MACHINE LEARNING IN CREDIT RISK ASSESSMENT: ADVANCES, CHALLENGES, AND IMPLICATIONS FOR EMERGING MARKETS

Sarvar Umarkhonovich Rejabbaev
Lecturer, Department of Finance and Financial Technologies, TSUE
s.rejabbaev@tsue.uz
Ozodbek Abdurayim o'g'li Sodiqov
Student, Tashkent State University of Economics
ozodbeksodiqov089@gmail.com

Abstract: Credit risk assessment is a critical function for banks and financial institutions, as it *directly influences financial stability, profitability, and access to credit. While traditional credit scoring approaches—such as rule-based systems and logistic regression—continue to be widely adopted due to their interpretability and alignment with regulatory and supervisory requirements, they are often constrained by linear assumptions and challenges in capturing complex, non-linear relationships in borrower data.*

The increasing availability of large-scale financial datasets, combined with advances in artificial intelligence, has led to growing adoption of machine learning (ML) techniques to enhance credit risk modeling. This paper provides a comprehensive review of ML applications in credit risk assessment, examining prominent model families—including tree-based ensembles (e.g., random forests), gradient boosting methods (e.g., XGBoost and LightGBM), and neural networks—alongside key performance metrics such as AUC-ROC and Precision-Recall curves.

The review also addresses major implementation challenges, including data quality and bias, model explainability (through explainable AI techniques), fairness considerations, regulatory compliance (e.g., Basel frameworks and supervisory guidelines), and the need for robust governance, validation, and human oversight. Drawing on academic literature, industry practices, and regulatory publications (e.g., BIS Working Papers), the analysis demonstrates that ML models can deliver superior predictive accuracy and improved risk segmentation compared to traditional methods, particularly when leveraging richer or alternative data sources.

Special attention is given to the opportunities and considerations for emerging markets, such as Uzbekistan, where ML-based approaches—supported by responsible use of alternative data—hold significant potential to promote financial inclusion for borrowers with limited traditional credit histories, provided that strong data governance, ethical safeguards, transparency, and alignment with local regulatory expectations are maintained.

Keywords*:* *credit risk assessment, credit scoring, machine learning, explainable AI, regulatory compliance, financial inclusion, emerging markets, alternative data.*

INTRODUCTION

The possibility that a borrower will not repay a loan in accordance with the terms agreed upon is known as credit risk. Managing this risk is a central responsibility of banks and other lenders. Loan approval, pricing, portfolio management, and regulatory capital calculations are all made possible by credit risk models. Strong risk measurement and adequate capital are emphasized by supervisory frameworks like Basel II and III [1, 2]. Historically, credit risk assessment relied on statistical methods (e.g., logistic regression, discriminant analysis) and expert rule-based systems. These approaches are transparent and relatively simple to validate, but they often assume linear or additive relationships and may underperform when borrower behavior is complex or when high-dimensional data is available.

ML algorithms became usable for credit scoring as financial data sources and computing power increased. ML models can learn non-linear patterns and interactions among variables, potentially improving the prediction of defaults and losses. In regulated lending environments, however, model risk management, explainability, and fairness constraints are crucial. This paper aims to summarize leading ML approaches to credit risk, compare them to traditional methods, and discuss practical considerations for responsible deployment—especially in emerging markets such as Uzbekistan.

LITERATURE REVIEW

1. Traditional credit risk models

Logistic regression and related statistical techniques remain widely adopted in traditional credit scoring due to their strong interpretability and seamless integration into established model governance processes [1] (Basel Committee on Banking Supervision, 2006). Under the Basel II Internal Ratings-Based (IRB) approach, banks are required to estimate key risk parameters, such as probability of default (PD) and loss given default (LGD), which serve as foundational inputs for regulatory capital calculations and broader risk management practices [1, 3] (Basel Committee on Banking Supervision, 2006; European Banking Authority, 2020).

2. Machine learning for credit scoring

Empirical research consistently demonstrates that machine learning (ML) techniques can substantially enhance credit risk prediction by effectively leveraging high-dimensional datasets and capturing complex non-linear relationships among variables [5] (Khandani, Kim and Lo, 2010). For instance, Khandani, Kim and Lo (2010) illustrate how ML approaches significantly improve consumer credit risk modeling when integrated with richer behavioral and transaction-level data, outperforming conventional linear models in forecasting delinquencies and defaults.

