



Article

A Next-Generation Integrated Data-Driven Framework for Corporate Creditworthiness Assessment Using Financial Technologies, Artificial Intelligence, and Government Data

Bekhzod Yusupov¹

1. Independent Researcher, Tashkent State University of Economics, Uzbekistan

* Correspondence: bekzod130@gmail.com

Citation: Yusupov B. A Next-Generation Integrated Data-Driven Framework for Corporate Creditworthiness Assessment Using Financial Technologies, Artificial Intelligence, and Government Data. Journal of Economics and Business Management 2026, 9(7), 211-220.

Received: 10th Apr 2026

Revised: 11th May 2026

Accepted: 11th Jun 2026

Published: 06th July 2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>)

Abstract: This article develops and empirically tests an integrated data-driven framework for corporate creditworthiness assessment. The proposed framework, named FinTech-Based Corporate Credit Assessment Framework (FT-CCAF), combines internal bank data, credit bureau records, tax authority information, pledge registry data, sectoral benchmarks, transactional cash-flow indicators and explainable artificial intelligence. The framework is designed for emerging and bank-based financial systems where corporate lending still relies heavily on documents supplied by the borrower and on relationship-based judgement. The empirical base is a pilot implementation in a commercial bank in Uzbekistan, used here not as a local case description, but as a test environment for a transferable methodology. The study applies a pre/post quasi-experimental design, two-proportion tests, chi-square statistics and comparison of analytical preparation time. The results indicate that integrated official data and automated financial analytics can materially improve early risk detection, reduce calculation errors and shorten the corporate lending cycle while preserving the role of underwriters and credit committees. The proposed model can be adapted by banks in other emerging markets if equivalent public registries, credit bureau records and digital consent mechanisms are available.

Keywords: corporate creditworthiness, credit risk, FinTech, artificial intelligence, government data, credit bureau, alternative data, explainable AI, underwriting, Uzbekistan.

INTRODUCTION

Corporate credit assessment is still shaped by a simple but persistent problem: lenders rarely observe the full economic life of a borrowing firm. Financial statements may arrive late, cash flows may be split across several banks, collateral may already be pledged elsewhere, and tax discipline may be visible only through public or semi-public registries. This creates information asymmetry and, in markets with imperfect information, can lead to credit rationing or mispriced risk [1].

Traditional relationship banking partially reduces this gap because repeated interaction allows a lender to accumulate soft information. Yet relationship data are local, sometimes subjective and often locked inside one institution. When firms use several banks, a single lender sees only a part of the borrower's cash-flow cycle. The problem is sharper in emerging banking markets, where corporate reporting practices and data exchange infrastructures are still developing [2].

The international literature has already moved from a narrow view of collateral and balance-sheet ratios to a broader understanding of SME and corporate finance. Berger and Udell describe SME finance as a set of lending technologies, not a single technique, and this point is useful for corporate credit scoring as well. Modern corporate underwriting needs a layered design: verified official data, internal bank data, transaction analytics, statistical modelling and human judgement should work together rather than replace each other [3].

Credit reporting systems and public registries can reduce the blind spots of banks. The World Bank treats credit information sharing as part of financial infrastructure, because it improves lenders' ability to assess borrower quality and discipline repayment behavior. Cross-country evidence also links stronger credit information institutions with deeper private credit markets. The unresolved issue is methodological. Banks need a practical architecture that converts fragmented official data into a coherent and explainable risk profile [4].

This article proposes such an architecture. The FT-CCAF framework integrates four ideas. First, corporate creditworthiness is assessed as a data completeness problem before it becomes a scoring problem. Second, official and semi-official sources are organized into comparable risk blocks. Third, AI is used as a decision-support layer, not as an autonomous substitute for the underwriter. Fourth, the credit process is divided into a fast digital pre-screen and a deeper expert stage. The empirical setting is Uzbekistan, but the aim is broader. The proposed approach can be tested in any banking system with credit bureau data, tax or fiscal data, pledge registry data and digital transaction records [5].

Research aim and contribution

The aim of the study is to develop and validate a FinTech-oriented model for integrated corporate creditworthiness assessment. The contribution is threefold. The first contribution is conceptual: the article formalizes corporate credit assessment as a data-integration and decision-governance problem. The second contribution is methodological: it proposes a modular framework, SFA-Engine and DRSEP, that may be implemented without replacing a bank's core system. The third contribution is empirical: the model is tested on real implementation data from an emerging banking market [6].

