Machine Learning Driven Recovery Recommendations for Defaulted Loans: A SHAP-Based Decision Support Framework

MSc Research Project
FinTech

Sreeram Natarajan
Student ID: 23292661

School of Computing
National College of Ireland

Supervisor: Faithful Chiagoziem ONWUEGBUCHE

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sreeram Natarajan
Student ID:	23292661
Programme:	FinTech
Year:	2025
Module:	MSc Research Project
Supervisor:	Faithful Chiagoziem ONWUEGBUCHE
Submission Due Date:	11/08/2025
Project Title:	Machine Learning Driven Recovery Recommendations for Defaulted Loans: A SHAP-Based Decision Support Framework
Word Count:	6255
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	9th August 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Machine Learning Driven Recovery Recommendations for Defaulted Loans: A SHAP-Based Decision Support Framework

Sreeram Natarajan
23292661

Abstract

Loan defaults and recovery are vital factors that directly affect the profitability of financial institutions. Identifying the right recovery strategy for each borrower remains challenging with traditional methods. This research proposes a machine learning (ML)-driven recovery decision support framework aimed at assisting the recovery teams of financial institutions in making informed and transparent decisions for each defaulted loan. Since publicly available datasets with recovery action details are scarce, this study uses a dataset derived from Lending Club via Kaggle and creates a rule-based recovery action as the target variable. This target variable enabled the supervised models to learn recovery patterns based on various borrower characteristics and loan statuses. To ensure transparency and enhance trust in decision making, the framework integrates SHAP (SHapley Additive exPlanations) to provide borrower-centric explanations of the key features influencing recovery outcomes. Multiple classification ML models such as Logistic Regression, Random Forest, and XGBoost were evaluated, alongside data resampling techniques like SMOTE, class balancing, and undersampling to address data imbalance. Among these, XGBoost with class balancing was identified as the best performing model with an Accuracy of 0.9986, Balanced Accuracy of 0.9927, Macro ROC AUC of 0.9999 and a perfect Macro F1-score of 1.00. Random Forest with class balancing also demonstrated good performance across metrics. Fit status analysis was carried out by comparing training and test performance, which showed no signs of underfitting or overfitting across all models. While this study focuses on the evaluation of ML models and explainability using SHAP, future work could involve developing a front-end loan recovery dashboard, deploying the framework in real-time banking environments and testing its scalability with real-world data from financial institutions.

1 Introduction

Non-performing loans (NPLs) are one of the serious issues faced by financial institutions, impacting profitability, provisioning requirements, and regulatory compliance. While most existing studies have focused on loan default prediction using Machine Learning techniques, there has been limited research on post-default loan management. Gamba-Santamaria et al. (2024) decomposed NPL dynamics using Colombian credit vintage data, identifying how borrower payment capacity and bank risk-taking behaviors evolve with credit cycles. However, their work remains aggregate in scope and does not offer borrower-specific recovery strategies.

Traditional recovery methods are generalised for all loans and do not take into account individual borrower characteristics like home ownership, annual income, and debt-to income ratio. Moreover, they often do not consider the various stages of delinquency and default such as Grace Period, Late, Default, and Charged Off when determining recovery actions. Although some studies have focussed on modelling recovery rates and optimising debt collection via recommendation systems, they often lack transparency at the borrower level. Bellotti et al. (2021) demonstrated that borrower behaviour and debt collector actions significantly impact recovery rate estimation, though their work focuses more on aggregate predictors and less on borrower-specific guidance.

The advancement of explainable AI (XAI) tools, such as SHAP, offers a way to address this gap by revealing the variables that influence recovery outcomes for each borrower. However, the incorporation of explainability into recovery recommendations remains underexplored in both research and practice. Nwafor and Nwafor (2023) applied SHAP and ensemble neural models to improve transparency in credit risk prediction, but did not extend their work to recovery recommendations. Lin et al. (2025) used SHAP in the context of automobile loan assessment and emphasized the interpretability benefits of explainable ML, although their work remained focused on default classification. Models such as those by Liu et al. (2023) and Chang et al. (2022) enhanced default recall and credit scoring performance respectively in P2P lending, yet offered limited interpretability and no borrower-level recovery action insights.

This research addresses the question: How can machine learning models, combined with explainability techniques like SHAP, be utilized to recommend effective recovery actions for defaulted or delinquent loans based on borrower profiles and loan status stages? To answer this question, the study sets out to:

1. Develop a Machine Learning (ML) framework that maps borrower characteristics and loan statuses to appropriate recovery actions.

2. Implement SHAP to provide transparent, borrower-level insights into the features influencing recovery recommendations.

3. Evaluate multiple machine learning models such as Logistic Regression, Random Forest, and XGBoost, while applying balancing strategies like SMOTE, class weighting, and undersampling.

4. Carry out a comprehensive performance comparison of the models using standard evaluation metrics like accuracy, balanced accuracy, macro F1-score, and ROC AUC.

This study goes beyond mere default prediction by focussing on practical post-default recovery actions, an area often overlooked in financial analytics.

Key contributions include:

1. A novel end-to-end ML framework for generating customized recovery recommendations.

2. Integration of SHAP for enhancing model interpretability and decision transparency.

3. A comprehensive performance analysis across multiple algorithms and data balancing strategies.

Since publicly available datasets, including our selected Lending Club dataset, lack explicit recovery action details, we created a rule-based target variable for supervised learning. While this approach facilitates effective model training, deploying the framework across institutions may demand modifications due to varying recovery policies and data resources.

The remainder of the report is structured as follows: Section 2 presents the related work, reviewing existing research and machine learning techniques relevant to credit default prediction, recovery modelling, and explainable AI. Section 3 outlines the methodology, describing the scientific process followed for dataset preparation, model development, evaluation, and SHAP-based explainability. Section 4 details the design specification, explaining the underlying framework and logic supporting the recovery recommendation system. Section 5 discusses the implementation of the proposed solution, including tools used, models developed, and outputs generated. Section 6 provides a comprehensive evaluation of the results, supported by visualizations and analysis of academic and practical implications. Finally, Section 7 concludes the study, summarizing the key findings, discussing limitations, and proposing meaningful directions for future research.

2 Related Work

The existing works related to credit risk, non-performing loans (NPLs) and machine learning in finance are mostly focused on default prediction, with limited attention given to post-default recovery actions. Several studies have explored the macroeconomic and institutional causes of NPLs through econometric and exploratory analyses (Gamba-Santamaria et al. (2024); Nwafor and Nwafor (2023)), alongside advancements in ML-based borrower-level default prediction. Machine Learning (ML) models such as Logistic Regression, Random Forest, and XGBoost have been widely applied to credit scoring in both traditional and peer-to-peer lending environments (Chang et al. (2022); Liu et al. (2023). At the same time, explainability tools like SHAP (SHapley Additive exPlanations) have gained prominence in enhancing the transparency of model predictions, yet their application to post-default recovery decisions remains limited (Lin et al. (2025)). Although research has addressed aspects such as recovery rate modelling and ML-driven loan recovery (Bellotti et al. (2021)), customised, stage-sensitive recovery approaches remain insufficiently studied. This review critically examines four key areas: default prediction models, explainable machine learning in finance, recovery modeling and NPL risk analysis, and advanced machine learning-based techniques, to identify research gaps and justify the proposed SHAP-based machine learning framework for delinquent and defaulted loans.

2.1 Default Prediction Models

Even though there are very few studies on machine learning–driven post-default recovery, several works have explored ML-based default prediction across various lending contexts.

Blanco et al. (2024) introduce a logistic regression model to predict the one-year-ahead probability of default (PD) for Spanish non-financial corporations (NFCs), redefining default having non-performing loans (NPLs) for a three-month period within a year, rather than traditional bankruptcy. Using panel data spanning 1996 to 2019, the model integrates firm-level financial ratios and macroeconomic indicators, assessing performance across various economic cycles. While it proves useful for early detection of financial distress, it does not address recovery strategies at the borrower level.

