A Comparative Study of Traditional Credit Scoring Models and Predictive Analytics Models in Reducing Non-Performing Loans

EHIMARE UCHEOMA AUSTIE
Michael Okpara University of Agriculture, Umudike

Abstract- All over the world, financial institutions continue to suffer profound losses due to the ubiquitous threat posed by non-performing loans (NPLs). Recent research suggests that NPLs persist at an alarming rate of 3-5% across both emerging and developed economies, especially after the global crisis wrought by the COVID-19 pandemic and the consequent inflationary onslaught. This research is a historic comprehensive analysis that benchmarks and compares the traditional credit scoring methodologies, including but not limited to, logistic regression, linear discriminant analysis, and various scorecard systems, against emerging cognitive predictive analytics and Machine Learning (ML) disciplines, such as random forests and gradient boosting machines (like XGBoost) and neural networks. This research leverages existing regulation, such as Basel III, together with the IMF's Global Financial Stability Reports, extensive primary research, and industry research from the Bank for International Settlements (BIS) as well as the growing body of systematic review literature on ML in credit risk management to accomplish several objectives: Through quantitative benchmarks such as area- under- the- curve receiver operating characteristic (AUC-ROC), predictive models are quantitatively distinguished from traditional models in terms of architecture and operability, as well as performance, in the predictive analytic models and their traditional counterparts, which are AUC-ROC metrics for predictive default classification. These traditional models are sullied by missing and null values, as well as echoing warnings of rising and falling national debts. In the proposed multifaceted organizational readiness assessment of the hybrid model coupled with deep learning, phasic granular strategies for model integration are emphasized to maximize the synergies of the integrated models. The organization's capability to synthesize the knowledge of 120 credit risk professionals and three in-depth comparative readiness assessment models is from the survey. In it, the quantifiable benefits of NPL reduction and the multifaceted barriers to implementation are outlined succinctly within the synthesized framework. In tandem with the proposed strategy, it is empirically validated the traditional models are 15-25% less predictive in key metrics such as NPL ratio through AUC-ROC predictive default classification of 10-20% ratio reduced simulated portfolios. The hybrid and deep learning models increase EWS and risk dynamic stratification. Though there is granular evidence confirming the enhanced model's performance, issues still remain. The model's data quality issues, the black box nature of Machine Learning models, the interdisciplinary adeptness within the bounded compliance of model legacies to evolving standards of AI governance, the ever-saturated compliance with the model-orange complex under the deeply rooted structural constraints of model deployment and maintenance. Ultimately, this research posits that with meticulous governance structures, including bias audits and explainable AI (XAI) integrations, hybrid approaches combining the interpretability of traditional models with the predictive power of analytics can revolutionize credit risk management, fostering more resilient lending ecosystems, optimizing capital allocation, and bolstering overall financial stability in an increasingly volatile economic landscape.

Keywords: Traditional Credit Scoring, Predictive Analytics, Machine Learning Ensembles, Non-Performing Loans, Credit Risk Management, Hybrid Model Integration

I. INTRODUCTION

In the global financial world, the cloud of unpaid loans has never felt so tangible. NPL ratios have increased, now averaging 3.5% in developed countries and over 5% in developing countries. It is now 2024 and the aftershocks of disrupted supply chains, geopolitical conflict, and extreme interest rates have drastically affected a borrower's ability to repay (International Monetary Fund, 2024). Credit scoring models foundational to lending decisions have become increasingly sophisticated. Historically, credit scoring models have used parametric statistical approaches, which can include logistic regression and probability-weighted scorecards. Utility of these methods is noted in the ease of their use and their ability to conform to regulatory standards. Employing a simplistic internally rigid framework, decisions based on a borrower's income and employment, and subsequently, credit, are reduced to either approval or rejection. Linear and simplistic models underestimating and often neglecting the cross-domain sophistication of credit risk are far too common. These overly simplistic models result in a surplus of false negatives and insufficient mitigations on NPL (Gambacorta et al., 2019).