2. Literature background

The literature on bank lending distinguishes hard information from soft information. Hard information is numerical, transferable and easier to verify; soft information is contextual and often held by loan officers. Liberti and Petersen argue that the organizational value of information depends not only on its content, but also on the distance between the person collecting it and the person making the decision. In corporate lending this matters because a digital model can standardize hard information, while underwriters remain responsible for interpretation and exceptions [7].

Recent FinTech studies add another dimension. Digital footprints, platform data and transaction traces may contain useful signals about borrower quality. Alternative data and machine learning have also been linked with broader access to credit, although the benefits depend on data quality, model governance and fairness controls. Evidence from non-traditional data in credit scoring suggests that new sources can improve predictive power, but they do not remove the need for supervision and validation [8].

Machine learning studies in credit scoring repeatedly find that ensemble methods, including gradient boosting, often outperform classical statistical models on structured datasets. Yet better prediction is not the whole objective in bank credit decisions. Credit models influence access to finance and can create unequal outcomes if they learn from biased data or if their decisions cannot be audited. For this reason, explainability techniques such as SHAP values are relevant to corporate lending, where a credit committee needs a defensible rationale for approving, restructuring or rejecting a loan [9].

The proposed framework follows this balanced view. It does not treat AI as a final decision-maker. It treats AI as an analytical layer that improves the completeness, consistency and early-warning capacity of credit assessment. This position is close to modern supervisory thinking. Risk data aggregation standards emphasize accuracy, completeness, timeliness and adaptability of bank risk data. Supervisory discussions of FinTech also stress that innovation should be connected with governance, accountability and risk controls [10].

3. Materials and Methods

3.1. Empirical setting

The empirical base is a pilot implementation in a commercial bank in Uzbekistan. The setting is useful for international research because it has the features found in many emerging banking systems: expanding corporate lending, fragmented data flows, reliance on borrower-submitted documents, developing credit information infrastructure and growing demand for faster underwriting. The IMF's 2025 financial stability assessment of Uzbekistan also places the country in a broader reform context where banking-sector modernization and risk management are central issues.

The implementation package contains three instruments. FT-CCAF is the integrated methodology. SFA-Engine is the automated analytical module for collecting, consolidating and calculating financial indicators. DRSEP is the two-stage digital protocol for pre-screening and expert evaluation. The bank implementation materials describe the use of internal banking data, credit bureau data, tax authority data, pledge registry data and sectoral benchmarks.

Table 1. Data domains in the FT-CCAF framework

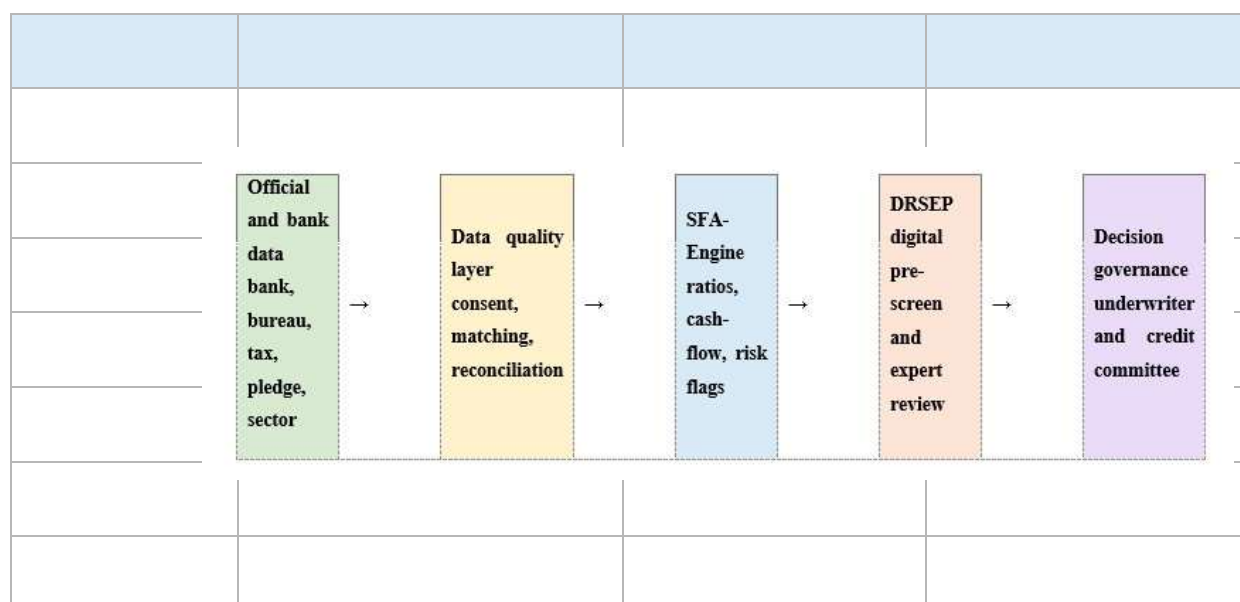
Data domain	Typical source	Main risk signal	International equivalent
Internal bank records	Core banking, loan administration, current accounts	Existing exposure, repayment discipline, account turnover inside the lender	Core banking systems in any commercial bank
Multi-bank cash-flow data	Client consent, credit information infrastructure, payment systems	Real turnover, seasonality, concentration of inflows, hidden liquidity stress	Open banking, account aggregation, payment data
Credit bureau records	KATM or credit bureau	Credit history, overdue debt, total debt burden, exposure in other banks	Public/private credit bureau or credit registry
Tax authority data	Fiscal records, VAT/sales data, tax arrears	Revenue consistency, fiscal discipline, hidden liabilities	Tax administration or e-invoicing data
Pledge registry data	Unified collateral/pledge registry	Existing encumbrances, priority of claims, legal availability of collateral	Movable collateral registry or land/asset registry