Liu et al. (2023) propose a stacking ensemble model combining LightGBM and XGBoost to improve recall in P2P default prediction using Lending Club data. Their model increases recall rate of defaulting customers by 24.43% and improves AUC by 6.71% compared to standalone XGBoost, demonstrating strong performance for imbalanced data. However, the accuracy of the proposed model is slightly lower than the base models, and the approach lacks interpretability and does not address recovery decisions after default.

Li et al. (2025) use the Lending Club dataset and apply LightGBM to predict defaults on social lending platforms, achieving an accuracy of 0.87. Their model incorporates borrower and loan-level features to support early credit risk identification. However, the study focuses solely on default prediction and does not extend to recovery strategies.

Gao et al. (2023) use Lending Club and U.S. climate data to assess the impact of severe weather on farmer defaults in P2P lending. Applying Artificial Neural Networks (ANNs), Gradient Boosting, and Random Forest, they achieve accuracies of 70%, 74%, and 81%, respectively. SHAP highlights weather variables as key predictors, underscoring their economic relevance. However, the study focuses solely on default prediction, not recovery.

Wahab et al. (2024) compare ML and DL models on credit card default prediction using the UCI dataset. AdaBoost and Decision Trees yield the best performance at 82% accuracy, surpassing the ANN model. While preprocessing and tuning are well executed, the study lacks interpretability and provides no insights into post-default strategy.

Tareaf et al. (2024) conducted a comparative study of XGBoost, LightGBM, and CatBoost for credit risk prediction using the American Express dataset. CatBoost performed best on large data chunks but required more computational resources. The study emphasized the influence of data volume, missing value handling, and hyperparameter tuning. While the findings support the use of gradient boosting in finance, the work lacks recovery-focused insights and explainability analysis.

2.2 Explainable Machine Learning in Finance

While machine learning has improved credit risk prediction, many models remain opaque. To address this, recent studies use explainable AI techniques like SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) to make model decisions more transparent and trustworthy.

Li et al. (2025) applied SHAP and LIME alongside LightGBM for default prediction on Lending Club data, achieving 87% accuracy. Their explainability approach revealed both global and local feature contributions, helping lenders understand why predictions were made for individual borrowers. This transparency supports more accountable and trustworthy decision-making. However, as noted in Section 2.1, their study does not address borrower-specific recovery recommendations.

M and M U (2024) applied SHAP and LIME with Logistic Regression and Decision Tree models to enhance individual credit risk assessment. Their analysis identified salary, employment duration, and interest rate as key factors, showing the value of XAI in improving transparency. However, the models showed poor recall on defaults and were tested on a small, balanced dataset. The study suggests that future work should address data imbalance, scalability, and ethical deployment in real-world settings.

Nallakaruppan et al. (2024) proposed a credit evaluation framework using a Random Forest model combined with SHAP and LIME for explainability. The model provided both local and global interpretations of loan approval decisions. It achieved high accuracy (0.998), sensitivity (0.998), and specificity (0.997). While the framework enhances transparency using XAI, its extension into futuristic contexts like Industry 5.0 and the metaverse limits its immediate real-world applicability, though it offers promise for future financial systems.

Lin et al. (2025) assessed automobile loan credit risk using logistic regression and XGBoost, with SHAP used for model interpretability. Using data from a Chinese auto finance firm, SHAP identified credit level (0.67), credit score (0.337), and disbursed amount (0.34) as top contributors to default. The model offered improved predictive accuracy and interpretability, supporting risk-based decisions. However, the authors note that SHAP may not precisely quantify individual feature effects and may be prone to bias or misinterpretation. They recommend combining SHAP with causal inference to enhance robustness in future credit risk evaluations.

Aruleba and Sun (2024) applied ensemble classifiers including Random Forest, XGBoost, AdaBoost, and LightGBM, combined with SMOTE-ENN, to improve credit risk prediction on imbalanced datasets. SHAP was used for model explanation, identifying key features influencing default predictions. XGBoost achieved a recall of 93.0 percent and specificity of 84.6 percent on the German dataset, while Random Forest achieved 90.7 percent recall and 92.2 percent specificity on the Australian dataset. SHAP enabled feature-level transparency, supporting informed credit decisions. However, the study is limited by its focus on classification accuracy without exploring fairness, interpretability depth, or post-default recovery actions.

Stevens et al. (2020) investigated fairness and explainability in lending using an XGBoost-based recommendation model on Kiva loan data. They compared four bias mitigation techniques: Learning Fair Representations, Reweighing, Equality of Odds, and Reject Option based Classification. SHAP analysis identified repayment interval, world region, loan amount, and gender as key drivers. While the baseline had marginally better accuracy, the Reweighing method achieved the best trade-off, with 74.11 percent accuracy, 63.58 percent recall, and 80.54 percent precision, while satisfying fairness metrics. However, the study is limited by reliance on a single classifier, lack of in-processing bias mitigation, and equal misclassification cost assumptions.

2.3 Recovery Modelling and NPL Risk Analysis

Recovery modelling focuses on predicting repayment after default, while NPL risk analysis examines the factors that affect recovery outcomes. Recent studies apply machine learning and reinforcement learning to improve debt collection strategies and estimate recovery rates.

Sivamayilvelan et al. (2024) introduced a flexible debt collection recommender system using Deep Reinforcement Learning (FRS-DRL) to optimize field agent actions based on defaulter risk categories. Defaulters were classified into five risk levels, and a Hybrid Actor-Critic algorithm guided personalized collection strategies. The model improved recovery rates by over 20 percent and reduced field visits by 16.30 percent. While the framework improved operational efficiency, it lacks model interpretability and generalizability, and relies heavily on transfer learning for optimal performance.

Chen et al. (2019) proposed a two-stage framework to predict recovery rates for leveraged loans using Moody's data. A three-split ensemble model with parallel combining yielded the best results, and junior debt share was found to be the most influential feature. The study's strengths include high predictive accuracy and focus on recovery rather than default. However, the approach lacks explainability, is limited to historical leveraged loan data, and does not offer actionable recovery recommendations.

Raşid Bakır et al. (2025) combined deep neural networks with causal inference methods to study macroeconomic drivers of non-performing loans (NPLs) in Türkiye. Their model used a triple-validation approach integrating DNNs, random forest, and the DoWhy framework. Results showed that foreign direct investment had the strongest mitigating effect, while interest rate tools were less effective. However, the study is limited by its macro-level scope, lack of model explainability, and focus on a single-country dataset.

Botha et al. (2021) developed a simulation-based expert system to optimize the timing of loan recovery by identifying delinquency thresholds that minimize portfolio losses. The model balances trade-offs between accumulating arrears and forfeiting future interest, and accommodates systematic or episodic delinquency. Results show that threshold-based strategies outperform arbitrary collection decisions. However, the approach does not incorporate machine learning or borrower-level personalization and relies heavily on calibrated assumptions for accuracy.

Chai and Sun (2024) explored the role of fintech in mitigating non-performing loan (NPL) risk among Chinese commercial banks. Using a panel dataset from 2011 to 2022 and a two-fixed-effects model, they showed that fintech reduces NPLs through improved pre-loan screening and post-loan cost efficiency. The study highlights the stabilizing role of digital innovation in managing credit risk. However, its findings are limited to 40 A-share listed banks, and the fintech measurement relied on text mining rather than actual investment or implementation data, which future research should address.

2.4 Advanced Machine Learning Techniques

Recent research has explored hybrid models and advanced techniques such as ensemble learning, deep learning, and model stacking to improve prediction accuracy in credit risk modelling.