3. Benchmarking ML vs. traditional models

Comparative studies frequently report that ensemble methods, such as random forests and gradient boosting machines, deliver superior predictive accuracy relative to traditional statistical models, although rigorous hyperparameter tuning and validation procedures are essential to mitigate risks such as overfitting [6] (Lessmann et al., 2015). A large-scale benchmarking analysis by Lessmann et al. (2015) evaluates a wide range of state-of-the-art classification algorithms across multiple credit scoring datasets, concluding that modern ensemble techniques generally achieve the strongest performance in terms of discriminatory power and overall predictive quality.

4. Explainability and ethical concerns

Increasing regulatory scrutiny and societal expectations place greater emphasis on transparency, accountability, and fairness in automated credit decision-making processes [3] (European Banking Authority, 2020). Explainable AI (XAI) techniques play a critical role in enabling compliance while facilitating interpretation of complex, non-linear models. Among the most commonly applied methods are Individual Conditional Expectation (ICE) plots and Partial Dependence Plots (PDP), which help visualize feature effects, marginal contributions, and heterogeneity across borrowers [10, 11] (Goldstein et al., 2015; Friedman, 2001). These tools support better understanding of model behavior and assist in identifying potential biases or unintended discriminatory outcomes.

METHODOLOGY

This study adopts a qualitative and analytical review approach to examine the application of machine learning (ML) in credit risk assessment. The review synthesizes and critically evaluates evidence drawn from three main categories of sources: academic literature, official regulatory publications, and industry reports and practitioner insights.

Data sources and temporal scope

The analysis primarily focuses on publications released between 2016 and 2025 in order to capture the most recent methodological advances, empirical findings, and regulatory developments in the field. Foundational studies published prior to 2016 are included selectively when they remain highly influential or provide essential conceptual background.

RESULTS

Applications of Machine Learning in Credit Risk Assessment

ML helps with a wide range of credit risk tasks, such as (i) predicting loss and default, (ii) dividing borrowers into risk classes, (iii) spotting fraud and unusual patterns, (iv) making decision-making possible in near real time, and (v) assisting portfolio monitoring early warning systems.

Official Figures and Interpretation

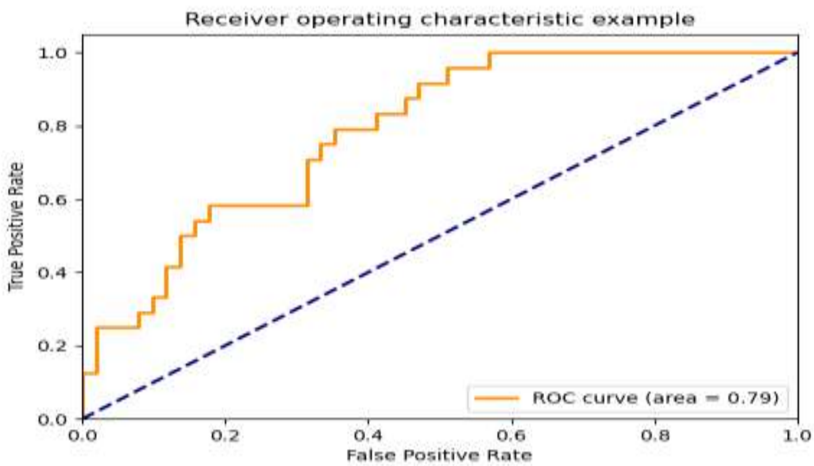
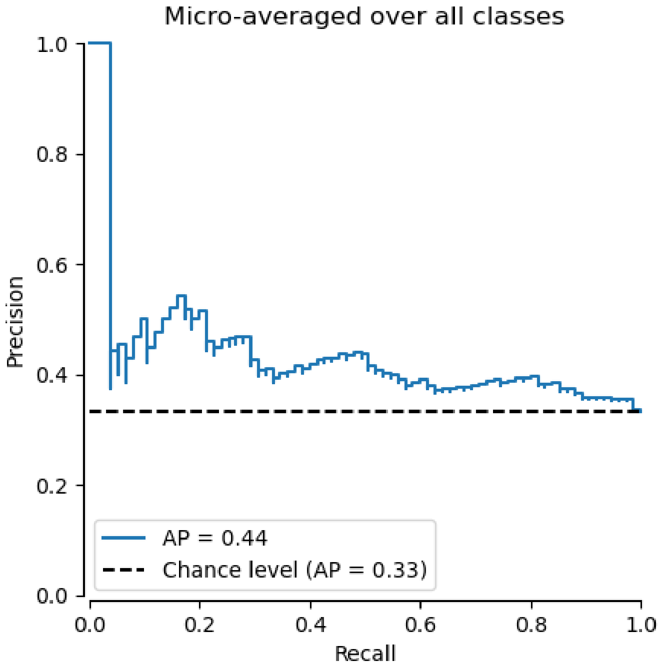


Figure 1. Multiclass ROC curve illustration (scikit-learn example output).