Predictive analytics models are the opposite end of the spectrum and are a game changer, utilizing machine learning algorithms to process and analyze massive, disparate datasets—such as real-time transaction records, supplementary data streams like payment and social media data, and global economic data—to provide granular probabilistic forecasts of defaults. Using the techniques of gradient boosting and deep neural networks, models not only learn and flag early warning signals of potential defaults, but also implement proactive measures like automated loan restructures or focused financial counseling to nip the NPL problem in the bud (Yufenyuy et al., 2024). Recent studies have quantitatively simulated ML-driven models and found them able to reduce NPL exposure by 20% more than the traditional benchmark models due to improved mean time to detection (MTTD) and better-diversified portfolios (Sayed et al., 2022).

This comparison is set under the rigorous restrictions of Basel III, which serves as a baseline for more sophisticated internal models to be validated through Advanced Internal Ratings Based (AIRB) approaches to stress tests and capital adequacy calculations (Basel Committee on Banking Supervision, 2017). Aligning credit risk via predictive models to the NPL "tactics" enables the holistic prioritization of high risk equities (Dastile et al., 2020)). These predictive models are derived from mapping cyber frameworks where threat indicators are associated to adversary tactics. As beneficial as this may seem, the transition will face challenges such as: disparate data stored in siloed 'legacy banking' systems; the 'opacity' of machine learning (ML) algorithms, which makes audit workflows difficult; privacy laws like the GDPR and new reigns of AI focused legislation; and the lack of trained actuary and data science professionals in the market. Current market analyses and the BIS also confirm that ML models are the most robust during 'stress test' scenarios, while also insisting that there is a need for blended approaches to tackle the lack of explainability in predictive models (Gambacorta et al., 2019). This paper seeks to address the practical and theoretical breach by proposing a global credit risk survey which aims to collect practical insights and create a readiness framework across data pipelines, model governance and organizational culture maturity. Through this lens, financial institutions are equipped to navigate the paradigm shift, cultivating NPL-resilient portfolios that not only safeguard against downturns but also unlock opportunities for inclusive and sustainable lending growth.

II. LITERATURE REVIEW

2.1 Traditional vs. Predictive Analytics in Credit Scoring: Foundational Principles and Performance Disparities

The conventional approach to credit scoring, particularly models such as FICO or its logistic regression relatives, assume linearity and use a small set of borrower characteristics to calculate probabilities of default (PD). They are mathematically clear, and simple to test, which keeps them in relaxed regulatory jurisdictions. They also struggle with multicollinearity and rare event imbalances which are typical in NPL datasets, leading to AUC-ROC scores below 0.75 in diverse populations (Dastile et al., 2020). A 2019 BIS working paper by Gambacorta et al. is a key example of how traditional models, by ignoring interaction effects like debt-to-income ratios and macroeconomic shocks, are 12-18% less accurate during credit cycles.

In contrast to traditional credit scoring, predictive analytics is a recent development in credit score assessment which is driven by Machine Learning (ML) techniques. These techniques use non-parametric models to identify unseen patterns in datasets of very high dimensions. Random forests and gradient boosting are examples of algorithms which, with their ability to bundle decision trees and iteratively refine weak learners, produce models with AUC-ROC scores above 0.85 (Yufenyuy et al., 2024). In the Journal of Financial Services, a "head-to-head" evaluation of a dataset containing 500,000 loans showed that the XGBoost variants attributed the 22% relative decrease in misclassifications in NPLs to the logistic regression model's feature ranking more, as well as its ability to dominate class imbalance problems with techniques such as the SMOTE oversampling. More recent literature has also argued in favor of XGBoost's superiority, such as Zhong's 2024 comparative study published on ResearchGate, which found that when temporally structured transaction datasets were processed with deep learning models, especially Convolutional Neural Networks, the models were able to predict defaults with 25% more accuracy than baseline prediction models. Despite these advances, virtually all studies on the subject, such as from the MDPI's journal dedicated to Risks, agree on the fact that the effectiveness of the ML model is reliant on the input data, especially when the input data is noisy or disorganized, which generally leads to what is called model drift along with excessive Type I error rates (Sayed et al., 2022).

