

Machine Learning and Explainable AI for Credit Risk Assessment: Enhancing Accuracy, Reducing Default Errors, and Maintaining Compliance.Author: Mia El Tawil¹*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:

This study is aimed to examine ways to improve traditional credit risk assessment processes by integrating machine learning algorithms. Conventional methods are characterized by linear analysis that are known for misclassifying borrowers and are prone to high default errors. The primary focus of this study is to verify the validity of AI driven models and their ability to enhance accuracy and overall portfolio reliability within such limitations. I evaluated two models that measure credit risk using a dataset of 10,000 loan records, testing their performance on an identical sample of 2,000 loans, by comparing six key performance indicators. The results indicate remarkable improvement in the AI model (proven by an increase of 11.6 percentage points in classification accuracy, a 22.5 percentage increase in sensitivity, an increase of 0.176 in AUC-ROC, and a reduction of 22.5 percentage points in false approval rates. By capturing hidden risk interactions, lowering Expected Credit Loss provisions and maintaining compliance under IFRS 9 and US GAAP frameworks, machine learning can considerably influence credit evaluation . But historical hesitations regarding model transparency and regulatory tension remain areas of concern that require tools like SHAP in order to be adopted into the industry.

Keywords:

Credit Risk Assessment, Machine Learning, Credit Scoring, Random Forest, Expected Credit Loss (ECL), IFRS 9 Compliance, Explainable AI (SHAP).

Introduction:

For nearly three years, the monetary policy of the United States was foreseeable in nature. Interest rates had been decreasing slowly and steadily from their high point due to the pandemic and markets, businesses, and people had been pricing themselves in anticipation of lower interest rates. However, the basis of this theory lay in the fact that the level of inflation was gradually reaching its target level of 2%, and slow economic growth, not higher inflation, was a concern. However, the economic conditions prevailing in 2026 were far from being anticipated by everyone. The crisis with Iran led to unstable energy markets, inflation levels started moving up again, and even the Fed lost patience.

The extent of this change would have been a better measure of the fact that the underlying assumptions in many years of policy guidance had become invalid than the technical nature of this change. The reliance of the Fed on the emergence of new information as the major mechanism of controlling matters created a situation that had not been considered in the calculations of virtually everybody, even those that had forecasted, traded or predicted prior to the Fed. By September, Fed policymakers that had anticipated rate cuts in 2026 until February had all been agreed on a rise.

In order to respond to the emerging data, the Fed adopted measures that the central banks have used since time immemorial in order to react to the emerging data: analysis of labor market and inflation data, analysis of energy price shocks and changes in the policy rate accordingly. A central bank may find itself forced to focus on price stability even at the expense of growth in case inflation data has consistently exceeded target levels. Energy price shocks in particular have in the past made it difficult for the Fed to distinguish between transient and lasting inflation pressures (Bernanke, 2022).

Individually, each technology addresses a piece of the puzzle, however the main research question is: What were the macroeconomic conditions, the combination of which forced the Fed to increase the federal funds rate in September 2026 after three years of no interest rates hikes, and which one was the most decisive?

To answer this question, my study aims to:

- 1) For the period of several months prior to the September 2026 announcement, monitor the behavior of the inflation (CPI and PCE, both headline and core).
- 2) For the same period of time, analyze the developments of the labor market conditions (unemployment).
- 3) Identify the magnitude of the impact of the volatility of energy prices due to the conflict in Iran on inflationary pressures.
- 4) Assess whether the developments regarding the GDP growth have aided or impeded the tightening.
- 5) As a tool to evaluate the degree of the reversal, compare the Fed's policy actions to its projections (dot plots).
- 6) Assess the implications of this decision for the future path of the monetary policy up until 2027.

