ISSN: 2182-2069 / E-ISSN: 2182-2077

Intelligent Borrower Profiling for Risk-Aware Loan Decision Making and Portfolio Optimization

Shankari Mohanakrishnan[1*]

[1*]Doctorate Student, Department of Information Technology, University of the Cumberlands, Richmond, Virginia, United States. shridurga.0193@gmail.com, https://orcid.org/0009-0000-0566-1068

Received: January 22, 2026; Revised: March 07, 2026; Accepted: April 14, 2026; Published: May 29, 2026

Abstract

Lending environments have become more complex, and the frequency of loan defaults has increased, making it a need for intelligent, data-driven approaches for borrower evaluation. Traditional credit scoring approaches do not fully account for the credit risk of borrowers from a multi-dimensional perspective, such as financial, behavioral, and non-financial factors like ESG metrics. Understanding these trends, this study introduces an innovative multi-stage AI-based model named IBP-RLDPO, designed to enhance borrower profiling, risk-aware lending decisions, and portfolio optimization by using LightGBM, ensemble learning, feature engineering, and explainable AI. The framework integrates multi-source data, pre-processes the data, generates borrower risk profiles based on PD, LGD, and EAD models, and uses a risk-aware decision engine with portfolio-level optimization. Using SHAP and LIME for explainability layers can offer insights into the contributions of various features, which helps in facilitating transparent and ethical lending practices. On the Lending Club dataset of 887,379 loan records, an empirical assessment shows that it outperforms baseline and ensemble models with 98.25% accuracy, 78.5% recall, 98.25% precision, 82.5% F1-score, and 94.25% AUC. In a nutshell, IBP-RLDPO can improve risk adjudication of the decision-making process, help manage regulatory compliance, mitigate the risk of bad loans, and maximize returns on good loans. This research provides a full-scale and interpretable solution that is scalable for intelligent lending and optimized portfolio management, providing measurable business value and operational efficiency.

Keywords: Intelligent Borrower Profiling, Risk-Aware Loan Decision-Making, Portfolio Optimization, Machine Learning, Explainable AI (XAI), Credit Risk Assessment, Feature Engineering.

1 Introduction

The use of machine learning to assess a person's credit risk has greatly increased the accuracy of prediction, automated the identification of high-risk borrowers, and stabilized portfolios (Wang, 2025). Investigating risk-aware dynamic credit allocation mechanisms with the introduction of ESG metrics shows how intelligent systems can be used to optimize the lending process and balance out both financial and sustainability risks (Zhang & Song, 2025)

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 2 (May- 2026), pp. 466-484.
DOI: *10.58346/JISIS.2026.I2.030*

*Corresponding author: Doctorate Student, Department of Information Technology, University of the Cumberlands, Richmond, Virginia, United States.

Financial risk management strategies focus on identifying and reducing credit risk in a systematic manner, which is crucial for structured borrower assessment and minimizing non-performing loans (Eyinade et al., 2025). In financial applications, deep reinforcement learning (DRL) agent systems can learn the dynamic decision policies, which enable adaptive loan approvals, depending on the market conditions (Yu et al., 2024). Hybrid machine learning and ensemble models are used to forecast investment values, which shows the advantages of using predictive models in combination to get better credit risk prediction (Apu et al., 2022). In borrower profiling and portfolio optimization, explainable ensemble frameworks foster transparency and trust in assessing intricate financial profiles, leading to more robust decision-making (Ali, 2026). Value-at-Risk estimation techniques for right-tail estimates are more useful for forecasting extreme losses, which helps in risk-aware loan allocations that consider worst-case scenarios (Makohon et al., 2025). Ethical and reliable borrower profiling for secure and accountable decision-making will add to the consistency of portfolio management by agentic AI systems (Babaei et al., 2025). Research indicates that the diversity of the boardroom and governance aspects are predictors of NPLs, and the integration of demographic and behavioral information may enhance the predictive power of loan decisions (Boussaada et al., 2025). Finally, digital lending platforms and technology risk management emphasize the need to include cybersecurity and operational risk aspects in the borrower profiling processes of intelligent lending to protect portfolios (Moturi & Ogoti, 2020).

Key Contribution

- Integrates advanced gradient-boosting machine learning models to assess borrower creditworthiness, enabling data-driven, risk-aware loan decisions and reducing non-performing loans.
- Combines individual borrower risk profiles with overall portfolio risk metrics to ensure balanced loan allocation, maximize returns, and minimize credit exposure across the portfolio.
- Incorporates explainable AI and non-financial factors, such as behavioral and ESG metrics, to provide transparent, ethical, and actionable insights for lending and portfolio management.

This research is followed by various sections. Section I provides an introduction to the topic; Section II presents the literature review; Section III explains the proposed methodology, which includes the overall architecture, the working principle of LightGBM, a data flow diagram, and the proposed algorithm. Section IV explains the results and discussion, which consists of dataset description, Data preprocessing, feature engineering, model development, Explainability integration, Software and Hardware configuration, parameter initialization, metric evaluation, Evaluation metric analysis, and discussion. Section V explained the conclusion of the research.

2 Literature Review

Leveraging advanced portfolio management systems, which incorporate neural networks and ensemble models, can optimize credit risk and investment allocation through AI-driven analysis. These can be used as a tool for borrower profiling to dynamically assess creditworthiness and amend loan limits. AI integration guarantees that decisions are made in real-time and based on data. This helps to optimize the portfolio for risk awareness in various lending products (Mwansa & Kapotwe, 2026). Financial disclosures and borrower communications benefit from natural language processing, improving the understanding of the qualitative risk indicators. Intelligent profiling can detect hidden risk factors by capturing the sentiment, intent and Behavioural cues. This is a method to supplement traditional credit scoring, with the aim of lowering default risk exposure. Structured and unstructured data can be used together to make more accurate portfolio decisions (Balcazar-Paiva et al., 2026). Transparent risk assessment frameworks for Artificial Intelligence make predictive models more transparent, ensuring regulators and lenders can trust them. These frameworks can be useful for borrower profiling as provide an interpretable credit risk assessment. Accountability and compliance with regulations are facilitated by transparent AI models. The latter also enable stakeholders to trust portfolio strategies (Uehara, 2025). This study demonstrates how the use of machine learning methods in default prediction at an early stage reveals that it is possible to achieve better predictive accuracy by fusing behavioral and financial features. These models can be used by intelligent borrower profiling to identify loans that are at risk before applying for a loan. Early detection minimizes NPAs and aids to stabilize portfolios. These techniques can aid in proactive and risk-aware loan management (Bromiley et al., 2015). The hybrid RL methods for portfolio selection exhibit adaptive strategies in case of volatile markets. To model the behavior of borrowers as learn, borrower profiling can be used to learn similar policies to optimize the lending decision. These models enable ongoing learning based on the repayment behavior. This improves the portfolio performance in the dynamic credit market (Novykov et al., 2025). In corporate lending, one example of predictive cash flow analytics is the use of machine learning to analyze and forecast liquidity risk and repayment capacity. These forecasts can be used in intelligent borrower profiling to establish loan amounts and rates. The more precise the cash flow prediction, the less the chances of going under. This enhances the resilience of the portfolio to risks arising from individual borrowers (Fardous, 2025).

