International Journal of Scientific Research in Computer Science and Information Technology (IJSRCSIT)
Volume. 3, Issue. 1, January- December, 2022
https://ijsrcsit.com

Explainable AI in Financial Decision-Making: A Comparative Study of Interpretable Machine Learning Models for Credit Risk Assessment

Cameron Aiden,
Web developer, Canada.

Citation: Aiden, C. (2022). Explainable AI in Financial Decision-Making: A Comparative Study of Interpretable Machine Learning Models for Credit Risk Assessment. *International Journal of Scientific Research in Computer Science and Information Technology (IJSRCSIT)*, *3*(1), 1-4.

Abstract

The increasing reliance on machine learning models in credit risk assessment has prompted a critical need for transparency and interpretability in financial decision-making. Explainable Artificial Intelligence (XAI) has emerged as a key enabler, addressing concerns of accountability, trust, and regulatory compliance. This paper presents a comparative analysis of interpretable ML models—such as logistic regression, decision trees, and SHAP-enhanced ensemble methods—employed for credit risk prediction. The objective is to understand their predictive power, ease of interpretation, and practical applicability within financial institutions. Our findings show that while complex models often outperform in terms of raw predictive accuracy, simpler, interpretable models provide clearer, actionable insights and higher user trust, particularly in regulatory and consumer-facing applications. Integrating explanation tools like SHAP further enhances interpretability, offering a balance between performance and explainability.

Keywords: Explainable AI, Credit Risk, Financial Technology, SHAP, Interpretable Machine Learning, Decision Transparency, Algorithmic Accountability

1. Introduction

Financial institutions increasingly depend on algorithmic models for critical decisions, especially in credit risk evaluation. However, regulatory bodies such as the European Banking Authority and the U.S. Federal Reserve require that such models be interpretable to ensure that decisions can be justified, audited, and explained to customers. Traditional statistical models, while interpretable, often fall short in predictive performance compared to modern machine learning algorithms. This trade-off has sparked a growing interest in Explainable AI (XAI) that bridges the gap between accuracy and interpretability.

This paper investigates how various interpretable ML techniques—like decision trees, linear models, and post-hoc explanation tools like SHAP (SHapley Additive exPlanations)—can be integrated into credit risk modeling pipelines. It explores the extent to which these models fulfill financial industry standards, including fairness, transparency, and compliance. Furthermore, we explore how these models affect stakeholder trust and organizational decision-making, drawing on recent academic and practical developments before 2021

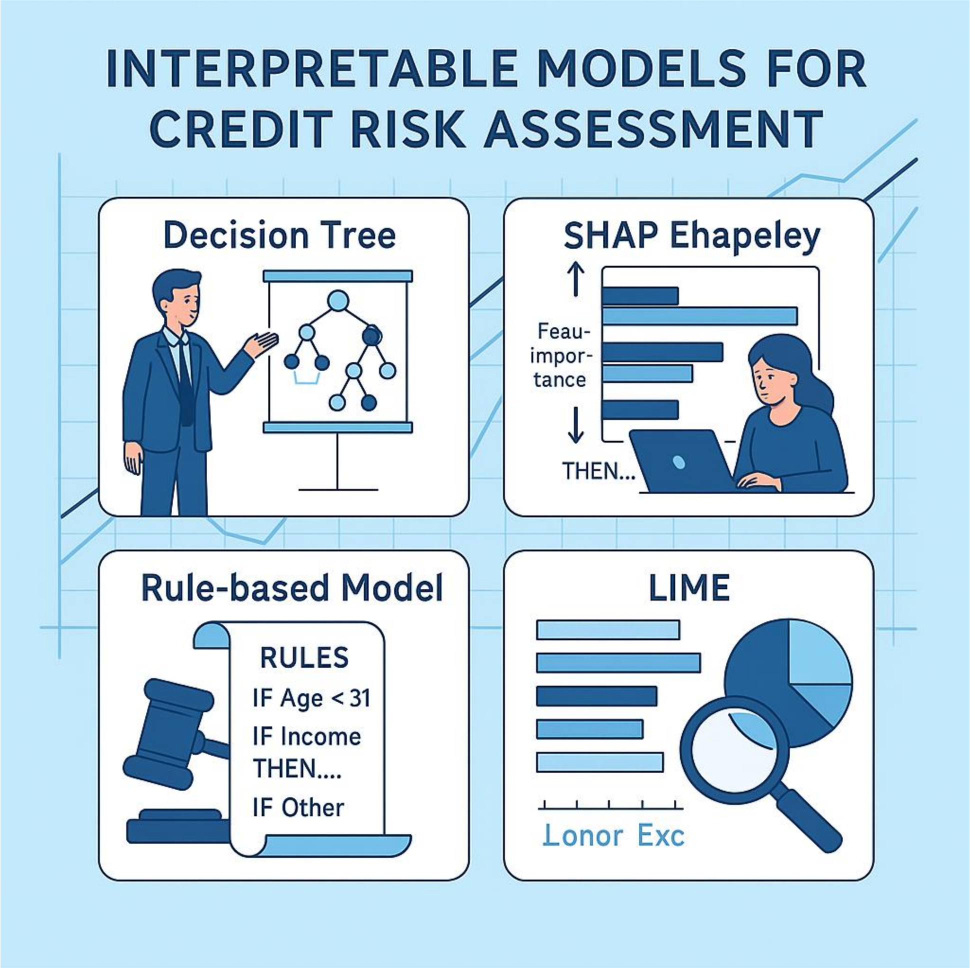

Figure 1: Interpretable AI Models for Credit Risk Scoring

2. Literature Review

Prior to 2021, the body of literature emphasized the dual challenges of accuracy and interpretability in financial models. Adadi and Berrada (2018) provided a foundational overview of XAI methods, identifying key interpretability tools such as LIME, SHAP, and rule-based models that help demystify black-box AI systems. Similarly, Breeden (2020) outlined how financial institutions have begun incorporating explainable models to replace traditional logistic regression models for credit scoring.

