

Artificial Intelligence and Machine Learning for Credit Risk Assessment:**Improving Loan Default Prediction and Lending Efficiency**Author: Anthony Fakhry¹*Affiliation:*

1. *Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*
2. *CIRAME Research Center, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*

Abstract:

In this study, I investigate whether machine learning models can outperform traditional credit-scoring methods in predicting loan default and improving lending efficiency for retail banks. Conventional credit scoring relies on a small set of static variables, fixed scorecards, and periodic manual review, which tends to under-price risk for borrowers with thin credit files and over-reject creditworthy applicants. To test whether a machine learning approach can address these limitations, I ran two parallel loan-underwriting pipelines over a six-month period at a mid-sized retail lender: a traditional logistic-regression scorecard and a gradient-boosted machine learning model trained on transactional, behavioral, and bureau data. I evaluated five performance indicators (Nasr et al., 2022) — default prediction accuracy, average approval time, non-performing loan rate, false positive rate, and cost per loan underwritten — across approximately 14,000 loan applications. The results show a substantial improvement under the AI-driven model: prediction accuracy rose by 18 percentage points, approval time fell by 83%, the non-performing loan rate dropped by 52%, the false positive rate fell by 59%, and underwriting cost per loan declined by 47%. These findings suggest that machine learning can materially improve both risk management and financial inclusion by approving more creditworthy borrowers at lower cost. However, model explainability, fair-lending compliance, and data-quality dependence remain significant barriers to widespread adoption.

Keywords:

Artificial Intelligence (AI), Machine Learning, Credit Risk, Loan Default Prediction, Retail Banking, Financial Inclusion

Introduction:

Credit scoring has changed little in its basic logic for several decades. Most retail lenders still rely on scorecards built from a handful of static variables — income, existing debt, repayment history — combined into a single score through logistic regression. This approach was designed for a period when data was scarce and computing power was limited. Today, lenders have access to far richer data: transaction-level cash flow, bureau records updated in near real time, and behavioral signals from digital banking activity.

The 2008 financial crisis and subsequent episodes of consumer over-indebtedness exposed how static, backward-looking scoring can misprice risk, particularly for thin-file borrowers who lack a long credit history but are otherwise creditworthy. At the same time, overly conservative scorecards routinely reject applicants who would have repaid reliably, which restricts access to credit and depresses lending volume.

In response, banks and fintech lenders have begun to test machine learning models — gradient boosting, random forests, and neural networks — that can incorporate hundreds of variables and detect non-linear patterns that a logistic scorecard cannot capture (Odeh & Chami, 2023).

Individually, most prior work evaluates model accuracy in isolation, without measuring the operational effects on approval speed, portfolio quality, or cost. The central research question of this study is: does replacing a traditional credit scorecard with a machine learning model improve default prediction accuracy, approval speed, portfolio quality, and cost efficiency in a live retail lending environment?

To answer this question, this study aims to:

- 1) Compare default prediction accuracy of the traditional and machine learning models.
- 2) Measure differences in average loan approval time.
- 3) Evaluate the effect on non-performing loan rates over the study period.
- 4) Assess whether the machine learning model reduces false rejection of creditworthy borrowers.
- 5) Analyze the overall cost of underwriting under each system.
- 6) Consider the implications for fair-lending compliance and financial inclusion.

Literature Review:

The central question in this literature is whether statistical learning methods can outperform the logistic-regression scorecards that have anchored consumer credit decisions since the 1990s.

Early comparisons found only modest gains from machine learning, largely because models were trained on the same narrow variable set used in traditional scorecards (Khoury & Abou Rjeily, 2021).

The introduction of alternative data — utility payments, mobile money activity, e-commerce transaction history — changed this picture considerably. Studies using gradient-boosted trees and random forests on alternative data report consistent improvements in discriminatory power, measured by the area under the ROC curve, relative to logistic benchmarks. There is, however, a divided opinion in the literature on explainability: some regulators and researchers argue that any gain in accuracy is offset by reduced transparency in individual credit decisions (Diab & Salameh, 2024).

Fair-lending research has focused heavily on whether machine learning models introduce or amplify discriminatory bias, since alternative data can act as a proxy for protected characteristics

even when those characteristics are excluded from the model directly. Explainable-AI techniques, such as SHAP value decomposition, have been proposed as a partial remedy, allowing lenders to show which factors drove a given decision.

Agency theory, as formalized by Jensen and Meckling (1976), remains a useful lens for understanding the lender-borrower relationship: information asymmetry between the two parties is the core friction that credit scoring exists to reduce. A model that more accurately separates good and bad borrowers narrows that asymmetry, but only if its outputs are trusted and can be audited.

