

# Research on Interpretable Loan Approval Identification Using Multi-Dimensional Features

Tiantian Lu

Economic and Social Statistics, Guizhou University of Finance and Economics, Guiyang-550025, China

Received: 09 March 2026/ Revised: 15 March 2026/ Accepted: 24 March 2026/ Published: 31-03-2026

Copyright © 2026 International Journal of Engineering Research and Science

This is an Open-Access article distributed under the terms of the Creative Commons Attribution

Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted

Non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract**— Loan approval is a pivotal component of financial risk control. Current research often relies on black-box prediction models and lacks in-depth exploration of multidimensional features and interpretability, resulting in shortcomings in model interpretability and robustness. To address this, this paper proposes an Interpretable Loan Approval Identification Model based on Multidimensional Features (ILA-MDF). The ILA-MDF model is constructed using the CatBoost algorithm and is compared with benchmark models such as Random Forest in experimental evaluations. The results indicate that the proposed model performs optimally across six metrics—Accuracy, Precision, Recall, F1-score, AUC, and MCC—achieving values of 93.33%, 87.19%, 82.03%, 84.53%, 97.77%, and 80.34%, respectively. Furthermore, the SHAP framework is introduced to analyze the key factors influencing model decisions. Interpretability analysis reveals that the probability of loan approval increases significantly when the loan interest rate exceeds 14% or when the previous loan default indicator is 0. The ILA-MDF model proposed in this paper demonstrates significant advantages in both predictive performance and decision interpretability, providing a reference for formulating loan approval and risk prevention strategies.

**Keywords**— CatBoost; Machine Learning; SHAP Framework; Interpretability.

## I. INTRODUCTION

In recent years, as demand for social financing has continued to grow, loans have become an indispensable financial tool for both individuals and businesses. For financial institutions, the quality of loan approval decisions directly affects asset security and operational stability. Against this backdrop, credit scoring models serve as the foundation of automated approval systems by quantifying borrowers' credit risk and providing an objective basis for approval decisions [1]. A robust and accurate model not only improves approval efficiency and reduces default losses but also plays a key role in advancing the intelligent transformation of financial services.

However, traditional loan approval processes have long relied on manual experience, which comes with limitations such as low efficiency and inconsistent standards. While machine learning techniques have been widely adopted to build automated approval models, existing research still has gaps: on one hand, most models focus narrowly on default prediction and fail to adequately support comprehensive approval decisions involving multi-dimensional rules; on the other, many high-performance models are "black boxes" with opaque decision-making logic [2], making it difficult to meet the strict regulatory requirements for fairness and interpretability, and leaving applicants unable to understand the rationale behind approval outcomes. Therefore, exploring loan approval models that combine high predictive performance with strong interpretability has become an important focus for both academia and industry.

In response to these challenges, extensive research has been conducted in academia on loan approval modeling, forming three main directions.

In terms of constructing ensemble learning and complex models, Kokate and Chetty employed a combination of machine learning methods such as gradient boosting, random forests, and decision trees to build a credit scoring model for automated approval, using feature selection techniques to improve model efficiency. Their model demonstrated superior risk discrimination ability and stability on real-world banking data, validating the effectiveness of combining ensemble learning with feature optimization [3]. Further, Uddin et al. proposed a hybrid approach integrating deep learning with ExtraTrees and

adopted an ensemble voting mechanism to combine the three best-performing base models, achieving an accuracy of 87.26% in bank loan default prediction, highlighting the significant improvement of ensemble strategies on classification performance [4]. In line with this, Perera and Premaratne developed a stacking ensemble model based on a voting mechanism, further confirming the practicality and stability of ensemble learning in credit risk assessment [5]. Lakshmi and Rao constructed a loan default prediction model by integrating algorithms such as naive Bayes, decision trees, and multi-layer perceptrons, training on Kaggle historical data and ultimately achieving 90% accuracy, further demonstrating that ensemble learning methods can effectively support banks in making loan approval decisions [6].

In terms of benchmarking and optimization of traditional machine learning models, a substantial body of research has focused on providing benchmarks for model selection through systematic comparisons. Singh employed logistic regression, random forests, and support vector machines to model loan approval, achieving an accuracy of 78.785% on historical data [7]. Nureni and Adekola systematically compared eight machine learning algorithms and found that logistic regression performed particularly well in terms of accuracy (up to 83.24%) and sensitivity (up to 97.76%) [8]. Tumuluru et al. pointed out that among models for default prediction based on historical customer data, the random forest algorithm achieved the best accuracy [9]. Viswanatha conducted a comprehensive comparison of algorithms including random forests, naive Bayes, decision trees, and KNN, with naive Bayes achieving the highest accuracy at 83.73% [10]. A systematic comparison by Sarkar et al. also indicated that random forest and logistic regression models slightly outperformed other compared algorithms in terms of overall performance. These extensive comparative studies provide empirical support for model selection across different scenarios [11].

In terms of model efficiency, system deployment, and interpretability, relevant research has shown a clear trajectory of development—starting from basic system construction, moving toward feature optimization and model comparison, and ultimately deepening into interpretability and ethical considerations. Gomathy et al. built an automated loan prediction system based on decision trees, aiming to reduce manual review errors and improve overall approval efficiency [12]. Karthikeyan and Ravikumar introduced the Boruta feature selection method combined with random forest models to screen key features, effectively enhancing the model's discriminative power [13]. Bhattad et al. integrated multiple machine learning algorithms into a loan prediction system to improve the responsiveness and accuracy of the approval process [14]. Nalawade et al. further incorporated multi-algorithm comparison in system development, finding that logistic regression achieved the highest accuracy (88.70%), and designed a user interface while also noting the limitations of such models in practice due to their reliance on multi-attribute comprehensive judgment [15]. Subsequently, studies by Sharma et al. and Krishnaraj et al. both emphasized the potential of logistic regression in balancing predictive accuracy with model interpretability [16, 17]. More recent research has placed greater emphasis on the synergy between system performance and ethical standards: Badhan et al. built an approval system with high accuracy and anti-fraud capabilities through algorithm comparison and ensemble techniques [18]; Raheem further pointed out that while improving model performance, attention must also be paid to decision transparency and ethical considerations, thereby offering an important perspective for the sustainable development of loan decision systems [19].

