FROM STATIC THRESHOLDS TO PREDICTIVE SCORING: AN ARTIFICIAL INTELLIGENCE MECHANISM FOR CREDIT RISK ASSESSMENT AND FORECASTING IN COMMERCIAL BANKS

Qurbonov Ilhom Jumaqulovich
Agency for Governance Efficiency
e-mail: ilhomkurbanov49@gmail.com

Annotatsiya. *Maqolada tijorat banklarida kredit riskini baholash va prognozlash mexanizmini sun'iy intellekt (mashinali o'qitish) usullari asosida takomillashtirish masalalari ko'rib chiqilgan. O'zbekiston bank tizimida 2021–2025-yillarda muammoli kreditlar hajmi 3,7 baravar oshgani va jismoniy shaxslar kreditlari ulushi ortib borayotgani asoslab berilgan. Xalqaro tadqiqotlar tahlili asosida ma'lumotlar, skoring, prognozlash va boshqaruv bloklaridan iborat to'rt darajali mexanizm taklif etilgan. Ushbu mexanizm qarz oluvchining defolt ehtimolini baholash, muammoli kreditlar ulushini segmentlar bo'yicha prognozlash va portfelning 2, 3 va 5 foizli stress-ssenariy chegaralariga yaqinlashishini oldindan aniqlash imkonini beradi.*

Kalit so'zlar: *kredit riski, sun'iy intellekt, mashinali o'qitish, muammoli kreditlar, XGBoost, SHAP, stress-test.*

Аннотация. *В статье рассматриваются вопросы совершенствования механизма оценки и прогнозирования кредитных рисков в коммерческих банках на основе методов искусственного интеллекта (машинного обучения). Показано, что в 2021–2025 годах объём проблемных кредитов в банковской системе Узбекистана вырос в 3,7 раза, а доля кредитов физическим лицам увеличивается. На основе анализа международных исследований предложен четырёхуровневый механизм, включающий блоки данных, скоринга, прогнозирования и управления, который позволяет оценивать вероятность дефолта заёмщиков, прогнозировать долю проблемных кредитов по сегментам и заблаговременно выявлять приближение портфеля к пороговым значениям стресс-сценариев в 2, 3 и 5 %.*

Ключевые слова: *кредитный риск, искусственный интеллект, машинное обучение, проблемные кредиты, XGBoost, SHAP, стресс-тестирование.*

Abstract. *The article examines how the credit risk assessment and forecasting mechanism of commercial banks can be improved using artificial intelligence*

(machine learning) methods. It shows that non-performing loans in Uzbekistan's banking system grew 3.7 times in 2021–2025 while the share of retail loans kept rising. Drawing on international evidence, the paper proposes a four-layer mechanism comprising data, scoring, forecasting and governance blocks. It enables banks to estimate borrowers' default risk, forecast the non-performing loan ratio by portfolio segment and detect early when a portfolio is approaching the 2%, 3% or 5% stress-scenario thresholds, so that provisions and limits can be adjusted in time.

Keywords: *credit risk, artificial intelligence, machine learning, non-performing loans, XGBoost, SHAP, stress testing.*

Introduction. Credit risk is the dominant risk of commercial banks: as of 1 January 2025, loans accounted for 66.9% of the assets of Uzbekistan's banking system [1]. Between 2021 and 2025 the loan portfolio almost doubled, from UZS 277.0 trillion to UZS 533.1 trillion, while non-performing loans (NPLs) grew 3.7 times, from UZS 5.8 trillion to UZS 21.2 trillion, and their share rose from 2.1% to 4.0% (Table 1) [2]. By 1 April 2025 NPLs had reached UZS 24.3 trillion, or 4.5% of the portfolio [1]. At the same time, the share of loans to individuals increased from 19.8% in 2020 to 33.0% in 2024 [2]. Retail lending consists of a large number of small, similar loans accompanied by rich repayment and transaction data, which is precisely the setting in which machine-learning models perform best.

Current practice relies on regulatory exposure limits, loan classification with provisioning, and stress tests with fixed thresholds. These instruments are largely backward-looking: they record deterioration after repayment discipline has already weakened. The purpose of this paper is to substantiate a mechanism that uses artificial intelligence (AI) to assess and forecast credit risk in commercial banks and to embed it in the existing risk-management framework.

