF	83.5%	75.6%	60.2%	67.1%
GB	85.4%	78.3%	65.4%	71.3%

The result of this objective offers valuable insights into the effectiveness of various ML models for predicting loan default. Notably, the random forest model achieved an F1 score of 67.1%, indicating its robustness compared to LR, which faced challenges with a recall rate of only 8.74%, as illustrated in Table 5.1. These findings suggest ensemble methods, such as RF, demonstrate better generalization across diverse borrower profiles and loan characteristics. These results are consistent with the result content by (Zhang, Li, Xiang, & Worthington, 2023), which identifies the DTI ratio as

a key predictor of loan default in comparable financial contexts. These empirical results are in line with Credit Risk Theory, according to which the financial history of a borrower, his/her ability to have income, and his/her records of repayment are the crucial factors of his/her possibility to default. Furthermore, the fact that such characteristics are prevalent in the performance of models, once more proves the Adverse Selection Theory as what can be truly said is that high-risk borrowers tend to stand out in various ways compared to those people with creditworthiness, hence the introduction of carefully constructed data-driven screening functions. The significant influence of DTI on loan default prediction is due to the strong performance of the RF models, which effectively managed this feature alongside others, including income level and loan amount.

Although Decision trees are suitable for interpretability, the study revealed that they exhibited poor accuracy, primarily due to overfitting a common issue with single tree models that struggle to generalize effectively; they also demonstrated better performance, particularly in recall, as it failed to identify a significant number of actual defaulters. In contrast, the proposed model showcased markedly improved performance metrics, underscoring its effectiveness in accurately predicting loan default. This enhanced capability can capture complex relationships within the data and its robustness against overfitting. The findings strongly suggest that more sophisticated models, such as RF and GB, are better suited for predicting loan default in the context of Pakistan MFBs. These results are in line with the existing study by Lu, Li, and Wu (2024), which indicated the strength of ensemble methods such as RF and GB to perform financial prediction tasks. On the same note, DTI ratio and income level emerged as the greatest predictors in the study by Zhang et al. (2023). Further, Rehman et al. (2019) accentuated the need of borrower-specific information within Pakistani microfinance contexts, which further supports the application of the models to the local lending environment. Practically, these findings suggest that RF or GB models might present large opportunities to upgrade the credit screening that MFBs in Pakistan can perform and, as a result, minimize non-performing loans and streamline the portfolio risk. By leveraging these advanced techniques, financial institutions can enhance risk assessment processes and make more informed lending decisions.

## 5.2 PRACTICAL IMPLICATION FOR MICROFINANCE BANKS

MFBs face significant challenges in managing loan defaults, and adopting ML techniques can provide effective strategies to mitigate these risks. The following outlines practical implications for MFBs regarding strategies to reduce loan defaults, the benefits of adopting ML, and the limitations they may encounter.

### 5.2.1 Strategies to Reduce Loan Defaults

1. **Enhanced Borrower Profiling Using ML Predictions:** By utilizing ML algorithms, MFBs can improve borrower summarising, allowing for more accurate assessments of loan aspirants (Ramya, Babu, Rangarajan, & Dixit, 2024). ML models can analyze historical data to identify patterns and characteristics associated with loan defaults, enabling institutions to make knowledgeable lending decisions and target borrowers less likely to default (Susana, 2024).
2. **Tailoring Loan Terms and Interest Rates Based on Risk Assessments:** ML predictions can facilitate the customization of loan terms and interest rates according to the assessed risk of individual borrowers (Sayed, Alabrah, Rahouma, Zohaib, & Badry, 2024). The higher-risk debtors may receive loans with adjusted terms that reflect their repayment capabilities, potentially reducing default rates while still providing access to credit.

### 5.2.2 Benefits of Adopting ML

Adopting ML-based systems can revolutionize risk management in Microfinance banks through several key benefits:

- **Enhanced Decision-Making:** ML enables better profiling high-risk borrowers by analyzing vast amounts of data, leading to more informed lending decisions. This improved decision-making process helps MFBs identify potential defaulters early on.

- **Operational Efficiency:** ML systems streamline operational processes by automating **assessments**, resulting in faster loan approvals. This efficiency enhances customer satisfaction and allows MFBs to allocate resources more effectively (Mondol et al., 2024).
- **Strategic Insights:** With data-backed strategies derived from ML analyses, MFBs can customize their loan products to meet the needs of different borrower segments better. Using predictive insights, MFBs are able to create financial products including smaller instalment loans or income-adjustment repayment plans fit to the borrower and less prone to be defaulted.

All of these empirical insights are not of mere academic value; the MFB operations are directly affected. As it can be observed in such studies as Sarfo et al. (2024), predictive ML tools can temporarily align the financial resilience of microfinance institutions by reducing human bias and approving loans faster. This in particular is very beneficial in Pakistan in the context of formal credit scoring systems being non-available. Hence, use of such interpretable and accurate models as GB can result in more intelligent credit decisions by both parties; deploying such models can be in the best interest of borrowers and lenders.

### 5.3 KEY BORROWER-LEVEL RISK FACTORS

This analysis compares various studies and approaches to understanding the factors influencing loan default prediction within MFBs in Pakistan. The focus is on the borrower’s characteristics, loan character sets, and the methods employed for prediction, as seen in Table 5.2.

Table 5.2 Borrower-level risk factors

Factor	Characteristics	Relevance to default prediction	Comparative insights
Loan Amount (loan_amt)	The total sum borrowed typically ranges	More significant loan amounts can increase the risk of default if	Studies indicate that larger loans correlate with higher default

	from pkr 60000 to 100000.	debtors cannot manage repayments.	rates due to repayment challenges (Jumaa, Saqib, & Attar, 2023).
Loan Term (term)	The duration of repayment is commonly set at 12 36 months.	Longer terms may reduce monthly payments but increase overall interest costs and risk.	Research shows mixed results; while longer terms may ease immediate repayment pressure, they can lead to higher cumulative debt (I. Ahmed, Diaz-Rainey, Roberts, & Nguyen, 2022).
Interest Rate (int_rate)	The cost of the loan is expressed as a percentage, generally between 18% and 35%.	Higher interest rates can strain debtors' finances, increasing default likelihood.	Studies consistently find that higher interest rates correlate with increased default rates due to affordability issues (Uwais & Khaleghzadeh, 2022).
Installment Size	Monthly payment amount, which can range from 3000 to 5000.	Larger installment sizes may exceed debtors' repayment capabilities, leading to defaults.	Analysis shows that borrowers with high monthly payments relative to income are at greater risk of defaulting on their loans (Anand, Velu, & Whig, 2022).
Loan Status	Current status of the loan (Fully Paid, Charged Off), indicating repayment outcome.	The status provides direct insight into borrower behavior and repayment outcomes over time.	Analysis of loan statuses reveals that early indicators of trouble often appear in payment patterns before formal defaults

			occur, allowing for proactive measures by lenders.
Purpose of the Loan (purpose)	Intended use of the loan funds (Car, business, education, and wedding).	Specific purposes may inherently carry more risk; for example, business loans may be riskier than personal loans due to market volatility.	Comparative studies indicate that personal loans for essential needs tend to have lower default rates than discretionary spending loans like vacations or luxury items (Fersi & Abbes, 2023).
Debt-to-Income Ratio (dti)	Ratio comparing monthly debt payments to income; lower values indicate better financial health.	A lower DTI suggests better financial health and a reduced likelihood of defaulting on loans.	Research consistently shows that a DTI above 40% is a significant predictor of potential defaults across various lending contexts, including microfinance sectors (L. Zhang, Wang, & Liu, 2023).

The comparative analysis highlights how each factor contributes uniquely to predicting loan default within MFBs in Pakistan. By understanding the characteristics ranging from borrowers' demographic and employment stability to loan attributes like amount and purpose, MFBs can enhance their risk assessment framework and develop more effective predictive models. Integrating these insights into lending practices not only aids in identifying high-risk borrowers but also supports informed decision-making processes that ultimately contribute to more excellent financial stability in the microfinance sector in Pakistan.

## 5.4 MACHINE LEARNING MODEL PERFORMANCE ACROSS STUDIES

This analysis compares the performance of various ML algorithms in predicting loan defaults, focusing on accuracy, precision, recall, and F1 scores. The data presented in the query is compared with findings from recent studies to provide a broader context for evaluating these MFI Pakistan models.