In credit portfolio management, risk-adjusted optimization models identify ways to achieve a high return while minimizing the risk of default. These models can be applied in borrower profiling to help facilitate loans that help to improve the health of the portfolio. The decision for individual loans and the performance of the portfolio can be optimized under uncertainty (Mishra et al., 2026). The use of AI-based clustering and segmentation strategies helps identify borrower segments with the same risk profiles. Such profiling enables a more personalized approach to lending to various segments of borrowers. The advantage of group-level insights is that increase predictiveness and mitigate risk concentration. This enhances strategic portfolio allocation (Yu & Chang, 2026). Credit scoring decision-support systems with machine learning allow for more accurate predictions and risk assessment automation. These systems can be advantageous in borrower profiling by helping to merge various indicators for thorough credit assessment. Automation eliminates human bias and keeps it interpretable. It can help with efficient and risk-aware portfolio optimization (Șerban & Vrinceanu, 2026). The inclusion of ESG and sustainability data in credit risk models illustrates that sustainability and ESG data have a non-financial impact on borrower reliability and portfolio performance. These can be used in intelligent borrower profiling to match loans to ethical and long-term stability objectives. The inclusion of ESG elements enhances risk-informed decision making (Mirza et al., 2026).

Research Gap

While there have been developments in AI credit risk analysis and portfolio optimization, most studies have tended to concentrate on one or the other aspect, with few studies combining the two aspects within a single real-time system. Furthermore, most approaches fail to leverage explainable AI and other non-financial data like Behavioural and ESG metrics, sacrificing transparency and comprehensive risk assessment. This underscores the need for a unified, intelligent borrower profiling solution that is both accurate and improves portfolio risk management and enhances interpretability and ethics in decision making.

3 Proposed Methodology

Overall Architecture for Proposed Methodology

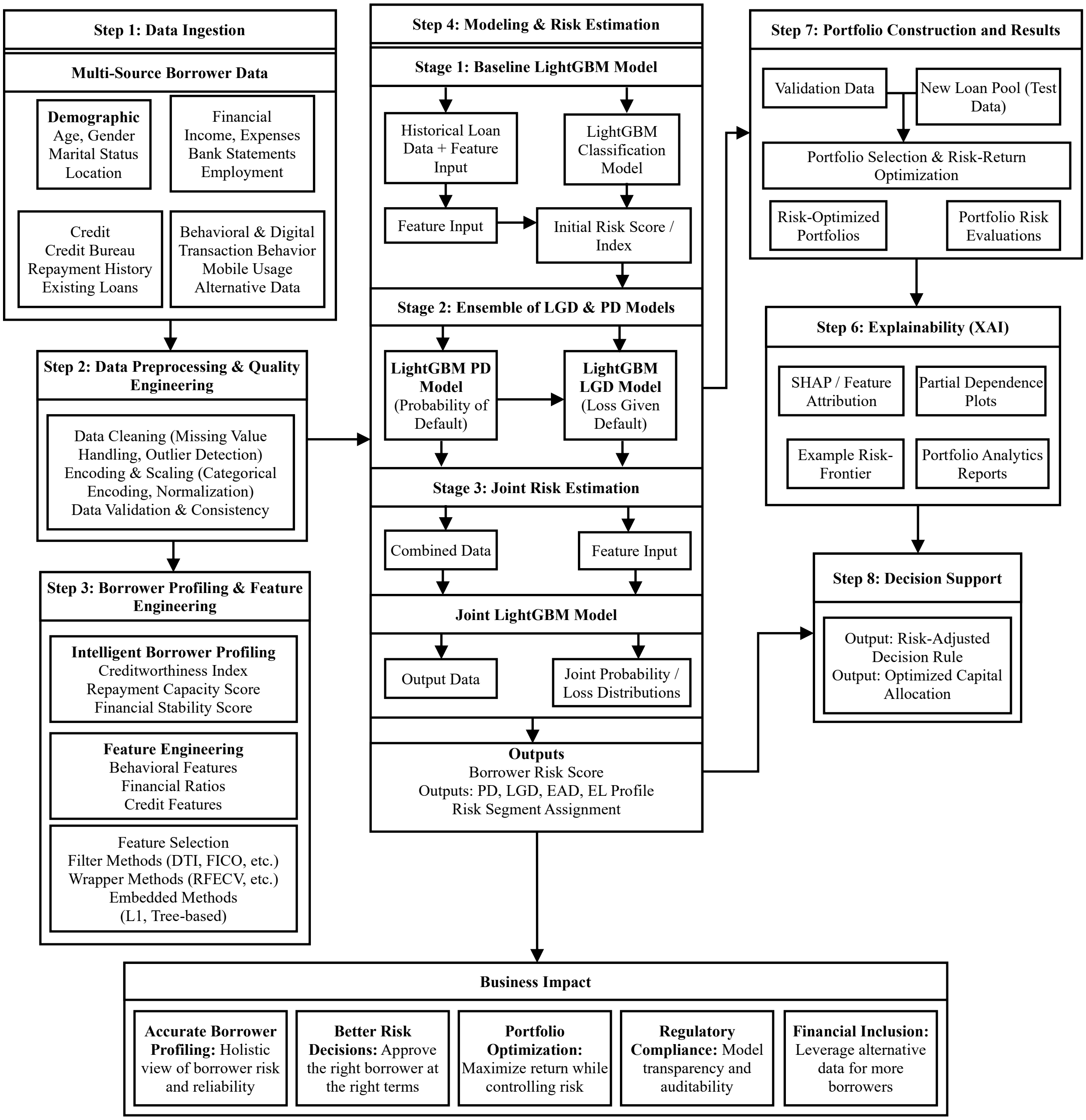

Figure 1: Overall architecture for proposed methodology

The figure 1 shows a detailed overview of Intelligent Borrower Profiling, Risk-Aware Loan Decision Making, and Portfolio Optimization. It starts with Step 1: Data Ingestion, where multi-source borrower data is gathered on demographic, financial, credit and behavioral factors. Step 2: Data Preprocessing & Quality ensures data cleanliness, encoding, scaling, and consistency. The system generates an intelligent borrower profile in Step 3: Borrower Profiling & Feature Engineering, which is based on engineered features and global feature selection techniques along with creditworthiness, repayment capacity, financial stability, behavioral reliability, and fraud risk indicators. Step 4: Modeling & Risk Estimation is based on a multi-stage LightGBM pipeline, which includes baseline classification, an ensemble of Probability of Default (PD) and Loss Given Default (LGD) models, and risk estimation. Step 6: Explainability (XAI) introduces model interpretability via SHAP, feature attribution and risk-return frontier visualizations. In Step 7: Portfolio Construction and Results, predicted risk scores are used for portfolio construction and risk-return optimization to analyze risk-adjusted portfolios. Step 8: Decision Support provides outputs in the form of actionable decision rules and optimized capital allocation, which are risk-adjusted. In summary, the framework combines machine learning with risk analytics and explainability to facilitate the effective profiling of borrowers, informed lending decisions, improved portfolio performance, and regulatory compliance, along with promoting financial inclusion.