Literature review. Statistical credit scoring began with Altman's discriminant analysis of corporate failure [3]; logistic regression later became the industry standard owing to its transparency. Machine learning shifted attention to non-linear interactions among many predictors: Khandani, Kim and Lo showed that models built on consumer transaction and credit bureau data can materially improve forecasts of consumer credit delinquencies [4]. Lessmann et al. benchmarked more than forty classifiers on several credit datasets and found that ensemble methods generally outperform individual classifiers and the logistic-regression benchmark [5]. Gradient boosting, notably XGBoost[20] [6], has since become a leading approach

[20] XGBoost (eXtreme Gradient Boosting) is a machine-learning algorithm that builds a prediction from many small decision trees added one after another, where each new tree is trained to correct the errors of the previous ones. It

for tabular credit data, while SHAP[21] values [7] make individual predictions explainable by attributing them to input features.

A systematic review of 63 studies by Ayari et al. confirms the advantage of ensemble and deep-learning models but points to persistent gaps: inconsistent evaluation metrics, class imbalance, limited use of macroeconomic variables and insufficient treatment of algorithmic bias [8]. The Financial Stability Board adds a supervisory perspective: limited explainability impedes model validation, opaque training data complicates data-quality assessment, and similar models trained on similar data may produce correlated predictions that amplify stress [9]. Domestic studies, including the author's own, concentrate on portfolio structure, NPL dynamics and stress-scenario classification [2]; the use of AI-based forecasting within the national regulatory framework remains at an early stage.

Research methodology. The study is conceptual and analytical. It combines (i) descriptive analysis of statistics of Central Bank of Uzbekistan on the banking system for 2021–2025, (ii) a structured review of recent international evidence on machine-learning credit scoring and supervisory guidance, and (iii) the design of a mechanism that links model outputs to a bank's limits, provisioning and stress-testing procedures. The mechanism is proposed for validation on bank-level loan data, no claims are made about its performance on Uzbek portfolios.

Analysis and results. Table 1 shows that credit quality deteriorated faster than the portfolio grew: the loan portfolio expanded 1.9 times over 2021–2025, whereas NPLs grew 3.7 times. The NPL ratio exceeded the 3% "stressed" threshold proposed in the author's earlier scenario classification (2% stable, 3% stressed, 5% crisis) in every year after 2021 and reached 5.2% in 2022 [2]. Such thresholds describe the current state of the portfolio but do not anticipate its movement; a forecasting layer is needed to indicate, months in advance, that a segment is approaching the next scenario class.

handles non-linear relationships among predictors and performs strongly on tabular data such as loan application and repayment records

[21] SHAP (SHapley Additive exPlanations) is a method for explaining the output of a machine-learning model by showing how much each input variable, such as income, loan amount or repayment history, contributed to an individual prediction. It is based on Shapley values from cooperative game theory and allows a credit decision to be traced back to its main drivers

Table 1. Loan portfolio and non-performing loans of the banking system (as of 1 January)

Year	Total loans, UZS bn	NPLs, UZS bn	NPL share, %
2021	276 975	5 785	2.1
2022	326 386	16 974	5.2
2023	390 049	13 992	3.6
2024	471 406	16 621	3.5
2025	533 121	21 185	4.0

Source: compiled by the author based on Central Bank data [2].

International evidence indicates a substantial accuracy gain from AI methods. In a recent comparison on the UCI German credit dataset [22](1,000 applications), the area under the ROC curve (AUC)[23] was 0.76 for logistic regression, 0.79 for a decision tree, 0.87 for a neural network and 0.89 for XGBoost; under a simulated 20% income decline, the AUC of XGBoost fell only to 0.83 while that of logistic regression fell to 0.68 (Table 2) [10]. These figures come from a small public dataset and a simulated shock, so they indicate the direction of the effect rather than the magnitude expected on domestic portfolios.

Table 2. Discriminatory power of credit scoring models (AUC)

Model	Baseline	Income shock (–20%)
Logistic regression	0.76	0.68
Decision tree	0.79	–
Neural network (MLP)	0.87	–
XGBoost	0.89	0.83

Source: compiled by the author based on Vakrani et al., "Evaluating AI-driven credit scoring models versus traditional statistical techniques" (2026) [10].

On this basis, a four-layer mechanism is proposed (Table 3). Its central principle is that AI complements rather than replaces the established scorecard: the logistic model is retained as a transparent benchmark, and a gradient-boosting model competes with it as a challenger and is adopted only if it outperforms the benchmark

[22] UCI German Credit dataset (Statlog German Credit Data) is a publicly available benchmark dataset from the UCI Machine Learning Repository, originally compiled by Hans Hofmann of the University of Hamburg. It contains 1,000 loan applications described by 20 attributes.

