ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

Explainable Artificial Intelligence for Credit Risk Assessment in Microfinance: An XGBoost-SHAP Framework for Transparent and Accountable Lending Decisions

[1] Dr.M. Suresh Kumar, [2] Bhuvaneshwaran A, [3] Brinesh Varshan M, [4] Idhika P

[1] [2] [3] [4] Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Kanchipuram, Sriperumbudur, Tamil Nadu, India

Corresponding Author Email: [1] babumrm@gmail.com, [2] bhuvanannamalai73@gmail.com, [3] brineshvarshan28@gmail.com, [4] idhikaprabakaran@gmail.com

***Abstract*— Machine learning tools are becoming more common in microfinance institutions for automatically assessing credit risk. These models often perform better than traditional statistical methods in predicting creditworthiness. However, their black-box nature creates important issues around transparency, fairness, regulation, and the rights of borrowers. In countries like India, where more than 190 million people still do not have access to financial services, unclear credit rejection decisions can worsen existing inequalities and reduce trust in digital lending systems. To address this, this paper introduces a complete Explainable AI framework that combines XGBoost with SHAP. The framework is tested using the "Give Me Some Credit" dataset (150,000 real-world loan applications). The system achieves a classification accuracy of 91.2% and an AUC-ROC of 0.943, outperforming Logistic Regression, Decision Tree, Random Forest, and LightGBM. SHAP analysis shows that credit card usage, frequency of late payments, and applicant age are the most important factors. The system provides detailed, easy-to-understand reports for each applicant, helping loan officers explain their decisions and allowing borrowers to improve their credit standing.**

Index Terms*— Explainable Artificial Intelligence (XAI), Credit Risk Assessment, XGBoost, SHAP, Microfinance, Machine Learning, Responsible AI, Financial Inclusion, Feature Importance, Borrower Empowerment.

I. INTRODUCTION

The fast growth of digital technology in financial services has changed how credit risk is assessed. Old ways, like using rule-based scorecards and simple models like linear regression, are being replaced quickly by more advanced machine learning algorithms. These new methods can handle many variables and find complex risk patterns with greater accuracy [1]. This change is especially important in the microfinance industry, where people with little or no credit history need help getting loans, requiring the use of different types of data and more powerful analytical tools.

But there is a big downside to these machine learning models: they are hard to understand. Techniques like Random Forests, Gradient Boosting, and deep neural networks work like black boxes. They take in data and give predictions but do not explain why they made those predictions. This lack of transparency is not just a theoretical problem. When someone's loan application is denied by an automated system, not being able to understand the reason has serious real-life effects. The applicant cannot show they are following the rules, and auditors cannot check if the model is treating people unfairly based on protected traits.

This problem is especially significant in India's microfinance environment. The Reserve Bank of India (RBI) reports that over 190 million people in India do not have access to formal banking services [1]. Many of these people are farmers, small business owners, or daily wage workers applying for small loans from Rs. 5,000 to Rs. 200,000. When their loan applications are rejected by automated systems without any explanation, it is not just a technical issue--it is a fairness issue. Borrowers have no way to improve their chances.

Regulators are also moving toward requiring more transparency. In the European Union, the General Data Protection Regulation (GDPR) gives people the right to know why automated decisions were made about them [2]. The upcoming EU AI Act labels credit scoring as a high-risk use of AI and requires strict transparency and accountability [3]. In India, the RBI has started guiding banks to use AI responsibly, with a focus on fairness, auditability, and decision explanation [3].

To tackle this issue, Explainable Artificial Intelligence (XAI) has become a primary solution. XAI aims to make machine learning predictions clear for people without losing their predictive ability. Among the many XAI methods, SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee [4], is popular because of its strong theoretical base, performance with tree-based models, and its ability to provide both global and individual explanations.