Table 5.3 Comparative Analysis of ML Performance

Model	Comparative Insights
<b>Decision Tree</b>	A study reported a significantly higher accuracy of 99.62% for Decision Trees, indicating that model tuning and dataset quality can substantially influence performance (Oghenekaro & Chimela, 2023).
<b>Logistic Regression</b>	Logistic Regression is often noted for its interpretability; however, it typically performs less than ensemble methods like Random Forest and XGBoost (Alvi & Arif, 2024).
<b>Random Forest</b>	Random Forest has achieved an accuracy of around 99.62% in various studies, making it one of the top performers in loan default prediction (Pradnyana & Rahadi, 2022).
<b>Gradient Boosting</b>	Gradient Boosting models, particularly XGBoost, have demonstrated superior performance with high accuracies and F1 scores, often exceeding those of Random Forest (Uphade, Muley, & Chalwadi, 2024).

The comparative analysis indicated that while the models in Table 5.3 listed in the query provide a baseline for understanding loan default prediction performance, they fall short compared to the findings. The significant difference in performance metrics underscores the importance of model selection and tuning in achieving accurate loan default prediction within MFI in Pakistan.

## 5.5 TOP PREDICTIVE FACTORS AND THEIR WEIGHT

Analyzing factors contributing to loan charge-offs provides critical insights into the dynamic of loan default within MFBs in Pakistan. Several key factors significantly influence the likelihood of a borrower defaulting on their loan with DTI, such as installment amount, annual income, total number of accounts, loan amount, and interest rate, which are emerging as the most impactful.

**Debt-to-Income Ratio (DTI)** is the most significant predictor of loan charge-offs, with a normalized importance of approximately 14%. This metric reflects a borrower's ability to manage monthly debt payments relative to their income. A high DTI indicates that a significant portion of a borrower's income is allocated to debt repayment, which can lead to financial strain and an increased risk of default. This finding aligns with existing literature that emphasizes the importance of DTI as a critical indicator of financial health and repayment capacity.

**Installment Amount**, ranked second with over 12% importance, highlights the financial burden that large monthly payments can impose on borrowers. When borrowers face high installment amounts relative to their income, they may struggle to meet their financial obligations, increasing the likelihood of default. This observation is consistent with studies indicating that manageable installment sizes are crucial for maintaining borrower solvency.

**Annual Income**, contributing around 11%, directly correlates with a borrower's ability to repay loans. Lower income levels are frequently associated with higher default rates, mainly when unexpected expenses arise. This relationship underscores lenders' need to consider income stability when assessing credit risk.

**Total Number of Accounts** also significantly predicts defaults, with an importance rating of about 10%. A higher number of active credit accounts may indicate that borrowers are juggling multiple debts, complicating their financial management and increasing the risk of missed payments.

**Loan Amount and Interest Rate**, each contributing around 9%, further illustrate how larger loans and higher borrowing costs can elevate default risks.

Borrowers with substantial loans may struggle to maintain repayment schedules, especially if interest rates significantly increase their financial burden.

In contrast, factors such as **Open Accounts**, **Public Records**, **Loan Terms**, and the Purpose of **Debt Consolidation** were identified as less influential in predicting loan defaults. While these factors remain relevant, more critical indicators like DTI and installment amounts overshadow their impact.

The high predictive power of the debt-to-income (DTI) ratio also points to the Credit Risk Theory that emphasised repayment capacity as the main indicator of borrower reliability (Altman Saunders, 1998). High DTI implies too much financial pressure which is more likely to fail (Zhang, Wang, & Liu, 2023). On the same note, low income levels mean that the cash flow will be low, thereby restricting the capacity of a borrower to cushion against financial shocks (Sarfo, Zhang, O'Kane, & O'Kane, 2024). In behavioral finance terms, unaffordable loan amounts and huge installments can be psychologically bullying to borrowers, more so in lower-income communities with the consequences being an increase in the default propensity (Bhagat & Reddy, 2022; Banerjee, Karlan, & Zinman, 2015).

To apply the principles of duplication in factors of comparisons which do not overlap, some major points in the previous ML performance tables (Tables 5.2 and 5.3) will be condensed here. The idea is to simplify the argument and concentrate on the best predictors that come out consistently in the models and empirical studies.

With Gradient Boosting, a unit change in DTI is linked with an approximate surge in default possibility of around 12 to 15 percent thus showing its high marginal impact. Likewise, 10,000 PKR rise in loan value raises the risk of default approximating to 8 per cent. These effect sizes establish operational importance of each of the predictors in the risk dynamics of borrowers.

Table 5.4 Comparative analysis synthesizes findings from various studies

Factor	Comparative Findings
<b>Debt-To-Income Ratio (DTI)</b>	Consistent across multiple studies, high DTI is often cited as a leading indicator for predicting defaults (Uphade et al., 2024).

<b>Factor</b>	<b>Comparative Findings</b>
<b>Installment Amount</b>	Similar findings were reported in other research where high installment amounts were linked to higher default rates (Akinjole, Shobayo, Popoola, Okoyeigbo, & Ogunleye, 2024).
<b>Annual Income</b>	Across studies, income remains critical; lower income levels are consistently associated with increased defaults (Owusu, Quainoo, Mensah, & Appati, 2023).
<b>Total Number of Accounts</b>	This is supported by literature indicating that borrowers with numerous accounts face more significant challenges in repayment (Uddin et al., 2023).
<b>Loan Amount</b>	Research shows that larger loan sizes are linked to increased default probabilities.
<b>Open Accounts</b>	While relevant, its impact is often overshadowed by more critical metrics like DTI and income.
<b>Public Records</b>	Studies confirm that while public records matter, they are less predictive than DTI and income.
<b>Loan Terms</b>	Generally acknowledged but often ranked lower than other metrics in predictive models (Krasovytskyi & Stavvytsky, 2024).

Identifying factors contributing to the loan default provides valuable insights for MFI in Pakistan. Lenders can assess borrowers' risk profiles by focusing on key indicators such as DTI and installment amount while considering annual income and total accounts.

Although the machine learning models carried very good scores in terms of their empirical performance, it was important to provide the domain expert to proofread such results gained. Thus, in the next section, expert interviews of bank loan officers on the applicability and interpretation of the top predictive factors are offered.

## **5.6 COMPARISON OF OBJECTIVE 04: TO VALIDATE THE TOP FACTORS INFLUENCING LOAN DEFAULT THROUGH BANK LOAN OFFICERS**

Based on insights from bank loan officers, the comparative analysis of the factors influencing loan default can be structured around the key theme derived from the provided search results. This analysis will address the relevance of factors, additional considerations, practical implications, and suggestions for improvement in loan assessment models.

**Marital Status and Income:** Studies highlight that marital status and income levels significantly influence loan repayment capabilities. Higher income generally correlates with lower default rates, while marital status may impact financial stability.

**Loan Diversion and Monitoring:** The diversion of funds for non-productive uses is frequently cited as a critical factor leading to default. Proper monitoring by loan officers can mitigate this risk.

**Economic Conditions:** Macroeconomic factors such as interest rates and market volatility also play a crucial role. For instance, rising interest rates or economic downturns can increase default likelihood.

**Critical Factors:** Employment status, age, and loan amount are often seen as more critical than others. For example, younger borrowers may exhibit better repayment performance due to their potential for income growth.

**Behavioral Factors:** Beyond traditional metrics, financial literacy, and risk perception can influence loan default intentions. Young adults, in particular, show varying default intentions based on these behavior factors.

**Sector-Specific Characteristics:** Certain sectors, such as manufacturing and agriculture, have higher default rates due to inherent risks associated with those industries. Understanding these sector-specific dynamics is essential for accurate risk assessment.

The observations made by bank loan officers support the strength of the machine learning results provided in the previous sections. It is worth noting that some of the most important variables, including debt-to-income ratio, loan amount, income level,

and loan diversion, were repeatedly cited by both the predictive models and domain experts. This congruence enhances the validity of the study indicating that the models did not only work out well in a statistical sense but also reflected the financial realities as seen by the practitioners. In addition, the feedback on the behavior of the borrowers, sector-specific risks, and monitoring issues offer important qualitative background that facilitates the practical usefulness of the predictive features. In this way, the incorporation of expert validation contributes to the theoretical soundness and the practical stability of the research results.