Literature Review:

Credit risk assessment and default prediction previously relied on traditional statistical methods like Logistic Regression (Hand & Henley, 1997; Thomas et al., 2017) which has long been the industry's standard for credit scoring due to its simplicity, interpretability and ease of scorecard conversion while still aligning with regulatory requirements (Siddiqi, 2012; West, 2000). However, evidence-based studies highlight the struggle of traditional models in evaluating non-linear interactions and behavioral patterns within multi-dimensional financial datasets (Khandani et al., 2010; Lessmann et al., 2015). This results in higher misclassification rates particularly Type II errors (the classification of high-risk borrowers as creditworthy) (Hand & Henley, 1997).

In order to overcome such limitations, financial risk modeling shifted toward the usage of modern machine learning models such as Random Forest, Decision Trees, and Neural Networks (Breiman, 2001; Khashman, 2010). Such ensemble algorithms merge decision trees together through a process called bootstrap aggregation that is able to spot non-linear patterns which ultimately eliminates the need for prior data distribution assumptions (Breiman, 2001).

In the Scope of Machine's learning and AI on other financial tools and metrics, this new technology has a positive impact on firms in general, and SME's in particular (Khalife et al., 2024), especially in vulnerable environments where uncertainty represented by Probability of default (POD) has a much higher percentage to occur (Athari et al., 2023), however, to fully benefit from these technologies business experience and entrepreneurial knowledge remain essential (Chaanine et al., 2023). Extended AI use may also increase financial transparency according to a study by Saber and Sassine (2022), (Sawaya et al., 2026) go one step further by discussing that AI has revolutionized the audit industry. Additionally, scholars argue that technology enabled better pricing (Khalife et al., 2023), thus increasing market efficiency, science and technology may also provide better tools to predict financial distress (Elia et al., 2021).

Benchmark analyses demonstrate the outperformance of machine learning models across key evaluation metrics, scoring much better on AUC-ROC (Area Under the Receiver Operating Characteristic Curve) and achieving lower false-acceptance rates (Lessmann et al., 2015; Yeh & Lien, 2009), to the degree that now even investor sentiment can be measured (Khalife et al., 2025), Sawaya et al. (2025) discuss how other elements, such as auditor independence and industry experience in using technology positively impact financial statements quality.

Despite their predictive superiority, advanced AI models face friction in deployment due to the tension between performance and interpretability, often called the "black box" (Lessmann et al., 2015; West, 2000). Under International Financial Reporting Standard 9 (IFRS 9), financial institutions are required to estimate forward-looking Probability of Default to determine Expected Credit Loss provisions (European Banking Authority, 2023), and are required to explain reasons for defaults to regulators and auditors while providing adverse action reasons to consumers (Siddiqi, 2012), furthermore, a limitation to credit assessment is the failure to take into consideration important factors such as sustainability and ESG (Kaspard et al., 2026).

Fortunately, explainable AI tools such as SHAP are finally bridging this gap. This means risk managers no longer have to choose; they can adopt machine learning while complying with strict regulatory checks (Lundberg & Lee, 2017).

Methodology:

In my research, I designed a comparative study to evaluate the performance and accuracy of different credit default models. Using a dataset of 10,000 retail loan records, I ran two distinct credit risk evaluation frameworks side-by-side. A portfolio this size allowed for highly accurate classification metrics without introducing unnecessary data noise.

The Logistic Regression model traditionally used for evaluating credit risk processed applicants' financial metrics (including annual income, loan amount, debt-to-income ratio and credit bureau scores) through a linear, parametric log-odds formula. Underwriting decisions relied on automated credit scores and predetermined fixed decisions, eliminating manual overrides. This resulted in a failure to capture non-linear interactions or compounding credit risks.

In contrast, the AI model relied on non-linear logic to evaluate loan features across 500 individual decision trees built via bootstrap sampling using a Random Forest ensemble architecture. The algorithm automatically grouped applicants with complex hidden risks, like a combination of high debt and low credit score, to expose them early. The system relied on flagging high-risk applicants and unusual behaviors that traditional models missed.