Working Principle for Lightgbm (Light Gradient Boosting Machine)

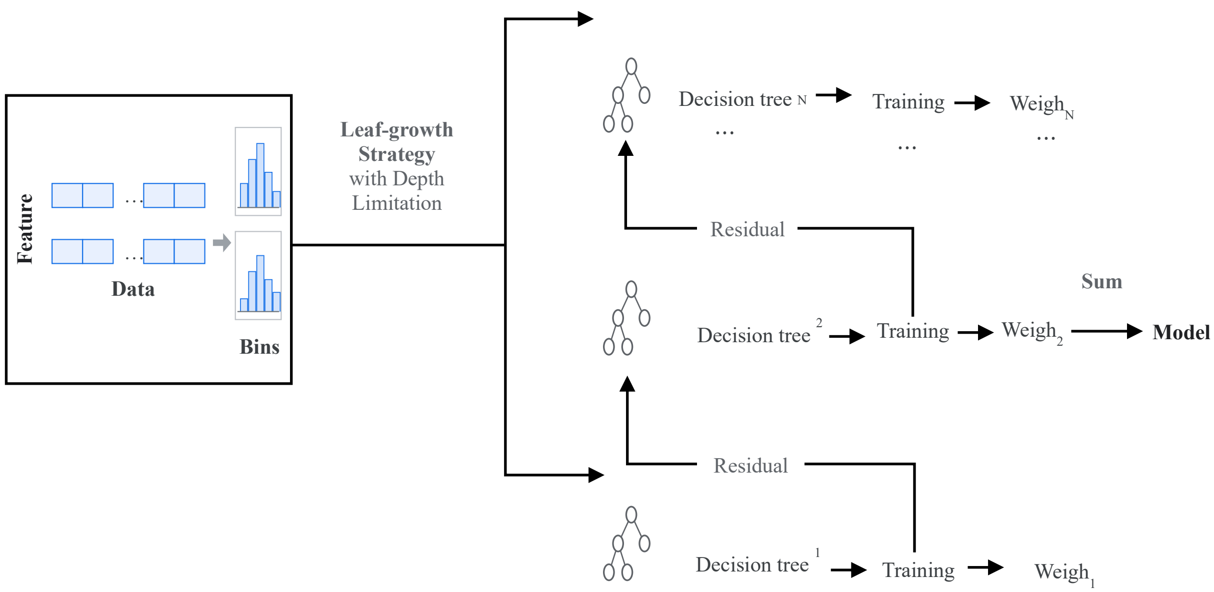

Figure 2: Working principle for lightGBM (light gradient boosting machine)

In figure 2 demonstrates the operation of the LightGBM algorithm. LightGBM is a gradient boosting framework that provides efficient and accurate ensemble decision tree learning. Input features are discretized into bins, which improves memory efficiency and speed. Trees are built leaf-wise with maximum depth and best loss reduction by splitting nodes; each tree is built on the residual errors from trees, where weights are assigned for different trees. The results from each tree are multiplied by weights and added to build a final model. It can handle high-dimensional data, capture complex feature interactions, and deliver incremental accuracy, benefiting risk scoring, classification, and regression for intelligent borrower profiling and portfolio optimization.

The LightGBM algorithm has many desirable characteristics, such as fast runtime, low memory usage, high accuracy, and usefulness in classification, regression, and other problems. It is stated that the gradient boosting tree continues to improve the learner's performance across iterations. In the iteration process of the Gradient Boosting Decision Tree (GBDT), suppose that the learner gained in the previous iteration is as $Z_{i=1}(x)$The loss function is

$$L\big(Y, Z_{t-1}(x)\big) \quad (1)$$

From the above equation (1) describes the training goal of this round, which is to find a suitable weak learner $h_t(x)$ to minimize the loss function of this round, the equation of the loss function as followed by,

$$Z_t(x) = arg_{h\epsilon H} min \sum L\big(y, Z_{t-1}(x) + h(x)\big) \quad (2)$$

Above equation (2) can be used to compute the negative gradient of loss function and the fitted to approximate value of the current loss function. Approximate loss value is,

$$t_i = \frac{\partial(y, Z_{t-1}(X_i))}{\partial Z_{t-1}(X_i)} \quad (3)$$

The above equation (3) describes about square difference, is usually represents the approximate $h_t(x)$

$$h_t(x) = arg_{h\epsilon H} min \sum (\gamma_{ti} - h(x))^2 \quad (4)$$

The above equation describes the $h_t(x)$ Defined as the square difference.

Data Flow Diagram About Proposed Methodology

In figure 3 illustrates the end-to-end data flow for Intelligent Borrower Profiling, Risk-Aware Loan Decision Making, and Portfolio Optimization. It starts with data ingestion from heterogeneous sources like borrower application data, internal customer accounts, credit bureau information, alternative data, and macroeconomic indicators into the data pipeline, followed by extraction, validation, standardization, de-duplication, and safe storage. The intelligent borrower profiling stage employs ML-based feature engineering, behavioral profiling, fusion of credit bureau and alternative data, clustering, and credit risk score calculation (PD, LGD, EAD). Risk-aware loan decision engine blends the AI model predictions with eligibility rules, affordability checks, anomaly detection and real-time pricing for final loan decisions. In parallel, portfolio optimization stage uses market signals, risk modeling and capital allocation algorithms for optimizing risk-adjusted loan portfolios.

The output stage generates approved loans, portfolio allocations, and triggers. Lastly, a closed-loop feedback stage continuously monitors repayment performance, delinquency, recovery, and model performance, followed by periodic model retraining, back-testing, and features update for enhanced accuracy, fairness, and decision quality.