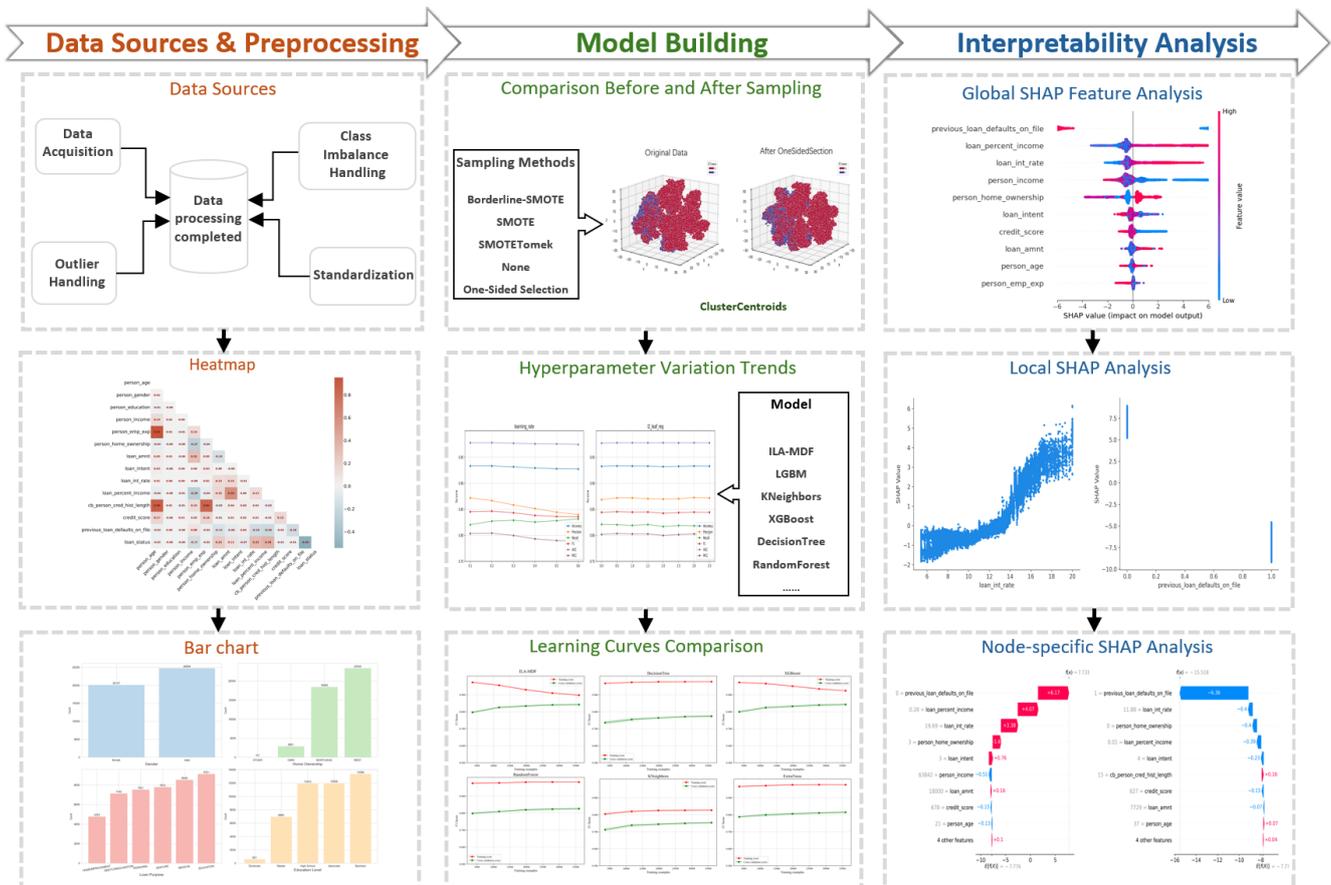
Looking at the body of research discussed above, existing studies have laid a solid foundation for loan approval identification, yet there remain two issues that warrant further exploration. First, most approaches rely on a single type of data and fail to effectively integrate multi-dimensional applicant information, which limits the discriminative power and robustness of the models. Second, current research generally overlooks model interpretability—an aspect that is critically important for regulatory compliance and risk control in the context of loan approval identification.

To address these issues, this paper focuses on the task of loan approval identification and proposes an Interpretable Loan Approval Identification Model Using Multi-Dimensional Features (ILA-MDF). This study differs from existing research in the following two aspects. First, at the identification level, it builds a high-precision model for approval outcome prediction that outperforms existing benchmark models in classification performance while demonstrating strong generalization ability. Second, at the interpretability level, it introduces the SHAP framework to reveal the influence of key features on model decisions, significantly enhancing model interpretability and providing a reference for the formulation of loan approval risk management strategies.

## II. MODEL AND THEORY

### 2.1 ILA-MDF Framework:

This study focuses on the issue of loan approval identification and divides the overall modeling process into three core stages: data source and processing, model building, and interpretability analysis, as illustrated in Figure 1.



**FIGURE 1: ILA-MDF Modeling Flowchart**

Figure 1 shows the three-stage framework of the proposed ILA-MDF model. Stage 1 (Data Source and Processing) involves outlier detection, missing value handling, standardization, and One-Sided Selection undersampling. Stage 2 (Model Building) involves CatBoost model construction, grid search with cross-validation for hyperparameter tuning, and comprehensive performance evaluation. Stage 3 (Interpretability Analysis) involves SHAP-based global feature importance analysis, dependence plots for feature impact analysis, and waterfall plots for individual sample decision path analysis.\*

- 1) **Data Source and Processing:** The data is sourced from the Kaggle platform, consisting of 45,000 records with 14 features. In terms of data preprocessing, outlier and missing value detection are first performed to ensure data integrity, followed by standardization. Subsequently, to address class imbalance, One-Sided Selection is introduced for undersampling to improve the identification performance of minority class instances.
- 2) **Model Building:** This paper constructs the ILA-MDF model based on CatBoost and evaluates its performance comprehensively across six dimensions: accuracy, precision, recall, F1 score, AUC, and Matthews correlation coefficient. On this basis, grid search combined with five-fold cross-validation is further employed to fine-tune the model's hyperparameters, enhancing its generalization ability and robustness.
- 3) **Interpretability Analysis:** This paper introduces the SHAP interpretability framework to analyze the model's decision mechanism from both global and local perspectives. Specifically, a summary plot is used to reveal the importance ranking of features in predicting outcomes; dependence plots focus on representative features, illustrating the relationship between their value variations and SHAP values; furthermore, waterfall plots are employed to analyze the prediction process of individual positive and negative samples, depicting the contribution and direction of influence of features on the final outcome.