This paper makes the following original contributions to the XAI and credit risk literature:

- A complete end-to-end XAI pipeline for credit risk prediction, integrating data preprocessing, XGBoost classification, SHAP-based explanation, and borrower-facing report generation;
- A rigorous empirical evaluation against five baseline classifiers on a large-scale real-world financial dataset,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

demonstrating superior AUC-ROC and F1-Score for XGBoost;

- A SHAP-based global feature importance analysis identifying the primary drivers of default risk in consumer lending data;
- A structured per-applicant explanation report framework that translates SHAP values into actionable, plain-language credit improvement recommendations.

The remainder of this paper is organized as follows. Section II reviews relevant prior work. Section III describes the dataset and preprocessing pipeline. Section IV presents the proposed methodology. Section V reports experimental results. Section VI provides discussion of findings, limitations, and implications. Section VII concludes the paper and outlines future work.

II. RELATED WORK

Machine learning has been widely studied for credit scoring over the last twenty years. Baesens et al. [5] conducted a major study comparing twenty-seven different classification methods using eight actual credit datasets. They found that ensemble methods, especially bagging and boosting techniques, perform better than traditional logistic regression models. XGBoost, developed by Chen and Guestrin [12], is currently the leading model for gradient boosting on tabular data. It uses a regularized objective function, second-order gradient calculations, and handles missing values automatically, making it ideal for financial data with skewed distributions, missing income information, and imbalanced classes [16].

A. Explainability in Machine Learning

Ribeiro et al. [6] created LIME, which uses a simple linear model to explain predictions by examining the model's behaviour near a specific prediction. Although LIME works with many types of models, it can be inconsistent--similar inputs might get different explanations due to random sampling. Lundberg and Lee [4] improved on this by introducing SHAP, based on game-theoretic Shapley values, which ensures more reliable explanations by meeting fairness and consistency standards. For tree-based models, Lundberg et al. [13] developed the TreeSHAP algorithm, which calculates SHAP values in O(TLD^2) time, enabling efficient deployment at scale.

B. XAI in Financial Services

Bhatt et al. [7] showed that models with explainable decisions are more likely to follow regulations during audits. Ariza-Garzon et al. [8] found that using XGBoost with SHAP explanations in peer-to-peer lending improved AUROC and increased borrower trust compared to logistic regression. However, a major gap exists: most studies use data from Western countries, which differ significantly from the Indian microfinance context in terms of income patterns, loan-taking behaviour, and credit record completeness. Most existing research also focuses on explainability from the lender's side rather than providing borrower-facing feedback. This paper addresses all three gaps.

III. DATASET AND PREPROCESSING

This study uses the "Give Me Some Credit" dataset from Kaggle [9], derived from historical loan records of a major U.S. financial institution. It includes missing data, imbalanced classes, and non-normal feature distributions. The dataset has 150,000 entries, each with 11 input features and a binary outcome variable indicating whether the borrower had 90 or more days of serious late payments within a two-year period. The class distribution is approximately 6.68% positive (high-risk) and 93.32% negative (low-risk).

A. Data Preprocessing Pipeline

A multi-stage preprocessing pipeline was created. First, missing values were handled: MonthlyIncome (18.84% missing) was filled using K-Nearest Neighbour (KNN) imputation with k=5 [10], and NumberOfDependents (2.62% missing) used median imputation. Second, extreme values in DebtRatio and RevolvingUtilizationOfUnsecuredLines were capped at the 99th percentile (winsorization) [5]. Third, class imbalance was addressed using SMOTE [11], applied only to the training set to prevent data leakage, producing a 1:3 ratio of positive to negative classes. StandardScaler normalization was applied for baseline models; XGBoost does not require feature scaling.

IV. PROPOSED METHODOLOGY

XGBoost [12] is a gradient boosting framework that builds an ensemble of decision trees sequentially, where each tree corrects the mistakes made by previous ones. The model optimizes a regularized objective function:

L(phi) = SUM l(y_hat_i, y_i) + SUM Omega(f_k)

where l is a differentiable loss function, and Omega(f_k) = gamma*T + (1/2)*lambda*||w||^2 is a regularization term penalizing tree complexity through the number of leaves T and the L2 norm of leaf weights w. This regularization is important for XGBoost's ability to handle noisy financial data without overfitting.