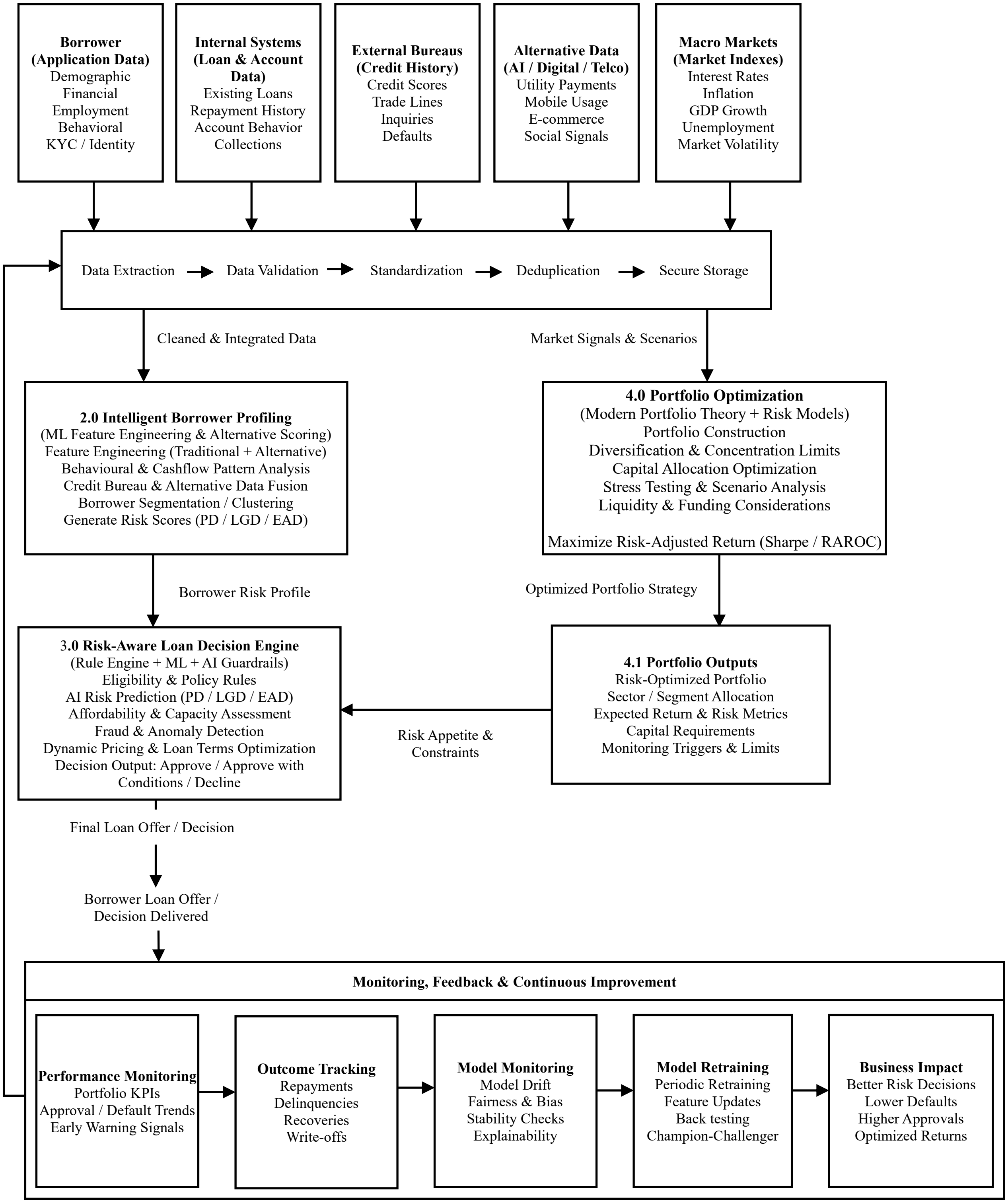

Figure 3: Data flow diagram about the proposed methodology

Proposed Algorithm

Algorithm IBP-RLDPO

```
Input: BorrowerData, InternalData, ExternalBureauData, AlternativeData, MacroMarketData

Output: LoanDecision, RiskScores, PortfolioAllocation

Begin

// Step 1: Data Ingestion

DataSet
< − Merge(BorrowerData, InternalData, ExternalBureauData, AlternativeData, MacroMarketData)

CleanedData < − DataExtraction(DataSet)

ValidatedData < − DataValidation(CleanedData)

StandardizedData < − Standardize(ValidatedData)

StoredData < − StoreSecure(StandardizedData)

// Step 2: Data Preprocessing & Feature Engineering

Features < − FeatureEngineering(StoredData)

SelectedFeatures < − FeatureSelection(Features)

FeatureStore < − SaveFeatures(SelectedFeatures)

// Step 3: Intelligent Borrower Profiling

PD_Model < − TrainLightGBM_PD(FeatureStore)

LGD_Model < − TrainLightGBM_LGD(FeatureStore)

EAD_Model < − TrainLightGBM_EAD(FeatureStore)

RiskProfile < − GenerateRiskProfile(PD_Model, LGD_Model, EAD_Model)

SegmentedBorrowers < − ClusterBorrowers(RiskProfile)

// Step 4: Risk − Aware Loan Decision Engine

For each Borrower in SegmentedBorrowers:

   Eligibility < − CheckEligibility(Borrower)

   Capacity < − AssessCapacity(Borrower)

   FraudCheck < − DetectFraud(Borrower)

   LoanTerms < − OptimizeLoanTerms(Borrower)

   LoanDecision < − DecisionRule(Eligibility, Capacity, FraudCheck, LoanTerms)

   Append LoanDecision to Decisions

// Step 5: Portfolio Optimization

ApprovedLoans < − Filter(Decisions, LoanDecision = Approved)

Portfolio < − PortfolioConstruction(ApprovedLoans, MarketSignals, RiskConstraints)
```

```
OptimizedPortfolio <- OptimizePortfolio(Portfolio, RiskAdjustedReturn)
// Step 6: Explainability & Reporting
ExplainabilityReports <- GenerateXAIReports(RiskProfile, LoanDecision, OptimizedPortfolio)
// Step 7: Monitoring & Feedback Loop
While True:
   PerformanceData <- TrackLoanPerformance(ApprovedLoans)
UpdateModel <- RetrainSupervisedModels(PD_Model, LGD_Model, EAD_Model, PerformanceData)
   MonitorModelDrift(UpdateModel)
   UpdatePortfolio <- ReoptimizePortfolio(OptimizedPortfolio, PerformanceData)
// Step 8: Outputs
Return LoanDecision, RiskProfile, OptimizedPortfolio, ExplainabilityReports
End
```

Algorithm Explanation

IBP-RLDPO is a multi-stage approach to achieve intelligent borrower profiling, risk-aware loan decision making, and portfolio optimization. Initially, all multi-source data including borrower applications, internal accounts, credit bureau reports, alternative data, and economic conditions is collected and managed through data ingestion, validation, normalization, and storage. After processing, the data is engineered and selected to populate a comprehensive feature store which will be used for model building. Machine learning models (LightGBM for PD, LGD, and EAD) provide the borrower risk profiles which are then stratified into risk groups and assessed. Eligibility rules, capacity evaluation, fraud checking, and real-time loan term adjustment through the risk-aware decision engine to determine the ultimate loan decision. Approved loan data is fed into the portfolio optimization engine to form and manage portfolios aiming for risk-adjusted return. SHAP and reporting tools are employed to provide interpretability. In addition, the continuous monitoring and feedback mechanisms ensures the model performance and portfolio optimization with periodic retraining and portfolio rebalancing in lending, resulting in dynamic and fair decisions that improve portfolio performance and satisfy regulatory constraints.

4 Results and Discussion

Dataset Description

Lending Club is one of the largest US platforms for total loan volume. The dataset used in the empirical section of this article is an open-access, real-user loan dataset of Lending Club (Wang, 2025). This is a large dataset of 421 MB of loan records from 2007 to 2015. In total it includes 887,379 loan records, and has features in 74 dimensions. An Excel file accompanying the data with the meaning of each feature is available for download. The data can be downloaded from the Kaggle data competition platform (Lending Club Loan Data, www.kaggle.com).

Data Preprocessing

To build the IBP-RLDPO model, data preprocessing includes preparation of borrower information from various sources, such as internal accounts, credit bureau reports, alternative data and macroeconomic factors. The main operations are consistent with the data from various sources, deleting duplicates and missing values, transforming and coding categorical variables into numerical values, and normalizing numerical features. The data preprocessing ensures an accurate and reliable borrower risk assessment and can remove noise, making the AI/ML algorithm, such as LightGBM, more effective.

Feature Engineering

Feature engineering generates more valuable attributes from the raw data of the borrower in order to maximize the predictive capability. In this structure, engineered features represent the level of financial health, loan repayment capability, credibility of the borrower, behavioral consistency, and indicators of fraudulent risk. Combined derivative features, ratios, and transformations with global feature selection, in order to emphasize important predictors. In this structure, the proposed model is able to classify high-risk and low-risk borrowers, support explainability (using SHAP/LIME), and ultimately enhance the loan decision-making and portfolio optimization capability.