## **5.7 CRITICAL REVIEW OF THE WORK**

This study critically examines ML models' application in predicting loan default within Microfinance Banks in Pakistan, a context marked by unique economic and dynamic challenges. ML models, mainly RF and GB, effectively predict loan defaults in Pakistani MFBs, outperforming traditional models like LR. DTI, installment ratio, and annual income are significant predictors, with RF at 67.1% and GB at 71.3% achieving better F1 scores than LR at 8.74%.

## **5.8 THEORETICAL CONTRIBUTION**

This paper is relevant to the theoretical literature because it combines Credit Risk Theory and Adverse Selection Theory in the predictive modeling of loan defaults with machine learning (ML). Although the literature has examined credit scoring or ML performance, not many have anchored their results to the existing financial theory. The study empirically confirms that the characteristics of a borrower, including DTI ratio, income, and installment amount, have been found to be consistent predictors of default, and therefore, the study offers evidence that reinforces and builds on these underlying theories. Moreover, the use of ensemble ML models illustrates how such theories can be implemented into data science to improve the credit risk classification in microfinance institutions (Rehman et al., 2019; Zhang et al., 2023).

The paper validates the premises of the Credit Risk Theory concerning risk indicators on the borrower level and expands the Adverse Selection Theory by proving

empirically that data-based screening systems can be more effective than subjective credit evaluation in the context of high-risk lending, such as microfinance.

## **5.9 PRACTICAL CONTRIBUTION**

In practice, the research provides practical implications to microfinance banks (MFBs) in situations of resource-limited and underserved environments like Pakistan. The determination of high-impact borrower-level variables can help MFBs to enhance the credit screening procedures and reduce the defaults. Furthermore, the fact that Gradient Boosting and Random Forest models have shown to be superior to the traditional methods (e.g., Logistic Regression) indicates that the scalable solution to improving lending decisions can be achieved. Such insights are especially important as the industry depends on non-traditional borrowers and there is no formal credit history. The study offers a guide to data-driven, inclusive lending by matching predictive accuracy with feedback of loan officers in the real world (Sarfo et al., 2024).

## **5.10 METHODOLOGICAL CONTRIBUTION AND APPLIED CONTRIBUTION**

Methodologically, the study is a contribution because it compares several ML algorithms in a systematic way with a real-world dataset of Pakistani MFBs. It proposes a repeatable assessment system founded on precision, recall, and F1-score, which allows comparing the performance of various models. Moreover, the contribution of the implemented method is the confirmation of the model forecasts by the expert opinion of bank loan officers, which is a new combination of quantitative and qualitative data, which enhances the practical credibility of the results. The mixed-methods validation strategy enhances the application of ML in the financial risk assessment (Ahmed et al., 2024; Hoque, 2025).

## **5.11 CONCLUSION**

The combination of sophisticated ML models and financial characteristics of borrowers provides an inclusive approach to improving credit risk evaluation in the microfinance market in Pakistan. This study connects the model performance with the actual decision-making in financial institutions rather than perceiving predictive modeling as a technical task. The main conclusions of all objectives confirm that debt-to-income ratio, income stability, and the size of installments are stable predictors of default, and ensemble techniques, including GB and RF, are the most appropriate to model these non-linear dependencies. These findings are also reinforced by validation with bank officers regarding the practical reliability of these findings. This combination of quantitative modeling and qualitative understanding can give a basis of real-world application.

### **5.11.1 Model Performance Summary**

Among the models assessed, GB emerged as the leading performer with an F1 score of 71.3%, closely followed by RF at 67.1%. In contrast, despite its interpretability, LR struggled to accurately identify true defaulters, as indicated by its low probability of detection scores of 8.74%. These results emphasize the need for advanced ML techniques to tackle complex prediction challenges like forecasting loan defaults effectively.

The GB model has an F1-score of 71.3%, which means that it is well balanced in terms of identifying real defaulters and false positives. In the case of Microfinance Banks, this means better credit decisions, better portfolio quality and less write-offs. Such improvement in performance can facilitate more inclusive and stable lending in a developing economy such as Pakistan.

### **5.11.2 Research Contributions**

Building on the model comparisons and factor analysis, this subsection highlights the study's practical impact. This study effectively bridges the gap between traditional lending practices and advanced ML methodologies, providing Microfinance banks with

scalable tools to mitigate loan default risk. The application of these tools not only enhances the accuracy of credit risk assessments but also improves methods for evaluating borrower risks.

### **5.11.3 Future Directions**

The future directions proposed here stem directly from observed limitations and predictive gaps in existing approaches. To enhance the implementation of ML in predicting loan defaults, several recommendations can be made:

#### ***5.11.3.1 Use of Advanced Ensemble Techniques***

Leveraging advanced ensemble methods such as stacking or blending can improve predictive performance by combining multiple models to capture different aspects of borrower behavior (Bhoite, Thatte, More, & Ruikar, 2023).

#### ***5.11.3.2 Incorporating Alternative Data Sources***

Utilizing alternative data sources, such as transaction histories, social media activity, or mobile payment patterns, can provide a more comprehensive view of borrowers and enhance profiling accuracy (Tigges, Mestwerdt, Tschirner, & Mauer, 2024).

#### ***5.11.3.3 Addressing Data Quality and Privacy Concerns***

Ensuring high data quality is crucial for practical model training. Additionally, addressing privacy concerns related to borrower data is essential to maintain trust and comply with regulations (Kowalewski, Pisany, & Ślęzak, 2022).

#### ***5.11.3.4 Longitudinal Studies***

Conducting long-term studies to analyze how changes in borrower behavior and economic conditions affect loan default rates over time (Ranjan, 2025).

#### ***5.11.3.5 Real-time Predictive Models***

Developing a real-time predictive system integrated with digital lending platforms will allow MFBs to dynamically monitor borrower behavior, enabling proactive interventions (Omowole, Urefe, Mokogwu, & Ewim, 2024b). By leveraging real-time data, such as historical transactions or income changes, MFI can identify signs of distress early and take steps to mitigate the risk of default.

#### ***5.11.3.6 Impact of Policy Changes***

Analyzing how regulatory changes in the microfinance sector influence default rates and borrower behavior (Durango-Gutiérrez, Lara-Rubio, & Navarro-Galera, 2023).

#### ***5.11.3.7 Training for Loan Officers***

Implementing training programs for loan officers based on research findings will improve their assessment skills and methodologies in evaluating loan applications (Hoque, 2025).

#### ***5.11.3.8 Sector-specific Models***

A promising avenue for future research involves tailoring ML models to specific sectors, such as agriculture or manufacturing, where unique risk factors exist. For example, income seasonality could significantly affect repayment behavior in agriculture, requiring a specialized risk model to improve prediction exactness. By integrating sector-specific variables into predictive models, MFBs can better assess risk and offer more precise loan products (Jonas, Christian, Ntombela, & Letsoalo, 2025).

The findings of this study highlight the critical factors influencing loan defaults within MFBs in Pakistan and demonstrate the effectiveness of ML algorithms in predicting these defaults. By identifying key indicators such as DTI, annual income, and installment amounts, the research provides valuable insights for lenders seeking to enhance their risk assessment frameworks. This study significantly contributes to the microfinance sector and the ML field by bridging the gap between traditional lending practices and modern predictive analytics. The insights gained from this research inform better lending decisions and pave the way for future advancements in risk management strategies within microfinance institutions. By integrating ML techniques into their operations, MFBs can improve their ability to serve underserved populations while minimizing financial risks associated with loan defaults. The highlighted sections represent the newly incorporated content regarding real-time models, data privacy and ethics, and sector-specific models.

#### ***5.11.3.9 Macroeconomic Factors***

Future research should integrate macroeconomic variables such as GDP growth, inflation, and interest rates into predictive models to assess their impact on loan defaults. Examining the relationship between borrower-specific factors and economic conditions will provide a more comprehensive risk assessment framework. Longitudinal datasets should also be explored to understand the temporal effects of monetary changes on repayment behaviors, enabling a more dynamic and holistic understanding of credit risk (Omowole, Urefe, Mokogwu, & Ewim, 2024a).

The macroeconomic variables are very important in determining the behaviour of borrowers and the risk of default. As an example, inflation causes a reduction in real income thus, it becomes difficult for the borrowers to repay the loans. On the same note, the increase in unemployment rates lessens the stability of household income, making it more probable that the payments will be missed. Researchers like Durango-Gutierrez et al. (2023) and Omowole et al. (2024) stress that the risk of loan default is usually cyclical, and it is highest during economic recessions. Within the framework of the unstable economy of Pakistan, such indicators can be used to improve the capability of MFBs to be proactive in risk management. Thus, ML models which dynamically

combine GDP growth, inflation rates, and monetary policy changes provide a more realistic and complete picture of the credit risk over the time.