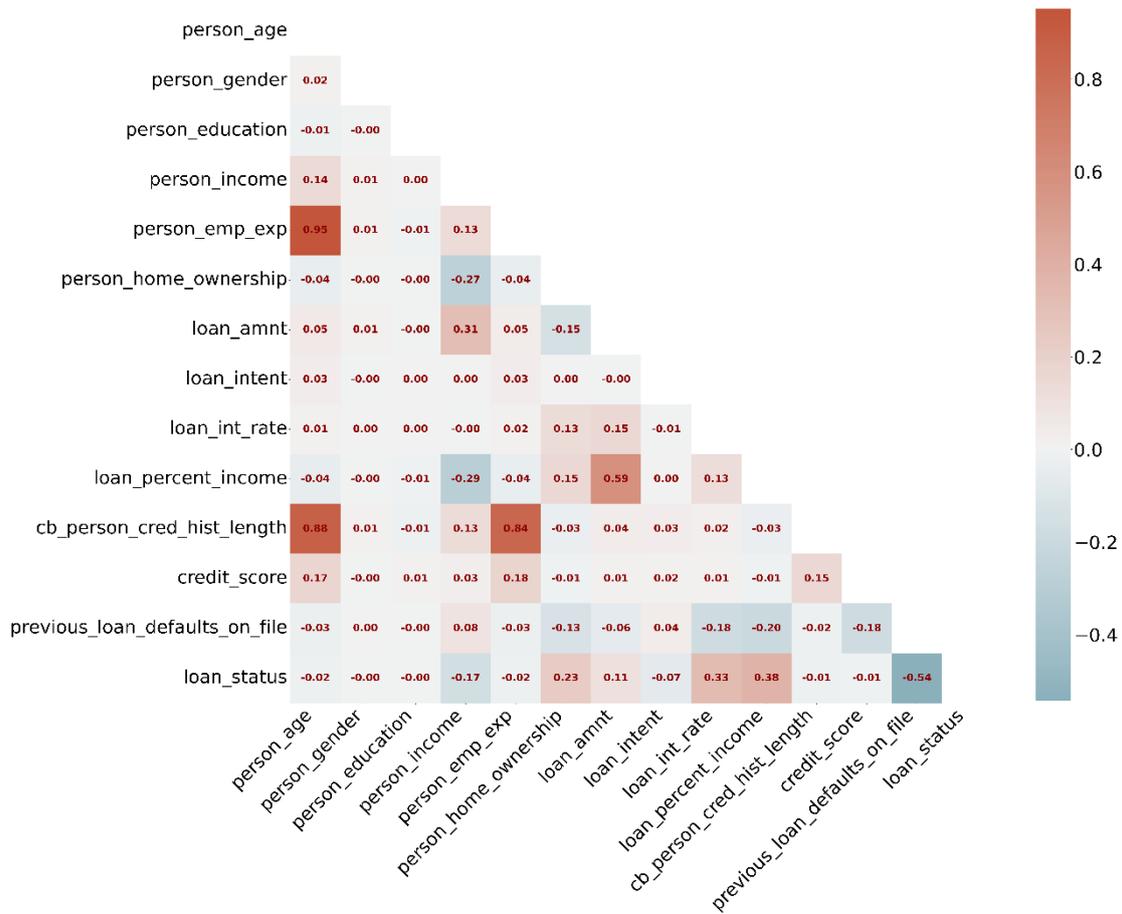


FIGURE 2: Correlation among Features

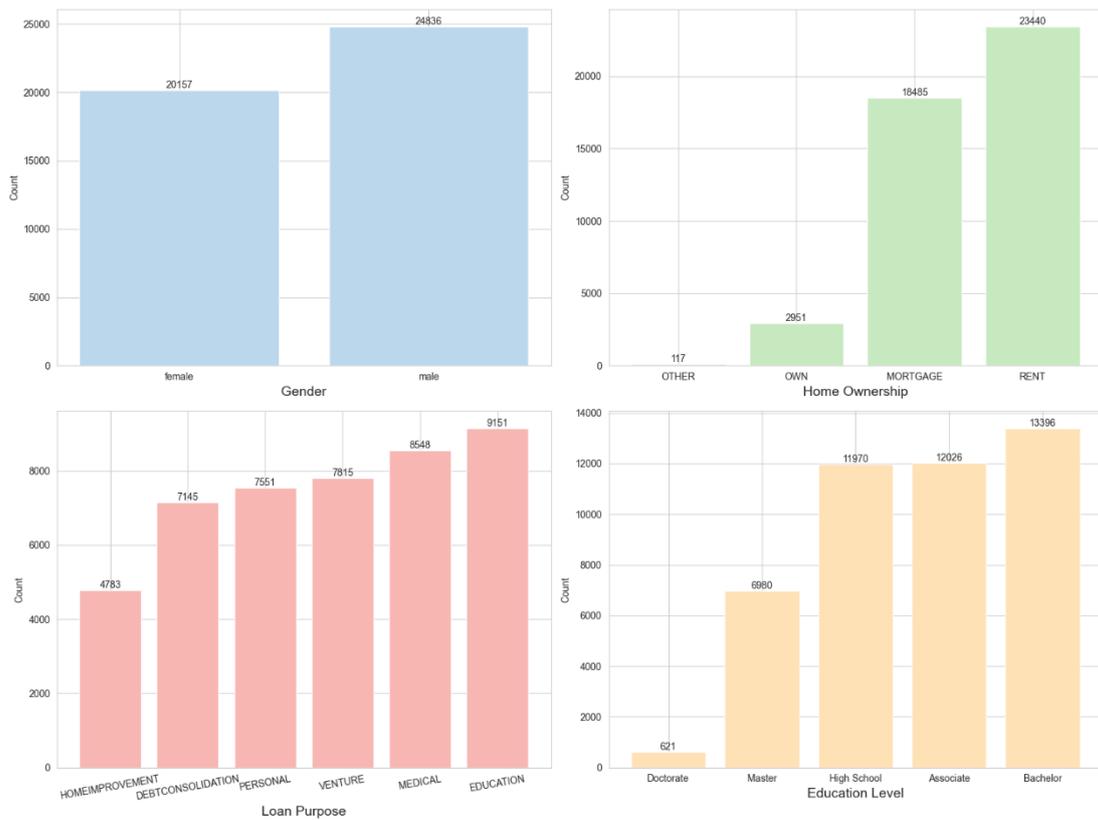


FIGURE 3: Proportion of Categorical Variables

## 2.2 CatBoost Algorithm:

This paper constructs the ILA-MDF model based on CatBoost. CatBoost is an open-source machine learning algorithm based on gradient boosting decision trees, proposed by Prokhorenkova et al. [20] in 2018. Compared with traditional GBDT, CatBoost introduces two key improvements: first, through "ordered target statistics" and "ordered boosting" mechanisms, it effectively alleviates target leakage in categorical feature encoding and reduces gradient estimation bias; second, it uses "symmetric trees" as base learners, which not only improves inference efficiency and generalization performance but also enhances parallelization in storage and computation. Based on these mechanisms, this paper further divides the derivation and optimization of the objective function into the following three steps:

### 2.2.1 Objective Function Definition:

The objective function consists of a loss function and a regularization term, as shown in Equation (1):

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (1)$$

where  $l$  is the loss function,  $\hat{y}_i^{(t-1)}$  is the predicted value from the previous  $t - 1$  iterations,  $f_t(x_i)$  is the tree added at the  $t$ -th iteration, and  $\Omega(f_t)$  is the regularization term. Its specific form is:

$$\Omega(f) = \gamma T + \lambda \sum_{j=1}^T w_j^2 \quad (2)$$

where  $T$  is the number of leaf nodes,  $\gamma$  controls the penalty on the number of leaves,  $\lambda$  is the L2 regularization coefficient, and  $w_j$  is the weight of leaf node  $j$ .

### 2.2.2 Ordered Boosting Strategy:

To avoid target leakage, CatBoost introduces an ordered boosting strategy in gradient computation. Specifically, for a sample  $x_i$ , its gradient depends only on historical information from the first  $i - 1$  samples:

$$g_i = \frac{\partial l(y_i, \hat{y}^{(t-1)})}{\partial \hat{y}^{(t-1)}} \quad (3)$$

$$h_i = \frac{\partial^2 l(y_i, \hat{y}^{(t-1)})}{\partial (\hat{y}^{(t-1)})^2} \quad (4)$$

### 2.2.3 Symmetric Tree Design and Leaf Weight Optimization:

CatBoost adopts a symmetric tree structure, where nodes at the same level share splitting rules, thereby improving computational efficiency while ensuring model stability. For leaf node  $j$ , the optimal leaf weight  $w_j^*$  can be obtained by minimizing the objective function:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (5)$$

where  $I_j$  is the set of samples in leaf node  $j$ . Let  $G_j = \sum_{i \in I_j} g_i$  and  $H_j = \sum_{i \in I_j} h_i$ . Substituting these into the objective function yields:

$$Obj^{(t)} = - \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (6)$$

In summary, CatBoost automatically adjusts category weights to improve the identification of minority class samples in loan approval identification. Its gradient computation mechanism based on historical samples effectively suppresses target leakage and reduces prediction bias. The symmetric tree structure enhances generalization performance and inference efficiency while controlling overfitting.