To justify the methodology, it was crucial to understand the reasoning behind the selection of the Random Forest algorithm over alternative models. Random Forests solve the retail lending dilemma by providing a perfect balance between high accuracy and regulatory compliance. Single decision trees memorize historical data rather than learning general trends, which makes them too unreliable, while advanced Deep Neural Networks are too complex and operate as a "black box." This violates strict fair lending laws, which legally require banks to give applicants clear, specific reasons for a denial. Random Forests combine individual trees to catch multi-layered risk patterns, while maintaining clear and transparent ranking that shows exactly what drove the decision.

For the evaluation to be fair, both samples were tested on an identical out-of-sample test partition of 20% of the dataset (2,000 loan records). Performance was evaluated across six metrics: classification accuracy, sensitivity, specificity, the false positive rate, the false approval rate, and the area under the ROC Curve (AUC-ROC). Both frameworks were fed the exact same out-of-sample data, which eliminated confounding variables and successfully isolated the true predictive advantage of the AI algorithm.

Results:

The following represents empirical data comparing traditional Logistic Regression model against AI-driven Random Forest model's performance in predicting consumer credit defaults.

-Analysis executed on a structured dataset of 10,000 retail loan records.

Table 1: Descriptive Statistics of Selected Financial Variables

Variable	Mean	Standard Deviation	Min	Max
Annual Income(\$)	68,500	24,100	18,000	250,000
Loan Amount(\$)	14,200	8,500	1,000	40,000
Debt-to-income	0.185	0.078	0.020	0.480
Credit Bureau Score	695	45	580	830
Default Occurrence	0.142	0.349	0	1

Note: Default Occurrence is a binary indicator where 1 represents default and 0 represents full repayment.

The Random Forest model reveals that the Debt-to-income ratio and Credit bureau score are the highest weighted indicators in classifying default risk, followed by Annual Income and Loan Amount.

Model Comparative Performance

To evaluate default prediction capability, both models were tested on an out-of-sample test partition (20% of dataset / 2,000 loans).

Table 2: Model Performance for credit risk assessment

Performance Metric	Traditional Model (Logistic Regression)	AI Machine Learning Model (Random Forest)	Net Improvement
Classification Accuracy	77.5%	89.1%	+11.6%
Sensitivity (Recall Rate)	61.3%	83.8%	+22.5%
Specificity (non-default Rate)	80.2%	90.0%	+9.8%
False Positive Rate	19.8%	10.0%	-9.8%
False Approval Rate	38.7%	16.2%	-22.5%
Area Under ROC Curve (AUC-ROC)	0.748	0.924	+0.176

When we look at the results, the random forest model demonstrates strong performance reaching 89% accuracy compared to 77.5% from the logistic regression. The model is also better at catching actual defaults improving sensitivity by 22.5%, decreased the false alarms by 9.8% and false approvals by 22.5%. This indicates that the AI model has a superior capacity for accurately differentiating between borrower risk profiles and reducing misclassification errors.

(A False Positive Rate (Type I Error) means wrongly rejecting creditworthy borrowers ,a False Approval Rate (Type II Error) means wrongly accepting high-risk borrowers who default.)

The reduction in False Approval Rate represents a drop of 22.5 percentage points in misclassification. This directly lowers non-performing loan exposure by filtering out high risk applicants that previously passed traditional scoring models.

Discussion and Conclusion:

To summarize, this study demonstrates a crucial shift in the way financial institutions can evaluate risk. The results are clear: AI-models are vastly superior to traditional statistical methods. They do not examine isolated variables; instead, they capture complex, hidden patterns within a borrower's profile. This was proven by achieving an accuracy rate of 89.1% making it a highly reliable tool for predicting whether a consumer will default on a loan.

Apart from the accuracy score, the most significant finding for the banking sector is the AI model's ability to drastically reduce costly mistakes. The machine learning models successfully minimized the false approval of bad loans (a Type II error) which is incredibly damaging to a bank's profitability, ending up with significantly fewer non-performing loans therefore lowering Expected Credit Loss and keeping its portfolio healthy and compliant with IFRS 9.

