IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INTERPRETABLE MACHINE LEARNING FOR CUSTOMER DEFAULT PREDICTION USING ENSEMBLE MODELS AND EXPLAINABLE AI

[1] Mr. K. Raveendra Chaitanya, [2]Mr. Y. V. Ramesh

[1] Assistant Professor, [2] Assistant Professor

[*]Department of Computer Science and Engineering

Geethanjali Institute of Science and Technology (Autonomous), Nellore, Andhra Pradesh, India

Abstract: Accurate assessment of customer financial risk is a critical requirement for lending institutions, as incorrect default predictions can result in significant financial losses and unstable credit portfolios. Traditional rule-based risk evaluation systems are limited in their ability to capture complex, non-linear relationships present in large-scale financial datasets and often fail to adapt to evolving customer behavior. This paper presents an interpretable machine learning framework for customer default prediction that combines ensemble gradient boosting models, namely XGBoost, LightGBM, and CatBoost, with Isolation Forest for unsupervised anomaly detection. The dataset used is the Home Credit Default Risk dataset, comprising 307,511 customer records described by 137 engineered features spanning demographic, financial, credit history, employment, and property attributes. Data preprocessing techniques including median and mode-based missing value imputation, standard scaling, and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE) are applied to improve data quality, given that defaulting customers represent approximately 8% of the dataset. To classify customers into Low, Medium, and High risk categories, ensemble models are trained and evaluated using accuracy, precision, recall, and ROC-AUC. Experimental results show that LightGBM achieves the highest accuracy of 92.04%, closely followed by XGBoost at 92.01%, while CatBoost attains 83.40%. The Isolation Forest model independently flags 8.03% of test records as anomalous, complementing the supervised models by surfacing unusual application patterns. To ensure transparency, Explainable Artificial Intelligence (XAI) using SHAP (Shapley Additive Explanations) is integrated to interpret individual predictions and quantify feature contributions, allowing analysts to understand why a specific customer is classified into a particular risk category. The proposed system demonstrates that combining high-performing ensemble models with explainability techniques can deliver accurate, transparent, and trustworthy customer risk assessment for real-world financial applications.

Index Terms - Customer Default Prediction, Credit Risk Assessment, Ensemble Learning, XGBoost, LightGBM, CatBoost, Isolation Forest, Explainable Artificial Intelligence, SHAP, Anomaly Detection.

I. INTRODUCTION

Financial institutions and lending organizations routinely handle large volumes of customer financial data to assess the likelihood that a borrower will repay a loan or other credit obligation on time. Accurately identifying customers who are likely to default is essential, as misjudged risk assessments can lead to direct financial losses, weakened lending portfolios, and reduced confidence in credit decision-making processes.

Traditional credit risk evaluation systems are largely rule-based and rely on fixed thresholds and predefined conditions derived from historical experience. While such systems are simple to implement, they are not well suited to capturing complex, non-linear relationships that exist among financial attributes such as income, credit history, employment status, and repayment behavior. As customer behavior and economic conditions continue to evolve, static rule-based systems struggle to adapt, resulting in less accurate and less reliable risk predictions.

Machine learning offers an alternative approach by learning patterns directly from historical data rather than relying on manually defined rules. However, many high-performing machine learning models, particularly ensemble and boosting-based methods, operate as "black boxes," making it difficult for analysts and decision-makers to understand the reasoning behind a given prediction. In financial applications, this lack of transparency is a significant limitation, since regulatory and business requirements often demand clear justification for credit decisions.

This paper proposes an interpretable machine learning approach for customer default prediction that combines ensemble learning models, namely XGBoost, LightGBM, and CatBoost, with Isolation Forest for detecting unusual or anomalous customer behavior. To address the lack of transparency commonly associated with ensemble models, Explainable Artificial Intelligence (XAI) techniques, specifically SHAP (Shapley Additive Explanations), are incorporated to interpret model predictions and quantify the contribution of individual features. The system classifies customers into Low, Medium, and High risk categories based on their financial data, supporting more informed and data-driven lending decisions.

The remainder of this paper is organized as follows: Section II presents a literature survey of related work; Section III describes the proposed methodology; Section IV presents implementation details; Section V discusses the experimental results; and Section VI concludes the paper with directions for future work.

II. LITERATURE SURVEY

Customer risk prediction using machine learning has become an active area of research in financial risk management, as organizations increasingly deal with large volumes of customer financial data that are difficult to analyze manually. Traditional risk evaluation methods depend on fixed rules and predefined conditions, which are not flexible enough to handle complex and dynamic data and frequently fail to identify hidden relationships between financial factors such as income, credit history, transaction patterns, and repayment behavior. To overcome these limitations, researchers have introduced machine learning techniques capable of automatically learning patterns from data, with algorithms such as XGBoost, LightGBM, and CatBoost being widely used because of their ability to handle large datasets and capture complex feature relationships.