[23] The area under the ROC curve (AUC) is a measure of how well a model distinguishes good borrowers from bad ones across all possible decision thresholds. It equals the probability that the model assigns a higher risk score to a randomly chosen defaulting borrower than to a randomly chosen non-defaulting one. Its value ranges from 0.5 (no better than random guessing) to 1.0 (perfect discrimination).

on an out-of-time sample. The probability of default (PD) estimated by the model, together with loss given default (LGD) and exposure at default (EAD), yields expected loss and supports timely formation of provisions, which the author's previous study identified as a necessary condition of bank stability [2].

Table 3. Proposed AI-based credit risk assessment and forecasting mechanism

Layer	Content	Control indicators
1. Data	Repayment history, credit bureau data, transactional and behavioural data; data-quality rules and lineage	Completeness, timeliness, consistency
2. Scoring	Logistic scorecard as benchmark versus gradient boosting as challenger; 12-month PD, LGD and EAD	AUC, Gini, KS on out-of-time sample
3. Forecasting and stress testing	Segment-level NPL forecasts (retail, industry, agriculture, other); macro scenarios; mapping of the projected NPL ratio to the 2%, 3% and 5% scenario classes [2]	Back-testing of forecasts; early-warning triggers
4. Governance	SHAP explanations for credit officers and clients; drift monitoring; bias testing; independent validation; human override	PSI (alert levels 0.10 and 0.25 are common in practice); approval-rate parity; override rate

Source: developed by the author.

Given the rise of the retail share up to 33%, the mechanism should first be piloted on the retail portfolio, where the volume of homogeneous data is the greatest. Corporate lending, with smaller samples and greater reliance on financial statements and industry factors, is better served by hybrid models that combine statistical output with expert judgment. The governance layer addresses the risks identified by the FSB [9]: explainability tools and independent validation reduce model risk, while monitoring of input drift and of third-party model dependence limits correlated errors across institutions.

Conclusion. First, the rapid growth of NPLs against a portfolio dominated by loans and an expanding retail segment makes static, backward-looking risk assessment insufficient. Second, international evidence supports the use of gradient-boosting and related AI models as challengers to logistic scorecards, provided that explainability, bias testing and independent validation are built in; the four-layer mechanism proposed here links model outputs to provisioning and to the 2%, 3% and 5% stress-scenario classes. Third, the next step is empirical validation on bank-

level retail loan data using out-of-time samples and back-testing, in cooperation with the Central Bank and commercial banks, before wider adoption.

References

1. Central Bank of the Republic of Uzbekistan. Banking sector statistics. – Tashkent. – https://cbu.uz/oz/statistics/bankstats/

2. Qurbonov I.J. Tijorat banklarida kredit riskini boshqarish va muammoli kreditlar muammosi: tahlil va yechimlar // “Yashil” iqtisodiyot va taraqqiyot. – 2026. – № 7, I-qism. – B. 244–247.

3. Altman E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy // The Journal of Finance. – 1968. – Vol. 23, No. 4. – P. 589–609.

4. Khandani A.E., Kim A.J., Lo A.W. Consumer credit-risk models via machine-learning algorithms // Journal of Banking & Finance. – 2010. – Vol. 34, No. 11. – P. 2767–2787.

5. Lessmann S., Baesens B., Seow H.-V., Thomas L.C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research // European Journal of Operational Research. – 2015. – Vol. 247, No. 1. – P. 124–136.

6. Chen T., Guestrin C. XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – New York: ACM, 2016. – P. 785–794.

7. Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems 30. – 2017. – P. 4765–4774.

8. Ayari H., Guetari R., Kraïem N. Machine learning powered financial credit scoring: a systematic literature review // Artificial Intelligence Review. – 2025. – Vol. 59, No. 1. – https://doi.org/10.1007/s10462-025-11416-2

9. Financial Stability Board. The Financial Stability Implications of Artificial Intelligence. – Basel: FSB, 14 November 2024. – https://www.fsb.org/uploads/P14112024.pdf

10. Vakrani et al. Evaluating AI-driven credit scoring models versus traditional statistical techniques // Discover Artificial Intelligence. – 2026. – Vol. 6. – https://doi.org/10.1007/s44163-025-00772-1