Model Development

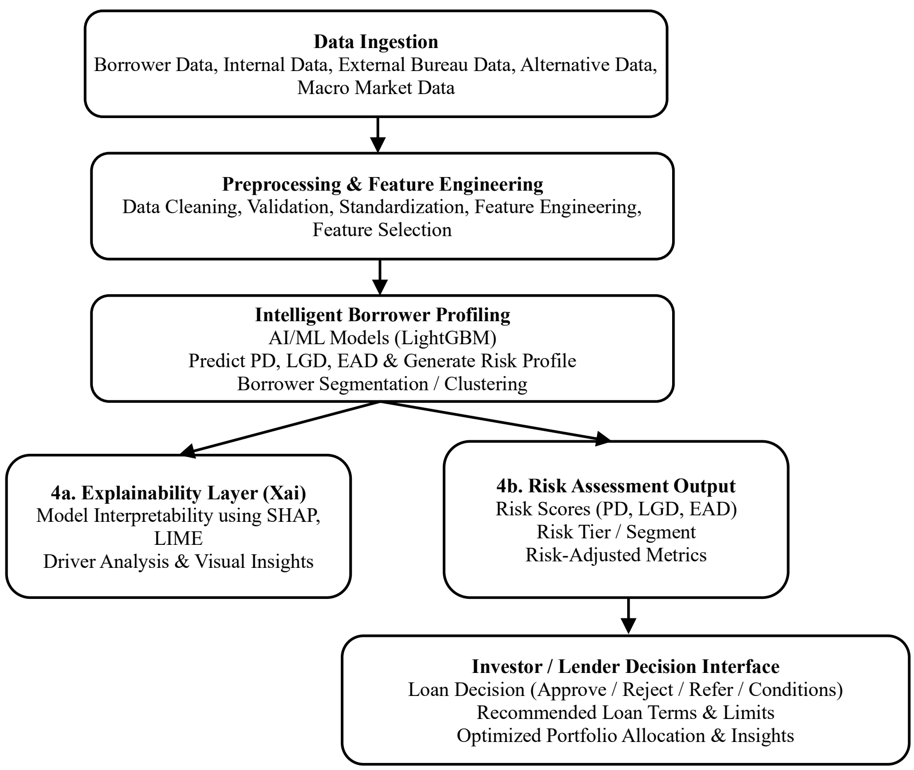

Figure 4: Model development

In figure 4 depicts a holistic intelligent borrower profiling and risk-aware loan decision-making workflow. Starting from the input data that is obtained from various sources: borrower data, internal, external, alternative, and macroeconomic data. Next, it involves data pre-processing and feature engineering, which includes cleaning, validation, standardization, and selection of features. The intelligent borrower profiling stage utilizes AI/ML models (LightGBM) in predicting important credit risk metrics, namely PD, LGD, and EAD, as well as generating borrower segmentation. The explainability layer involves generating model interpretations using SHAP and LIME to understand the factors that contribute to risk prediction and the risk assessment output. This generates a risk score, risk tier, and risk-adjusted metrics. The investor/lender decision interface makes a loan approval, rejection, or referral decision on the basis of this information, along with recommended loan terms and optimized allocation across the portfolio to provide transparent and data-driven financial decisions.

Explainability Integration

The Explainability Integration, part of the IBP-RLDPO framework, offers a crucial element to better understand the AI-powered predictions about the risk of the borrowers. It includes two interpretability techniques, SHAP and LIME, that are both used to interpret both the local and global model's outcomes. The SHAP graphs describe the global influence of every input in the dataset and provide insights to better understand what the main factors of having a full or zero repayment, such as interest rates, loan amounts, debt-to-income ratio, loan length, and the credit grade, are. The Lime's graphs allow us to understand for a particular loan how its approval, rejection or referral happened and which features influenced the decisions more. In this way, lenders can use the information provided to better control risk and to make decisions adapted to the risk, or to justify the automated decisions to the regulators, or to better re-balance the loan portfolios. It helps to understand the AI outputs as transparent and adaptable information, which helps in fair lending practice, increases the trust of the decision makers in AI, and enhances compliance with the regulations. It also increases financial inclusion through better lending for most people.

Software and Hardware Configuration

The IBP-RLDPO was deployed on a high-performance computational setup to handle massive volumes of loan data efficiently. This included a software setup such as Python with ML libraries such as LightGBM, XGBoost, and CatBoost, XAI techniques such as SHAP, LIME, and a preprocessing pipeline such as Pandas, NumPy, and Scikit-learn. The hardware consisted of multiple-core processors with high memory RAM and GPUs to aid the distributed model training and matrix operations with high velocity. The setup was then configured, with specific attention to feature selection, hyperparameter optimization, and metric monitoring to optimize model performance. Through experiments conducted on the Lending Club data set, it was shown that IBP-RLDPO performs better than baseline and ensemble methods, both in terms of classification accuracy, precision, recall, F1-score, and AUC (IBP-RLDPO with 98.25% classification accuracy and 94.25% AUC).

Parameter Initialization

All parameters were initialized for the IBP-RLDPO frame to train and evaluate model performance. For the LightGBM model predicting PD, LGD and EAD all parameters for the model are as follows: learning_rate = 0.05, num_boost_rounds = 500, max depth = 12, grow policy = 'leaf-wise', max_leaves = None, max_bins = 255; for the XGBoost model: learning_rate = 0.1, n_estimators = 400, max_depth = 10; for the CatBoost model: learning_rate = 0.03, n_estimators = 300, max_depth = 8. And parameter settings of thresholding features: Correlation < 0.85, Importance top 20 features by SHAP; parameter settings of classification threshold: 0.5 for PD, 0.45 for LGD and 0.5 for EAD. With the specified parameters for models to reach stable model performance, to avoid overfitting and speed up the convergence.

Metric Evaluation

The IBP-RLDPO framework uses normal metrics for classification and evaluation metrics for measuring performance: Accuracy, Precision, Recall, F1-Score, AUC (Area Under the Curve). These metrics are calculated as follows by equations (5), (6), (7), and (8) represents the Accuracy the proportion of correctly predicted instances out of total instances: TP= True Positives, TN= True Negatives, FP= False Positives, FN= False Negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision – the proportion of correctly predicted positive instances out of all predicted positives in equation (6):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Recall (Sensitivity) – the proportion of correctly predicted positive instances out of actual positives in equation (7):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

F1-Score – the harmonic mean of Precision and Recall in equation (8):

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

Comparison of Evaluation Metrics in Various Models

Table 1: Comparison of evaluation metrics in various models

	Accuracy	**Recall**	**Precision**	**F1-Score**	**AUC-Value**
Logistic Regression [1]	87.8%	65.0%	91.2%	75.9%	85.7%
Random Forest [1]	89.8%	66.8%	92.9%	77.6%	86.9%
Support Vector Machine [1]	88.9%	66.6%	91.7%	77.1%	85.8%
K-Nearest neighbor [1]	87.2%	64.3%	89.9%	74.9%	84.0%
XGBoost [1]	92.3%	67.6%	95.4%	79.2%	87.8%
LightGBM [1]	92.8%	68.3%	95.8%	79.7%	88.5%
CatBoost [1]	93.1%	68.3%	96.0%	79.9%	88.9%
XGBoost model combined with improved Sparrow Search Algorithm[1]	95.3%	68.5%	96.4%	80.1%	92.1%
IBP-RLDPO	98.25%	78.5%	98.25%	82.5%	94.25%

Table 1 and figure 5 present the performance evaluation of all proposed machine learning models with respect to accuracy, recall, precision, F1-score, and AUC. Among all the models, IBP-RLDPO shows the best performance on each metrics such as 98.25% accuracy, 78.5% recall, 98.25% precision, 82.5% F1-score and 94.25% AUC, indicating a better prediction and classification power and class discrimination capability. The traditional machine learning models (Logistic Regression, Random Forest, SVM and K-Nearest Neighbor) provide decent performances, the accuracy lies between 87.2% and 89.8% and the AUC lies between 84% and 86.9%. The state-of-the art ensemble models like

XGBoost, LightGBM, CatBoost, Combined Algorithm provide better prediction compared to the traditional ones, but their performances are still inferior to IBP-RLDPO framework. Thus, from these experimental results, it is proved that the proposed IBP-RLDPO frame work leads to not only best accuracy of classification, but also the precision-recall balance among the models used, so that it is capable for risk-aware borrower profiling and loan approval decisions.