Muslim et al. (2023) proposed a stacking ensemble framework for P2P loan default prediction, combining SMOTE, LightGBM for feature selection, and a meta-learner stacking KNN, SVM, and Random Forest into XGBoost. Tested on two datasets, it achieved 99.98 percent and 91.43 percent accuracy. While effective for handling imbalance and improving prediction, the study lacks model interpretability and real-world deployment considerations.

Yu et al. (2024) proposed a hybrid credit risk model combining LightGBM, XGBoost, and TabNet, with SMOTEENN for class balancing and PCA and T-SNE for dimensionality reduction. The model outperformed traditional approaches in accuracy and robustness on high-dimensional credit data. However, its effectiveness depends heavily on data quality and ongoing retraining, which limits practical deployment without continuous monitoring.

Xia et al. (2025) proposed a two-stage credit scoring model (LLM-FP-CatBoost) combining large language models and FocalPoly loss with CatBoost to tackle class imbalance and data scarcity in fintech lending. ERNIE 4.0, GPT-4, and BERT were used to extract narrative features, with ERNIE 4.0 showing the best performance. While SHAP improved interpretability, the model is limited by dataset size, lack of external validation, and reliance on a fixed loss function.

Wang et al. (2023) proposed a credit default prediction model combining Bayesian Whale Optimization Algorithm and CatBoost, with feature selection via Random Forest and Gradient Boosting. The BWOA-CatBoost model outperformed Logistic Regression, Random Forest, and LightGBM in both accuracy and generalization. However, its applicability across sectors and robustness on diverse datasets remain untested, and future work is needed to explore alternative optimizers and feature selection methods.

Yu (2024) proposed an AE-CatBoost model combining AutoEncoder and CatBoost to improve bank credit risk control. Tested on a public loan dataset, the model achieved an AUC roughly 0.034 higher than CatBoost alone. The approach demonstrates the strength of hybrid deep learning and boosting. However, the paper lacks model interpretability and external dataset validation, limiting its practical adoption.

Zhang and Wang (2022) proposed a hybrid credit scoring model combining Transformer networks and CatBoost to leverage both unsupervised and supervised learning. Using a Chinese banking dataset, the model outperformed baselines in AUC and KS metrics. While the architecture improves feature representation and scoring accuracy, it relies heavily on manual feature engineering and complex data processing, limiting scalability.

In conclusion, while significant progress has been made in applying machine learning and explainable AI to credit risk prediction, most existing studies remain focused on default classification rather than offering borrower-specific recovery guidance. Despite the growing use of tools like SHAP and LIME to enhance transparency, their application to post-default decision-making is still limited. Recovery-related research has explored macroeconomic factors and operational strategies but lacks interpretability and customization at the individual borrower level. To address this gap, the present study proposes a SHAP-based machine learning framework that delivers transparent, borrower-level recovery recommendations for delinquent and defaulted loans, bridging a critical missing link in the credit risk lifecycle.

3 Methodology

This section outlines the structured scientific process followed in this research project, encompassing dataset selection, exploratory data analysis, data cleaning, label engineering, model development, evaluation, and integration of explainability techniques. The methodology was grounded in practical recovery strategies, supported by domain knowledge, and implemented using reproducible machine learning practices. Emphasis was placed on handling class imbalance, selecting interpretable features, and generating borrower-level explanations using SHAP.

Figure 1 illustrates the complete methodology followed in this study, covering data preparation, model development, evaluation, and explainability stages.

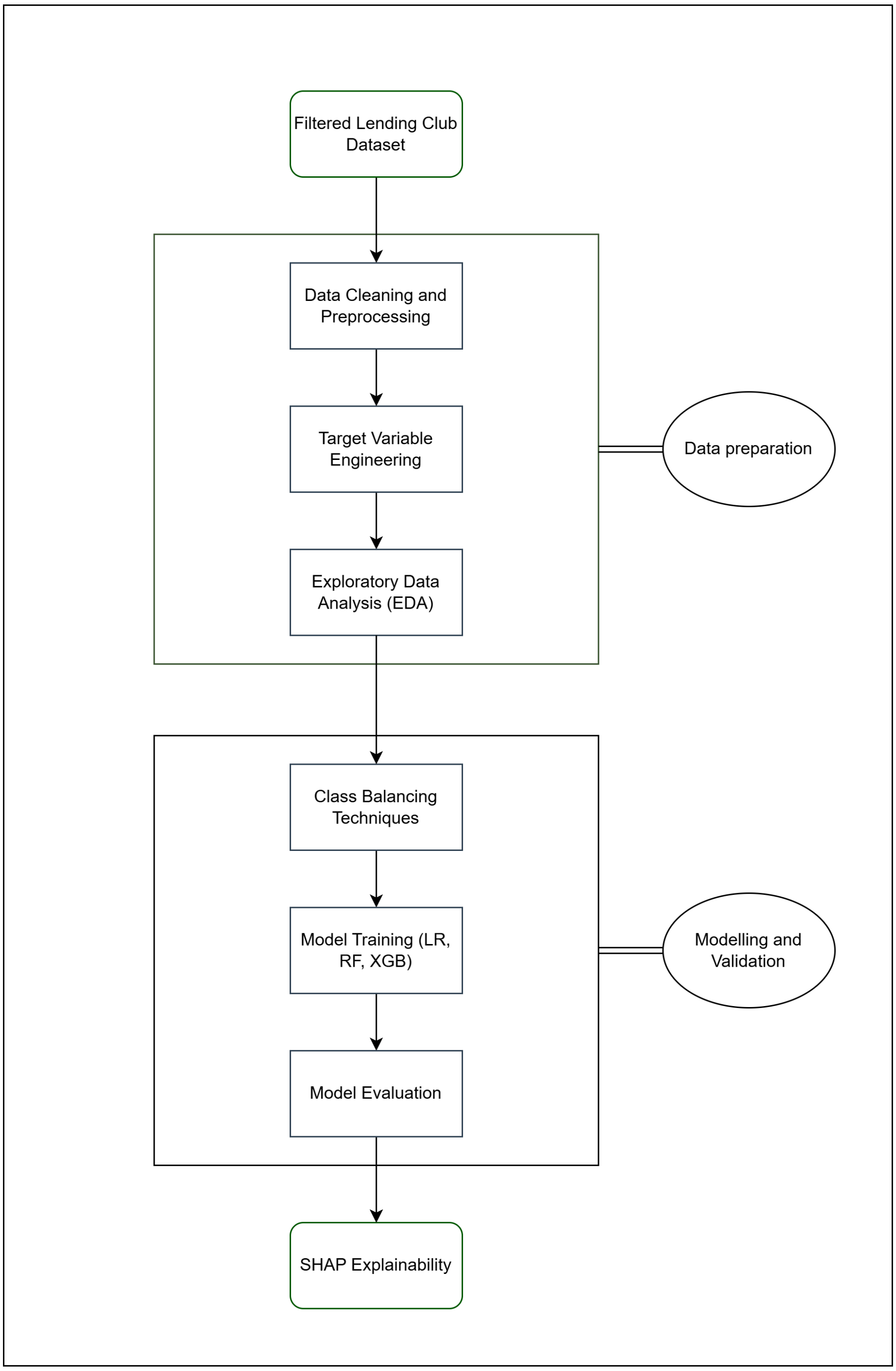

Figure 1: End-to-end machine learning pipeline for recovery recommendation

3.1 Dataset Selection

The dataset used in this study was obtained from Kaggle, titled *Lending Club 2007–2020Q3* bojrick.eth (2020). It contains loan originations from 2007 to 2020. After filtering to retain only loans that were either delinquent, in grace period, or defaulted, a total of 135,883 records were selected for analysis. These included loans with statuses such as Late (16-30 days), Late (31-120 days), Default, Charged Off, and In Grace Period. These categories represent distinct stages of loan arrears and provide the foundation for stage-sensitive recovery modelling.