Anomaly detection forms another important strand of research in this domain. Isolation Forest is an unsupervised machine learning algorithm specifically designed to identify data points that deviate from normal observations without requiring labeled data. The algorithm isolates anomalies by randomly selecting features and recursively splitting the data, where points that require fewer splits to isolate are treated as anomalies. In financial systems, such anomalies may correspond to customers exhibiting unusual transaction patterns or inconsistent income profiles, and identifying them helps to surface hidden risk cases that supervised models may not directly capture.

Feature engineering and data preprocessing are recognized as essential steps in building effective risk prediction systems, since financial datasets are often large, contain missing values, and exhibit significant class imbalance between defaulting and non-defaulting customers. Median and mode imputation are commonly used to handle missing values, while the Synthetic Minority Over-sampling Technique (SMOTE) is widely applied to balance minority default cases and improve model learning. Engineered features such as the credit-to-income ratio and payment behavior indicators have been shown to provide additional insight into customer risk levels.

Ensemble learning models such as XGBoost, LightGBM, and CatBoost are considered state-of-the-art techniques for predictive analysis in financial risk prediction because they combine multiple learning algorithms to improve accuracy and reliability. XGBoost builds decision trees sequentially and reduces error at each step, making it efficient and capable of handling complex data patterns. LightGBM is designed for faster training on large datasets through a leaf-wise growth strategy, while CatBoost is specifically optimized to handle categorical features efficiently, reducing the need for extensive preprocessing.

A persistent challenge in applying advanced machine learning models to financial decision-making is the lack of interpretability, since many models act as black boxes that obscure the reasoning behind individual predictions. Explainable Artificial Intelligence (XAI) techniques address this gap, with SHAP (Shapley Additive Explanations) being a widely used method grounded in cooperative game theory. SHAP assigns importance values to individual features, indicating whether each feature increases or decreases the predicted risk, thereby improving trust, transparency, and accountability in automated credit risk assessment systems.

Table 1: Existing Customer Risk Prediction Approaches

Model	Description	Limitations
Logistic Regression / Decision Trees	Basic classification of customers as risky or non-risky.	Limited ability to model complex non-linear relationships
Random Forest	Ensemble of decision trees for risk classification.	Computationally expensive on large datasets
XGBoost / LightGBM (rule-augmented)	Gradient boosting models used for default prediction.	Operate as black-box models with limited transparency
Standalone Isolation Forest	Unsupervised detection of unusual customer records.	Does not directly predict default probability

Research gaps remain in achieving multi-level risk classification beyond simple binary outcomes, in effectively handling severe class imbalance, and in providing transparent, feature-level explanations for individual predictions. Most existing approaches focus on improving raw predictive accuracy without adequately addressing interpretability, which limits their adoption in regulated financial environments. The present work addresses these gaps through a framework that combines ensemble gradient boosting models with anomaly detection and SHAP-based explainability.

III. PROPOSED METHODOLOGY

The proposed system predicts customer default risk using an interpretable machine learning framework that integrates ensemble gradient boosting models with unsupervised anomaly detection and explainable AI. The methodology consists of several stages: data collection, data preprocessing, handling of class imbalance, feature engineering, model training, model evaluation, and explainability, as illustrated in the system architecture.

A. Data Collection and Dataset Description

The dataset used in this project is the Home Credit Default Risk dataset obtained from Kaggle, which contains real-world loan application data representing customer financial and behavioral information such as income, credit amount, credit history, employment details, and repayment patterns. The dataset comprises 307,511 total customer records described by 137 engineered features, split into 246,008 training samples and 61,503 test samples. Key attributes of the dataset are summarized in Table 2.

Table 2: Dataset Attributes Used

Attribute	Data Type	Description
Demographic Features	Categorical / Numerical	Age, gender, family status, education level
Financial Features	Numerical	Income, credit amount, annuity, goods price
Credit History	Numerical / Categorical	Previous applications, credit bureau records
Employment Features	Categorical	Employment type, organization type, occupation
Engineered Ratios	Numerical	Credit-to-income ratio, annuity-to-income ratio
TARGET	Categorical	Indicates Default or Non-Default outcome

B. Data Preprocessing

Several preprocessing techniques are applied to improve dataset quality before model training. Missing values are handled using median imputation for numerical features and mode imputation for categorical features, ensuring that the dataset remains complete without discarding important records. Duplicate and irrelevant records are removed, categorical variables are converted into numerical representations through appropriate encoding, and all numerical features are normalized using standard scaling to maintain consistent feature ranges across the dataset.

