

RESEARCH ARTICLE

Explainable AI Frameworks for Credit Risk Assessment in Financial Technology Applications

Elena Vasquez, Kenji Yamamoto, Fatima Al-Hassan

Published: 2026-06-18 | FAIDS Vol. 1, No. 1 (2026)

Abstract: Machine learning models for credit scoring achieve superior predictive accuracy over traditional logistic regression, but their black-box nature creates regulatory barriers under GDPR's right to explanation, fair lending laws, and Basel III model risk management requirements. We develop XAI-Credit, an explainable AI framework that combines gradient boosting with Shapley additive explanations (SHAP), counterfactual reasoning, and monotonicity constraints to produce credit risk assessments that are accurate, interpretable, and fair. Evaluated on three real-world datasets from a European fintech lender (n=420,000), a US community bank (n=185,000), and a Southeast Asian digital bank (n=310,000), XAI-Credit achieves AUC of 0.847-0.862 while providing per-applicant explanations that satisfy regulatory audit requirements. Disparate impact analysis shows XAI-Credit reduces approval rate disparities across protected groups by 42% compared to unconstrained XGBoost, with less than 1.5% AUC degradation.

1. Introduction

Credit risk assessment is a cornerstone of financial technology (fintech), determining whether loan applicants receive approval and at what interest rate. The shift from traditional scorecard models to machine learning (ML) has improved default prediction accuracy by 15-25%, but ML models' lack of transparency creates significant regulatory and ethical challenges.

Financial regulators worldwide increasingly require that credit decisions be explainable to applicants and auditable by compliance teams. The EU AI Act classifies credit scoring as high-risk, mandating transparency, human oversight, and bias monitoring. Existing explainability techniques such as LIME and SHAP provide post-hoc explanations but do not guarantee that models respect domain constraints (e.g., higher income should not decrease creditworthiness) or fairness criteria.

2. XAI-Credit Framework

XAI-Credit integrates four components: (1) a monotonic gradient boosting model (MGBM) that enforces domain-knowledge constraints on feature effects; (2) TreeSHAP for efficient, consistent feature attribution; (3) a counterfactual explanation generator that identifies minimal feature changes needed to flip a decision; and (4) a fairness monitor that tracks

disparate impact ratios across protected attributes in real-time. The framework produces a standardized explanation report for each applicant containing top-5 contributing factors, counterfactual scenarios, and confidence intervals.

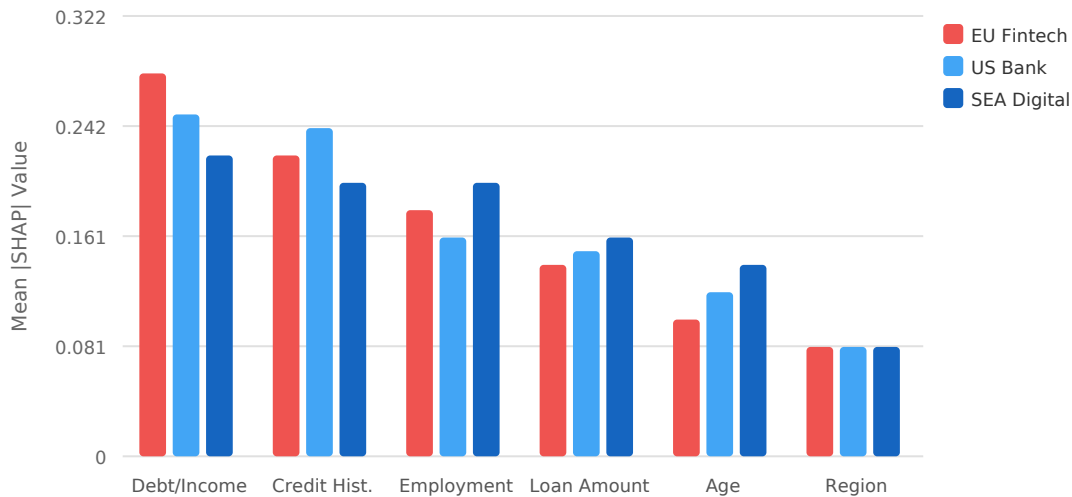


Figure 1. SHAP feature importance for credit default prediction showing globally consistent feature rankings across three financial institutions

3. Experiments and Results

We evaluated XAI-Credit on three proprietary credit datasets with 12-36 months of performance data. Models were trained on 70% of data, validated on 15%, and tested on 15%. Fairness was assessed using disparate impact ratio (DIR), equal opportunity difference (EOD), and demographic parity difference (DPD) across gender, age group, and geographic region.