Hyperparameter optimization was performed using 5-fold stratified cross-validation with grid search. The optimal configuration (Table II) was selected based on AUC-ROC on the validation fold, the appropriate metric for imbalanced classification tasks.

Table I: Xgboost Hyperparameter Configuration

Parameter	Value	Description
n_estimators	200	Number of boosting rounds
max_depth	6	Maximum depth of each tree
Learning rate	0.05	Step size shrinkage (eta)
subsample	0.80	Row sampling ratio per tree
colsample_bytree	0.80	Feature sampling ratio per tree

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

Parameter	Value	Description
scale_pos_weight	7.0	Class imbalance compensation
reg_alpha	0.01	L1 regularization term
reg_lambda	1.0	L2 regularization term

A. SHAP Explainability Framework

SHAP values are calculated as the weighted average contribution of each feature across all possible feature coalitions, drawing on Shapley values from cooperative game theory. For feature i, the SHAP value is:

*phi_i = SUM [|S|!(n-|S|-1)!/n!] * [f(S+{i}) - f(S)]*

This method adheres to three key principles: (1) Efficiency--the sum of all SHAP values equals the difference between the model prediction and its average output; (2) Consistency--if a feature's contribution increases in a model, its SHAP value should not decrease; and (3) Missingness--absent features receive a SHAP value of zero [4]. The TreeSHAP algorithm [13] calculates exact SHAP values efficiently, making it practical for real-time credit decisions. The framework provides global explanations (SHAP summary plots aggregated across the test set) and local explanations (SHAP force plots per applicant).

B. System Architecture

The proposed architecture has five layers, as shown in Fig. 1. The Data Ingestion Layer accepts raw applicant data (CSV or database). The Preprocessing Layer applies the pipeline from Section III. The Model Training Layer trains an XGBoost classifier and saves it using joblib. The XAI Explainability Layer creates a SHAP TreeExplainer, calculates SHAP values for all test cases, and uses LIME for cross-validation. Finally, the Output Generation Layer combines the binary prediction, a continuous risk score, ranked SHAP feature importance, and a rule-based template to generate a structured JSON explanation and a plain-language report for each applicant.

LAYER 1 -- DATA INGESTION

Raw Applicant Records | Kaggle Dataset (150,000 rows) | 11 Financial Features | Binary Default Label

v

LAYER 2 -- PREPROCESSING

KNN/Median Imputation | SMOTE Oversampling | Outlier Capping (99th pct) | StandardScaler Normalization

v

LAYER 3 -- MODEL TRAINING

XGBoost Classifier | 5-Fold Stratified Cross-Validation | Grid Search Tuning | Model Persistence (joblib)

v

LAYER 4 -- XAI EXPLAINABILITY

TreeSHAP Global Analysis | SHAP Force Plot (Per Applicant) | LIME Validation | Fairness Audit

v

LAYER 5 -- OUTPUT GENERATION

Credit Decision (Approved/Rejected) | Risk Score (0.00-1.00) | SHAP Report (Top 5 Factors) | Recommendations

Figure 1. Proposed Five-Layer XGBoost-SHAP Architecture

V. EXPERIMENTAL RESULTS AND ANALYSIS

All experiments were conducted using Python with scikit-learn [5], XGBoost, SHAP, and imbalanced-learn. The dataset was divided 80/20 (120,000 training, 30,000 test records) by stratified random sampling. SMOTE was applied only within each cross-validation fold to prevent data leakage. Experiments ran on an Intel Core i7 11th Gen, 16GB RAM, no GPU. Results are the average from five cross-validation folds.

A. Classification Performance

Table I shows how well the XGBoost-SHAP framework performs compared to four other classifiers. The proposed model achieves an accuracy of 91.2% and an AUC-ROC of 0.943, improvements of 11.9 and 13.1 percentage points over Logistic Regression respectively. The model also outperforms LightGBM by 3.5 percentage points in AUC-ROC. The F1-Score of 0.889 demonstrates effective identification of the minority (high-risk) class, confirming that SMOTE combined with the scale_pos_weight parameter successfully addressed class imbalance.