Figure 1 illustrates Receiver Operating Characteristic (ROC) curves, which compare the true positive rate (sensitivity) against the false positive rate across classification thresholds. The Area Under the Curve (AUC) summarizes discriminative power: values closer to 1 indicate better separation between default and non-default outcomes. ROC/AUC is widely used in credit risk because it is threshold-independent.



However, calibration checks and business-driven cutoffs should be used in conjunction with it.

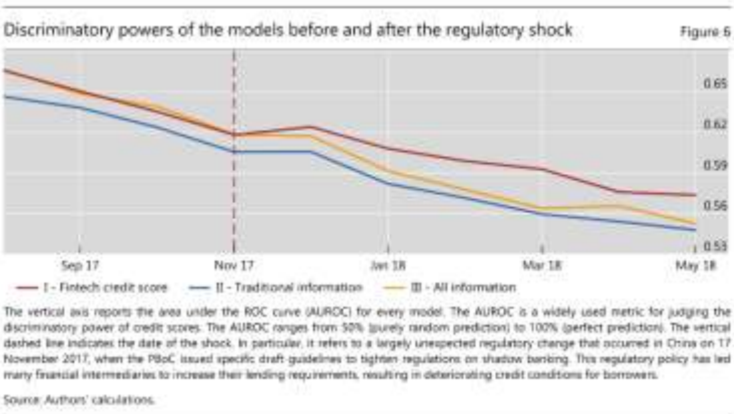


Figure 2. Precision–Recall curve illustration (scikit-learn example output).

Figure 2 shows Precision–Recall curves, which are especially informative when defaults are rare (class imbalance). Precision measures the proportion of correctly predicted defaults, while recall measures the proportion of correctly detected defaults. Because it focuses on performance on the positive (default) class, precision–recall analysis can be more relevant to decisions for highly imbalanced loan portfolios than ROC.

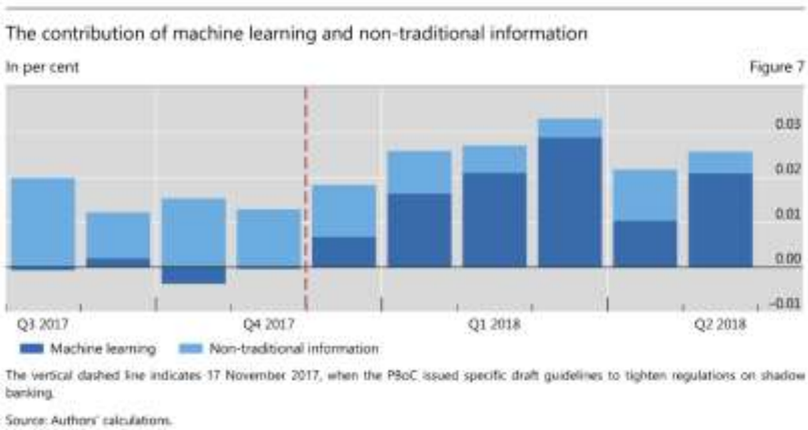


Figure 3. Discriminatory power gap by model family (BIS Working Paper No 834) [4].

Differences in discriminatory power between model types and information sets are depicted in Figure 3, which comes from BIS Working Paper No. 834. It highlights how ML methods can provide improved separation between outcomes compared to simpler approaches, particularly when richer (non-traditional) information is included. Practice should focus on the fact that predictive gains are influenced not only by the algorithm but also by the quantity and quality of input data.

Figure 4. Contribution of machine learning and non-traditional information (BIS Working Paper No 834) [4].

Figure 4 breaks down the advantages of using non-traditional information versus machine learning techniques. The figure suggests that algorithm selection and feature set expansion can both be important. In operational credit scoring, this supports a two-track strategy: improve modeling techniques while also investing in reliable, permissioned data pipelines that enhance borrower signals (e.g., verified transactions, repayment history, and behavioral indicators).