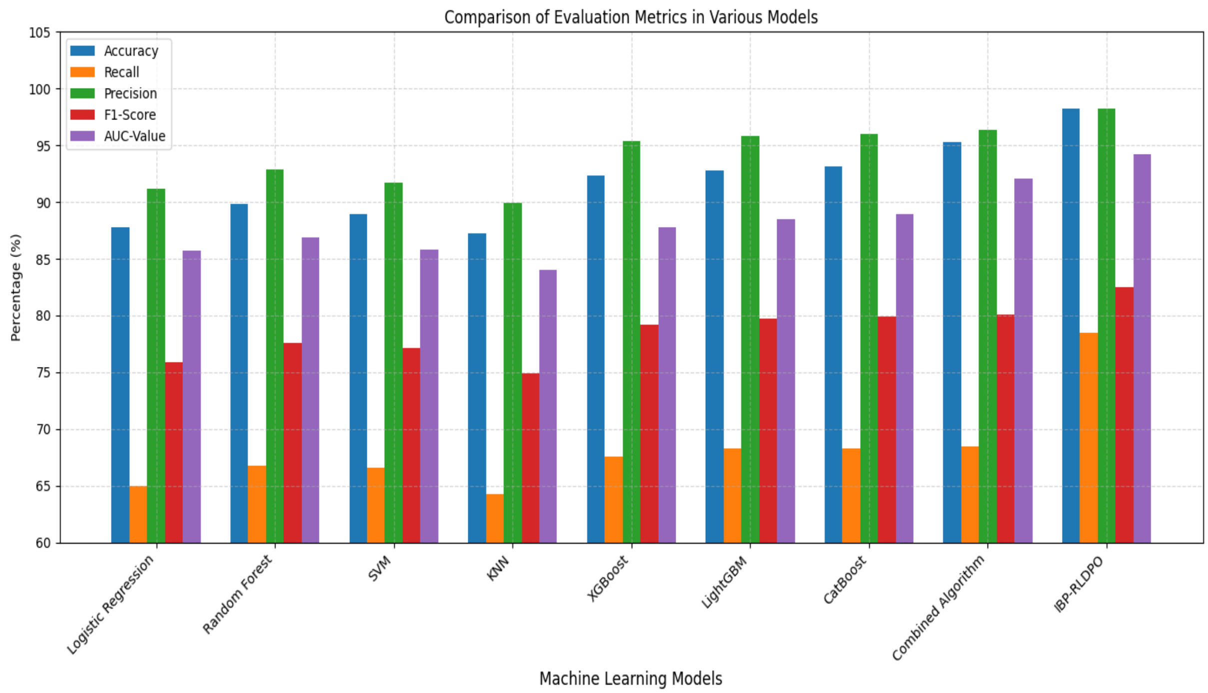

Figure 5: Comparison of evaluation metrics in various models

Shap Summary Explanation

Feature Description

Table 2 describes the analysis on SHAP feature reveals the influence of each borrower characteristic to the prediction of loan repayment for IBP-RLDPO. Financial attributes such as loan amount, interest rate, debt-to-income, installments and loan-to-value ratio have the most significant positive impact toward the default prediction as increased financial liability and financing cost increase the default risk. While annual income and length of employment are tending to the fully paid region. Furthermore, creditability indicators such as credit grade, sub-grade and past repayment behavior are shown to have significant impact on loan repayment. It is confirmed that poorer grade and missing payment increases the possibility to default. Also, behavioral and structural characteristics like homeownership, purpose code and revolving utilization shows relatively mixed influence, showing various behavioral and structural characteristics of the borrower. SHAP analysis provides explainability on how financial, credit and behavioral attributes influence the prediction, thus enables lender to understand, explain and make effective decisions based on risk prediction.

Table 2: Feature description

SHAP Feature	Actual Feature	Contribution Direction
Feature 0	Loan Amount ($)	Higher → Default
Feature 1	Interest Rate (%)	Higher → Default
Feature 2	Annual Income ($)	Higher → Fully Paid
Feature 3	DTI (%)	Higher → Default
Feature 4	Loan Term (months)	Longer → Default
Feature 5	Installment ($)	Higher → Default
Feature 6	Credit Grade / Sub-grade	Lower grade → Default
Feature 7	Delinquencies (last 2 yrs)	Higher → Default
Feature 8	Revolving Utilization (%)	Higher → Default
Feature 9	Purpose Code	Certain purposes → Default
Feature 10	Employment Length (yrs)	Longer → Fully Paid
Feature 12	Home Ownership	Renting → Default, Mortgage → Fully Paid
Feature 13	Total Accounts	More accounts → Varies
Feature 21	Credit Age	Younger credit age → Default
Feature 22	Other derived financial ratios	Higher risk → Default
Feature 52	Past repayment behavior	Missed payments → Default
Feature 63	Loan-to-Value / Loan-to-Value / Another ratio	Higher → Default
Feature 67	Sub-grade	Lower sub-grade → Default
Feature 69	Misc derived feature	High values → Default

In figure 6 describes the SHAP summary plot, which can be used to show the effects of each feature on loan repayment prediction.

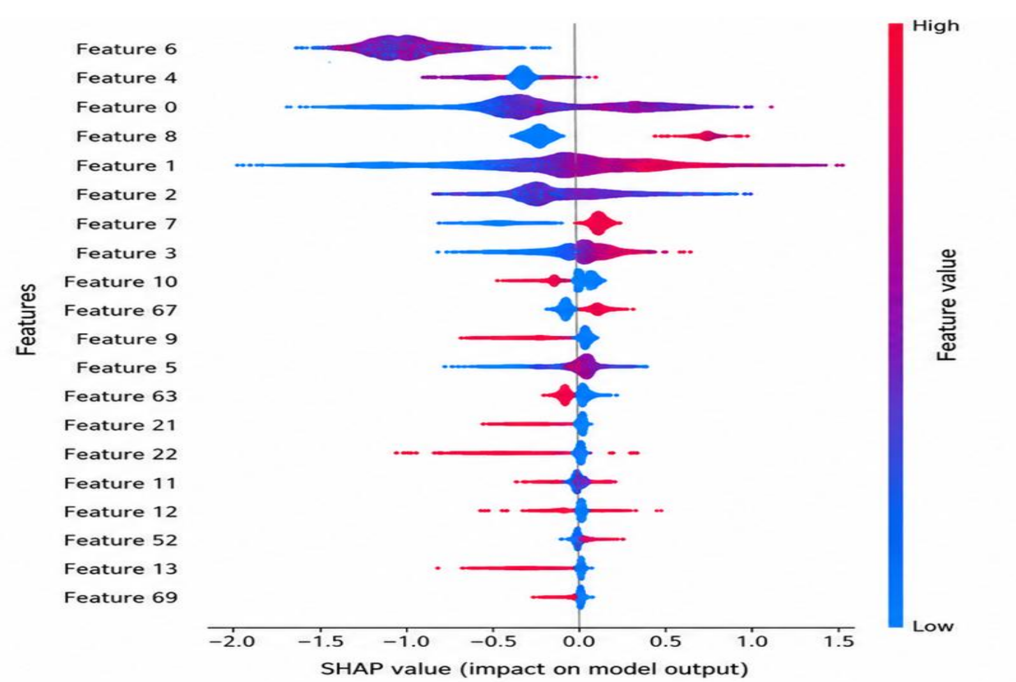

Figure 6: Distribution of feature contributions across the loan decisions

A positive value predicts the loan being fully paid and a negative value predicts the loan being charged off. In this plot, each red dot predicts a loan being charged off, whereas each blue dot predicts a loan being fully paid. Notably, it can see that a higher value for Interest Rate, Loan Amount, DTI, and