### 2.3 SHAP Framework:

SHAP (SHapley Additive exPlanations) is a unified framework proposed by Lundberg et al. [21] in 2017 for explaining predictions made by black-box machine learning models. The framework is based on the concept of Shapley values from cooperative game theory, and it converts the output of black-box models into interpretable numerical results by fairly distributing the marginal contribution of each feature to the prediction, thereby meeting interpretability requirements while preserving predictive performance. The formula for SHAP values is as follows:

$$f(x_i) = f_0 + f(x_{i,1}) + f(x_{i,2}) + \dots + f(x_{i,k}) \tag{7}$$

where  $x_{i,k}$  represents the  $k$ -th feature of the  $i$ -th sample;  $f(x_{i,k})$  represents the Shapley value of  $x_{i,k}$ ;  $f_0$  represents the baseline prediction of the overall model; and  $f(x_i)$  represents the predicted value for the  $i$ -th sample. When  $f(x_{i,k}) > 0$ , the feature has a positive contribution to the prediction; when  $f(x_{i,k}) < 0$ , the feature has a negative contribution. Therefore, introducing the SHAP framework in this study not only enables the identification of key features influencing classification outcomes in loan approval based on the magnitude of feature SHAP values, helping to pinpoint potential risk factors, but also provides deeper insight into the model's decision-making mechanisms and working principles, offering a reliable basis for model optimization.

## III. RESULTS AND DISCUSSION

### 3.1 Data Source:

The dataset used in this study is sourced from the Kaggle platform, consisting of 45,000 records with 14 features. The specific details are presented in Table 1. In terms of data preprocessing, after data inspection, outliers were identified and removed, and categorical variables were one-hot encoded, resulting in a complete dataset that laid the foundation for subsequent model training and analysis.

**TABLE 1**  
**DESCRIPTION OF DATASET FEATURES**

| Feature                        | English explanation                     |
|--------------------------------|---|
| person_age                     | Age                                     |
| person_gender                  | Gender                                  |
| person_education               | Highest Education Level                 |
| person_income                  | Annual Income                           |
| person_emp_exp                 | Years of Work Experience                |
| person_home_ownership          | Home Ownership Status                   |
| loan_amnt                      | Loan Amount Applied                     |
| loan_intent                    | Loan Purpose                            |
| loan_int_rate                  | Loan Interest Rate                      |
| loan_percent_income            | Loan Amount to Annual Income Percentage |
| cb_person_cred_hist_length     | Credit History Length                   |
| credit_score                   | Credit Score                            |
| previous_loan_defaults_on_file | Previous Loan Default Indicator         |
| loan_status                    | Loan Approval Status                    |

### 3.2 Feature Correlation Analysis:

To gain deeper insight into the relationship between various features and loan approval status, this study conducted a systematic analysis of linear correlations among variables using a heatmap. The results reveal the differentiated roles that different features play in the approval decision.

Specifically, the ratio of loan amount to annual income (`loan_percent_income`) shows a moderate positive correlation with approval status ( $r = 0.38$ ), indicating that this metric serves as an important predictor of approval outcomes—that is, the higher the debt-to-income ratio, the lower the likelihood of loan approval. A stronger negative correlation was observed for previous loan defaults on file ( $r = -0.54$ ), further confirming the significant impact of this feature on approval decisions: applicants with

a history of default face a considerably higher risk of rejection. In addition, home ownership status (person\_home\_ownership) exhibits a certain positive correlation with approval status ( $r = 0.23$ ), suggesting that applicants who own a home are generally perceived as having a more stable credit background and are therefore more likely to be approved.

It is worth noting that credit score shows no significant linear correlation with approval status ( $r = -0.01$ ), indicating that, within the current modeling framework, this variable offers limited explanatory power for approval outcomes. Its potential influence may need to be further explored through nonlinear modeling approaches.

### 3.3 Analysis of Categorical Variables:

The dataset includes four main categorical variables: gender, home ownership, loan purpose, and default history. Gender is roughly balanced, reducing the risk of bias. Home ownership is concentrated among renters and mortgage holders, with owners and other categories making up only a small share—a pattern that reflects the link between housing status and financial stability. Loan purpose covers a range of categories such as education, personal consumption, and medical expenses, capturing diverse real-world borrowing scenarios. Default history, however, is highly imbalanced: only a small fraction of applicants have prior defaults, which calls for sampling techniques in modeling to improve identification of this minority but critical group.

### 3.4 Experimental Environment and Evaluation Metrics:

The experiments were conducted in the following software and hardware environment: Intel(R) Core(TM) i5-10200H processor (base frequency 2.40 GHz), 16GB RAM, and Windows 10 operating system. The development environment was PyCharm 2023.2 with Python 3.11.

TABLE 2  
 CONFUSION MATRIX

|                    | Actual Positive     | Actual Negative     |
|--------------------|---------------------|---------------------|
| Predicted Positive | TP (True Positive)  | FP (False Positive) |
| Predicted Negative | FN (False Negative) | TN (True Negative)  |

Six metrics were selected to evaluate model performance: accuracy, precision, recall, F1 score, AUC, and Matthews correlation coefficient (MCC). These metrics were calculated based on the confusion matrix by comparing predicted results with actual labels.

Accuracy measures the proportion of samples correctly predicted by the model:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

Precision focuses on how many of the samples predicted as positive are actually positive:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{9}$$

Recall measures how many of the actual positive samples are correctly identified:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{10}$$

The F1 score is the harmonic mean of precision and recall:

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP+FP+FN} \tag{11}$$

AUC evaluates the overall performance of the model by plotting the true positive rate (TPR) against the false positive rate (FPR); the larger the area, the stronger the model's ability to distinguish between positive and negative classes.

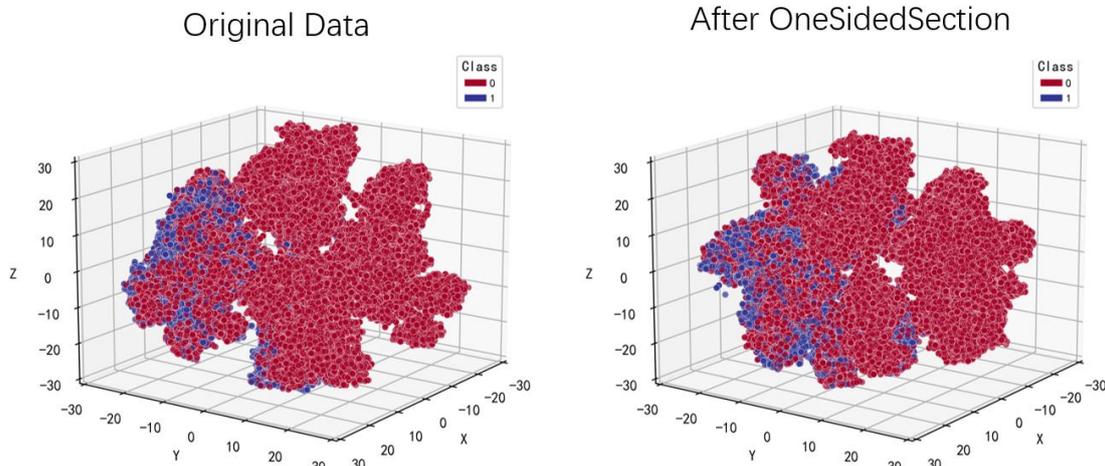
$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{FP+TN} \tag{12}$$

MCC takes into account all elements of the confusion matrix (TP, TN, FP, FN) and provides a comprehensive measure of classification quality.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{13}$$

### 3.5 Imbalance Handling:

The dataset in this study suffers from class imbalance. To mitigate its impact on model training, the One-Sided Selection method was applied for sampling. Figure 4 shows a comparison of the data distribution before and after sampling. The left panel shows a clear class imbalance in the original data, with samples of class 0 significantly outnumbering those of class 1. The right panel shows that after sampling, the two classes become more balanced, which helps improve the model’s ability to identify the minority class.