C. Handling Imbalanced Data using SMOTE

In the Home Credit Default Risk dataset, defaulting customers represent a clear minority, accounting for approximately 8.07% of records compared to 91.9% non-default cases in both the training and test sets. To address this imbalance and prevent the models from being biased toward the majority class, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate synthetic samples for the minority default class, enabling the models to learn discriminative patterns associated with high-risk customers more effectively.

D. Feature Engineering

Feature engineering is performed to create additional meaningful attributes that enhance the model's ability to capture customer risk. Engineered features include the credit-to-income ratio, annuity-to-income ratio, age group categorization, and payment behavior indicators derived from previous instalment records. The resulting feature set spans demographic features, financial features, credit history, employment features, property features, engineered ratios, and normalized external source scores, providing a comprehensive representation of customer financial behavior.

E. Model Training: Ensemble Learning and Anomaly Detection

The proposed system trains multiple complementary models on the preprocessed and balanced dataset. XGBoost, a gradient boosting algorithm that builds decision trees sequentially and corrects the errors of preceding trees, is used to classify customers based on complex financial patterns. LightGBM, which uses a leaf-wise tree growth strategy and histogram-based learning, is employed for faster training and efficient handling of the large-scale dataset. CatBoost, which uses ordered boosting and natively handles categorical variables without extensive preprocessing, is included to improve performance on categorical financial

attributes. In parallel, Isolation Forest, an unsupervised algorithm that isolates data points through random feature splits, is used to flag customer records exhibiting unusual or anomalous patterns that may not be captured by the supervised models.

F. Explainability using SHAP

To address the lack of transparency commonly associated with ensemble models, SHAP (Shapley Additive Explanations) is integrated into the system. SHAP is based on cooperative game theory and computes the contribution of each input feature to a specific prediction, indicating whether a feature increases or decreases the predicted default risk. This allows analysts to understand, on a case-by-case basis, why a particular customer was classified into a given risk category, improving the overall trust and accountability of the system.

G. Risk Classification

Based on the predicted default probability, the system classifies customers into Low Risk, Medium Risk, and High Risk categories. This multi-level classification provides more granular insight than a simple binary risky/non-risky outcome, helping financial institutions prioritize manual review and intervention for customers identified as higher risk.

Table 3: Models and Techniques Used

Component	Technique	Purpose
Missing Value Handling	Median / Mode Imputation	Maintains dataset completeness
Class Balancing	SMOTE	Balances default vs. non-default classes
Supervised Classification	XGBoost, LightGBM, CatBoost	Predicts customer default risk
Anomaly Detection	Isolation Forest	Flags unusual customer patterns
Explainability	SHAP	Explains feature contributions to predictions

IV. IMPLEMENTATION

A. Software Requirements

The system is implemented using the Python programming language, which provides extensive support for machine learning and data analysis. The ensemble models are built using the XGBoost, LightGBM, and CatBoost libraries, while Scikit-learn is used for the Isolation Forest implementation, preprocessing, and evaluation metrics. SHAP is used for model explainability. Data processing is handled using Pandas and NumPy, while Matplotlib is used for generating visualizations. The interactive application interface is implemented using Streamlit, and the system is designed to run on a standard 64-bit machine with at least 8 GB of RAM.

B. System Architecture

The overall system architecture consists of sequential components: (i) Data Collection, which loads the Home Credit Default Risk dataset; (ii) Data Preprocessing, responsible for cleaning, imputing missing values, and balancing the dataset using SMOTE; (iii) Feature Engineering, which derives additional informative attributes; (iv) Model Training, where XGBoost, LightGBM, CatBoost, and Isolation Forest are trained; (v) Model Evaluation, which measures performance using accuracy, precision, recall, and ROC-AUC; (vi) Explainability, where SHAP generates feature-level explanations; and (vii) Risk Prediction, which classifies customers into Low, Medium, and High risk categories and displays results to the user.

C. System Modules

The implementation is organized into modular components. The dataset collection module loads and structures the Home Credit Default Risk data. The preprocessing module performs missing value imputation, duplicate removal, encoding, and normalization. The imbalance-handling module applies SMOTE to the training data. The feature engineering module derives ratios and behavioral indicators. The model training module trains XGBoost, LightGBM, CatBoost, and Isolation Forest on the processed data. The evaluation module computes performance metrics for each model and selects the best-performing configuration. The explainability module applies SHAP to generate global and local feature importance explanations, and the prediction module outputs the final risk category along with the associated default probability for each customer record.