#### **5.11.4 Policy and Practical Implications**

This study has a number of important policy and practical implications to the Microfinance Banks (MFBs), financial regulators and future research stakeholders working in the credit environment of Pakistan.

To begin with, the fact that ensemble-based machine learning models like Gradient Boosting and Random Forest perform better than other models highlights the possibility of data-driven credit evaluation systems in the microfinance industry. These models are more effective than the traditional logistic regression as they can identify nonlinear relationships and complicated borrower patterns, as suggested by Lu et al. (2024) and Pradnyana & Rahadi (2022). Hence, it is suggested that MFBs implement ensemble algorithms in the process of risk profiling borrowers in order to enhance predictive performance and efficiency.

Second, the paper emphasizes the significance of the loan product customization according to the individual risk scores. The interest rates, the sizes of installments, and the terms of repayment can be customized to each borrower according to his or her ability to repay based on ML-generated scores, a strategy that is also supported by Uphade et al. (2024) and Ahmed et al. (2024) who state that personalized risk-based lending can help decrease default rates and increase financial inclusion.

Third, it is of utmost importance that MFBs and regulatory bodies like the State Bank of Pakistan (SBP) work together. Regulators should facilitate AI-based lending activities by innovating policies and supervisory structures in a way that makes automated credit decisions explainable, reasonable, and ethically acceptable. According to Kowalewski et al. (2022), financial algorithm governance frameworks are necessary to ensure transparency, prevent algorithmic bias, and align ML implementation with national financial inclusion objectives.

Fourth, the results emphasize the need to build capacity of frontline loan officers. Training sessions that would increase the knowledge of the officers about ML insights

and their practical application in the credit evaluation are necessary. This reinforces a mixed model of decision-making, which is the integration of human judgment and algorithmic suggestions, as suggested by Hoque (2025).

Lastly, to implement the ML tools, there is a need to standardize the practices of collection of borrower data among MFBs. Poor or incomplete data sets decrease the reliability of the model and limit the applicability of the information. The creation of a common digital framework, possibly in the form of a public-private partnership, would enhance the quality of data and increase the fairness of access to predictive credit systems (Tigges et al., 2024).

Overall, these implications support a technologically advanced but ethically conscious modernization of the microfinance industry in Pakistan, based on machine learning that would strike the right balance between financial risk and social development goals.

#### **5.11.5 Ethical Considerations and Alignment with Islamic Finance**

When machine learning (ML) is incorporated in credit risk assessment, it brings up a number of ethical issues, especially in data-sensitive settings such as microfinance. The most important concerns are the possibility of algorithmic bias, explainability, and data privacy breaches. Without being resolved, these problems may end up strengthening the financial exclusion of already vulnerable groups. Consequently, Microfinance Banks (MFBs) and regulatory agencies must guarantee that the model design is transparent, privacy-preserving data methods are used, and ML systems are audited on a regular basis, as recommended by international ethical AI principles (Kowalewski et al., 2022; Tigges et al., 2024).

In Islamic finance terms, application of ML should be considered in light of the concepts of justice (adl), fairness (ihsan) and non-exploitation (gharar). The practical implementation of ML in the given research follows these principles as it encourages unbiased, data-driven lending practices, minimizes human bias, and increases access to credit among the underserved groups. As long as the models do not entail interest-based lending systems and unfair discrimination, their application in credit assessment does

not necessarily contravene the Islamic finance ethics (Ahmad, 2023). In this way, ML tools can be ethically and theologically aligned with Islamic microfinance systems with proper protection.

## **5.12 SUMMARY**

This research provides a comprehensive framework for mitigating loan defaults within Pakistani MFBs. By demonstrating the effectiveness of ML algorithms, mainly RF and GB, and identifying key factors influencing loan defaults, this study offers valuable insights for lenders seeking to enhance their risk assessment frameworks. The integration of insights from bank loan officers further strengthens the practical relevance of these findings. This study bridges the gap between traditional lending practices and modern predictive analytics, paving the way for future advancements in risk management strategies and promoting financial inclusion in the microfinance sector. Further research should explore advanced ensemble techniques, incorporate alternative data sources, and address data quality and privacy concerns to improve the implementation of ML in predicting loan defaults.

## REFERENCES

- Abbas, F., Iqbal, S., & Aziz, B. (2019). The impact of bank capital, bank liquidity and credit risk on profitability in postcrisis period: A comparative study of US and Asia. *Cogent Economics & Finance*, 7(1), 1605683.
- Abbott, D. (2014). Applied predictive analytics: Principles and techniques for the professional data analyst. *John Wiley & Sons*.
- Abubecker, F., Sherwani, F. K., Nomani, A., & Shafeeq, I. Y. (2020). Comparative Analysis of Conventional and Non-Conventional Microfinance. *International Journal of Management, Technology and Engineering*, 8(10), 481-488.
- Abdul Rahman, R., & Dean, F. (2013). Challenges and solutions in Islamic microfinance. *Humanomics*, 29(4), 293–306.
- Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks*, 6(2), 38.
- Adewusi, A. O., Oyedokun, T. B., & Bello, M. O. (2016). Application of artificial neural network to loan recovery prediction. *International Journal of Housing Markets and Analysis*, 9(2), 222-238.
- Ahmed, F., Nizam, K., Sajid, Z., & Qamar, S. (2024). Striking a Balance: Evaluating Credit Risk with Traditional and Machine Learning Models. *Bulletin of Business and Economics (BBE)*, 13(3), 30-35.
- Ahmed, I., Diaz-Rainey, I., Roberts, H., & Nguyen, D. T. T. (2022). Expand or avoid: microfinance credit risk and climate vulnerability. Available at SSRN 4261881.
- Ahmed, M., Tahir, S. H., & Ahmad, G. (2023). Fostering inclusive growth: exploring gender dynamics and loan sizes in microfinance institutions. *Annals of Social Sciences and Perspective*, 4(2), 285-305.
- Akhtar, S. (2007). Building an effective Islamic financial system. *BIS Review*, 38, 1-7.

- Akhtar, S., & Pearce, D. (2010). *Microfinance in the Arab World: The challenge of financial inclusion*. (pp. 1-4). World Bank, Washington, DC.
- Akinjole, A., Shobayo, O., Popoola, J., Okoyeigbo, O., & Ogunleye, B. (2024). Ensemble based Machine learning Algorithm for Loan Default Risk Prediction. *Mathematics*, 12(21), 3423.
- Alagic, A., Zivic, N., Kadusic, E., Hamzic, D., Hadzajlic, N., Dizdarevic, M., & Selmanovic, E. (2024). Machine learning for an Enhanced Credit Risk Analysis: A Comparative Study of Loan Approval Prediction Models Integrating Mental Health Data. *Machine learning and Knowledge Extraction*, 6(1), 53-77.
- Alam, M. N. (2008). A comparative study of financing small and cottage industries by interest-free banks in Turkey, Cyprus, Sudan and Bangladesh. *Humanomics*, 24(2), 145-161.
- Alam, T. M., Shaikat, K., Hameed, I. A., Luo, S., Sarwar, M. U., Shabbir, S., . . . Khushi, M. (2020). An investigation of credit card default prediction in the imbalanced datasets. *IEEE Access*, 8, 201173-201198.
- Ali, A., & Sajid, A. (2020). Towards Inclusive Growth: Financial Sector Dynamics and Poverty Reduction in Pakistan. *Journal of Business and Economic Options*, 3(4), 129-140.
- Al-Khazaleh, M. J., Alian, M., & Jaradat, M. A. (2024). Sentiment analysis of imbalanced Arabic data using sampling techniques and classification algorithms. *Bulletin of Electrical Engineering and Informatics*, 13(1), 607-618.
- Alonso Robisco, A., & Carbo Martinez, J. M. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction. *Financial Innovation*, 8(1), 70.
- Alpaydin, E. (2016). *Machine learning: the new AI*. MIT press.
- Alvi, J., & Arif, I. (2024). Default prediction modeling (DPM) with machine learning algorithms: case of non-financial listed companies in Pakistan. *Kybernetes*.