**FIGURE 4: Comparison Before and After Sampling**

As shown in Figure 4, the One-Sided Selection sampling method effectively alleviates the class imbalance issue. To further validate its impact on model performance, this study compares different sampling methods experimentally.

The results are shown in Table 3. The One-Sided Selection method performs best across three key metrics: recall, F1 score, and MCC. In subsequent experiments, One-Sided Selection is adopted as the default sampling strategy for further modeling and evaluation.

**TABLE 3  
 COMPARISON OF IMBALANCE HANDLING RESULTS**

| Sampling            | Accuracy | Precision | Recall | F1     | AUC    | MCC    |
|---------------------|----------|-----------|--------|--------|--------|--------|
| Borderline-SMOTE    | 93.22%   | 87.46%    | 81.15% | 84.18% | 97.75% | 79.97% |
| SMOTE               | 93.26%   | 87.67%    | 81.10% | 84.25% | 97.77% | 80.08% |
| SMOTETomek          | 93.28%   | 87.82%    | 81.02% | 84.28% | 97.78% | 80.12% |
| None                | 93.36%   | 89.80%    | 79.12% | 84.12% | 97.81% | 80.20% |
| One-Sided Selection | 93.30%   | 87.23%    | 81.84% | 84.45% | 97.77% | 80.25% |

### 3.6 Comparison with Mainstream Machine Learning Models:

To comprehensively evaluate the performance of the ILA-MDF model, this study selects a variety of mainstream machine learning models as baseline methods for comparative analysis. All models are trained and tested under the same feature engineering and data preprocessing pipeline, with evaluation metrics including accuracy, precision, recall, F1 score, AUC, and MCC.

As shown in Table 4, ILA-MDF ranks first in accuracy, precision, F1 score, and MCC, demonstrating strong classification performance and stability. Although XGBoost achieves higher recall and AUC, ILA-MDF maintains a clear advantage in overall performance. With an accuracy of 93.30%, ILA-MDF exhibits excellent classification capability. In contrast, traditional models like KNeighbors, while acceptable in accuracy, often underperform in recall and F1 when handling nonlinear, high-dimensional, and imbalanced data, limiting their applicability in complex real-world scenarios. Overall, ILA-MDF performs well across multiple metrics, maintaining high predictive accuracy and generalization ability, particularly in high-dimensional imbalanced settings.

**TABLE 4**  
**COMPARISON WITH MAINSTREAM MACHINE LEARNING MODELS**

| Model                | Accuracy | Precision | Recall | F1     | AUC    | MCC    |
|----------------------|----------|-----------|--------|--------|--------|--------|
| KNeighbors           | 89.06%   | 75.49%    | 75.18% | 75.33% | 92.73% | 68.31% |
| LogisticRegression   | 88.89%   | 74.22%    | 76.67% | 75.42% | 94.95% | 68.27% |
| DecisionTree         | 89.70%   | 75.28%    | 79.88% | 77.51% | 86.19% | 70.89% |
| GradientBoosting     | 91.99%   | 84.44%    | 78.39% | 81.30% | 97.02% | 76.29% |
| ExtraTrees           | 91.94%   | 83.59%    | 79.33% | 81.40% | 96.92% | 76.31% |
| RandomForest         | 92.58%   | 86.27%    | 79.22% | 82.59% | 97.30% | 78.00% |
| HistGradientBoosting | 93.06%   | 86.61%    | 81.36% | 83.90% | 97.68% | 79.55% |
| LGBM                 | 93.09%   | 86.62%    | 81.49% | 83.97% | 97.70% | 79.64% |
| XGBoost              | 93.21%   | 86.08%    | 82.82% | 84.42% | 97.79% | 80.10% |
| ILA-MDF              | 93.30%   | 87.23%    | 81.84% | 84.45% | 97.77% | 80.25% |

### 3.7 Comparison with Existing Studies:

To highlight the performance advantages of the ILA-MDF model in loan approval classification tasks, this study compares it with representative models from existing literature. The comparison covers traditional machine learning methods, deep learning models, and mainstream ensemble algorithms to ensure the validity of the results.

As shown in Table 5, the ILA-MDF model achieves the best performance across multiple key metrics, including accuracy, precision, F1 score, AUC, and MCC, significantly outperforming all benchmark models. It demonstrates not only stable classification capability but also strong generalization in handling imbalanced data and complex pattern recognition. Therefore, the model not only outperforms existing methods in a statistically significant manner but also shows strong potential and robustness for real-world applications.

**TABLE 5**  
**COMPARISON WITH EXISTING STUDIES**

| Author                         | Model    | Accuracy | Precision | Recall | F1     | AUC    | MCC    |
|--------------------------------|----------|----------|-----------|--------|--------|--------|--------|
| Uddin et al. <sup>[4]</sup>    | Ensemble | 82.76%   | /         | /      | /      | /      | /      |
| Perera et al. <sup>[5]</sup>   | Stacking | 78.00%   | 78.00%    | 92.00% | 84.00% | 75.00% | /      |
| Nureni et al. <sup>[8]</sup>   | LR       | 78.16%   | 79.00%    | 75.96% | 75.96% | /      | /      |
| Natasha et al. <sup>[22]</sup> | DNN      | /        | /         | /      | /      | 63.80% | /      |
| Shinde et al. <sup>[23]</sup>  | LR       | 72.14%   | /         | /      | 82.79% | /      | /      |
| Ndayisenga <sup>[24]</sup>     | GB       | 81.11%   | 82.14%    | 80.98% | 81.47% | /      | /      |
| Yang <sup>[25]</sup>           | LR       | 76.00%   | /         | /      | /      | /      | /      |
| Ours                           | ILA-MDF  | 93.30%   | 87.23%    | 81.84% | 84.45% | 97.77% | 80.25% |

### 3.8 Hyperparameter Perturbation Analysis:

To enhance the robustness and generalization ability of the model, a grid search strategy was employed for hyperparameter optimization. The tuning focused on four key parameters, with the F1 score used as the performance evaluation metric. The search range and final optimal values for each parameter are shown in Table 6.

**TABLE 6**  
**HYPERPARAMETER TUNING RESULTS**

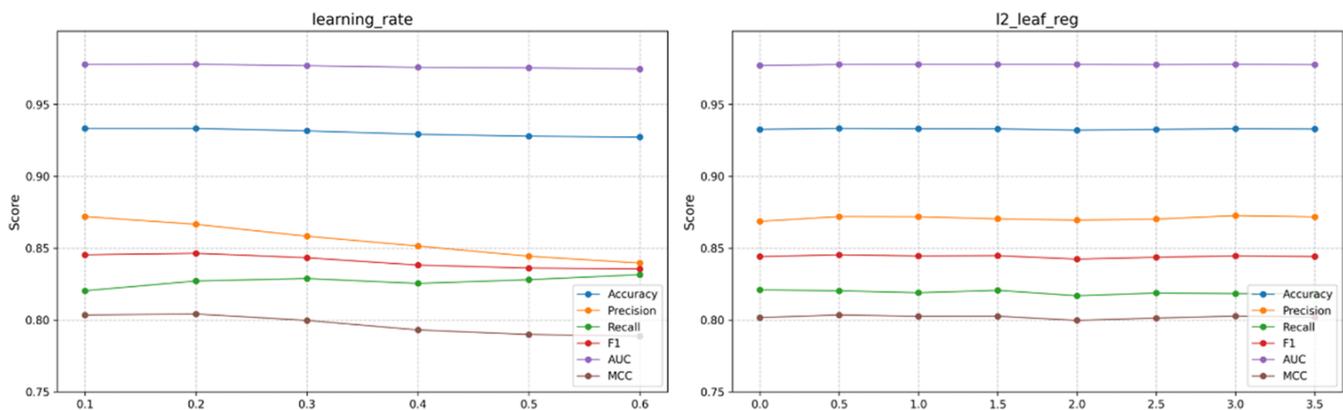
| Parameter     | Search Range      | Optimal Value |
|---------------|-------------------|---------------|
| Iterations    | [300, 400, 500]   | 500           |
| Learning_rate | [0.01, 0.05, 0.1] | 0.1           |
| Depth         | [4, 6, 8]         | 6             |
| L2_leaf_reg   | [0.5, 1, 1.5]     | 0.5           |