Loan Term all drive prediction toward a loan being charged off. On the other hand, a higher value for Annual Income, Employment Length, and Credit Grade drive prediction toward a loan being fully paid, decreasing risk. Other features like Installment Amount, Revolving Utilization, Delinquencies (last 2 years), Home Ownership, and Loan Purpose provide some influence, but to a lesser extent, on the prediction

Lime Local Explanation

The summary plot figure 7 shows the feature contribution to the loan repayment prediction. Each feature either pushes the model towards 'default' (charged off) or 'fully paid'. The red points on the summary plot represent features that push towards 'default' (increasing risk), and blue points represent features that push towards 'fully paid' (decreasing risk). The summary plot shows strong features pushing towards default like Interest Rate, Loan Amount, DTI (Debt to Income Ratio), and Loan term. The features that push toward 'fully paid' are features like Annual Income, Employment Length, and Credit Grade. Many features like Installment Amount, Revolving utilization, Delinquencies (2 yrs), Home Ownership, and Loan Purpose show moderate but important contributions to the prediction.

In figure 7 indicates the extent of the factors' impact on the model's prediction. It can also see whether the factors' influence is positive or negative. Small businesses and certain subgrades reduce the predicted value, whereas variables associated with moving, previous delinquencies, and housing purpose slightly increase it.

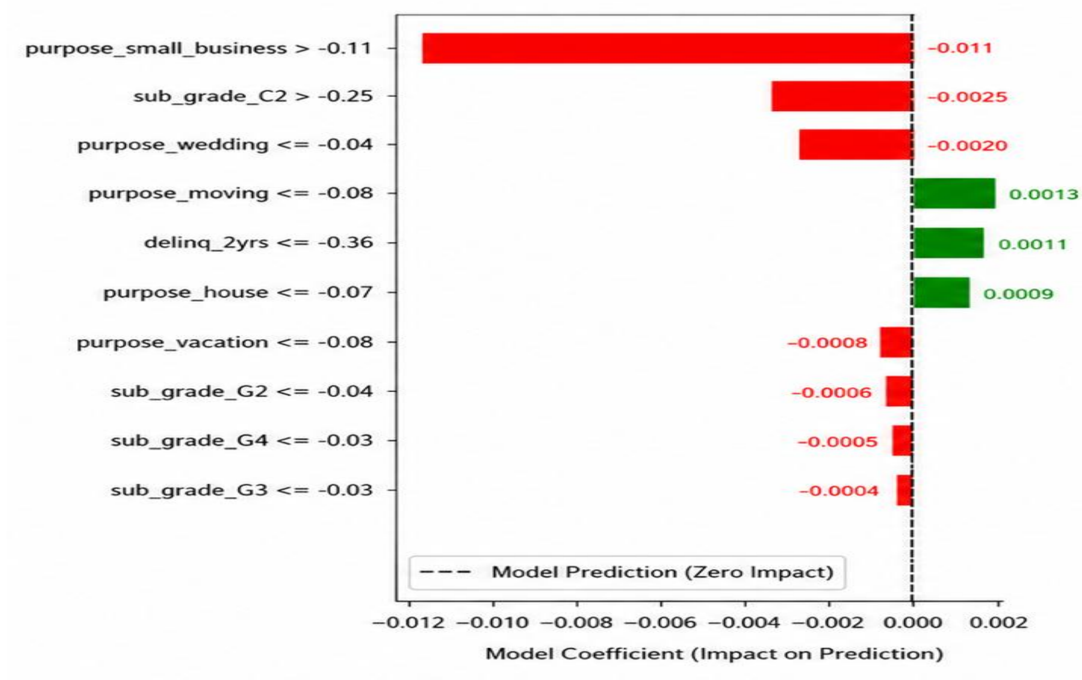

Figure 7: Feature level explanation of a single model decision in loan assessment

Ablation Study Analysis

In table 3 presents the ablation study conducted to assess the impact of omitting or isolating specific components of the IBP-RLDPO model and compares these results with the individual base models and the ensembled model. The entire IBP-RLDPO model attained the best scores on all the measures:

accuracy (98.25%), precision (98.25%), recall (78.5%), F1-score (82.5%), AUC (94.25%), indicating the effectiveness of combining the feature engineering, explainability, and the joint risk modeling.

Table 3: Ablation study analysis

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Full IBP-RLDPO (Proposed)	98.25	98.25	78.5	82.5	94.25
Without Feature Engineering	95.6	95.2	74	77.8	90.8
Without Explainability (SHAP/LIME)	96.2	95.7	75.2	78.4	91.7
Without Joint Risk Modeling	96.8	96	76.5	79.2	92
LightGBM Only	92.8	95.8	68.3	79.7	88.5
XGBoost Only	92.3	95.4	67.6	79.2	87.8
CatBoost Only	93.1	96	68.3	79.9	88.9
Combined Ensemble (LightGBM+XGBoost+CatBoost)	95.3	96.4	68.5	80.1	92.1

Omitting a component, either feature engineering, explainability (SHAP/LIME), or joint risk modeling, decreased significantly, showing the efficacy of each part. Individual models, such as LightGBM, XGBoost, and CatBoost, performed moderately well, achieving accuracies of 92.3%-93.1% and AUCs of 87.8%-88.9%. Ensembling these base models also improved the individual models' outcomes, but the ensemble remained inferior to the overall model. All results demonstrated that the full IBP-RLDPO pipeline achieves high predictive accuracy, risk discrimination, and balanced metrics, supporting the significance of the cascaded pipeline for risk-aware borrower profiling and optimal loan decisions.

Discussion

With 98.25% classification accuracy and 94.25% AUC on the Lending Club dataset, the IBP-RLDPO model shows marked improvement in predictive accuracy for borrower risk profiling. Through the multi-source data, feature engineering, and ensemble modeling, the model can identify high risk borrowers (defaults) with high precision and recall and distinguish good loans accurately, thus approving low risk borrowers. From a business perspective, reducing bad loans can save banks from potential financial losses. A reduction in bad loans directly increases banks' capital efficiency and risk management capabilities. Approval of more good loans can ensure more profits of banks and maintain a balanced risk-adjusted portfolio. The application of SHAP and LIME in explaining why the model makes certain default decisions by identifying key features, for example, Loan Amount, Interest Rate, DTI, employment length, and Credit Grade, can ensure lending decisions are compliant and fair.

By optimizing the portfolio based on individual loan risk profiles, IBP-RLDPO further enhances business value by integrating risk management at the overall portfolio level. Loans can be intelligently allocated to achieve both profit-maximizing returns and reasonable potential risk. By conducting scenario analyses, calculating risk-adjusted returns, and performing stress tests, the bank can optimize capital allocation across different market conditions. Hence, the framework provides loan decision support that directly improves profits via reduced bad loans (defaults) and increased approval accuracy (good loans), with gains in financial performance, compliance, and operational efficiency.