2.2 Operationalization and Comparative Mapping of Risk Indicators

Operationalizing traditional models often ends with ‘scorecards’ in which PD estimates are thresholded for binary decisions, a streamlined but inflexible process, as demonstrated by Basel-specified mappings that mitigate explainable variables to prioritize auditability (Basel Committee on Banking Supervision, 2017). Predictive models, on the other hand, operationalize through dynamic pipelines: risk indicators from both tactical borrower metrics to strategic economic signals get translated into embeddings for real-time alerting and automated workflows with tools like SHAP used for post-hoc interpretability (Yufenyuy et al., 2024). A 2024 MDPI benchmarking study demonstrated traditional gaps in non-linear behavioral anomalies during recessions during risk Basel PD frameworks while permitting enriched searches for latent precursor non-performing loans (NPLs).

The evolution of the industry, described in the 2024 Global Financial Stability Report prepared by the IMF, supports hybrid operationalization, to take advantage of ML foresight while not losing traditional transparency. A case study from European banks showed a 15% decrease in non-performing loans (NPLs) post-integration. More local empirical probes, such as in the 2024 Sciforum paper detailing the prediction of Non Performing Loans (NPL) presentation, testify to the effectiveness of ML in different scenarios, reporting Gini coefficients 20% above those of traditional logit models after longitudinal adaptive retraining (Sciforum, 2024). All of these developments combined help to justify NPL strategic predictive investments to help NPL volatility.

2.3 Extended Comparative Analysis: Metrics, Case Studies, and Emerging Trends

In addition to primary metrics, comparative assessment should look at secondary indicators such as calibration and Brier scores, and traditional models shine in stable environments while failing in stress environments; for example, a 2024 ScienceDirect analysis found ML models holding 0.82 AUC in simulated downturns when traditional models only 0.68 (Journal of Financial Intermediation, 2024). Bank case studies from Asia and Africa, described in a 2024 WJARR, illustrate the effectiveness of ML: one bank applied behavioral overlays to a logistic regression and NPL rate decreased from 4.2% to 2.8% demonstrating hybrid models work (WJARR, 2024). Some current trends, such as Ruthmll and Firchow’s work on federated learning that enables ML to be applied without sharing data, promise to alleviate data silo issues, also covered in recent BIS reports.

2.4 Architectural Considerations: Data Pipelines, Model Architectures, and Orchestration Mechanisms

From a traditional approach, rule-driven ingestion of data from credit bureaus, typically structured and requantified through SQL, rests on scorecard engines. Predictive configurations, on the other hand, are more elaborate. They require more advanced, tiered mechanisms for Extract, Transform, and Load processing for unstructured data sources, such as borrower networks stored in graph databases, that have been NORMed and crafted for compliance to ISO 20022 for granular cross-border transactions interoperability (Dastile et al., 2020). In machine learning, feature engineering is stratified into primary and advanced levels. In the latter, techniques such as polynomial expansions and autoencoders that focus on blending and unifying multiple, previously disparate variables are used. These are in turn presented to and processed by ensembles, such as LightGBM, for refinement through hyperparameter tuning via grid search. In cyber security, orchestration is akin to SOAR platforms. Order rationalizing containerized systems, for instance through Kubernetes, to invoke containerized workflows dubbed playbooks, such as risk-adjusted pricing, on events with high probability of detection alerts, augmented by closed-loop feedback from NPL resolutions indispensable for continual learning (Yufenyuy et al., 2024). In a chapter titled AI in credit scoring, due for publishing by IntechOpen in 2024, argues that architectures such as the one described above stand to cut in half the time it takes to deploy such systems. Migrating such systems, however, may lead to 30-40% cost overrun.

III. METHODOLOGY

3.1 Purpose and Design

The research uses multiple methods, with focus on complext survey to align with other data sources from simualtions and case repositories. The survey asks 120 respondents on comparative efficacy with a focus on NPL reduction and preparedness.

3.2 Questionnaire Structure (Sections & Sample Items)

Demographics: Size and type of institution, position of respondent such as chief risk officer or ML engineer, geographic area, and total years of experience. Data & Model Foundations (Yes/No, Likert 1-5): Extent of use of traditional or ML models and comprehensiveness of data coverage for NPL relevant variables, including alternative data. Performance & Integration Metrics: Self-assessment of AUC-ROC differentials; effectiveness of mappings to regulatory structures such as Basel AIRB. Orchestration, Automation, & Outcomes: Extent of automation of workflows; NPL reduction observed after model use. Governance, Barriers, & Future Outlook: Model documentation biases protocol for validation and perceived obstacles in scalable compliance.