After training the model with the optimal hyperparameter combination, the performance on the test set was compared before and after tuning. The results are shown in Table 7. All metrics improved after tuning. Specifically, recall increased by 0.19%, F1 score by 0.08%, and MCC by 0.09%, indicating that hyperparameter optimization effectively enhanced model classification performance and stability.

**TABLE 7**  
**MODEL PERFORMANCE BEFORE AND AFTER TUNING**

|               | Accuracy | Precision | Recall | F1     | AUC    | MCC    |
|---------------|----------|-----------|--------|--------|--------|--------|
| Before Tuning | 93.30%   | 87.23%    | 81.84% | 84.45% | 97.77% | 80.25% |
| After Tuning  | 93.33%   | 87.19%    | 82.03% | 84.53% | 97.77% | 80.34% |

To further analyze the impact of hyperparameters on model performance, this study examined the effect of varying the learning rate and L2 regularization coefficient on key metrics (see Figure 5). The left panel shows that when the learning rate is set to 0.1, all metrics reach their highest values, and as the learning rate continues to increase, performance declines across the board, indicating that an excessively high learning rate suppresses the model's learning ability and leads to underfitting. Therefore, the learning rate was ultimately set to 0.1.



**FIGURE 5: Hyperparameter Variation Trend Chart**

The right panel illustrates the effect of the L2 regularization coefficient on model performance. The results show that as the coefficient increases, all metrics except recall show an upward trend at 0.5. However, beyond 0.5, precision, recall, F1, and MCC exhibit a fluctuating pattern of initial decline followed by a subsequent rise. Considering overall performance, the optimal effect is achieved at a coefficient of 0.5, which was ultimately chosen.

### 3.9 SHAP Interpretability Analysis:

To enhance the interpretability of the model, this study introduces the SHAP framework, employing summary plots, dependence plots, and waterfall plots to visually analyze the model's decision-making process. The analysis explores the key

factors influencing loan approval identification from three perspectives: global feature contribution, feature impact mechanisms, and individual sample decision paths.

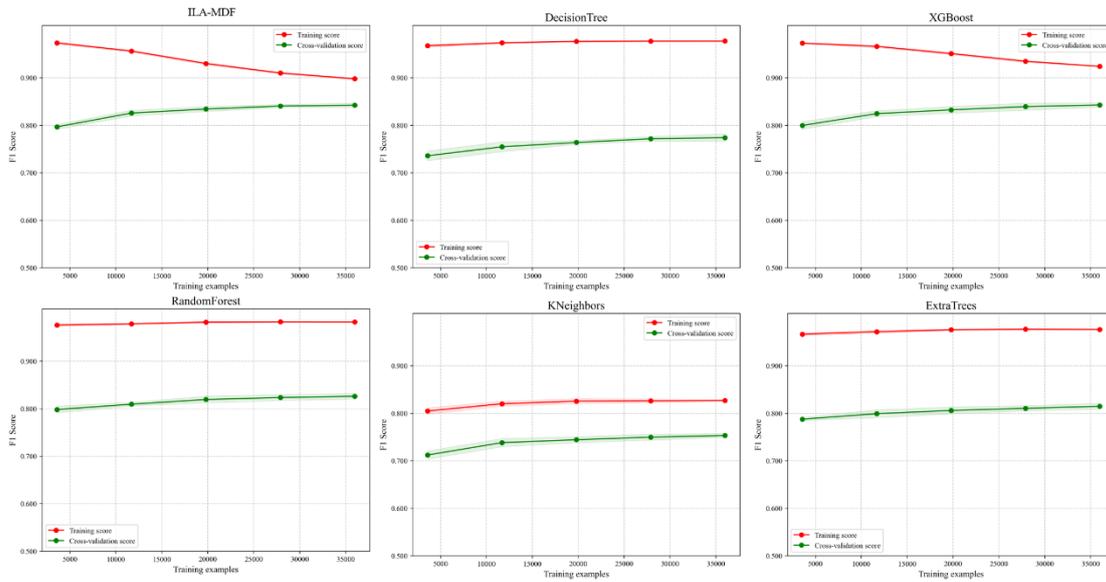


FIGURE 6: Comparison of Model Learning Curves

3.9.1 Global Feature Importance Analysis:

To examine the importance and direction of influence of each feature in the model’s decisions, a global interpretability analysis was conducted using the SHAP method. Based on the SHAP summary plot shown in Figure 7, the following conclusions can be drawn:

The indicator of previous loan defaults is the most influential feature on model predictions. When its value is high, it corresponds to a consistently negative SHAP value, indicating that a borrower’s history of default systematically reduces the likelihood of loan approval. This underscores the strong inhibitory effect of this feature on approval decisions.