3.2 Data Cleaning and Preprocessing

The raw dataset initially contained 142 columns, including both input features and the target variable. A multistage cleaning process was conducted. Columns with more than 80% missing values were dropped. A total of 16 relevant columns were retained for modelling based on relevance to loan recovery strategies, interpretability, and insights gained through exploratory data analysis. These included variables such as loan amount (`loan_amnt`), annual income (`annual_inc`), loan term (`term`), debt-to-income ratio (`dti`), and home ownership status (`home_ownership`). A complete list of excluded columns, along with the justification for their removal, is provided in Appendix A of the Configuration Manual. Percentage strings such as interest rate and revolving utilization were converted to numeric format, and the `term` field was cleaned to extract numerical values. Missing values in the debt-to-income ratio (`dti`) and revolving credit utilization (`revol_util`) fields were imputed using median values. Categorical variables were encoded using one-hot encoding, and numeric variables were scaled using the `StandardScaler` to normalize their distribution. To further simplify the classification task and address class sparsity, related recovery actions such as Restructure-Moratorium, Restructure-EMI Reduction, and Restructure-Term Extension were grouped under a unified class label: Restructure. This process ensured that the dataset was clean, balanced, and suitable for machine learning.

3.3 Target Variable Engineering

The original dataset did not contain labelled recovery actions. To enable supervised learning, a custom rule-based logic was implemented to generate a surrogate target variable, `recovery_action`. This logic reflects realistic recovery strategies used in financial institutions and considers `loan_status`, `debt_settlement_flag`, `loan_amnt`, `annual_inc`, `dti`, `home_ownership`, and `term`. Based on these attributes, each borrower was assigned a recovery action such as Legal Action, Reminder, Restructure-Moratorium, Outsource to Collection Agency, or Settled-No Further Action. The complete logic and associated decision criteria are presented in Section 4, Table 4.1. This step enabled the transformation of the dataset into a multi-class classification problem, grounded in domain-specific knowledge.

3.4 Exploratory Data Analysis (EDA)

An in-depth exploratory data analysis was conducted to understand the underlying structure of the dataset and validate the features selected for modelling. Descriptive statistics and distribution plots were generated for key variables such as loan amount, annual income, debt-to-income ratio, interest rate, loan term, installment amount, home ownership, and loan grade. Boxplots were used to identify outliers, and correlation heatmaps helped detect highly correlated features. Loan distributions across different delinquency stages were examined, revealing class imbalances. These insights helped validate the rule-based logic used for generating recovery actions, by confirming that borrower and loan characteristics such as income, debt ratio, term, and home ownership exhibited meaningful variation across loan status categories. This supported the practical relevance of the selected features and the conditions defined in the recovery framework. The EDA was conducted using Python libraries including pandas, seaborn, matplotlib, and ydata-profiling.

3.5 Handling Class Imbalance

The recovery action classes were not evenly distributed, resulting in a highly imbalanced dataset. To mitigate this, three balancing techniques were employed: Synthetic Minority Oversampling Technique (SMOTE), class weighting (via `class_weight='balanced'`), and random undersampling of the majority classes. Each of these techniques was independently applied during model training to compare their impact on classifier performance.

3.6 Model Development

Three machine learning algorithms were selected based on their performance, popularity in credit modelling, and interpretability: Logistic Regression, Random Forest, and XGBoost. Each model was trained separately using the resampled datasets produced by SMOTE, class weighting, and undersampling, resulting in nine distinct experiments. An 80:20 train-test split was used, with stratification to preserve class proportions.

3.7 Evaluation Metrics

To assess model performance, multiple evaluation metrics were used. These included accuracy, balanced accuracy, macro-averaged F1-score, and macro-averaged ROC AUC. Confusion matrices, ROC curves, and classification reports were also generated to evaluate class-wise performance in detail. Model evaluation was conducted on both training and test sets, enabling a fit-gap analysis to determine whether models were overfitting, underfitting, or generalizing well.

3.8 Explainability with SHAP

To ensure interpretability and support transparent decision-making, SHAP (SHapley Additive Explanations) was integrated into the framework. SHAP was used to analyze global and local feature importance across all models. Beeswarm plots and bar charts were generated to illustrate which borrower features most strongly influenced model predictions. This approach improves trust, supports compliance, and helps financial institutions justify automated decisions to regulators.

3.9 Tools and Environment

All implementation and analysis tasks were conducted using Python within Jupyter Notebook. Key libraries included pandas, numpy, scikit-learn, xgboost, imbalanced-learn, SHAP, seaborn, matplotlib, and ydata-profiling. These tools supported the end-to-end data processing, model development, evaluation, and explainability components of the framework.

4 Design Specification

The proposed recovery recommendation framework is designed to turn borrower-level data into meaningful recovery strategies using machine learning and explainable AI. The main goal is to support recovery teams in making informed decisions after a loan goes into default, by combining practical rules with predictive models and clear, understandable outputs. This section explains the overall design of the system, how recovery actions are decided, and the tools used to build it.

The project uses data from Lending Club, which includes loans issued between 2007 and 2020. This dataset was filtered to include only loans that were delinquent, defaulted, or charged off. Since the original data didn't include recovery outcomes, we created a new target column called `recovery_action` using a set of rule-based conditions. These rules are based on real-world banking experience and use features like loan status, loan amount, borrower income, debt-to-income ratio (DTI), home ownership, loan term, and whether a debt settlement had already been made. For example, borrowers with high unpaid amounts are assigned legal action, while early-stage defaulters are simply sent reminders. Some are recommended for restructuring based on income and DTI levels. The complete set of rule conditions is shown in Table 1.

Table 1: Rule-Based Logic for Assigning Recovery Actions Based on Loan Status and Borrower Profile

Loan Status	**Condition**	**Recovery Action**	**Reason**
debt_settlement_flag	debt_settlement_flag == 'Y'	Settled - No Further Action	Loan already settled; no further action required
Charged Off	loan > 10000	Legal Action	High-value loss; initiate legal process
Charged Off	3000 ≤ loan ≤ 10000 AND 20000 ≤ income ≤ 40000 AND home_ownership NOT IN ('OWN', 'MORTGAGE')	One-Time Settlement Offer	Mid-sized loan, moderate income, no property ownership; partial recovery is practical
Charged Off	income < 15000 AND loan < 3000 AND dti > 60	Outsource to Collection Agency	Low income, small loan, high DTI; internal recovery not viable
Charged Off	loan ≤ 10000	Legal Notice	Small loan; send formal legal warning
Default	loan > 20000	Legal Action	Large default; pursue legal recovery
Default	else	Outsource to Collection Agency	Small default; better handled by third-party agency
Late (31-120) or Late (16-30 days)	dti > 40 AND income < 35000	Restructure - Moratorium	High DTI and low income; temporary payment pause to avoid escalation
Late (31-120) or Late (16-30 days)	dti > 40 AND income ≥ 35000	Restructure - EMI Reduction	Moderate financial stress; reduce payment load
Late (31-120) or Late (16-30 days)	home_ownership == 'RENT' AND dti > 35	Restructure - EMI Reduction	Financially vulnerable renters; reduced EMI support
Late (31-120) or Late (16-30 days)	30 < dti ≤ 40 AND term contains "36"	Restructure - Term Extension	Medium term loans; extended terms ease payment burden
Late (31-120) or Late (16-30 days)	else	Reminder	Early-stage delinquency; send repayment reminders
In Grace Period	always	Reminder	Minimal delay; gentle reminder
Others	always	Reminder	Default fallback
Error Handling	any exception	Unknown	Unknown reason due to error

Once these recovery actions were assigned, the data was used to train machine learning models to learn from borrower patterns. Three popular algorithms were chosen: Logistic Regression, Random Forest, and XGBoost. Since the recovery actions were not evenly distributed, we used three balancing methods during training: SMOTE, class weighting, and random undersampling. This resulted in nine different versions of the models. An 80:20 train-test split was used to make sure class distribution stayed balanced during evaluation.