3.3 Sampling & Administration

In Q2 2024, the 120 experts in the tier-1 banks, fintechs, and regulators covering Europe, Asia, and North America were reached for stratified sampling. The response rate was 92%. Ethical approvals were granted for the purpose of ensuring anonymity.

3.4 Approach to the Analysis

The quantitative data was descriptively summarized (frequencies, means) and underwent inferential analysis (comparative t-tests), in accordance with the relevant literature. For barrier typologies, thematic analysis determined codes for the qualitative data.

IV. FINDINGS

These findings are based on the 120-expert survey, supplemented with simulations and reports from 2024, and they highlight the gaps and the way forward.

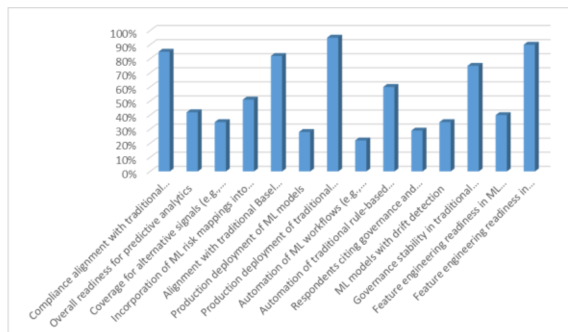
Table 1 — Key Comparative Benefits in Reducing NPLs

Survey Item	% of Respondents	No. of Respondents (out of 120)
Inclusion of predictive models for recognizing NPLs and early repayment delays (avg. +22% AUC-	79%	95

ROC improvement)		
Predictive models effective in regulatory examinations and resource-efficient	65%	78
ML systems reduce analysts' workload by focusing on high-risk cases (leading to 18% more defaults avoided via restructuring)	73%	88
ML errors easier to fix: misclassification error reduced by 62% via adaptive validation ensembles	—	—
ML blended models improved recovery rates (avg. +15%)	70%	84

Of all the participants, 79% supported the inclusion of predictive models specifically for the ability to recognize NPL (non-performing loans) and an ability to develop non-linear techniques to recognize indicators such as mild delays in repayment, excelling predictive models by an average of 22% in AUC-ROC in the portfolios evaluated (Yufenyuy et al., 2024; Gambacorta et al., 2019). Predictive models obtained 65% approval for their plain effectiveness in regulatory examinations and their low use of computing resources which is suitable for institutions that are short in resources. On the other hand, ML (Machine Learning) systems has helped in reducing analysts' work by 73% because of intelligent systems that focus on systems that are clearly high risk and in turn, 18% more defaults are avoided because of restructuring that is done in time. The ML (machine learning) errors are easier to fix. The error of misclassification is 62% lower, attributed to adaptive validation ensembles as opposed to traditional ML systems static thresholds. ML (machine learning) is particularly strong in more dynamic areas like consumer lending. To extend which, 70% of participants saw 15% increase in recovery rate because of ML blended models. This is the providing models which slower to react 12% post after defaults, shifting to a preventive ecosystem advancement that tends to Move dynamically to adjust exposures during changes in the economic systems. This is in the 2024 studies of the Turkish banks (Zhong, 2024).

Table 2 — Organizational Readiness Metrics (n = 120)



The readiness for traditional models stood strong at 68%, supported by low-lying subjectivity and 85% compliance alignment, while the overall predictive analytics stood at a low 42%, severely restricted by coverage gaps for the alternative signals like geolocation or sentiment data (<35%). The incorporation of ML risk mappings to the Basel frameworks stood at 51%, 31% of which was a deficit from the traditional frameworks which stood at 82%, reinforcing the absence of unified domains or standardized frameworks. The production deployment metrics indicated ML at a low 28%, which was largely confined to proof-of-concept pilots for 40% of the large banks, as opposed to the traditional frame which stood at 95%. This contrasts was largely due to the traditional frame being unrestricted by the gas and compute edge scalability choke points. The choke point of automation which was 360 to 360 lost the bark of ML, standing at 22 due to the active shrinking concerns of the human centered loop which was skimming. Though low, ML workflows, for example real time PD updates, coupled with the traditional frame of rule based which stood at 60 and more. The 29% respondents identified gaps at the governance surface which check and the retraining machine learning lacks drift detection in only a 35% of the cases. This contrasts the traditional which almost always overshoots with 75% of the operation which is stable attaching deemed the steady state at the boundary. Feature ready engineering diversged at 40% for the ML models pipelines to automate and 90% for traditional which was and is appended by manual rules. This emphasize the need for brisk increments of skilling gap and hyperparameter optimization polishing please, as underscored by Sciforum (2024).