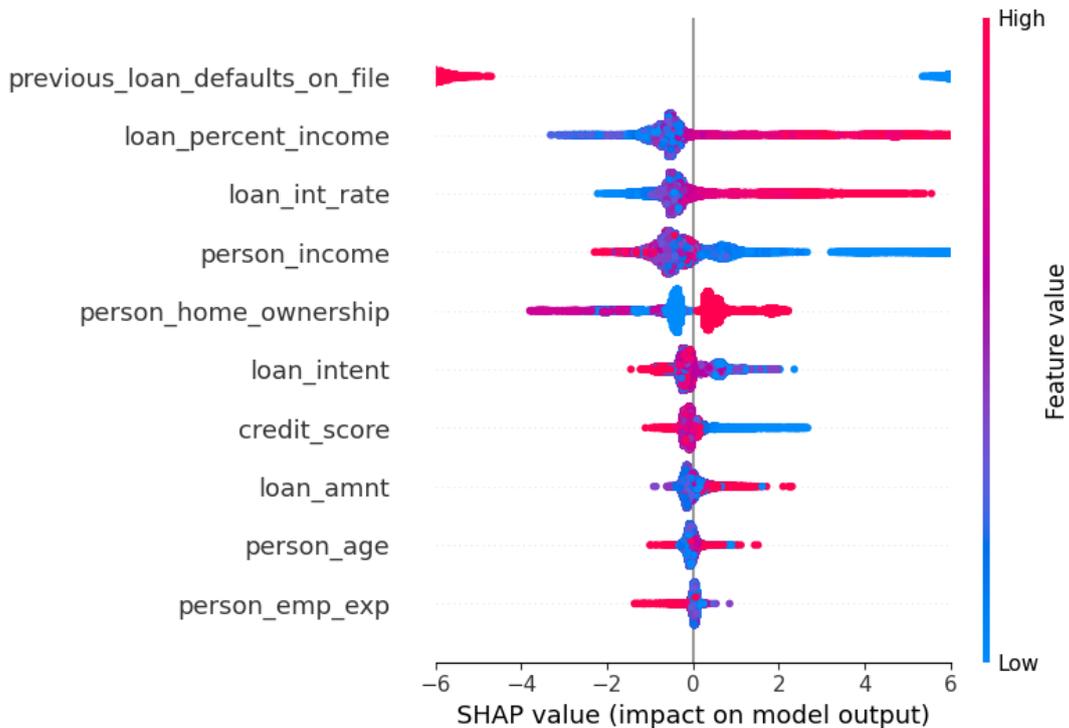


FIGURE 7: SHAP Summary Plot

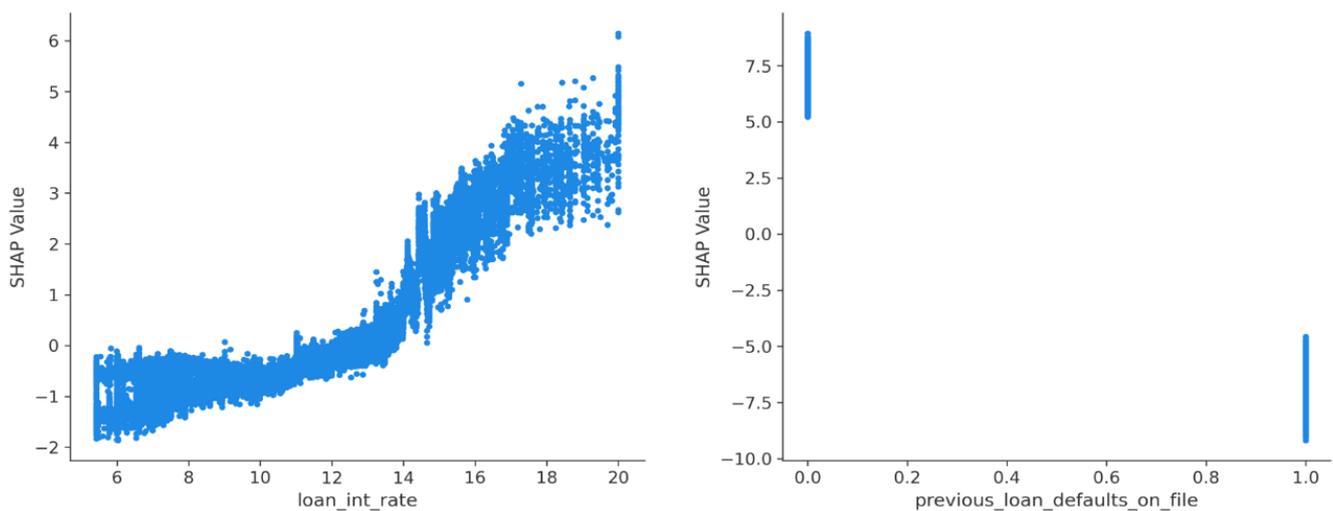
When the ratio of loan amount to annual income is low, the SHAP value is clearly negative. This suggests that when the loan amount is relatively small compared to the borrower’s annual income, the repayment burden is lighter, and the model tends to classify the applicant as “low risk,” thereby significantly increasing the probability of approval.

Similarly, when the interest rate is low, the SHAP value is also noticeably negative. Since lower interest rates are typically offered to high-quality borrowers, a low rate itself serves as a signal of lower risk, positively influencing the final approval outcome.

In contrast, annual income generates a positive SHAP value when its value is low. This means that when a borrower’s income level is low, this feature clearly pushes the model’s prediction toward the “high-risk” direction, significantly increasing the likelihood of rejection. This reflects the model’s sensitivity to insufficient repayment capacity.

### 3.9.2 Analysis of Feature Impact on Prediction Results:

To gain deeper insight into how key features influence model predictions, this study uses SHAP values to generate dependence plots (shown in Figure 8), revealing the relationship between feature values and their corresponding contributions.



**FIGURE 8: SHAP Dependence Plot**

The results show that the loan interest rate (`loan_int_rate`) exhibits a significant positive correlation with model risk prediction. When the interest rate exceeds 14%, the SHAP value begins to rise rapidly, forming a stable positive driving effect in the range of 17% to 20%. This indicates that a higher interest rate significantly increases the probability of the model classifying an applicant as high-risk. This aligns with real-world practice: the higher the loan rate, the more inclined financial institutions are to approve the loan application.

For the indicator of previous loan defaults (`previous_loan_defaults_on_file`), when the value is 0—indicating no history of default—the corresponding SHAP value is positive, reflecting that having no default record has a positive effect on loan approval. Conversely, when a default exists, the SHAP value is negative, indicating that a record of default negatively influences the likelihood of loan approval.

### 3.9.3 Analysis of Sample Decision Path:

To better understand the model’s decision-making mechanism at the sample level, this study further examines how individual features collectively contribute to the final prediction from a single-sample perspective. The SHAP waterfall plot shown in Figure 9 visually illustrates the cumulative contribution of each feature to the prediction for a given sample.

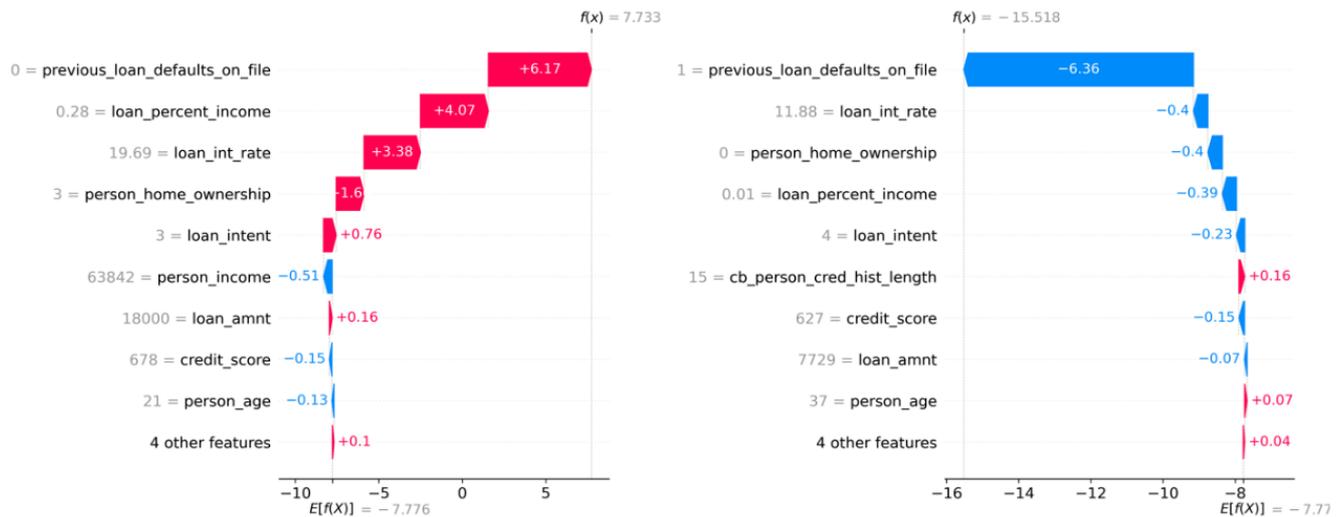


FIGURE 9: SHAP Waterfall Plot

In the approved sample (left panel), the indicator of previous loan defaults is the most significant positive contributor, indicating that the applicant's clean credit history plays a key role in the approval decision. The loan-to-income ratio and loan interest rate also show negative contributions, suggesting that a lower debt burden and a more favorable interest rate together point to the borrower's strong repayment capacity and low credit risk, thereby significantly increasing the likelihood of approval.