To make the results easy to understand and explain, we used SHAP (SHapley Additive Explanations). SHAP helps explain how each feature (like income, loan size, or DTI) influenced the model's decision, both overall and for individual borrowers. This transparency makes it easier for recovery teams and auditors to trust and justify the model's recommendations. We also used beeswarm plots, bar charts, and ROC curves to visualize model behavior and performance.

The entire framework was built using Python, with libraries like `scikit-learn`, `imbalanced-learn`, `XGBoost`, `SHAP`, and `pandas`. The design is flexible, so it can be adapted to different banks or lenders in the future. It can also be extended to work in real-time systems, depending on the institution's setup and data availability.

5 Implementation

This section presents the final stage of the solution development process, focusing on the key outputs produced and the tools used to implement the recovery recommendation framework.

After completing data preprocessing and defining recovery logic, the cleaned dataset was used to train machine learning models. The final dataset included 135,883 records with relevant borrower and loan attributes, along with a newly engineered column named `recovery_action`, which served as the target variable for supervised classification. The logic used to derive this column was rule-based and has been explained in Section 4.

Three machine learning algorithms Logistic Regression, Random Forest, and XGBoost were used, each combined with three class imbalance handling techniques: SMOTE, class weighting, and random undersampling. This resulted in a total of nine trained model variants, as listed in Table 2. An 80:20 stratified train-test split was used to maintain class proportions across both sets.

Table 2: Machine Learning Models Developed with Class Imbalance Strategies

Model ID	**Algorithm**	**Balancing Method**
1A	Logistic Regression	SMOTE
1B	Logistic Regression	Class Weighting
1C	Logistic Regression	Random Undersampling
2A	Random Forest	SMOTE
2B	Random Forest	Class Weighting
2C	Random Forest	Random Undersampling
3A	XGBoost	SMOTE
3B	XGBoost	Class Weighting
3C	XGBoost	Random Undersampling

All models were evaluated using standard classification metrics, including Accuracy, Balanced Accuracy, Macro F1-score, and ROC AUC, as listed in Table 3. During implementation, both training and test results were monitored to detect any signs of overfitting or underfitting. This helped ensure that the models were generalizing well and not simply memorizing patterns from the training data.

Table 3: Evaluation Metrics Used for Model Assessment

Metric	**Purpose**
Accuracy	Overall correctness of predictions
Balanced Accuracy	Average recall across classes, handling imbalance
Macro F1-score	Harmonic mean of precision and recall (macro-averaged)
ROC AUC (macro)	Area under ROC curve for multi-class classification

To ensure transparency and interpretability, SHAP (SHapley Additive Explanations) was integrated into the framework to analyse the influence of individual features on model predictions. SHAP visualisations were generated for all models, with class-wise feature importance and interaction plots produced for the best-performing model, XGBoost with class weighting, to illustrate the top contributing features across different recovery actions.

The implementation was carried out using Python in Jupyter Notebook, leveraging libraries such as `pandas`, `scikit-learn`, `xgboost`, `imbalanced-learn`, `SHAP`, `matplotlib`, and `seaborn`, all of which supported different stages of data processing, modeling, and interpretability.

6 Results

This section presents the key findings from the nine machine learning experiments conducted to predict recovery actions for defaulted loans. The primary objective is to evaluate how well each model performed in a multiclass imbalanced setting. A summary of performance on the test set is presented first, with Model 3B visualised in detail, as it outperformed the other models across the majority of evaluation metrics.

Table 4: Performance Metrics of All Models on the Test Set

ID	Model	Technique	Accuracy	Balanced Accuracy	Macro ROC AUC	Macro F1-score
1A	Logistic Reg.	SMOTE	0.9585	0.9634	0.9953	0.91
1B	Logistic Reg.	Class Balanced	0.9450	0.9653	0.9962	0.88
1C	Logistic Reg.	Undersampled	0.8289	0.8843	0.9826	0.69
2A	Random Forest	SMOTE	0.9722	0.9638	0.9984	0.90
2B	Random Forest	Class Balanced	0.9993	0.9835	0.9999	0.99
2C	Random Forest	Undersampled	0.9722	0.9638	0.9984	0.90
3A	XGBoost	SMOTE	0.9768	0.9832	0.9989	0.89
3B	XGBoost	Class Balanced	0.9986	0.9927	0.9999	1.00
3C	XGBoost	Undersampled	0.9768	0.9832	0.9989	0.89

Table 4 summarizes the test accuracy, balanced accuracy, macro F1-score, and macro ROC AUC for all nine experiments. These metrics are used to evaluate performance in a multiclass imbalanced scenario.

All models were trained using one of three class balancing methods: SMOTE, class weighting, or random undersampling. The resulting experiments are referred to as 1A-1C for Logistic Regression, 2A-2C for Random Forest, and 3A-3C for XGBoost. Across the three algorithms, performance varied depending on the class balancing method applied. For Logistic Regression (1A-1C), the SMOTE variant (1A) achieved higher accuracy and macro F1-score, while the class-weighted version (1B) delivered better balanced accuracy and macro ROC AUC. Both 1A and 1B performed considerably better than the undersampled model (1C), which showed the lowest performance across all metrics. Random Forest models (2A-2C) displayed consistently strong results, with the class-weighted variant (2B) outperforming the other two configurations across all evaluation metrics. The SMOTE (2A) and undersampled (2C) versions showed identical results, delivering moderately strong performance overall. XGBoost models (3A-3C) recorded the strongest performance across the majority of evaluation metrics, with the class-weighted variant (3B) achieving the highest scores in balanced accuracy, macro ROC AUC, and macro F1-score. The SMOTE (3A) and undersampled (3C) variants produced identical results and also demonstrated high overall performance, but fell short of 3B in these key metrics.

Model interpretability was examined using SHAP, with both global and class-specific plots generated for the best-performing model (3B). These visualisations highlight the most influential features driving each recovery action prediction. In addition, a comparison of train and test accuracies was performed for all models to assess generalization capability and detect potential overfitting.

The following figures and tables present detailed performance visualisations, SHAP-based feature importance plots, and the train-test accuracy comparison for all models.

	precision	recall	f1-score	support
Legal Action	1.00	1.00	1.00	13859
Legal Notice	1.00	1.00	1.00	6185
One-Time Settlement Offer	0.99	1.00	1.00	1538
Outsource to Collection Agency	1.00	0.95	0.98	21
Reminder	1.00	1.00	1.00	2760
Restructure	1.00	1.00	1.00	213
Settled – No Further Action	1.00	1.00	1.00	2601
accuracy			1.00	27177
macro avg	1.00	0.99	1.00	27177
weighted avg	1.00	1.00	1.00	27177

✅ Accuracy: 0.9986385546601906
✅ Balanced Accuracy: 0.9927110894976348
✅ Multiclass ROC AUC Score (macro average): 0.9999644046161544

Figure 2: Classification Report - Model 3B (XGBoost with Class Weighting)

Figure 2 presents the classification report for Model 3B. The model achieved near-perfect precision, recall, and F1-scores across all recovery action classes, demonstrating highly consistent predictive performance.