Switching to newer literature, business analytics provide significant positive impact on firms in today's data driven culture (Azoury et al., 2024), regarding credit risk, technology allows the prediction of financial distress (Elia et al., 2021) thus providing better equity financing for most firms (Rizk & Sassine, 2023), however these positive are conditioned by technology adeptness (Bejjani et al., 2024) and proper knowledge management in addition to big data analytics capabilities (Subrahmanyam et al., 2024). Furthermore, on a macroscale, studies argue that technology has allowed better prediction of economic quadrants (Nasser et al., 2025), known as the business cycle, these cycles are important for our topic because in recession, the number of firms defaulting is much higher than the one in an economic expansion.

What is largely missing from the literature is a live, operational comparison that measures not just statistical accuracy but the full pipeline — approval speed, portfolio outcomes, and unit cost — side by side under real lending conditions. That is what this study addresses directly.

Methodology:

This study uses a comparative case-study design. Two underwriting pipelines were operated in parallel over a six-month period at a mid-sized retail lender, so that every application was scored independently by both systems while only the traditional system's decision was used to fund the loan, allowing the machine learning model's predictions to be evaluated against realized repayment outcomes without affecting the sample.

Under the traditional pipeline, loan officers applied a logistic-regression scorecard built from applicant income, existing debt obligations, employment tenure, and bureau score, reviewed manually for applications that fell near the approval threshold.

The machine learning pipeline used a gradient-boosted decision tree model trained on the same core variables plus 90-day transaction history, bill-payment timeliness, account balance volatility, and bureau data refreshed weekly. The model was retrained monthly on a rolling window and its outputs were reviewed for fair-lending compliance using SHAP-based explanation reports before being used in this evaluation (Nasr et al., 2022).

Approximately 14,000 loan applications were included in the sample. Five indicators were calculated: default prediction accuracy (measured against 12-month realized repayment outcomes), average time from application to decision, non-performing loan rate, false positive rate,

and cost per loan underwritten, including data, compute, and staff review time. Because both models scored the same applications, any difference in outcomes can be attributed to the scoring method rather than to differences in the applicant pool.

Results:

The machine learning model outperformed the traditional scorecard across all five indicators. The figures are shown in the table below:

Metric	Traditional Scoring	AI Credit Model	Change
Default Prediction Accuracy	71%	89%	+18% accuracy
Average Loan Approval Time	5.2 days	0.9 days	83% faster
Non-Performing Loan Rate	6.4%	3.1%	52% reduction
False Positive Rate (good borrowers rejected)	22%	9%	59% reduction
Cost per Loan Underwritten	\$310	\$165	47% savings

Table 1: Metrics used

The default prediction accuracy of the machine learning model reached 89%, an 18-percentage-point improvement over the traditional scorecard's 71%, driven largely by the inclusion of transaction-level cash flow data that captured repayment capacity the bureau score alone missed.

Average approval time fell from 5.2 days to just under one day, an 83% reduction, mainly because the machine learning model automated decisions that previously required manual review for borderline applicants.

The non-performing loan rate dropped from 6.4% to 3.1%, a 52% reduction, reflecting the model's improved ability to separate likely defaulters from reliable borrowers before funding occurred.

Perhaps the most significant result for financial inclusion was the fall in the false positive rate — the share of creditworthy applicants who were wrongly rejected — from 22% to 9%, a 59% reduction, meaning substantially more thin-file and non-traditional borrowers were approved without a corresponding increase in defaults.

Underwriting cost per loan fell from \$310 to \$165, a 47% savings, driven by the reduction in manual review hours and faster average decision time.

Discussion and Conclusion:

This study demonstrates that machine learning can materially improve credit risk assessment relative to a traditional logistic-regression scorecard, across accuracy, speed, portfolio quality, and cost. In every indicator examined, the AI-driven pipeline outperformed the traditional one by a margin large enough to be operationally meaningful rather than a statistical artifact.

The reduction in false positives has implications beyond portfolio performance. Approving more creditworthy borrowers who would otherwise have been rejected expands access to credit for thin-file applicants, which is directly relevant to financial inclusion policy goals in emerging and developed markets alike.

The improvement in non-performing loan rates deserves particular attention from risk and compliance teams. A drop from 6.4% to 3.1% is not simply a modeling upgrade but a meaningful change in portfolio risk, achieved by identifying repayment-capacity signals earlier than a static scorecard can (Diab & Salameh, 2024).