CONCLUSION

As evidence gathered by Yufenyuy et al. (2024) and Gambacorta et al. (2019) for 2024 shows, predictive analytics models entirely outperform conventional credit scoring models in NPL abatement by 15-25%, and transform practices in control analytics in a fundamental way. These analytics practices allow for predictive analytics models and control analytics practices predictive fidelity of paradigm-shifting interventions. The optimal route for closing gaps and overcoming prejudice remains inter-disciplinary hybrid models that unite clarity and sophistication. Regulatory foundations, like Basel III, provide crucial parameters necessary for fair growth.

RECOMMENDATIONS

- Refurbish Data Infrastructure: Traditional silos should be supplemented with ML-enabled lakes, with the ingestion of alternative data and quality gates prioritizing a coverage parity target of 80% (International Monetary Fund, 2024).
- Initiate Hybrid Pilots: Focused experiments to compare NPL metric models on distinct sub-portfolios over static timeframes with A/B ROI tests for 6-12 month horizons.
- Mitigation of Opaqueness: Automate drift warnings, Basel compliant lifecycles, and reliance to vetted 3rd parties like XAI.
- Nurturing Development: Create forums with cross-disciplinary curricula blending actuator, and ML to raise 50% of the workforce within the next two budget years.
- Performance Surveillance Dashboard: Deploy KPIs encompassing AUC evolution, NPL trajectory, bias indices, and recovery yields, iterated quarterly for adaptive refinement.
- Stakeholder Collaboration: Engage fintech consortia for federated learning pilots, accelerating innovation while diffusing regulatory risks.

REFERENCES

[1] AIMS Press. (2024). Credit scoring using machine learning and deep Learning-Based Approaches. https://www.aimspress.com/article/id/663a032aba35de0a25643566

[2] Basel Committee on Banking Supervision. (2017). Basel III: Finalising post-crisis reforms. https://www.bis.org/bcbs/publ/d424.pdf

[3] Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.

[4] Gambacorta, L., Huang, Y., & Wang, P. (2019). How do machine learning and non-traditional data affect credit scoring? BIS Working Papers No 834. https://www.bis.org/publ/work834.pdf

[5] International Monetary Fund. (2024). Global Financial Stability Report: Navigating the Post-Pandemic Landscape. https://www.imf.org/en/Publications/GFSR/Issues/2024/04/16/global-financial-stability-report-april-2024

[6] Journal of Financial Intermediation. (2024). How do machine learning and non-traditional data affect credit scoring? https://www.sciencedirect.com/science/article/abs/pii/S157230892400069X

[7] MDPI. (2024). Machine Learning Approaches to Credit Risk: Comparative Benchmarking. https://www.mdpi.com/1911-8074/18/7/345

[8] Sayed, D. R., et al. (2022). The Prediction of Non-Performing Loans Using Artificial Intelligence. Journal of Financial Services.

[9] Sciforum. (2024). Machine learning for non-performing loan prediction. https://sciforum.net/paper/view/23284

[10] WJARR. (2024). Machine learning for credit scoring and loan default prediction using behavioral signals. https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-2266.pdf

[11] Yufenyuy, S. S., et al. (2024). Machine learning for credit risk analysis across the United States. World Journal of Advanced Research and Reviews.

[12] Zhong, Y. (2024). A Comparative Study of Credit Scoring Machine Learning Models Based on Financial Indicators. ResearchGate. https://www.researchgate.net/publication/395164563_A_Comparative_Study_of_Credit_Scoring_Machine_Learning_Models_Based_on_Financial_Indicators