Interpretable Credit Risk Modeling Using Machine Learning

A Research-Style Technical Report

Author:
Atul Sharma

Degree Program:
Master of Science in Big Data Science

Institution:
Potomac University, Washington, D.C.

Date:
February 2026

Introduction

Credit risk assessment is a central problem in financial decision-making, particularly for lending institutions that must evaluate the likelihood of borrower default. Accurate prediction of loan default risk enables financial institutions to manage losses, price loans appropriately, and maintain financial stability. Traditional credit scoring methods often rely on linear models and manually designed rules, which may fail to capture complex relationships present in large-scale financial data.

In recent years, machine learning techniques have been increasingly adopted in credit risk modeling due to their ability to handle high-dimensional data and nonlinear interactions between variables. However, the use of complex models introduces concerns regarding interpretability, transparency, and trust—especially in regulated financial environments where decisions must be explainable.

In this project, I investigate the application of machine learning models to predict loan default using real-world LendingClub loan data." Baseline models, including Logistic Regression and Decision Trees, are compared against a gradient-boosted decision tree model (XGBoost). Model performance is evaluated using the ROC-AUC metric, and feature importance analysis is conducted to improve interpretability. The objective of this study is to balance predictive performance with interpretability in the context of practical credit risk modeling.

Dataset

The dataset used in this study consists of publicly available LendingClub loan records spanning from 2007 to 2018. LendingClub is a peer-to-peer lending platform that provides detailed information on borrower characteristics, loan attributes, and repayment outcomes. The dataset includes both numerical and categorical features such as loan amount, interest rate, borrower income, debt-to-income ratio, credit grade, and FICO score ranges.

Due to the large size of the original dataset, a random subset of approximately 200,000 observations was selected for analysis to ensure computational feasibility while preserving representative data distributions. The target variable was defined as loan default status, derived from the loan outcome labels provided in the dataset. Loans classified as "Fully Paid" were treated as non-defaults, while charged-off or defaulted loans were treated as defaults.

Missing values were present in several features and were handled during preprocessing using appropriate imputation strategies. Categorical variables were encoded using one-hot encoding to enable compatibility with machine learning models

During data preparation, I explored multiple sampling strategies and selected a random subset of approximately 200,000 observations to balance computational feasibility with representativeness.

Methodology

Missing values were handled during preprocessing using median imputation for numerical features and most-frequent imputation for categorical features, based on empirical inspection of feature distributions . Numerical features were scaled where appropriate, and missing values were imputed using median or most-frequent strategies depending on the feature type. Categorical variables were transformed using one-hot encoding. The dataset was split into training and testing sets using an 80–20 split.

Three models were implemented and compared:

1. Logistic Regression
   Logistic Regression was used as a baseline model due to its simplicity, interpretability, and widespread use in traditional credit scoring. The model estimates the probability of loan default as a linear function of the input features.

2. Decision Tree Classifier
   A Decision Tree model was trained to capture nonlinear relationships between features. While more flexible than Logistic Regression, Decision Trees can be prone to overfitting and were used primarily as a secondary baseline.

3. XGBoost (Gradient Boosted Decision Trees)
   XGBoost was used as the final model due to its strong performance in structured tabular data. Gradient boosting combines multiple weak learners to produce a robust predictive model capable of capturing complex feature interactions.

Model performance was evaluated using the Receiver Operating Characteristic – Area Under the Curve (ROC-AUC) metric, which is well-suited for binary classification problems with imbalanced classes. Feature importance scores from the XGBoost model were analyzed to assess model interpretability.

Results

The performance of the models was evaluated on the held-out test set using ROC-AUC. Logistic Regression achieved a ROC-AUC of approximately 0.70, while the Decision Tree achieved a slightly lower performance. The XGBoost model achieved the highest performance, with a ROC-AUC of approximately 0.70, indicating improved discrimination between defaulted and non-defaulted loans.

These results demonstrate that gradient-boosted decision trees can capture nonlinear relationships that are not fully modeled by simpler baseline approaches, while maintaining competitive interpretability through feature importance analysis.

Feature Importance and Interpretation

Feature importance analysis from the XGBoost model revealed that borrower credit quality and loan characteristics were the strongest predictors of default risk. Features related to credit grade, sub-grade, interest rate, and debt-to-income ratio were consistently ranked among the most influential variables. Higher credit grades and FICO score ranges were associated with lower default risk, while higher interest rates and debt-to-income ratios were associated with increased risk.

These findings align with established financial intuition and confirm that the model captures meaningful economic relationships rather than spurious correlations. The use of feature importance analysis enhances model transparency and supports responsible deployment in real-world lending environments

Discussion

The results of this study highlight the trade-off between model complexity and interpretability in credit risk modeling. While simpler models such as Logistic Regression provide transparency, they may fail to capture complex interactions present in real-world data. Gradient boosting improves predictive performance while retaining a degree of interpretability through feature importance measures.

From a practical perspective, the model demonstrates how machine learning techniques can support data-driven lending decisions without sacrificing explainability. This balance is particularly important in financial applications subject to regulatory oversight.

Limitations

This study has several limitations. First, macroeconomic variables such as unemployment rates or interest rate trends were not included. Second, probability threshold optimization and cost-sensitive evaluation were not explored. Finally, fairness and bias analysis across demographic groups was outside the scope of this project but represents an important direction for future work.

Conclusion

This project demonstrates the application of machine learning techniques to credit risk prediction using real-world lending data. By comparing baseline models with a gradient-boosted decision tree approach, the study illustrates the benefits of non-linear modeling while emphasizing interpretability through feature importance analysis. The results suggest that machine learning models can effectively support transparent and data-driven credit risk assessment, with potential extensions in future research.

References

1. Breiman, L. (2001). *Random forests*. **Machine Learning**, 45(1), 5–32.
   → Foundational work on tree-based models and interpretability.

2. Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. **Annals of Statistics**, 29(5), 1189–1232.
→ Core theory behind gradient boosting methods.

3. Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, 785–794.
→ The **official XGBoost paper** (very important).

4. Hand, D. J., & Henley, W. E. (1997). *Statistical classification methods in consumer credit scoring: A review*. **Journal of the Royal Statistical Society**, Series A, 160(3), 523–541.
→ Classic reference for credit risk and scoring models.

5. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). *Benchmarking state-of-the-art classification algorithms for credit scoring*. **European Journal of Operational Research**, 247(1), 124–136.
→ Shows why ML models outperform traditional methods in credit risk.

6. Fawcett, T. (2006). *An introduction to ROC analysis*. **Pattern Recognition Letters**, 27(8), 861–874.
→ Justifies your use of ROC-AUC as an evaluation metric.

7. LendingClub Corporation. (2018). *LendingClub loan data*.
→ Public dataset documentation (you can cite as data source).