Table II: Classification Performance Comparison Across Models

Model	Accuracy	AUC-ROC	Precision	F1-Score
Logistic Regression	79.3%	0.812	0.761	0.752
Decision Tree	81.1%	0.834	0.779	0.771
Random Forest	84.7%	0.871	0.823	0.817
LightGBM	88.4%	0.908	0.861	0.856
XGBoost+SHAP (Proposed)	**91.2%**	**0.943**	**0.897**	**0.889**

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

B. SHAP Feature Importance Analysis

Table III shows global SHAP feature importance rankings. RevolvingUtilizationOfUnsecuredLines is the most important predictor (mean SHAP = 0.312), meaning borrowers who are nearly or fully using their credit limits are most likely to miss payments. NumberOfTimes90DaysLate ranks second (mean SHAP = 0.287), while Age (mean SHAP = 0.201) has a negative impact--older borrowers are viewed as less risky, consistent with Baesens et al. [5].

Table III: Global Shap Feature Importance Rankings

Rank	Feature Name	Mean |SHAP|	Direction	Impact
1	RevolvingUtilizationOfUnsecuredLines	0.312	Positive	High
2	NumberOfTimes90DaysLate	0.287	Positive	High
3	Age	0.201	Negative	Medium
4	DebtRatio	0.178	Positive	Medium
5	MonthlyIncome	0.143	Negative	Medium
6	NumberOfOpenCreditLinesAndLoans	0.098	Positive	Low
7	NumberOfDependents	0.074	Positive	Low

C. Individual-Level Explanation Report

To demonstrate local explanations, consider a rejected applicant: 34 years old, revolving credit utilization of 91%, twice 90 days late, debt ratio of 62%, monthly income of Rs. 28,000, three dependents. The model assigns a risk score of 0.79 (cutoff = 0.50). SHAP breakdown: revolving credit utilization +0.43 (strong risk increase); 90-days-late count +0.35; debt ratio +0.19; age -0.18 (risk reduction); monthly income -0.09 (risk reduction). System recommendation: "Your revolving credit at 91% greatly raises your default risk. Reducing this below 30% could decrease your risk score by approximately 18-22%. Your history of two late payments is also a major negative factor. Maintaining payments for 12 months without delay before reapplying is strongly recommended."

D. SHAP vs. LIME Cross-Validation

To verify the reliability of SHAP explanations, LIME was applied to 500 test cases. For the top three most important features, SHAP and LIME agreed 84.2% of the time; for the top five, they matched 76.8%. Differences mainly occurred when features were closely correlated, such as DebtRatio and MonthlyIncome (Pearson correlation = 0.67). In such cases, SHAP's consistency gives more stable, well-supported explanations compared to LIME's perturbation-based method.

VI. DISCUSSION

A common misunderstanding is the idea that there is a natural trade-off between accuracy and interpretability--that more explainable models are always less accurate [14]. The findings here challenge that belief in credit risk. The XGBoost-SHAP framework outperforms all other models tested and also provides detailed explanations. The SHAP method is applied post-hoc and does not alter predictions; it helps people understand the model's decision-making. This separation between prediction and explanation is a major benefit of the TreeSHAP approach.

A. Fairness and Bias Considerations

The proposed framework increases transparency but does not automatically guarantee fairness. Age is the third most important factor with a negative impact--older applicants are seen as less risky. While actuarially reasonable, this could raise age discrimination concerns for younger borrowers. In the Indian context, MonthlyIncome might also indirectly reflect geographic location or social background. Future versions should include fairness checks (demographic parity, equalized odds, counterfactual fairness) and test adversarial debiasing or prejudice remover regularization.

B. Computational Feasibility

TreeSHAP can generate explanations for 1,000 applicants in approximately 0.31 seconds using a regular 8-core CPU (about 3,226 explanations per second). This is sufficient for both overnight batch processing and near-real-time digital lending decisions. The model and SHAP explainer together require approximately 18MB of memory, enabling deployment even on resource-constrained systems.