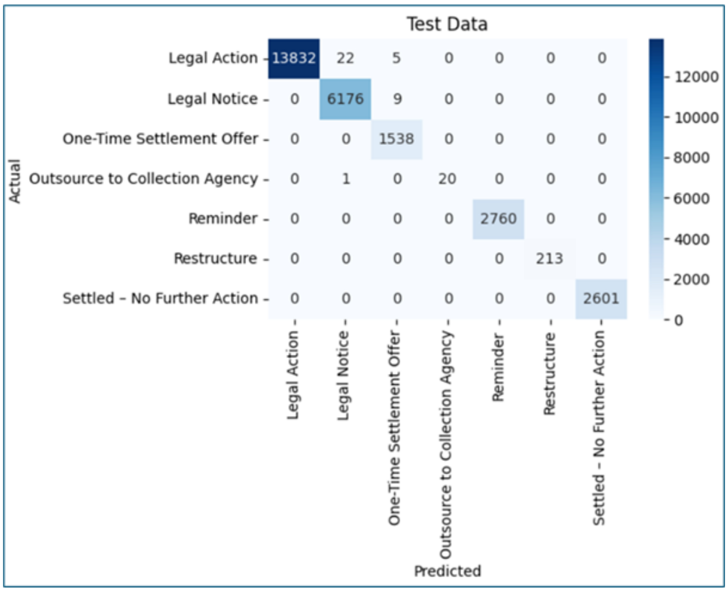

Figure 3: Confusion Matrix - Model 3B (XGBoost with Class Weighting)

Figure 3 shows the confusion matrix for Model 3B on the test set. All classes are classified with exceptionally high accuracy, indicating strong generalization and low misclassification rates across recovery action categories.

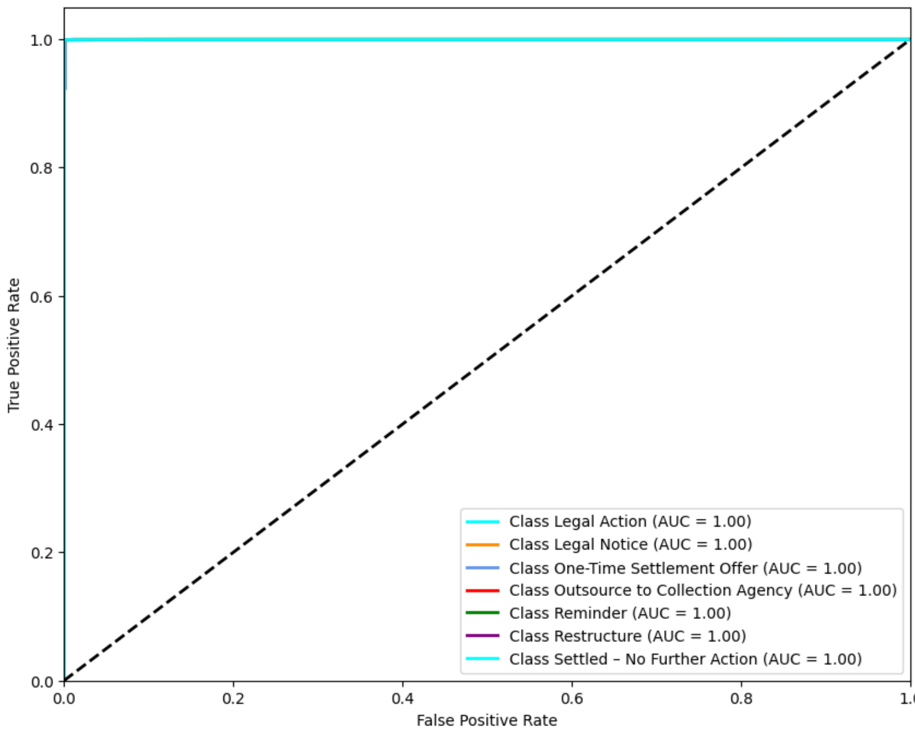

Figure 4: ROC Curve - Model 3B

Figure 4 illustrates the multi-class ROC curve for Model 3B on the test data. The model achieved perfect AUC scores of 1.00 across all classes, indicating flawless discriminative performance.

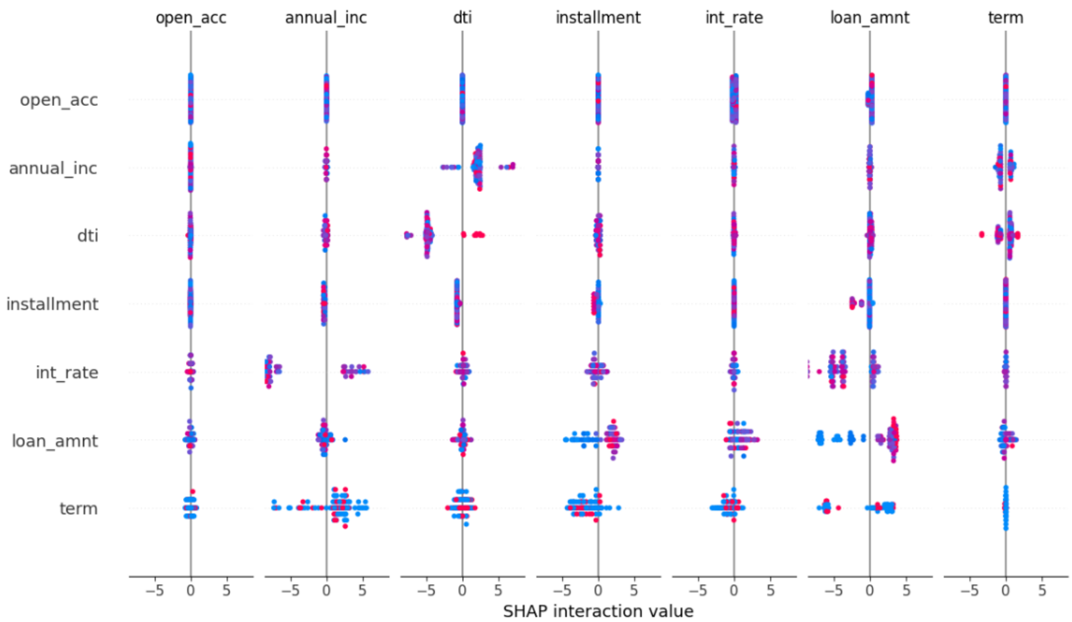

Figure 5: SHAP Beeswarm Plot - XGBoost with Class Weighting

Figure 5 visualises the distribution and impact of feature values across predictions for Model 3B. Features such as `dti`, `annual_inc`, and `loan_amnt` show significant variability, indicating their influence changes across different recovery action classes.