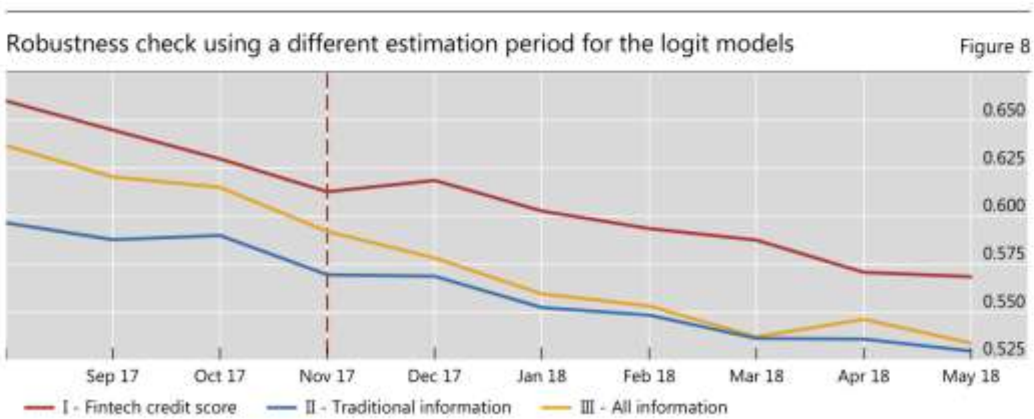


Figure 5. Robustness check (AUROC) under an economic/regulatory shock (BIS Working Paper No 834) [4].

Figure 5 depicts AUROC over time for various shock-related information and model configurations. It demonstrates that models with a wider range of information sources can be more resilient and that model performance can vary under stress. For banks, this underscores the need for stress-testing, drift monitoring, and contingency plans for retraining or recalibrating models when macro conditions shift.

Building and Operating ML Credit Risk Models

Data collection and preprocessing, feature engineering and selection, model training with cross-validation, evaluation on holdout and out-of-time samples, calibration and threshold setting aligned with business costs, documentation and independent validation, deployment with monitoring, and periodic review, retraining, and governance approvals are all typical components of a practical ML credit risk workflow. For challenger testing, segmentation, or decision support, banks frequently use hybrid strategies that combine ML models with interpretable baseline models (such as logistic regression scorecards). This satisfies regulatory demands for transparency while permitting innovation when supported by controls and evidence.

Challenges and Limitations

Data quality and bias. Data that is representative, complete, and accurate are necessary for ML performance. Biased training data can produce discriminatory outcomes, so institutions should conduct fairness testing, bias audits, and continuous monitoring.

Interpretability and governance. Customers and regulators may find it challenging to comprehend complex models. Overrides and exceptions should be clearly accountable for, and explainability tools for model risk management should be documented.

Regulatory and operational requirements. Supervisory expectations (such as Basel frameworks) and consumer protection requirements must be met by credit risk models. Banks need operationally secure data pipelines, model drift monitoring, and incident response procedures.

Discussion and Implications for Emerging Markets (Uzbekistan)

By assessing borrowers with limited traditional credit history, ML-based credit scoring can expand financial inclusion in developing markets. Alternative data can be of assistance, but it must be used responsibly: it should be legally collected, based on consent, and protected by national privacy laws. Regulators can support safe experimentation through supervisory guidance and, where available, regulatory sandboxes. For banks, the key is balancing predictive performance with transparency, fairness, and robust governance.

Conclusion

Due to its potential for improved prediction accuracy, risk segmentation with greater granularity, and operational efficiency, machine learning has emerged as an essential component of credit risk assessment. Model performance can be improved using ML techniques and a wider range of data sources, including under stressful conditions, according to official publications and research. However, strong data governance, explainability, fairness controls, independent validation, and human oversight are necessary for responsible adoption. In emerging markets, ML can support financial inclusion when implemented with appropriate regulatory compliance and ethical safeguards.

REFERENCES:

1. Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards*: A Revised Framework (Comprehensive Version) (Basel II). Bank for International Settlements.

2. Basel Committee on Banking Supervision. (2011). Basel III: A global regulatory framework for more resilient banks and banking systems (rev. June 2011). Bank for International Settlements.

3. European Banking Authority. (2020). Final Report – *Guidelines on loan origination and monitoring* (EBA/GL/2020/06).

4. Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. (2019). How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm (BIS Working Paper No 834). Bank for International Settlements.

5. 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.

6. 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.

7. Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Klamargias, A. (2018). A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting. In Ninth IFC Conference on "Are post-crisis statistical initiatives completed?" (IFC Bulletin No 49). Bank for International Settlements.

8. scikit-learn developers. (n.d.). *Multiclass Receiver Operating Characteristic* (ROC) – example. scikit-learn documentation.

9. scikit-learn developers. (n.d.). Precision–Recall – example. scikit-learn documentation.

10. Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics, 24(1), 44–65.

11. Friedman, J. H. (2001). Greedy function approximation: *A gradient boosting machine*. Annals of Statistics, 29(5), 1189–1232.