- Ampountolas, A., Nyarko Nde, T., Date, P., & Constantinescu, C. (2021). A machine learning approach for micro-credit scoring. *Risks*, 9(3), 50.
- Anand, M., Velu, A., & Whig, P. (2022). Prediction of loan behaviour with machine learning models for secure banking. *Journal of Computer Science and Engineering (JCSE)*, 3(1), 1-13.
- Anderson, R. (2007). *The credit scoring toolkit: theory and practice for retail credit risk management and decision automation*. Oxford University Press.
- Aniceto, M. C., Barboza, F., & Kimura, H. (2020). Machine learning predictivity applied to consumer creditworthiness. *Future Business Journal*, 6(1), 1-14.
- Asrat, H. (2018). Assessment of credit management practice at United Bank SC. *Unpublished PhD Thesis. Addis Ababa University College of Business and Economics, Ethiopia*.
- Assef, F. M., & Steiner, M. T. A. (2020). Machine Learning Techniques in Bank Credit Analysis. *International Journal of Economics and Management Engineering*, 14(7), 517-520.
- Attigeri, G., Pai, M. M., & Pai, R. M. (2017, September). Analysis of feature selection and extraction algorithm for loan data: A big data approach. In *2017 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 2147-2151). IEEE.
- Bader, M., Abdelwanis, M., Maalouf, M., & Jelinek, H. F. (2024). Detecting depression severity using weighted random forest and oxidative stress biomarkers. *Scientific reports*, 14(1), 16328.
- Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, 128, 301-315.
- Behr, A., & Weinblat, J. (2017). Default patterns in seven EU countries: A random forest approach. *International Journal of the Economics of Business*, 24(2), 181-222.

- Behr, A., & Weinblat, J. (2017). Default prediction using balance-sheet data: A comparison of models. *The Journal of Risk Finance*, 18(5), 523-540.
- Bekhet, H. A., & Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance*, 4(1), 20-28.
- Ben-David, A., & Frank, E. (2009). Accuracy of machine learning models versus “hand crafted” expert systems—a credit scoring case study. *Expert Systems with Applications*, 36(3), 5264-5271.
- Bhoite, S., Thatte, S., More, A., & Ruikar, D. (2023). Loan Eligibility Verification by Using Ensemble ML Techniques. In *Machine Learning and Optimization for Engineering Design* (pp. 121-134): Springer.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
- Boughaci, D., & Alkhaldeh, A. A. S. (2018). Three local search-based methods for feature selection in credit scoring. *Vietnam Journal of Computer Science*, 5, 107-121.
- Boz, Z., Gunec, D., Birbil, S. I., & Öztürk, M. K. (2018). Reassessment and monitoring of loan applications with machine learning. *Applied Artificial Intelligence*, 32(9-10), 939-955.
- Brandsma, J., & Hart, L. (2004). Making microfinance work better in the Middle East and North Africa. *Washington, DC.: World Bank*.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453.

- Brownlee, J. (2016). *XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn*. Machine Learning Mastery.
- Brownlee, J. (2020). *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery.
- Brygała, M., & Korol, T. (2024). Personal bankruptcy prediction using machine learning techniques. *Economics and Business Review*, 10(2).
- Cermeño, R., León, J., & Varas, G. M. (2011). Determinantes de la morosidad: Un estudio panel para el caso de las cajas municipales de ahorro y crédito del Perú, 2003-2010.
- Chakure, A. (2019). Random forest regression. *Towards Data Science*. Dostupno na.
- Charreire Petit, S., & Huault, I. (2008). From practice-based knowledge to the practice of research: Revisiting constructivist research works on knowledge. *Management learning*, 39(1), 73-91.
- Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter*, 6(1), 1-6.
- Chen, C. W., Dong, M. C., Liu, N., & Sriboonchitta, S. (2019). Inferences of default risk and borrower characteristics on P2P lending. *The North American Journal of Economics and Finance*, 50, 101013.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Chen, W., Li, Z., & Guo, J. (2020). A VNS-EDA algorithm-based feature selection for credit risk classification. *Mathematical Problems in Engineering*, 2020, 1-14.
- Chizi, B., Rokach, L., & Maimon, O. (2009). A survey of feature selection techniques. *Encyclopedia of Data Warehousing and Mining, Second Edition*, 1888-1895.

- Chowdhry, S. (2006). Creating an Islamic Microfinance Model-The Missing Dimension. *Dinar Standard: Business Strategies for the Muslim World*, 4.
- Condori-Alejo, H. I., Aceituno-Rojo, M. R., & Alzamora, G. S. (2021). Rural micro credit assessment using machine learning in a peruvian microfinance institution. *Procedia Computer Science*, 187, 408-413.
- Cornelissen, M. A. P. (2018). *Applying machine learning to the prediction of defaults in loans*. (Master's thesis, University of Twente).
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Cubiles-De-La-Vega, M. D., Blanco-Oliver, A., Pino-Mejías, R., & Lara-Rubio, J. (2013). Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques. *Expert systems with applications*, 40(17), 6910-6917.
- Cull, R., Demirgüç-Kunt, A., & Morduch, J. (2007). Financial performance and outreach: A global analysis of leading microbanks. *The Economic Journal*, 117(517), F107-F133.
- Dahiya, S., Handa, S. S., & Singh, N. P. (2017). A feature selection enabled hybrid-bagging algorithm for credit risk evaluation. *Expert Systems*, 34(6), e12217.
- Dansana, D., Patro, S. G. K., Mishra, B. K., Prasad, V., Razak, A., & Wodajo, A. W. (2024). Analyzing the impact of loan features on bank loan prediction using Random Forest algorithm. *Engineering Reports*, 6(2), e12707.
- De Cnudde, S., Moeyersoms, J., Stankova, M., Tobback, E., Javalry, V., & Martens, D. (2019). What does your Facebook profile reveal about your creditworthiness? Using alternative data for microfinance. *Journal of the Operational Research Society*, 70(3), 353-363.
- De Vaus, D. (2001). Research design in social research. *Research design in social research*, 1-296.

- Deng, M., Chen, J., Huang, J., & Niu, W. (2018). Agricultural drought risk evaluation based on an optimized comprehensive index system. *Sustainability*, 10(10), 3465.
- Dm, O., & Mm, M. (2018). Comparison of Accuracy of Support Vector Machine Model and Logistic Regression Model in Predicting Individual Loan Defaults. *Am. J. Appl. Math. Stat*, 6(6), 266-271.
- Durango-Gutiérrez, M. P., Lara-Rubio, J., & Navarro-Galera, A. (2023). Analysis of default risk in Microfinance banks under the Basel III framework. *International Journal of Finance & Economics*, 28(2), 1261-1278.
- Ekinci, R., & Poyraz, G. (2019). The effect of credit risk on financial performance of deposit banks in Turkey. *Procedia Computer Science*, 158, 979-987.
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. *Machine learning in radiation oncology: theory and applications*, 3-11.
- Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, 47(1), 54-70.
- Eze, C. C., & Ibekwe, U. C. (2007). Determinants of loan repayment under indigenous financial system in Southeast, Nigeria. *Medwell journal of social sciences*,
- Fan, Z., Sohail, S., Sabrina, F., & Gu, X. (2024). Sampling-Based Machine learning Models for Intrusion Detection in Imbalanced Dataset. *Electronics*, 13(10), 1878.
- Feki, A., Ishak, A. B., & Feki, S. (2012). Feature selection using Bayesian and multiclass Support Vector Machines approaches: Application to bank risk prediction. *Expert Systems with Applications*, 39(3), 3087-3099.
- Ferhi, A. (2018). Credit risk and banking stability: a comparative study between Islamic and conventional banks. *International Journal of Law and Management*, 60(4), 1009-1019.