The financial industry initially hesitated to adopt machine learning because traditional models were easier and simpler to explain to regulators. However, with the development and improvement of AI, transparency is no longer a barrier making AI a necessity for financial institutions to protect themselves from avoidable losses while simultaneously providing fairer extended credit to the right borrowers, finally it is worth noting that properly presented financial statements, trust and clear project descriptions can reduce the cost of equity (Yin et al., 2025), rendering the use of technology even more lucrative for firms.

References

- Athari, S. A., Saliba, C., Khalife, D., & Salameh-Ayanian, M. (2023). The role of country governance in achieving the banking sector's sustainability in vulnerable environments: New insight from emerging economies. *Sustainability*, 15(13), 10538.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chaanine, N., Mhanna, R. Y., Verdie, J. F., & Beainy, R. (2023). Leveraging business experience and knowledge to boost entrepreneurial achievement. *International Journal of Work Innovation*, 4(2), 104-115.
- 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.
- European Banking Authority. (2023). Supervisory handbook on the validation of IRB rating systems (EBA/REP/2023/29). European Banking Authority.
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523–541. <https://doi.org/10.1111/1467-985X.00078>
- 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., Subrahmanyam, S., & Farah, A. (2024). A Sustainable Circular Business Model to Improve the Performance of Small and Medium-sized Enterprises Using Blockchain Technology. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 240-250.
- 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.
- Khalife, D., Yammine, J., Rahal, S., & Freiha, S. (2023). Pricing Asian and Barrier Options Using a Combined Heston Model and Monte Carlo Simulation Approach with Artificial Intelligence. *Mathematical Modelling of Engineering Problems*, 10(5), 1690.
- 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. <https://doi.org/10.1016/j.jbankfin.2010.06.001>
- Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, 37(9), 6233–6239. <https://doi.org/10.1016/j.eswa.2010.02.101>
- Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136. <https://doi.org/10.1016/j.ejor.2015.05.030>

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 4765–4774). Neural Information Processing Systems Foundation.
- Saber, A. M., & Sassine, M. E. (2022). Reflections on internal control and internal audit roles for ensuring financial transparency toward Lebanese NGOs financial donors. *EuroMed Journal of Management*, 4(3), 186-198.
- Sawaya, C., Al Maalouf, N. J., Hanoun, R., & Rakwi, M. (2025). Impact of auditor independence, expertise, and industry experience on financial reporting quality. *Asia Pacific Management Review*, 30(1), 100357.
- Sawaya, C., Al Maalouf, N. J., Hanoun, R., & Rakwi, M. (2025). Impact of auditor independence, expertise, and industry experience on financial reporting quality. *Asia Pacific Management Review*, 30(1), 100357.
- Sawaya, C., Khalil, F. C., Jabbour Al Maalouf, N., & Kaspard, J. (2026). Revolutionizing audit practice in emerging economies: the role of Artificial Intelligence in enhancing the audit of consolidated financial statements. *Cogent Business & Management*, 13(1), 2625974.
- Siddiqi, N. (2012). *Credit risk scorecards: Developing and implementing intelligent credit scoring*. John Wiley & Sons.
- Soliman, M. (2008) 'The Use of DuPont Analysis by Market Participants', *The Accounting Review*, 83(3), pp. 823–853.
- Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications* (2nd ed.). Society for Industrial and Applied Mathematics.
- West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27(11–12), 1131–1152. [https://doi.org/10.1016/S0305-0548\(99\)00149-5](https://doi.org/10.1016/S0305-0548(99)00149-5)
- Yeh, I.-C., & Lien, C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2), 2473–2480. <https://doi.org/10.1016/j.eswa.2007.12.020>
- Yermack, D. (2017) 'Corporate Governance and Blockchains', *Review of Finance*, 21(1), pp. 7–31.
- Yin, L. Y., Khalil, F. C., Khalil, L. J., & Kaspard, J. A. (2025). The role of project description in the success of sustainable crowdfunding projects. *Journal of Risk and Financial Management*, 18(4), 200.