C. Applicability to Indian Microfinance

While this study uses a U.S. benchmark dataset, the approach can be readily applied to Indian microfinance data. Key adaptations would include incorporating mobile payment records, utility bill payments, and agricultural income patterns. The preprocessing pipeline should handle incomplete income records common in informal-sector workers. The output generation should also support local languages. The SHAP framework works with any dataset and requires no structural changes.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

VII. CONCLUSION AND FUTURE WORK

This paper introduces a thorough Explainable AI framework for microfinance credit risk assessment, combining XGBoost with SHAP TreeSHAP. The system achieves an AUC-ROC of 0.943, accuracy of 91.2%, and F1-score of 0.889 on the Give Me Some Credit dataset. It provides clear explanations at both the global model level and for individual applicants, helping borrowers understand approval or rejection decisions and suggesting specific improvements.

The framework addresses three key challenges in automated lending: regulatory compliance (GDPR, RBI guidelines), enabling loan officers to review and justify model decisions, and empowering borrowers with a clear path to improve their credit access. The computational efficiency of TreeSHAP supports real-world deployment in microfinance branches and mobile lending platforms.

Future work will focus on: (1) Testing on Indian microfinance datasets including mobile transaction records and agricultural income; (2) Incorporating fairness constraints during training to reduce age and income-related biases; (3) Creating a multilingual interface in Tamil, Hindi, Telugu, and Kannada; and (4) Expanding the XAI approach to ongoing credit monitoring, using SHAP for early warning when borrowers transition from low to high risk.

VIII. ACKNOWLEDGMENT

The authors express their gratitude to the Department of Computer Science and Engineering for providing access to the computational resources needed for this research. The "Give Me Some Credit" dataset was acquired from the Kaggle platform. No external funding was received for this study.

REFERENCES

[1] Reserve Bank of India, "Report of the Committee on Financial Inclusion," RBI Publications, Mumbai, India, Technical Report, 2019.

[2] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 General Data Protection Regulation (GDPR)," Official Journal of the European Union, L 119, April 2016.

[3] Reserve Bank of India, "Guidelines on Model Risk Management and Responsible AI/ML in Financial Services," RBI Circular RBI/2021-22/172, February 2022.

[4] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Conf. Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017.

[5] B. Baesens, T. Van Gestel, S. Viaene, M. Stepnova, J. Suykens, and J. Vanthienen, "Benchmarking state of the art classification algorithms for credit scoring," J. Operational Research Society, vol. 54, no. 6, 2003.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 1135-1144.

[7] U. Bhatt, A. Weller, and J. M. F. Moura, "Evaluating and aggregating feature-based model explanations," in Proc. 29th Int. Joint Conf. Artificial Intelligence (IJCAI), Yokohama, Japan, 2020, pp. 3016-3022.

[8] M. J. Ariza-Garzon, J. Arroyo, A. Caparrini, and M. Segovia-Vargas, "Explainability of a machine learning granting scoring model in peer-to-peer lending," IEEE Access, vol. 8, pp. 64873-64890, 2020.

[9] J. Ke, "Give Me Some Credit," Kaggle Competition Dataset, 2011. [Online]. Available: https://www.kaggle.com/c/GiveMeSomeCredit. [Accessed: Feb. 2025].

[10] S. van Buuren, Flexible Imputation of Missing Data, 2nd ed. Boca Raton, FL, USA: CRC Press / Taylor & Francis, 2018.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785-794.

[13] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nature Machine Intelligence, vol. 2, no. 1, pp. 56-67, Jan. 2020.

[14] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Machine Intelligence, vol. 1, no. 5, pp. 206-215, May 2019.

[15] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in Proc. 30th Conf. Neural Information Processing Systems (NeurIPS), Barcelona, Spain, 2016, pp. 3315-3323.

[16] X. Dastile, T. Celik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature review," Expert Systems with Applications, vol. 161, Art. no. 113676, Dec. 2020.