D. Testing

A structured testing strategy is employed to validate the system, comprising unit testing of individual components such as data preprocessing and feature engineering, integration testing of combined modules including model training and prediction, system testing of the complete pipeline, and acceptance testing to confirm that the system meets its stated objectives. Table 4 summarizes the testing levels applied during system validation.

Table 4: Levels of System Testing

Level	Scope	Outcome
Unit Testing	Individual components (preprocessing, feature engineering)	Pass
Integration Testing	Combined modules (preprocessing + model training + prediction)	Pass
System Testing	Complete end-to-end pipeline	Pass
Acceptance Testing	Final system validated against project objectives	Pass

V. RESULTS AND DISCUSSION

The proposed system is evaluated on the Home Credit Default Risk dataset, which contains 307,511 customer records and 137 engineered features, split into 246,008 training samples and 61,503 test samples. The class distribution in both the training and test sets shows that non-default customers account for approximately 91.9% of records, while default customers represent approximately 8.07%, confirming a significant class imbalance that is addressed through SMOTE during preprocessing.

Three ensemble models, namely LightGBM, XGBoost, and CatBoost, are trained and evaluated on the preprocessed dataset. The experimental results show that LightGBM achieves the highest accuracy of 92.04%, closely followed by XGBoost at 92.01%, while CatBoost achieves a comparatively lower accuracy of 83.40%. These results indicate that the leaf-wise, histogram-based learning strategy used by LightGBM and the sequential error-correction approach used by XGBoost are particularly effective for this dataset, while CatBoost, despite its strength in handling categorical features, achieves relatively lower overall accuracy in this configuration.

Table 5: Model Performance Comparison

Model	Accuracy
LightGBM	92.04%
XGBoost	92.01%
CatBoost	83.40%

A radar chart comparison across accuracy, precision, recall, F1-score, and ROC-AUC further illustrates that LightGBM and XGBoost maintain consistently strong performance across all evaluation dimensions, whereas CatBoost shows comparatively lower values for several metrics, suggesting that the boosting and gradient strategies of LightGBM and XGBoost are better aligned with the characteristics of the engineered feature set in this dataset.

The Isolation Forest model is applied independently to detect anomalous customer records. On the test dataset, it flags 4,939 records, corresponding to an anomaly rate of 8.03%. Among customers who actually defaulted, 8.62% are flagged as anomalies, while 7.98% of non-defaulting customers are similarly flagged, suggesting that these records exhibit patterns that deviate from typical customer behavior and may warrant additional manual review. This anomaly detection layer complements the supervised classification models by surfacing edge cases and potentially unusual applications that may not be strongly captured by the boosting models alone.

In an illustrative single-customer prediction, the system computes a default probability of 13.1%, corresponding to a Low Risk classification, with the Isolation Forest module marking the record's behavioral pattern as normal with an anomaly score of −0.3926. This demonstrates the system's ability to combine probabilistic risk scoring with anomaly-based validation for individual customer assessments.

To address model interpretability, SHAP is used to explain individual predictions. For the illustrative customer record, the SHAP waterfall explanation shows that the mean external source score contributes most strongly toward increasing the predicted risk, while factors such as gender, family status, and the credit-to-goods ratio contribute toward reducing the predicted risk relative to the baseline expected value. This feature-level breakdown allows analysts to understand precisely which financial and demographic attributes drive a given prediction, rather than relying solely on a numerical risk score.

Overall, the results demonstrate that the combination of high-accuracy ensemble models, SMOTE-based class balancing, Isolation Forest anomaly detection, and SHAP-based explainability produces a system that is both accurate and interpretable. The achieved accuracy levels of over 92% for LightGBM and XGBoost, together with transparent, feature-level explanations, support the system's applicability to real-world customer default prediction and credit risk assessment tasks.

VI. CONCLUSION

This paper presents an interpretable machine learning framework for customer default prediction that combines ensemble gradient boosting models, namely XGBoost, LightGBM, and CatBoost, with Isolation Forest for unsupervised anomaly detection. The system analyzes customer financial data drawn from the Home Credit Default Risk dataset to classify customers into Low, Medium, and High risk categories, supporting more informed lending decisions.