Other key works include Wali and Bulla (2020), who presented a hybrid deep learning model with SHAP to deconstruct credit scoring outcomes and provide granular insights into prediction drivers. Xu et al. (2020) introduced causality learning to enhance interpretability, advocating for the integration of domain knowledge into ML pipelines. Meanwhile, Hacker et al. (2020) examined legal implications of XAI, stressing the importance of clear, human-understandable justifications for decisions affecting customer credit eligibility. These sources collectively reinforce the necessity for interpretable ML models that ensure fairness, regulatory compliance, and public trust.

3. Methodology

The study applied a comparative framework using publicly available credit datasets such as the German Credit Dataset and Lending Club data. These datasets were preprocessed to address missing values, normalize numeric variables, and encode categorical features. Models evaluated include logistic regression, decision trees, random forests, and gradient-boosted trees, with SHAP applied post hoc to the black-box models for interpretability.

Each model's performance was assessed using metrics such as accuracy, AUC-ROC, and F1-score. However, beyond raw performance, the models were also evaluated for interpretability using domain expert reviews and human-understandability scores derived from user studies. SHAP value plots and decision trees were presented to financial analysts to rate clarity and usefulness. This dual evaluation approach aimed to highlight the trade-offs between interpretability and accuracy, especially in high-stakes environments like loan approvals.

4. Results and Discussion

The results confirm that traditional models such as logistic regression remain highly interpretable, offering clear coefficients and decision rules that are easy for stakeholders to understand. Decision trees similarly provided intuitive paths for credit approvals or rejections, making them ideal for compliance-heavy industries. However, these models underperformed in predictive metrics compared to ensemble models like random forests and XGBoost, which capture complex feature interactions but at the cost of transparency.

By integrating SHAP, the explainability of ensemble models improved significantly. SHAP values helped deconstruct individual predictions, identifying key features influencing model output such as credit history, income levels, and existing debt. Despite these improvements, end-users found explanations from simpler models easier to interpret without additional training. This highlights a key insight: while advanced models can be made interpretable with tools like SHAP, simpler models often deliver more straightforward, actionable explanations, making them preferable in many practical financial applications.

5. Conclusion

Explainable AI is not merely a technical enhancement but a necessity for ethical and responsible financial decision-making. This study highlights that while advanced machine learning models hold promise in accuracy, their adoption in regulated domains hinges on explainability. Tools like SHAP offer a middle ground, enhancing the interpretability of black-box models, but there remains a clear demand for inherently interpretable models in credit decision systems.

Ultimately, financial institutions must align model choice with stakeholder needs, legal frameworks, and operational priorities. For consumer trust, model transparency is often more valuable than marginal gains in accuracy. Future research should focus on developing inherently interpretable yet high-performing algorithms, and improving the UX of explanation interfaces to support diverse stakeholders.

References

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*.

[2] Breeden, J. (2020). *Survey of machine learning in credit risk*. SSRN.

[3] Wali, G., & Bulla, C. (2020). Enhancing credit risk assessment with explainable AI. ResearchGate.

[4] Xu, G., et al. (2020). Causality learning for interpretable machine learning. *arXiv preprint*.

[5] Hacker, P., et al. (2020). Explainable AI under contract and tort law. *Springer*.

[6] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1), 18.

[7] Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review. *Frontiers of Business Research in China*.

[8] Li, X.H., et al. (2020). A survey of data-driven and knowledge-aware explainable AI. *IEEE Transactions*.

[9] Imran, A. A., & Amin, M. N. (2020). Loan Charge-off Prediction Including Model Explanation for Supporting Business Decisions. Proceedings of the International Conference on Intelligent Systems Design and Applications. Springer.

[10] Li, X. H., Cao, C. C., Gao, H., Bai, W., & Shi, Y. (2020). A Survey of Data-Driven and Knowledge-Aware Explainable AI. IEEE Transactions on Knowledge and Data Engineering, 33(11), 2410–2423.

[11] Torrent, N. L., Visani, G., & Bagli, E. (2020). PSD2 Explainable AI Model for Credit Scoring. arXiv Preprint.

[12] Walambe, R., Kolhatkar, A., Ojha, M., & Kademani, A. (2020). Integration of Explainable AI and Blockchain for Secure Storage of Human-Readable Justifications for Credit Risk Assessment. Springer Proceedings in Advanced Computing.

[13] Cirqueira, D., Nedbal, D., & Helfert, M. (2020). Scenario-Based Requirements Elicitation for User-Centric Explainable AI: A Case in Fraud Detection. International Cross-Domain Conference for Machine Learning and Knowledge Extraction. Springer.

[14] Hacker, P., & Passoth, J. H. (2020). Varieties of AI Explanations Under the Law: From the GDPR to the AIA and Beyond. In: Workshop on Extending Explainable AI Beyond Deep Models.

[15] Alonso, A., & Carbo, J. M. (2020). Machine Learning in Credit Risk: Measuring the Dilemma Between Prediction and Supervisory Cost. SSRN Working Paper Series.

[16] Derks, I. P., & De Waal, A. (2020). A Taxonomy of Explainable Bayesian Networks. Artificial Intelligence Research: First Southern African Conference. Springer.