- Fersi, M., & Abbes, M. B. (2023). Comparative study of microfinance sustainability: Credit risk-taking and capital structure. *International Journal of Management, Economics and Social Sciences (IJMESS)*, 12(3), 184-216.
- Freeman, M., & Zeegers, M. P. (Eds.). (2016). *Forensic epidemiology: principles and practice*. Academic Press.
- Gao, X., Yang, X., & Zhao, Y. (2023). Rural micro-credit model design and credit risk assessment via improved LSTM algorithm. *PeerJ Computer Science*, 9, e1588.
- Ghalib, A. K. (2013). How effective is microfinance in reaching the poorest? Empirical evidence on programme outreach in rural Pakistan. *Journal of Business Economics and Management*, 14(3), 467-480.
- Gizaw, M., Kebede, M., & Selvaraj, S. (2015). The impact of credit risk on profitability performance of commercial banks in Ethiopia. *African journal of business management*, 9(2), 59.
- Goddard, G. J. (2011). *Islamic Finance: Principles and Practice* by Visser, H. (2010). Cheltenham UK, Northampton, MA: Edward Elgar, 184 pages. ISBN-10: 1849808821, ISBN-13: 978-1849808828. *Journal of Asia-Pacific Business*, 12(3), 304-312.
- Gonzalez, A. (2010). Analyzing microcredit interest rates: A review of the methodology proposed by Mohammed Yunus. *Mix data brief*, (4).
- Gonzalez, A. (2010). Is microfinance growing too fast?. *MIX data brief*, (5).
- Granström, D., & Abrahamsson, J. (2019). Loan default prediction using supervised machine learning algorithms. (Dissertation). Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-252312>.
- Grunow, D. (1995). The research design in organization studies: Problems and prospects. *Organization science*, 6(1), 93-103.
- Guerrero, A. M. G., Vargas, J. H., Sánchez, J. M. M., & Farías, F. J. M. (2025). A Bayesian Network Model to Evaluate the Credit Risk of Mexican Microfinance

- Institutions in 2023. *Revista Mexicana de Economía y Finanzas (REMEF): nueva época*, 20(1), 7.
- Guttman, J. M. (2008). Assortative matching, adverse selection, and group lending. *Journal of development Economics*, 87(1), 51-56.
- Harris, T. (2013). Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions. *Expert Systems with Applications*, 40(11), 4404-4413.
- Hassan, H., & Jebran, K. (2016). Challenges and Opportunities to Microfinance Institutions in Khyber Pakhtunkhwa Province, Pakistan. *Journal of Economics and Sustainable Development*, 7(17), 19-23.
- Hassan, S., Rahman, R. A., Bakar, N. A., Mohd, R., & Muhammad, A. D. (2013). Designing Islamic microfinance products for Islamic banks in Malaysia. *Middle-East Journal of Scientific Research*, 17(3), 359-366.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. (Vol. 2, pp. 1-758). New York: springer.
- Himberg, T. (2021). Loan Default Prediction with Machine Learning.
- Hoque, M. (2025). *Organizational Training Environment and Training Effectiveness: The Mediating Role of Motivation*. (Doctoral dissertation), University of Dhaka.
- Horton, J., Macve, R., & Struyven, G. (2004). Qualitative research: experiences in using semi-structured interviews. *The real life guide to accounting research*, 339-357.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
- Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision support systems*, 37(4), 543-558.

- Hunjra, A. I., Mehmood, A., Nguyen, H. P., & Tayachi, T. (2022). Do firm-specific risks affect bank performance?. *International Journal of Emerging Markets*, 17(3), 664-682.
- Huysmans, J., Baesens, B., Vanthienen, J., & Van Gestel, T. (2006). Failure prediction with self organizing maps. *Expert Systems with Applications*, 30(3), 479-487.
- Jadhav, S., He, H., & Jenkins, K. (2018). Information gain directed genetic algorithm wrapper feature selection for credit rating. *Applied Soft Computing*, 69, 541-553.
- Jonas, N., Christian, M., Ntombela, S., & Letsoalo, S. (2025). Empowering South African Smallholder Farmers: Integrating Climate Resilience into Credit Assessment. *Sustainability*, 17(1), 261.
- Jumaa, M., Saqib, M., & Attar, A. (2023). Improving credit risk assessment through deep learning-based consumer loan default prediction model. *International Journal of Finance & Banking Studies* (2147-4486), 12(1), 85-92.
- Karim, N., Tarazi, M., & Reille, X. (2008). Islam micro finance: an emerging market niche.
- Khan, I., & Gupta, R. (2024). Early depression detection using ensemble machine learning framework. *International Journal of Information Technology*, 1-8.
- Khashman, A. (2011). Credit risk evaluation using neural networks: Emotional versus conventional models. *Applied Soft Computing*, 11(8), 5477-5484.
- Khemakhem, S., & Boujelbene, Y. (2017). Artificial intelligence for credit risk assessment: artificial neural network and support vector machines. *ACRN Oxford Journal of Finance and Risk Perspectives*, 6(2), 1-17.
- Khemakhem, S., & Boujelbene, Y. (2018). Predicting credit risk on the basis of financial and non-financial variables and data mining. *Review of Accounting and Finance*, 17(3), 316-340.
- Khemakhem, S., & Boujelbene, Y. (2018). Support vector machines for credit risk assessment with imbalanced datasets. *International Journal of Data Mining, Modelling and Management*, 10(2), 171-187.

- Kim, K., Kim, J., Cho, K., Kim, J. G., & Hyun, S. (2018). Analysis of the resilience of common-pool resources during globalization: The case of Jeju common ranches in Korea. *Sustainability*, 10(12), 4346.
- Koch, T. W., & MacDonald, S. S. (2014). *Bank management*. Orlando: Dryden Press.
- Komba, G. V., & Komba, M. M. (2024). Examining microfinance loan repayment challenges and procedures by small business owners: The case of Vikenge Village, Morogoro, Tanzania. *Science Mundi*, 4(2), 107–116-107–116.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information sciences*, 275, 1-12.
- Kouser, R., TP, H. S., Sruthi, A., & Jadhav, A. (2024). Using Machine learning Models For Predicting Loan Status And Computation Of Interest Rate. In *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)* (pp. 1-5). IEEE.
- Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, 27, 11-23.
- Kowalewski, O., Pisany, P., & Ślęzak, E. (2022). Digitalization and data, institutional quality and culture as drivers of technology-based credit providers. *Journal of economics and business*, 121, 106069.
- Kozodoi, N., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., & Baesens, B. (2019). A multi-objective approach for profit-driven feature selection in credit scoring. *Decision support systems*, 120, 106-117.
- Krasovytskyi, D., & Stavvytskyy, A. (2024). Predicting Mortgage Loan Defaults Using Machine learning Techniques. *Ekonomika*, 103(2), 140-160.

- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221-232.
- Kumar, A., Sharma, S., & Mahdavi, M. (2021). Machine learning (ML) technologies for digital credit scoring in rural finance: A literature review. *Risks*, 9(11), 192.
- Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean journal of anesthesiology*, 70(4), 407-411.
- Laborda, J., & Ryoo, S. (2021). Feature selection in a credit scoring model. *Mathematics*, 9(7), 746.
- Larose, D. T. (2015). *Data mining and predictive analytics*. John Wiley & Sons.
- Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29.
- Lewis, J. C. (2008). Microloan sharks. *Stanford Social Innovation Review*, 6(3), 54-59.
- Li, Z., Li, A., Bellotti, A., & Yao, X. (2023). The profitability of online loans: A competing risks analysis on default and prepayment. *European Journal of Operational Research*, 306(2), 968-985.
- Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6), 321-332.
- Linh, T. N., Long, H. T., Chi, L. V., Tam, L. T., & Lebailly, P. (2019). Access to rural credit markets in developing countries, the case of Vietnam: A literature review. *Sustainability*, 11(5), 1468.
- Liu, J., & Chen, X. (2006). Credit risk assessment model for individual housing loan based on decision tree. *Computer engineering*, 32, 263-271.
- Liu, Y., Ghandar, A., & Theodoropoulos, G. (2019, June). Island model genetic algorithm for feature selection in non-traditional credit risk evaluation. In *2019 IEEE congress on evolutionary computation (CEC)* (pp. 2771-2778). IEEE.

- Lu, Z., Li, H., & Wu, J. (2024). Exploring the impact of financial literacy on predicting credit default among farmers: An analysis using a hybrid machine learning model. *Borsa Istanbul Review*, 24(2), 352-362.
- Luo, S. T., Cheng, B. W., & Hsieh, C. H. (2009). Prediction model building with clustering-launched classification and support vector machines in credit scoring. *Expert Systems with Applications*, 36(4), 7562-7566.
- Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications*, 200, 116889.
- Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012042). IOP Publishing.
- Maenuddin, Hamid, S. A., Fahlevi, M., Nassir, A. M., & Hashim, P. M. (2023). Predictors of microfinance sustainability: Empirical evidence from Bangladesh. *Cogent Economics & Finance*, 11(1), 2202964.
- Maenuddin, Hamid, S. A., Nassir, A. M., Fahlevi, M., Aljuaid, M., & Jermsittiparsert, K. (2024). Measuring the financial sustainability and its influential factors in microfinance sector of Pakistan. *Sage Open*, 14(3), 21582440241259288.
- Maldonado, S., Bravo, C., López, J., & Pérez, J. (2017). Integrated framework for profit-based feature selection and SVM classification in credit scoring. *Decision Support Systems*, 104, 113-121.
- Malhotra, D. K., Malhotra, K., & Malhotra, R. (2020). Evaluating consumer loans using machine learning techniques. *Applications of Management Science*, 20, 59-69.
- Malik, K., Meki, M., Morduch, J., Ogden, T., Quinn, S., & Said, F. (2020). COVID-19 and the Future of Microfinance: Evidence and Insights from Pakistan. *Oxford Review of Economic Policy*, 36(Supplement\_1), S138-S168.