Data preprocessing techniques including median and mode-based missing value imputation, standard scaling, and SMOTE-based class balancing are applied to address data quality issues and the significant class imbalance present in the dataset. Among the ensemble models evaluated, LightGBM achieves the highest accuracy of 92.04%, followed closely by XGBoost at 92.01%, while CatBoost achieves 83.40%. The Isolation Forest module independently flags 8.03% of test records as anomalous, complementing the supervised models by identifying unusual customer behavior.

A key contribution of this work is the integration of SHAP-based Explainable AI, which provides transparent, feature-level explanations for individual predictions and helps analysts understand the financial and demographic factors driving each risk classification. This addresses a major limitation of conventional black-box ensemble models and improves trust and accountability in automated credit risk assessment.

The developed system contributes to more accurate, efficient, and interpretable customer risk prediction, helping financial institutions manage risk more effectively. Future work will focus on extending the framework with deep learning architectures to capture more complex financial patterns, implementing real-time risk prediction on live financial data, deploying the system as a scalable web-based or cloud-based application, and incorporating additional Explainable AI techniques alongside SHAP to provide further insight into model behavior.

REFERENCES

[1] D. A. Dura and A. M. Cazacu, "The impact of artificial intelligence on credit risk assessment and business model transformation in the financial sector," Journal of Law and Economic Development, vol. 6, no. 1, pp. 199–203, 2024.

[2] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit risk prediction using machine learning and deep learning: A study on credit card customers," Risks, vol. 12, no. 11, p. 174, Nov. 2024.

[3] K. Liu and J. Zhao, "KACDP: A highly interpretable credit default prediction model," arXiv preprint arXiv:2411.17783, Nov. 2024.

[4] A. Hussain, M. A. Khan, A. A. K. S, and K. Kousarziya, "Enhancing credit scoring models with artificial intelligence: A comparative study of traditional methods and AI-powered techniques," in A Modern Approach to AI – Integrating Machine Learning with Agile Practices, Oct. 2024, pp. 99–107.

[5] W. A. Addy, C. E. Ugochukwu, A. T. Oyewole, O. C. Ofodile, O. B. Adeoye, and C. C. Okoye, "Predictive analytics in credit risk management for banks: A comprehensive review," GSC Advanced Research and Reviews, vol. 18, no. 2, pp. 434–449, Nov. 2024.

[6] M. K. Nallakaruppan, H. Chaturvedi, V. Grover, B. Balusamy, P. Jaraut, J. Bahadur, V. P. Meena, and I. A. Hameed, "Credit risk assessment and financial decision support using explainable artificial intelligence," Risks, vol. 12, no. 10, p. 164, Oct. 2024.

[7] A. S. Al-Afeef, "Credit risk prediction with and without weights of evidence using conventional predictive statistical models," Cogent Economics & Finance, vol. 12, no. 1, pp. 1–15, 2024.

[8] P. Widagdo, R. A. Pratiwi, H. Nurlinda, N. Nurbaeti, R. Ismiwati, F. R. Kurniawan, and S. Yusriani, "Artificial intelligence in credit risk: A literature review," in Proc. 6th Int. Seminar on Business, Economics, Social Science, and Technology (ISBEST), vol. 3, 2023.

[9] H. M. Blom, P. E. de Lange, and M. Risstad, "Estimating value-at-risk in the EUR/USD currency cross from implied volatilities using machine learning methods and quantile regression," Journal of Risk and Financial Management, vol. 16, no. 7, p. 312, Jul. 2023.

[10] L. O. Hjelkrem, P. E. de Lange, and E. Nesset, "The value of open banking data for application credit scoring: Case study of a Norwegian bank," Journal of Risk and Financial Management, vol. 15, no. 12, p. 597, 2022.

[11] M. Yusuff, "Credit risk modeling using predictive analytics," ResearchGate, Apr. 2021.

[12] P. M. Addo, D. Guegan, and B. Hassani, "Credit risk analysis using machine and deep learning models," Risks, vol. 6, no. 2, p. 38, 2018.

[13] M. S. Islam, L. Zhou, and F. Li, "Application of artificial intelligence (artificial neural network) to assess credit risk: A predictive model for credit card scoring," Master's thesis, Blekinge Institute of Technology, 2009.

[14] DNV, "Optimizing human-AI collaboration," DNV Research, Future of Digital Assurance.

[15] S. M. Lundberg and H. Chen, "SHAP: Beeswarm plot," SHAP Documentation, shap.readthedocs.io.

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[17] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems, 2017, pp. 3146–3154.

[18] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in Advances in Neural Information Processing Systems, 2018, pp. 6638–6648.

[19] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. 8th IEEE Int. Conf. Data Mining (ICDM), 2008, pp. 413–422.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.