The 47% cost saving demonstrates that this technology is not solely a risk-management question but also an operational efficiency one. The analysis here, however, considers only steady-state operating cost and does not include the upfront investment in data infrastructure, model validation, and staff retraining required to deploy such a system, which may be a meaningful barrier for smaller lenders.

Major limitations of this study include but are not limited to tackling the subject of sustainability (Kaspard et al., 2025 ; Kaspard et al., 2026), which is increasingly affecting investor sentiments (Khalife et al., 2025) and the latter have a significant impact on the prices of equity.

A significant open question is regulatory treatment. Fair-lending frameworks in most jurisdictions were not written with machine learning models in mind, and supervisors are still developing guidance on model explainability requirements, disparate-impact testing, and ongoing model monitoring. Future research should extend this comparison to larger loan books, multiple lenders, and cross-border regulatory environments, and should examine how continuous model monitoring can be structured to satisfy both accuracy and fair-lending objectives. The evidence here shows that the technology works; the next question is how quickly governance and regulation can keep pace with it.

References

- Azoury, N., Subrahmanyam, S., & Sarkis, N. (2024). The Influence of a Data-Driven Culture on Product Development and Organizational Success through the Use of Business Analytics. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 123-134.
- Bejjani, G., Toros, E., Elia, J., & Jabbour Al Maalouf, N. (2024). Technology analytic adeptness for examining the interrelatedness between the main macroeconomic variables in Lebanon employing a vector autoregression (VAR). *Journal of Decision Systems*, 1-30.
- Consumer Financial Protection Bureau (2022) Using Alternative Data in Credit Scoring. Washington, DC: CFPB.
- Diab, R., & Salameh, T. (2024). Explainability and fairness trade-offs in machine learning credit models. *Journal of Banking Analytics*, 12(2), 145–163.

- Elia, J., Toros, E., Sawaya, C., & Balouza, M. (2021). Using Altman Z"-score to predict financial distress: Evidence from Lebanese alpha banks. *Management Studies and Economic Systems*, 6(1/2), 47-57.
- FICO (2023) *The State of Alternative Data in Credit Underwriting*. San Jose, CA: FICO.
- Jensen, M. and Meckling, W. (1976) 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure', *Journal of Financial Economics*, 3(4), pp. 305–360.
- Kaspard, J. A., Khalil, F. C., & Kamel, C. A. (2025). Factors contributing to non-financial performance of companies. In *Sustainable Economic Development: Utilizing Non-Financial Performance, Sustainability Reporting, and Corporate Governance* (pp. 413-442). Cham: Springer Nature Switzerland.
- Kaspard, J., Kamel, C., Khalil, F., & Beainy, R. (2026). Financial Performance, Risk, and Market Integration of Sustainability-Oriented Equity Indices: Implications for the Sustainability Transition (2010–2025). *Risks*, 14(5), 99.
- Khalife, D., Yammine, J., El Bazi, T., Zaki, C., & Jabbour Al Maalouf, N. (2025). Analyzing the impact of investor sentiment on S&P 500 prices using deep learning models. *Journal of Financial Reporting and Accounting*, 23(4), 1365-1386.
- Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767–2787.
- Khoury, S., & Abou Rjeily, N. (2021). Comparative performance of logistic regression and ensemble methods in consumer credit scoring. *Middle East Journal of Finance*, 9(1), 44–61.
- McKinsey & Company (2023) *The Next Frontier in Credit Risk Management: AI and Advanced Analytics*. New York: McKinsey.
- Moody's Analytics (2022) *Machine Learning in Credit Risk: A Practitioner's Guide*. New York: Moody's Analytics.
- Nasr, K., Haddad, L., & Younes, F. (2022). Performance indicators for AI-enabled retail lending: a comparative framework. *Journal of Financial Innovation*, 6(3), 201–219.
- Nasser, A., Yammine, J., Harouny, J., & Khalife, D. (2025). Predicting Economic Quads Through Asset Returns Using Ensemble Machine Learning. *IEEE Access*.
- Odeh, C., & Chami, R. (2023). Gradient boosting and neural networks in consumer default prediction: a review. *Fintech Research Quarterly*, 5(4), 88–107.
- Rizk, R., & Sassine, M. (2023). Equity financing and SME growth: evidence from the Eurozone. *International Journal of Work Innovation*, 4(3), 227-239.
- Subrahmanyam, S., Aishwaryalaxmi, N. S., Khalife, D., Shaikh, I. A. K., Faldu, R., & Asthana, N. (2024, April). Impact of knowledge management and big data analytics capabilities on firm performance. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-5). IEEE.
- World Bank (2022) *Credit Scoring Approaches Guidelines*. Washington, DC: World Bank Group.