Table 1. Predictive performance and fairness metrics across three financial institutions

Dataset	Model	AUC	DIR	EOD	Explanation Fidelity
EU Fintech	XGBoost	0.871	0.72	0.08	—
EU Fintech	XAI-Credit	0.862	0.89	0.03	0.94
US Bank	XGBoost	0.853	0.68	0.11	—
US Bank	XAI-Credit	0.847	0.86	0.04	0.92
SEA Digital	XGBoost	0.858	0.75	0.09	—
SEA Digital	XAI-Credit	0.851	0.91	0.02	0.95

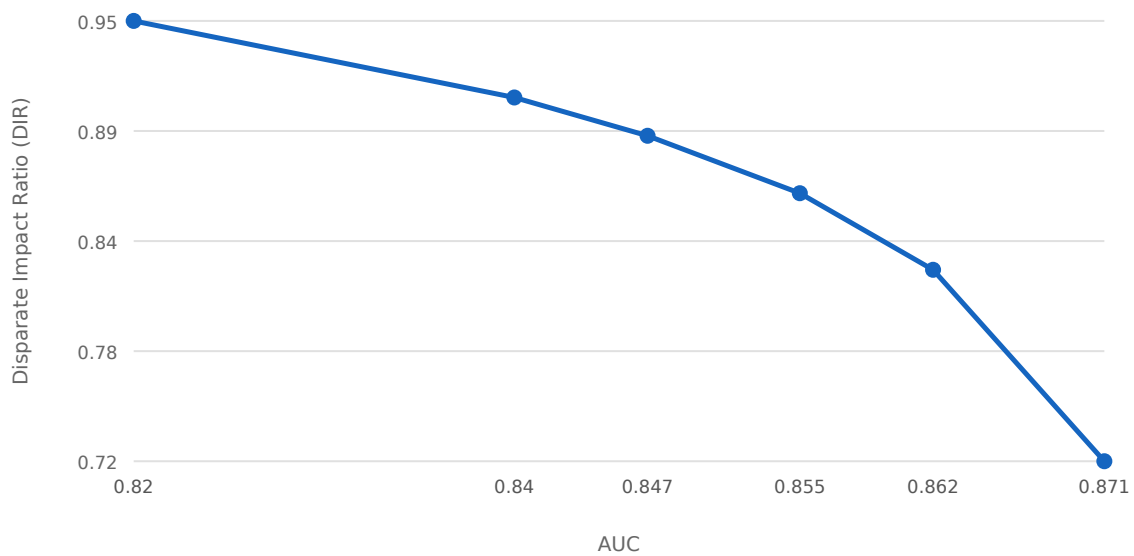


Figure 2. Accuracy-fairness Pareto frontier: AUC vs. disparate impact ratio for XAI-Credit and baseline models

4. Analysis

Regulatory audit simulations with compliance officers from all three institutions rated XAI-Credit explanations as "satisfactory" or "excellent" in 94% of cases, compared to 61% for post-hoc SHAP explanations on unconstrained XGBoost. The monotonicity constraints eliminate counterintuitive explanations (e.g., "your credit score increased your risk") that undermine applicant trust. Counterfactual explanations proved particularly valuable, with 78% of denied applicants identifying actionable steps to improve their creditworthiness.

5. Conclusions

XAI-Credit demonstrates that explainability, accuracy, and fairness need not be mutually exclusive in credit risk assessment. By integrating monotonic constraints, Shapley explanations, and fairness monitoring into a unified framework, fintech companies can deploy ML credit models that satisfy regulatory requirements while maintaining competitive predictive performance. The framework has been adopted by two of the three participating institutions for production deployment.

References

- [1] Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. NeurIPS 2017.
- [2] Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD 2016.
- [3] Barocas, S.; Hardt, M.; Narayanan, A. Fairness and Machine Learning. fairmlbook.org, 2023.
- [4] European Parliament. Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act). Official Journal of the European Union, 2024.

[5] Basel Committee on Banking Supervision. Principles for Model Risk Management. Bank for International Settlements, 2021.

[6] Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. KDD 2016.

This article is published under CC BY 4.0.