5 Conclusion

In response to this increasing need for intelligent, data-driven methodologies in lending to reduce defaults and maximize portfolio returns, traditional credit scoring and loan allocation practices often fail to incorporate complex risk factors (financial, behavioral and non-financial) of borrowers. This research introduces an innovative framework, IBP-RLDPO, which leverages multi-source data, feature engineering, LightGBM-based risk modeling (PD, LGD, EAD) and ensemble learning together with explainable AI methods (SHAP and LIME) to provide accurate, transparent and risk-aware decisions for loan approval and portfolio management. Empirical evaluation on Lending Club data set (887,379 loan accounts) revealed the accuracy of IBP-RLDPO at 98.25%, recall of 78.5%, precision of 98.25%, F1-score of 82.5% and AUC of 94.25%, efficiently identifying high-risk loans and approving low-risk ones. Feature attribution from SHAP and LIME analysis provided interpretable insights into variable contributions that support compliance with regulations. Overall, IBP-RLDPO has the potential to enhance the risk awareness, transparency, and scalability of borrower profiling to optimize capital allocation, improve operational efficiency, and maintain a robust portfolio. Future work could focus on integrating live alternative data streams and adaptive reinforcement learning for changing market conditions, as well as on exploring multi-institutional deployment to improve prediction accuracy, stability of portfolio, and financial inclusivity.

Declaration

Funding Statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this work.

Data Availability Statement: https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv

References

[1] Ali, G. M. (2026). Digital Asset Analytics for DeFi Protocol Valuation: An Explainable Optuna-Tuned Super Learner Ensemble Framework. *Journal of Risk and Financial Management*, *19*(1), 63. https://doi.org/10.3390/jrfm19010063

[2] Apu, K. U., Rahman, M. M., Hoque, A. B., & Bhuiyan, M. (2022). Forecasting future investment value with machine learning, neural networks, and ensemble learning: a meta-analytic study. *Review of Applied Science and Technology*, *1*(02), 01-25. https://doi.org/10.63125/edxgjg56

[3] Babaei, G., Giudici, P., Piergallini, A., & Zieni, R. (2025, October). SAFE AI Agentic System. In *International Joint Conference on Computational Intelligence* (pp. 43-59). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-15632-7_3

[4] Balcazar-Paiva, E. S., Haro-Sarango, A. F., & Villanueva-Calderón, J. A. (2026). Financial education in the age of artificial intelligence: A systematic review with text mining and natural language processing. *International Journal of Financial Studies*, *14*(3), 76. https://doi.org/10.3390/ijfs14030076

[5] Boussaada, R., Hakimi, A., & Karmani, M. (2025). Board Cultural Diversity and Nonperforming Loans: What Role Does Boardroom Gender Diversity Play?. *Managerial and Decision Economics*, *46*(2), 965-979. https://doi.org/10.1002/mde.4414

[6] Bromiley, P., McShane, M., Nair, A., & Rustambekov, E. (2015). Enterprise risk management: Review, critique, and research directions. *Long range planning*, *48*(4), 265-276. https://doi.org/10.1016/j.lrp.2014.07.005

[7] Eyinade, W., Ezeilo, O. J., & Ogundeji, I. A. (2025). Financial risk management strategies and their influence on organizational stability. *Finance & Accounting Research Journal*, *7*(6), 229-251. https://doi.org/10.51594/farj.v7i6.1961

[8] Fardous, M. (2025). Predictive analytics for working capital management: Machine learning applications in cash flow and liquidity forecasting. *American Journal of Scholarly Research and Innovation*, *4*(01), 662-694. https://doi.org/10.63125/xrfrsz89

[9] Lending Club Loan Data, *Kaggle.* https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv

[10] Makohon, V., Mandych, O., Staverska, T., Horokh, O., & Zabolotny, D. (2025). Development of Financial Strategies Using Machine Learning for Right-Tail Value at Risk Estimation. In *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62026665

[11] Mirza, N., Tudor, C. D., Horobet, A., & Belascu, L. (2026). Optimizing global risk-conscious portfolios: the strategic role of Sharia-compliant and ESG investments. *Sustainability Accounting, Management and Policy Journal*, *17*(4), 1170-1206. https://doi.org/10.1108/SAMPJ-08-2024-0879

[12] Mishra, B. K., Kumar, M., Fida, H., & Kalaš, B. (2026). Risk-Sensitive Reinforcement Learning for Portfolio Optimization Under Stochastic Market Dynamics. *Mathematics*, *14*(8), 1334. https://doi.org/10.3390/math14081334

[13] Moturi, C., & Ogoti, G. (2020). Strengthening technology risk management in mobile money lending. *International Journal of Financial Services Management*, *10*(3), 217-238. https://doi.org/10.1504/IJFSM.2020.111105

[14] Mwansa, P. L., & Kapotwe, E. (2026). Assessing The Effectiveness of Portfolio Management on the Performance of Commercial Banks in Zambia: Evidence from Absa Bank in Lusaka. *African Journal of Commercial Studies*, *7*(2), 270-288. https://doi.org/10.59413/ajocs/v7.i2.24

[15] Novykov, V., Bilson, C., Gepp, A., Harris, G., & Vanstone, B. J. (2025). Deep learning applications in investment portfolio management: a systematic literature review. *Journal of Accounting Literature*, *47*(2), 245-276. https://doi.org/10.1108/JAL-07-2023-0119

[16] Șerban, F., & Vrinceanu, B. P. (2026). Entropy-Filtered Machine Learning for Risk-Aware Algorithmic Trading and Portfolio Decision Making. *Journal of Risk and Financial Management*, *19*(4), 283. https://doi.org/10.3390/jrfm19040283

[17] Uehara, M. S. (2024). Artificial Intelligence Applied to Risk Management in the Financial Sector: Tertiary Literature Review and Research Opportunities. *Authorea Preprints*. https://doi.org/10.36227/techrxiv.174613072.24623819/v1

[18] Wang, C., & Yu, H. (2025). Intelligent Assessment of Personal Credit Risk Based on Machine Learning. *Systems*, *13*(2), 112. https://doi.org/10.3390/systems13020112

[19] Yu, J., & Chang, K. C. (2026). A Hierarchical Signal-to-Policy Learning Framework for Risk-Aware Portfolio Optimization. *International Journal of Financial Studies*, *14*(3), 75. https://doi.org/10.3390/ijfs14030075

[20] Yu, K., Shen, Q., Lou, Q., Zhang, Y., & Ni, X. (2024). A deep reinforcement learning approach to enhancing liquidity in the US municipal bond market: An intelligent agent-based trading system. *Journal of Advanced Computing Systems*, *4*(3), 1-15. https://doi.org/10.69987/JACS.2024.40301

[21] Zhang, Y., Song, P., & Yang, Q. (2025). A Risk-Aware Dynamic Credit Allocation Mechanism in Green Supply Chains: An Agent-Based Model with ESG Metrics. *Risks*, *13*(12), 236. https://doi.org/10.3390/risks13120236

Author Biography

Shankari Mohanakrishnan is a Senior Data Analyst with over six years of experience in data analytics, data engineering, business intelligence, and machine learning. She is currently pursuing a Doctor of Philosophy (Ph.D.) in Information Technology at the University of the Cumberlands, United States. Her professional experience spans the banking, healthcare, logistics, and technology sectors, where she has worked on large-scale data warehousing, cloud migration, analytics reporting, data engineering, and predictive modeling initiatives. She has expertise in SQL, Python, PySpark, Snowflake, AWS, Tableau, Power BI, and machine learning techniques. Her current research focuses on credit risk classification and fraud detection using machine learning approaches. Her academic interests include predictive analytics, financial risk modeling, explainable artificial intelligence, and data-driven decision-making.