In contrast, for the rejected sample (right panel), the loan-to-income ratio shows a strong negative contribution, indicating that a high debt burden is one of the main reasons the model denies the application. Meanwhile, annual income also exhibits a clear negative influence, further showing that low income combined with high debt pressure reinforces the applicant's credit risk, ultimately leading the model to make a rejection decision.

#### IV. CONCLUSION

This paper focuses on the problem of loan approval identification and proposes a model named ILA-MDF (Interpretable Loan Approval Identification Model based on Multidimensional Features).

First, outlier detection is performed on the acquired dataset, followed by standardization and One-Sided Selection undersampling, effectively improving data quality and class balance. One-Sided Selection was chosen for its ability to remove noisy majority class samples while preserving informative decision boundary samples.

Second, through a comprehensive comparison of multiple mainstream machine learning algorithms, CatBoost is selected as the base architecture, and grid search combined with five-fold cross-validation is used for hyperparameter tuning, significantly enhancing the model's generalization ability and robustness.

Third, multi-level analysis based on SHAP shows that features such as previous loan defaults, the ratio of loan amount to annual income, and loan interest rate have a significant impact on identification results. In particular, when the loan interest rate exceeds 14% or the indicator for previous loan defaults is 0, the probability of loan approval increases significantly. These findings provide a theoretical basis for risk control strategies in loan approval.

#### LIMITATIONS AND FUTURE WORK

This study has several limitations that should be acknowledged. First, the dataset used is cross-sectional, meaning the model does not account for temporal dynamics in borrower behavior or changing economic conditions. Second, while the model demonstrates strong predictive performance, potential fairness concerns—such as disparate impact across demographic groups—were not evaluated. Third, the computational cost and inference time of the CatBoost model compared to simpler baselines were not analyzed, which may be relevant for real-time deployment scenarios.

Future work will focus on addressing these limitations. Specifically, we aim to incorporate users' historical credit behavior data to construct time-aware models, enabling loan approval decisions to evolve from static assessments toward continuous monitoring and early warning based on dynamic risk profiling. Additionally, we plan to integrate fairness-aware machine

learning techniques to ensure equitable outcomes across diverse applicant groups and conduct computational efficiency analyses to support real-world deployment.

#### ACKNOWLEDGEMENTS

The author has no acknowledgements to declare.

#### CONFLICT OF INTEREST

The author declares no conflict of interest

#### CONFLICT OF INTEREST

The authors have no acknowledgements to declare.

#### REFERENCES

- [1] Sarathamani, S., & Kumar, S. (2024). Application of machine learning in credit scoring for loan approval systems. In *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* (pp. 1–7). IEEE.
- [2] Zou, Y., Xia, M., & Lan, X. (2025). Interpretable credit scoring based on an additive extreme gradient boosting. *Chaos, Solitons & Fractals*, *194*, Article 116216.
- [3] Kokate, S., & Chetty, M. S. R. (2021). Credit risk assessment of loan defaulters in commercial banks using voting classifier ensemble learner machine learning model. *International Journal of Safety and Security Engineering*, *11*(5), 565–572.
- [4] Uddin, N., Ahamed, M. K. U., Uddin, M. A., Memon, M. H., & Uddin, M. A. (2023). An ensemble machine learning based bank loan approval predictions system with a smart application. *International Journal of Cognitive Computing in Engineering*, *4*, 327–339.
- [5] Perera, C. L., & Premaratne, S. C. (2024). An ensemble machine learning approach for forecasting credit risk of loan applications. *WSEAS Transactions on Systems*, *23*, 31–46.
- [6] Lakshmi, M. V. N., & Rao, P. S. (2024). A prediction of modernized loan approval system based on machine learning approach. *International Journal for Modern Trends in Science and Technology*, *10*(6), 17–21.
- [7] Singh, V. (2021). Prediction of modernized loan approval system based on machine learning approach. In *2021 International Conference on Intelligent Technologies (CONIT)* (pp. 1–6). IEEE.
- [8] Nureni, A. A., & Adekola, O. E. (2022). Loan approval prediction based on machine learning approach. *FUDMA Journal of Sciences*, *6*(3), 41–50.
- [9] Tumuluru, P., Burra, L. R., Loukya, M., Sai, N. S., & Bhavani, K. (2022). Comparative analysis of customer loan approval prediction using machine learning algorithms. In *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 349–353). IEEE.
- [10] Viswanatha, V., Ramachandra, A. C., Vishwas, K. N., Kavyashree, B. S., & Nayana, G. (2023). Prediction of loan approval in banks using machine learning approach. *International Journal of Engineering and Management Research*, *13*(4), 7–19.
- [11] Sarkar, T., Rakhra, M., Sharma, V., Singh, S., & Sharma, P. (2024). An empirical comparison of machine learning techniques for bank loan approval prediction. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 137–143). IEEE.
- [12] Gomathy, C. K., Charulatha, M., Aakash, M., & Mani, P. (2021). The loan prediction using machine learning. *International Research Journal of Engineering and Technology*, *8*(10), 2395–0056.
- [13] Karthikeyan, S. M., & Ravikumar, P. (2021). A comparative analysis of feature selection for loan prediction model. *International Journal of Computer Applications*, *975*, 8887.
- [14] Bhattad, S., Bawane, S., Agrawal, S., & Tiwari, A. (2021). Loan prediction using machine learning algorithms. *International Journal of Computer Science Trends and Technology*, *9*(3), 143–146.
- [15] Nalawade, S., Andhe, S., Parab, S., & Gurav, S. (2022). Loan approval prediction. *International Research Journal of Engineering and Technology*, *9*(4), 669–673.
- [16] Sharma, H., Tyagi, I., Agarwal, G., & Sharma, A. (2023). An exhaustive investigation on loan prediction in banks using LRD. [*Journal name missing*].
- [17] Krishnaraj, P., Rita, S., & Jaiswal, J. (2023). Comparing machine learning techniques for loan approval prediction. In *Proceedings of the 1st International Conference on Artificial Intelligence, Communication, IoT, Data Engineering and Security (IACIDS)* (pp. 23–25).
- [18] Badhan, A., Rana, A., Malhi, S. S., & Singh, J. (2024). A comparative analysis for loan approval prediction using machine learning. In *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)* (Vol. 1, pp. 1–5). IEEE.
- [19] Raheem, M. (2024). Loan default prediction using machine learning: A review on the techniques. *Journal of Applied Technology and Innovation*, *8*(2), 1.
- [20] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *31*, 6638–6648.

- [21] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- [22] Natasha, A., Prastyo, D. D., & Suhartono. (2019). Credit scoring to classify consumer loan using machine learning. In *AIP Conference Proceedings* (Vol. 2194, No. 1, Article 020070). AIP Publishing.
- [23] Shinde, A., Patil, Y., Kotian, I., & Kharat, S. (2022). Loan prediction system using machine learning. In *ITM Web of Conferences* (Vol. 44, Article 03019). EDP Sciences.
- [24] Ndayisenga, T. (2021). *Bank loan approval prediction using machine learning techniques* [Master's thesis, [University name]].
- [25] Yang, C. (2024). Research on loan approval and credit risk based on the comparison of machine learning models. In *SHS Web of Conferences* (Vol. 181, Article 02003). EDP Sciences.