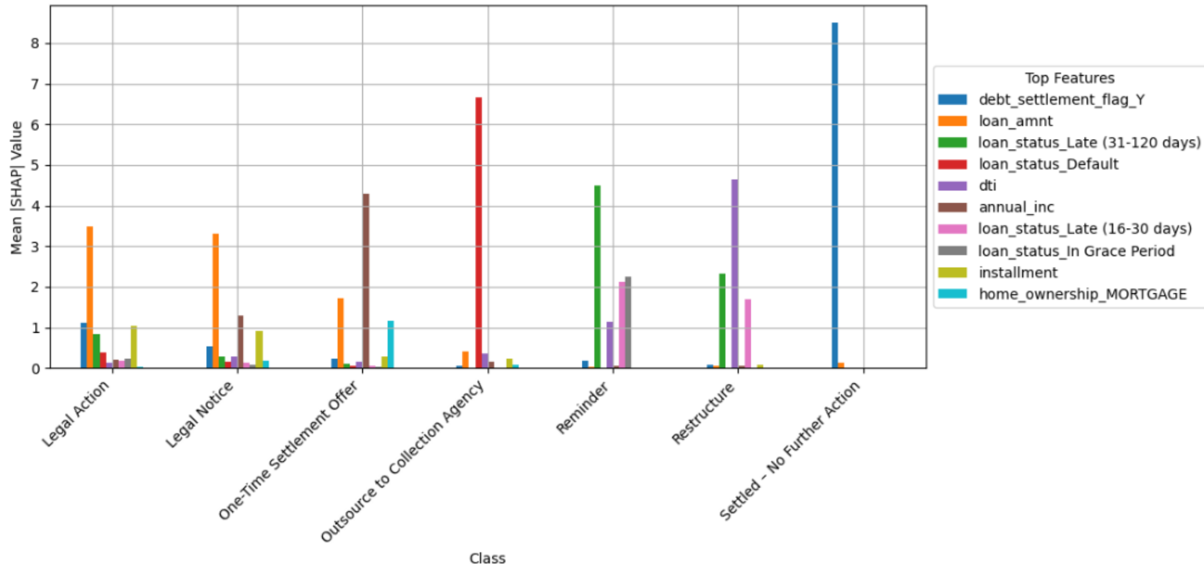

Figure 6: SHAP Bar Plot by Class - XGBoost with Class Weighting

Figure 6 presents the top contributing features for each recovery action class predicted by Model 3B. For example, `loan_status_Default` had the strongest impact on predicting 'Outsource to Collection Agency', while `debt_settlement_flag_Y` primarily influenced 'Settled - No Further Action'.

Table 5: Train vs Test Accuracy and Gap for All Models

Model	Technique	Train Acc	Test Acc	Accuracy Gap
Logistic Reg.	SMOTE	0.9725	0.9585	0.0140
Logistic Reg.	Class Balanced	0.9457	0.9450	0.0007
Logistic Reg.	Undersampled	0.9260	0.8289	0.0971
Random Forest	SMOTE	1.0000	0.9722	0.0278
Random Forest	Class Balanced	1.0000	0.9993	0.0007
Random Forest	Undersampled	1.0000	0.9722	0.0278
XGBoost	SMOTE	1.0000	0.9768	0.0232
XGBoost	Class Balanced	0.9994	0.9986	0.0008
XGBoost	Undersampled	1.0000	0.9768	0.0232

Table 5 shows that the accuracy gap across models was minimal, indicating no significant overfitting or underfitting and confirming that the models generalised well to unseen data.

7 Discussion

The experiments demonstrated that ensemble tree-based models, particularly XGBoost with class weighting (3B), delivered the most consistent overall performance across evaluation metrics, excelling in balanced accuracy, macro ROC AUC, and macro F1-score, while maintaining minimal train-test accuracy gaps. SHAP-based interpretability further enhanced transparency by identifying the most influential features for each recovery action, ensuring that predictions could be explained and validated in operational contexts.

The findings are consistent with prior research such as Lin et al. (2025) and Bellotti et al. (2021), where ensemble and gradient boosting methods outperformed linear models in credit risk prediction tasks. In Lin et al. (2025), XGBoost combined with SHAP improved both predictive accuracy and interpretability for automobile loan risk assessment, while Bellotti et al. (2021) demonstrated that tree-based and rule-based algorithms provided superior recovery rate forecasts compared to traditional regression approaches. These parallels reinforce the suitability of ensemble tree-based models for capturing complex borrower and loan relationships in credit risk contexts.

However, unlike studies such as Liu et al. (2023) and Li et al. (2025), which address default prediction as a binary classification problem, the present work extends the task to predicting multi-class borrower-specific recovery actions. While both of these studies employ ensemble methods and achieve strong predictive performance, their scope is limited to identifying whether a borrower will default. In contrast, the proposed framework not only determines likely recovery outcomes but also uses class-specific SHAP analysis to explain the drivers behind each recovery action. This broader approach enhances operational decision-making in post-default portfolio management, where tailored recovery strategies are critical.

Incorporating SHAP into the workflow also reflects similar efforts in explainable machine learning reported by M and M U (2024), and Aruleba and Sun (2024). M and M U (2024) demonstrated how SHAP and LIME could improve transparency in credit risk models, while Aruleba and Sun (2024) combined SHAP with ensemble classifiers and SMOTE-ENN to enhance both predictive performance and interpretability. While these studies confirm the usefulness of SHAP for understanding feature contributions, they do not explicitly link interpretability to specific recovery recommendations. In contrast, the present study applies SHAP at a class-specific level, enabling identification of key drivers for each recovery action. For example, `loan_status_Default` was the strongest predictor for 'Outsource to Collection Agency', `debt_settlement_flag_Y` primarily influenced 'Settled - No Further Action', and `loan_amnt` was critical in identifying 'Legal Action' and 'Legal Notice' cases. This level of granularity enhances the operational value of the model outputs.

When viewed alongside recovery-focused research such as Sivamayilvelan et al. (2024) and Botha et al. (2021), the proposed framework offers a notable advancement. Sivamayilvelan et al. (2024) used deep reinforcement learning to recommend collection actions based on customer risk categories, while Botha et al. (2021) optimised the timing of loan recovery to minimise losses. Both approaches address operational aspects of debt recovery but do not integrate feature-level interpretability into their models. The present study bridges this gap by combining high predictive accuracy with actionable, interpretable insights that directly support the choice of recovery strategy.

The similarities in feature importance across studies may be explained by the consistent role of core borrower and loan attributes (for example, loan amount, credit status, and debt settlement indicators) in determining repayment outcomes. Differences in performance can be attributed to methodological choices: the use of class weighting in XGBoost, the transformation of the prediction problem into multi-class recovery actions, and the application of SMOTE or undersampling in baseline comparisons. Furthermore, the dataset used in this study, Lending Club loan records, offers richer recovery-related attributes compared to datasets in some prior works, which may have contributed to the observed improvements.

From a practical standpoint, the combination of high performance and interpretability suggests that the framework could be deployed as a decision-support tool in loan recovery departments. Future research could focus on validating the model with real-world institutional datasets containing actual recovery outcomes, integrating it into core banking systems, and exploring the potential for automated recovery strategy optimisation based on evolving borrower profiles.

8 Conclusion and Future Work

This research addressed the core question: How can machine learning models, combined with explainability techniques like SHAP, be utilized to recommend effective recovery actions for defaulted or delinquent loans based on borrower profiles and loan status stages? To explore this, a practical and transparent decision support framework was developed that maps borrower characteristics and arrears stages to appropriate recovery strategies using machine learning (ML) and explainable AI.

The objectives of the study were fourfold: (1) to develop a rule-informed ML framework for recovery action prediction; (2) to generate a realistic target variable grounded in practical banking logic; (3) to train and evaluate three popular classification models (Logistic Regression, Random Forest, and XGBoost) under three balancing methods (SMOTE, class weighting, and undersampling); and (4) to integrate SHAP-based explainability to ensure transparency in model predictions.

The study successfully met all the objectives. A new target variable, recovery_action, was derived from Lending Club data using domain-informed rules based on borrower income, DTI, home ownership, loan amount, and loan status. The final dataset included 135,883 cleaned and labeled records, and nine ML models were trained using three balancing strategies. Evaluation was carried out using accuracy, balanced accuracy, macro F1-score, and ROC AUC. The best-performing model was XGBoost with class weighting, which achieved Accuracy = 0.9986, Balanced Accuracy = 0.9927, Macro F1 = 1.00, and Macro ROC AUC = 0.9999.

Key Findings

- Machine learning models can learn borrower-level patterns to recommend recovery actions with high accuracy.
- SHAP-based explainability offers both global and local insights into how features influence model decisions.
- Loan amount, debt settlement flag, DTI, and income were consistently the most influential features across models.
- Class weighting proved to be the most effective resampling strategy when used with XGBoost.

Implications and Limitations

The results show that combining interpretable machine learning with domain knowledge can bridge the gap between default prediction and recovery recommendations. The use of SHAP also supports transparency for regulators and internal audits.