- Marqués, A. I., García, V., & Sánchez, J. S. (2012). Two-level classifier ensembles for credit risk assessment. *Expert Systems with Applications*, 39(12), 10916-10922.
- Maté, D., Raza, H., & Ahmad, I. (2023). Comparative Analysis of Machine Learning Models for Bankruptcy Prediction in the Context of Pakistani Companies. *Risks*, 11(10), 176.
- Medeiros Assef, F., & Arns Steiner, M. T. (2020). Ten-year evolution on credit risk research: a Systematic Literature Review approach and discussion. *Ingeniería e Investigación*, 40(2), 50-71.
- Min, S. H., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert systems with applications*, 31(3), 652-660.
- Mohammed, S. J., Zubaidi, S. L., Ortega-Martorell, S., Al-Ansari, N., Ethaib, S., & Hashim, K. (2022). Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective. *Cogent Engineering*, 9(1), 2143051.
- Mondol, S., Satpathy, I., Patnaik, B., Nayak, A., Patnaik, A., & Khang, A. (2024). Impact of AI and data in revolutionizing microfinance in developing countries: improving outreach and efficiency. In *Data-driven modelling and predictive analytics in business and finance* (pp. 71-87): Auerbach Publications.
- Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation*, 5(1), 1-27.
- Mueller, J. P., & Massaron, L. (2016). *Data Science MIT Python für Dummies*. John Wiley & Sons.
- Muhammad, S. D. (2010). Microfinance challenges and opportunities in Pakistan. *European Journal of Social Sciences*, 14(1), 88.

- Munkhdalai, L., Munkhdalai, T., Namsrai, O. E., Lee, J. Y., & Ryu, K. H. (2019). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability, 11*(3), 699.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- Nasir, S. (2013). Microfinance in India: Contemporary issues and challenges. *Middle-east journal of scientific research, 15*(2), 191-199.
- Obaido, G., Mienye, I. D., Egbelowo, O. F., Emmanuel, I. D., Ogunleye, A., Ogbuokiri, B., Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine learning with Applications, 17*, 100576.
- Obaidullah, M. (2008). Introduction to Islamic microfinance. *Mohammed Obaidullah, Introduction to Islamic Microfinance, IBF Net Limited*.
- Obaidullah, M., & Khan, T. (2008). Islamic microfinance development: Challenges and initiatives. *Islamic Research & Training institute Policy Dialogue Paper, (2)*.
- Odegua, R. (2020). Predicting bank loan default with extreme gradient boosting. *arXiv preprint arXiv:2002.02011*.
- Oghenekaro, L., & Chimela, M. (2023). Design and implementation of a loan default prediction system using random forest algorithm. *Scientia Africana, 22*(3), 137-144.
- Omowole, B. M., Urefe, O., Mokogwu, C., & Ewim, S. E. (2024). Optimizing Loan Recovery Strategies in Microfinance: A Data-Driven Approach to Portfolio Management. *International Journal Of Engineering Research And Development, 20*(11), 1153-1161.
- Omowole, B. M., Urefe, O., Mokogwu, C., & Ewim, S. E. (2024b). Strategic approaches to enhancing credit risk management in microfinance institutions. *International Journal of Frontline Research in Multidisciplinary Studies, 4*(1), 053-062.

- OPOKU ANTWI, B. E. N. J. A. M. I. N. (2016). *The failure of Microfinance in Ghana: A case study of Noble Dream Microfinance Limited*. (Doctoral dissertation).
- Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert systems with applications*, 39(16), 12605-12617.
- Owusu, E., Quainoo, R., Mensah, S., & Appati, J. K. (2023). A deep learning approach for loan default prediction using imbalanced dataset. *International Journal of Intelligent Information Technologies (IJIT)*, 19(1), 1-16.
- Owusu-Nuamah, P. (2014). Collapse of microfinance companies: Companies shot themselves in the foot. Retrieved March, 1, 2018.
- Ozgun, O., Karagol, E. T., & Ozbugday, F. C. (2021). Machine learning approach to drivers of bank lending: evidence from an emerging economy. *Financial Innovation*, 7, 1-29.
- Papouskova, M., & Hajek, P. (2019). Two-stage consumer credit risk modelling using heterogeneous ensemble learning. *Decision support systems*, 118, 33-45.
- Parveen, J. A. (2009). Sustainability Issues of Interest-free Micro-finance institutions in rural development and poverty alleviation. The bangladesh perspective. *Theoretical and Empirical Researches in Urban Management*, 4(2 (11)), 112-133.
- Pradnyana, K. D., & Rahadi, R. A. (2022). Loan Default Prediction in Microfinance Group Lending with Machine Learning. *International Journal of Business and Technology Management*, 4(4), 83-95.
- Provenzano, A. R., Trifirò, D., Datteo, A., Giada, L., Jean, N., Riciputi, A., ... & Nordio, C. (2020). Machine learning approach for credit scoring. *arXiv preprint arXiv:2008.01687*.
- Pykes, K. (2020). Oversampling and Undersampling. *Medium, Towards Data Science*, 10.

- Quayes, S. (2012). Depth of outreach and financial sustainability of microfinance institutions. *Applied Economics*, 44(26), 3421-3433.
- Ramya, R., Babu, D. R., Rangarajan, K., & Dixit, N. Y. (2024). An AI-based approach for Microfinance banks serving underprivileged communities. In *Computer Science Engineering* (pp. 735-744): CRC Press.
- Ranjan, R. (2025). Behavioural Finance in Banking and Management: A Study on the Trends and Challenges in the Banking Industry. *Asian Journal of Economics, Business and Accounting*, 25(1), 374-386.
- Rauf, S. A., & Mahmood, T. (2009). Growth and performance of microfinance in Pakistan. *Pakistan economic and social review*, 99-122.
- Razzaq, S., Maqbool, N., & Hameed, W. U. (2019). Factors effecting the elasticity of micro credit demand in southern Punjab, Pakistan. *International Journal of Social Sciences and Economic Review*, 1(2), 46-53.
- Rehman, Z. U., Muhammad, N., Sarwar, B., & Raz, M. A. (2019). Impact of risk management strategies on the credit risk faced by commercial banks of Balochistan. *Financial Innovation*, 5, 1-13.
- Roslan, A. H., & Karim, M. Z. (2009). Determinants of microcredit repayment in Malaysia: The case of Agrobank. *Humanity & Social Sciences Journal*, 4(1), 45-52.
- Ruziq, A. (2013). The impact of credit and liquidity risk on bank financial performance: the case of Indonesian Conventional Bank with total asset above 10 trillion Rupiah. *International Journal of Economic Policy in Emerging Economies*, 6(2), 93-106.
- Saad, N. M. (2012). Microfinance and prospect for Islamic microfinance products: The case of Amanah Ikhtiar Malaysia. *Advances in Asian Social Science*, 1(1), 27-33.