However, there are some limitations. The dataset came from a single platform and lacked real-world recovery outcome labels. The target variable was engineered using predefined rules. As a result, deployment in practice would require adjustments to match institutional policies and borrower behaviour. Fairness and cost-benefit analysis were not explored in this work.

Future Work

Future research could apply the framework to banking datasets with real recovery outcomes to validate performance in operational settings. Additional enhancements may include optimisation layers that factor in recovery probability, cost, and efficiency, as well as fairness-aware techniques to detect and mitigate bias in recommendations. Combining SHAP with causal inference could help identify true drivers of recovery outcomes, moving beyond correlation-based explanations. Practical implementation could also involve developing a user-friendly front-end web interface to make the tool accessible to recovery officers, enabling interactive exploration of predictions and explanations. Finally, integration into real-time banking systems and simulations of alternative recovery strategies could further evaluate long-term impacts on loan portfolios.

This framework provides a strong starting point for intelligent, explainable recovery decision-making. With the proposed extensions, it can become even more reliable, ethical, and suitable for operational use in financial institutions.

References

Aruleba, I. and Sun, Y. (2024). Effective credit risk prediction using ensemble classifiers with model explanation, *IEEE Access* **12**: 115015–115025.

Bellotti, A., Brigo, D., Gambetti, P. and Vrins, F. (2021). Forecasting recovery rates on non-performing loans with machine learning, *International Journal of Forecasting* **37**(1): 428–444.
URL: *https://www.sciencedirect.com/science/article/pii/S016920702030100X*

Blanco, R., Fernández-Ortiz, E., García-Posada, M. and Mayordomo, S. (2024). A new estimation of default probabilities based on non-performing loans, *Finance Research Letters* **62**: 105149.
URL: *https://www.sciencedirect.com/science/article/pii/S154461232400179X*

bojrick.eth (2020). Lending club 2007–2020q3, `https://www.kaggle.com/datasets/ethon0426/lending-club-20072020q1`. Retrieved June 24, 2025, from Kaggle.

Botha, A., Beyers, C. and de Villiers, P. (2021). Simulation-based optimisation of the timing of loan recovery across different portfolios, *Expert Systems with Applications* **177**: 114878.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417421003195*

Chai, Y. and Sun, S. (2024). Can the development of fintech mitigate non-performing loan risk?, *Finance Research Letters* **67**: 105889.
URL: *https://www.sciencedirect.com/science/article/pii/S154461232400919X*

Chang, A.-H., Yang, L.-K., Tsaih, R. and Lin, S.-K. (2022). Machine learning and artificial neural networks to construct p2p lending credit-scoring model: A case using lending club data, *Quantitative Finance and Economics* **6**: 303–325.

Chen, X., Wang, G. and Zhang, X. (2019). Modeling recovery rate for leveraged loans, *Economic Modelling* **81**: 231–241.
URL: *https://www.sciencedirect.com/science/article/pii/S0264999318318339*

Gamba-Santamaria, S., Melo-Velandia, L. F. and Orozco-Vanegas, C. (2024). Decomposition of non-performing loans dynamics into a debt-servicing capacity and a risk taking indicators, *The Quarterly Review of Economics and Finance* **96**: 101860.
URL: *https://www.sciencedirect.com/science/article/pii/S1062976924000607*

Gao, W., Ju, M. and Yang, T. (2023). Severe weather and peer-to-peer farmers' loan default predictions: Evidence from machine learning analysis, *Finance Research Letters* **58**: 104287.
URL: *https://www.sciencedirect.com/science/article/pii/S1544612323006591*

Li, L.-H., Sharma, A. K. and Cheng, S.-T. (2025). Explainable ai based lightgbm prediction model to predict default borrower in social lending platform, *Intelligent Systems with Applications* **26**: 200514.
URL: *https://www.sciencedirect.com/science/article/pii/S2667305325000407*

Lin, S., Song, D., Cao, B., Gu, X. and Li, J. (2025). Credit risk assessment of automobile loans using machine learning-based shapley additive explanations approach, *Engineering Applications of Artificial Intelligence* **147**: 110236.
URL: *https://www.sciencedirect.com/science/article/pii/S0952197625002362*

Liu, Z., Zhang, Z., Yang, H., Wang, G. and Xu, Z. (2023). An innovative model fusion algorithm to improve the recall rate of peer-to-peer lending default customers, *Intelligent Systems with Applications* **20**: 200272.
URL: *https://www.sciencedirect.com/science/article/pii/S2667305323000972*

M, R. and M U, V. K. (2024). A study on application of explainable ai for credit risk management of an individual, *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, pp. 1–7.

Muslim, M. A., Nikmah, T. L., Pertiwi, D. A. A., Subhan, Jumanto, Dasril, Y. and Iswanto (2023). New model combination meta-learner to improve accuracy prediction p2p lending with stacking ensemble learning, *Intelligent Systems with Applications* **18**: 200204.
URL: *https://www.sciencedirect.com/science/article/pii/S2667305323000297*

Nallakaruppan, M., Balusamy, B., Shri, M. L., Malathi, V. and Bhattacharyya, S. (2024). An explainable ai framework for credit evaluation and analysis, *Applied Soft Computing* **153**: 111307.
URL: *https://www.sciencedirect.com/science/article/pii/S1568494624000814*

Nwafor, C. N. and Nwafor, O. Z. (2023). Determinants of non-performing loans: An explainable ensemble and deep neural network approach, *Finance Research Letters* **56**: 104084.
URL: *https://www.sciencedirect.com/science/article/pii/S1544612323004567*

Raşid Bakır, M., Atalay Çetin, M. and İbrahim Bakırtaş (2025). Revisiting the macroeconomic determinants of non-performing loans with a deep learning technique with causal inference: Evidence from türkiye, *Borsa Istanbul Review* **25**(3): 541–551.
URL: *https://www.sciencedirect.com/science/article/pii/S2214845025000420*

Sivamayilvelan, K., Rajasekar, E., Vairavasundaram, S., Balachandran, S. and Suresh, V. (2024). Flexible recommendation for optimizing the debt collection process based on customer risk using deep reinforcement learning, *Expert Systems with Applications* **256**: 124951.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417424018189*

Stevens, A., Deruyck, P., Veldhoven, Z. V. and Vanthienen, J. (2020). Explainability and fairness in machine learning: Improve fair end-to-end lending for kiva, *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1241–1248.

Tareaf, R. B., AbuJarour, M. and Zinn, F. (2024). Revolutionizing credit risk: A deep dive into gradient-boosting techniques in ai-driven finance, *2024 International Conference on Information Networking (ICOIN)*, pp. 322–327.

Wahab, F., Khan, I. and Sabada, S. (2024). Credit card default prediction using ml and dl techniques, *Internet of Things and Cyber-Physical Systems* **4**: 293–306.
URL: *https://www.sciencedirect.com/science/article/pii/S2667345224000087*

Wang, G., Liu, S. and Liu, Z. (2023). Credit default prediction based on adaptive strategy for bwoa-catboost, *2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML)*, pp. 543–551.

Xia, Y., Han, Z., Li, Y. and He, L. (2025). Credit scoring model for fintech lending: An integration of large language models and focalpoly loss, *International Journal of Forecasting* **41**(3): 894–919.
URL: *https://www.sciencedirect.com/science/article/pii/S0169207024000724*

Yu, C., Jin, Y., Xing, Q., Zhang, Y., Guo, S. and Meng, S. (2024). Advanced user credit risk prediction model using lightgbm, xgboost and tabnet with smoteenn, *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pp. 876–883.

Yu, T. (2024). Research on bank credit risk control based on ae-catboost, *2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*, pp. 1788–1791.

Zhang, Z. and Wang, Z. (2022). Research on credit scoring based on transformer-catboost network structure, *2022 IEEE 12th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 75–79.