- Sadatrassoul, S., Gholamian, M., & Shahanaghi, K. (2015). Combination of feature selection and optimized fuzzy apriori rules: the case of credit scoring. *International Arab Journal of Information Technology (IAJIT)*, 12(2).
- Saleh, I., & Abu Afifa, M. (2020). The effect of credit risk, liquidity risk and bank capital on bank profitability: Evidence from an emerging market. *Cogent Economics & Finance*, 8(1), 1814509.
- Sandhu, T. H. (2018). Machine Learning And Natural Language Processing-A Review. *International Journal of Advanced Research in Computer Science*, 9(2).
- Sarfo, C., Zhang, J. A., O'Kane, C., & O'Kane, P. (2024). Perceived value of microfinance and SME performance: The role of exploratory innovation. *International Journal of Innovation Studies*, 8(2), 172-185.
- Sayed, E. H., Alabrah, A., Rahouma, K. H., Zohaib, M., & Badry, R. M. (2024). Machine Learning and Deep Learning for Loan Prediction in Banking: Exploring Ensemble Methods and Data Balancing. *IEEE Access*, (12), 193997-194019.
- Schreiner, M. (2002). Aspects of outreach: A framework for discussion of the social benefits of microfinance. *Journal of international development*, 14(5), 591-603.
- Segrado, C. (2005). Case study "Islamic microfinance and socially responsible investments". *Italy: MEDA Project*.
- Serwadda, I. (2018). Impact of credit risk management systems on the financial performance of commercial banks in Uganda. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 66(6), 1627-1635.
- Sharma, A. K., Li, L. H., & Ahmad, R. (2022, November). Default Risk Prediction Using Random Forest and XGBoosting Classifier. In *2021 International Conference on Security and Information Technologies with AI, Internet Computing and Big-data Applications* (pp. 91-101). Springer International Publishing.

- Shehu, M. J., Abdulmajeed, T. I., Olaniyi, M. A., & Adebayo, H. U. (2022). Problem and Prospect of Microfinance Banks in Ilorin Metropolis, Kwara State, Nigeria. *Path of Science*, 8(2-3), 3001-3007.
- Shen, F., Wang, R., & Shen, Y. (2020). A cost-sensitive logistic regression credit scoring model based on multi-objective optimization approach. *Technological and Economic Development of Economy*, 26(2), 405-429.
- Siddiqi, N. (2012). *Credit risk scorecards: developing and implementing intelligent credit scoring*. John Wiley & Sons.
- Siddiqi, N. (2017). *Intelligent credit scoring: Building and implementing better credit risk scorecards*. John Wiley & Sons.
- Siddique, K., Akhtar, Z., Lee, H. G., Kim, W., & Kim, Y. (2017). Toward bulk synchronous parallel-based machine learning techniques for anomaly detection in high-speed big data networks. *Symmetry*, 9(9), 197.
- Sohn, S. Y., Kim, D. H., & Yoon, J. H. (2016). Technology credit scoring model with fuzzy logistic regression. *Applied Soft Computing*, 43, 150-158.
- Stevenson, M., Mues, C., & Bravo, C. (2021). The value of text for small business default prediction: A deep learning approach. *European Journal of Operational Research*, 295(2), 758-771.
- Susana, D. M. P. (2024). *Optimizing credit scoring models in face of global economic uncertainty: A comprehensive risk analysis in banking loans*. (Master's thesis, Universidade NOVA de Lisboa (Portugal)).
- Tchakoute-Tchuigoua, H., & Soumaré, I. (2019). The effect of loan approval decentralization on microfinance institutions' outreach and loan portfolio quality. *Journal of Business Research*, 94, 1-17.
- Teplý, P., & Polena, M. (2020). Best classification algorithms in peer-to-peer lending. *The North American Journal of Economics and Finance*, 51, 100904.
- Thiétart, R. A. (2014). *Méthodes de recherche en management-4ème édition*. Dunod.

- Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, 16(2), 149-172.
- Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications*. Society for industrial and Applied Mathematics.
- Tian, Z., Xiao, J., Feng, H., & Wei, Y. (2020). Credit risk assessment based on gradient boosting decision tree. *Procedia Computer Science*, 174, 150-160.
- Tigges, M., Mestwerdt, S., Tschirner, S., & Mauer, R. (2024). Who gets the money? A qualitative analysis of fintech lending and credit scoring through the adoption of AI and alternative data. *Technological forecasting and social change*, 205, 123491.
- Ting, T., Mia, M. A., Hossain, M. I., & Wah, K. K. (2024). Predicting the financial performance of microfinance institutions with machine learning techniques. *Journal of Modelling in Management*, 20(2), 322-347.
- Tuan, N. M. (2022). Machine Learning Performance on Predicting Banking Term Deposit. *ICEIS (1)*, 267-272.
- Twala, B. (2010). Multiple classifier application to credit risk assessment. *Expert systems with applications*, 37(4), 3326-3336.
- Uddin, M. H., Akter, S., Mollah, S., & Al Mahi, M. (2022). Differences in bank and microfinance business models: An analysis of the loan monitoring systems and funding sources. *Journal of International Financial Markets, Institutions and Money*, 80, 101644.
- Uddin, N., Ahamed, M. K. U., Uddin, M. A., Islam, M. M., Talukder, M. A., & Aryal, S. (2023). An ensemble machine learning based bank loan approval predictions system with a smart application. *International Journal of Cognitive Computing in Engineering*, 4, 327-339.

- Uphade, D., Muley, A., & Chalwadi, S. (2024). Identification of Most Preferable Machine Learning Technique for Prediction of Bank Loan Defaulters. *Indian Journal of Science and Technology*, 17(4), 343-351.
- Uwais, A. M., & Khaleghzadeh, H. (2022). Loan default prediction using spark machine learning algorithms. In *AIAI 29th Irish Conference on Artificial Intelligence and Cognitive Science* (pp. 118-129). CEUR Workshop Proceedings.
- Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, 26, 61-68.
- Wang, T., & Li, J. (2019, April). An improved support vector machine and its application in P2P lending personal credit scoring. In *IOP Conference Series: Materials Science and Engineering* (Vol. 490, No. 6, p. 062041). IOP Publishing.
- Wang, Y., Xu, X., Hu, L., Fan, J., & Han, M. (2024). A time series continuous missing values imputation method based on generative adversarial networks. *Knowledge-Based Systems*, 283, 111215.
- Wang, Y., Zhang, Y., Lu, Y., & Yu, X. (2020). A Comparative Assessment of Credit Risk Model Based on Machine Learning—a case study of bank loan data. *Procedia Computer Science*, 174, 141-149.
- Wijewardhana, U. A., Bandara, C., & Nanayakkara, T. (2018). A mathematical model for predicting debt repayment: A technical note. *Australasian Accounting, Business and Finance Journal*, 12(3), 127-135.
- Wójcicka, A. (2017). Neural networks vs discriminant analysis in the assessment of default. *Annales Universitatis Mariae Curie-Skłodowska, Sectio H Oeconomia*, 51(5), 339-349.
- Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications*, 78, 225-241.

- Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 38(7), 8336-8342.
- Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications*, 38(10), 13274-13283.
- Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert systems with applications*, 36(2), 2473-2480.
- Yoseph, Y. (2017). Assessment of Credit Risk Management In Micro Finance Institutions (In the case of Addis credit and saving institutions).
- Yu, L., Yue, W., Wang, S., & Lai, K. K. (2010). Support vector machine based multiagent ensemble learning for credit risk evaluation. *Expert Systems with Applications*, 37(2), 1351-1360.
- Yuan, G., Huang, S., Fu, J., & Jiang, X. (2022). Low rank representation and discriminant analysis-based models for peer-to-peer default risk assessment. *Journal of Systems and Information Technology*, 24(2), 96-111.
- Zhang, H., He, H., & Zhang, W. (2018). Classifier selection and clustering with fuzzy assignment in ensemble model for credit scoring. *Neurocomputing*, 316, 210-221.
- Zhang, L., Wang, J., & Liu, Z. (2023). What should lenders be more concerned about? Developing a profit-driven loan default prediction model. *Expert Systems with Applications*, 213, 118938.
- Zhang, X., Li, J., Xiang, D., & Worthington, A. C. (2023). Digitalization, financial inclusion, and small and medium-sized enterprise financing: Evidence from China. *Economic Modelling*, 126, 106410.

- Zhang, Z., Gao, G., & Shi, Y. (2014). Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. *European Journal of Operational Research*, 237(1), 335-348.
- Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y., & Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications*, 42(7), 3508-3516.
- Zhou, G., Zhang, Y., & Luo, S. (2018). P2P network lending, loss given default and credit risks. *Sustainability*, 10(4), 1010.
- Zhou, J., Li, W., Wang, J., Ding, S., & Xia, C. (2019). Default prediction in P2P lending from high-dimensional data based on machine learning. *Physica A: Statistical Mechanics and its Applications*, 534, 122370.
- Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, 162, 503-513.
- Zhu, X., Chu, Q., Song, X., Hu, P., & Peng, L. (2023). Explainable prediction of loan default based on machine learning models. *Data Science and Management*, 6(3), 123-133.
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). Business research methods.

## APPENDIX A