Data domain	Typical source	Main risk signal	International equivalent
Sectoral benchmarks	Statistical agency, industry databases	Profitability, margin, turnover and debt norms by sector	National statistics, commercial databases, industry associations

3.2. Model architecture

The framework transforms raw data into a single corporate risk profile. Each borrower is represented by six risk blocks: financial ratios, transaction behavior, credit-bureau profile, tax discipline, collateral/legal status and sectoral position. These blocks are not treated as mechanically equal. Their weights should be calibrated by historical default or problem-loan data when sufficient observations exist. When data are limited, an expert-constrained weighting scheme may be used at the first stage, with later recalibration.

Formally, the integrated corporate credit score can be written as $CCS_i = f(F_i, T_i, C_i, G_i, S_i, E_i)$, where F_i is the financial statement and ratio block, T_i is the transaction and cash-flow block, C_i is the credit bureau block, G_i is the government-data block, S_i



is the security and legal encumbrance block, and E_i is the sectoral benchmark block. The function f may be a transparent scorecard, logistic regression, gradient boosting model or a hybrid ensemble with monotonic constraints. The preferred model depends on the amount of historical data, regulatory expectations and internal validation capacity.

Figure 1. Native editable architecture of FT-CCAF

Table 2. Risk blocks, variables and analytical outputs

Risk block	Selected variables	Analytical output	Use in decision
Financial capacity	DSCR, ICR, Debt/EBITDA, current ratio, quick ratio, receivables turnover	Repayment capacity and leverage pressure	Limit, term, covenant and pricing calibration

Risk block	Selected variables	Analytical output	Use in decision
Cash-flow reliability	Monthly inflows, volatility, concentration of counterparties, seasonality, cash conversion cycle	Stability of operating cash flow	Stress testing and repayment schedule design
Credit history	Overdue days, number of active loans, total debt burden, restructurings	System-wide debt discipline	Eligibility and early-warning flags
Tax and fiscal discipline	Declared revenue, tax arrears, discrepancy between account turnover and tax base	Revenue verification and compliance risk	Request for clarification or enhanced due diligence
Collateral/legal status	Collateral value, pledge priority, encumbrances, legal disputes where available	Recoverability and legal enforceability	Collateral haircut and approval conditions
Sectoral position	Sector margin, benchmark leverage, turnover norms, macro sensitivity	Relative performance in peer group	Sector adjustment to rating and limit
Explainability layer	Top variables, adverse flags, SHAP-style local explanations	Decision rationale	Credit committee documentation and audit trail

3.3. Digital Risk Screening and Evaluation Protocol

DRSEP is a two-stage process. The first stage is a digital pre-screen. It quickly checks minimum criteria, data completeness, bureau red flags, tax arrears, encumbrance conflicts and basic repayment capacity. The second stage is a deeper expert evaluation. It includes business-model analysis, stress scenarios, collateral assessment and final structuring. This division reduces unnecessary committee workload because weak cases are identified before a full manual package is prepared.

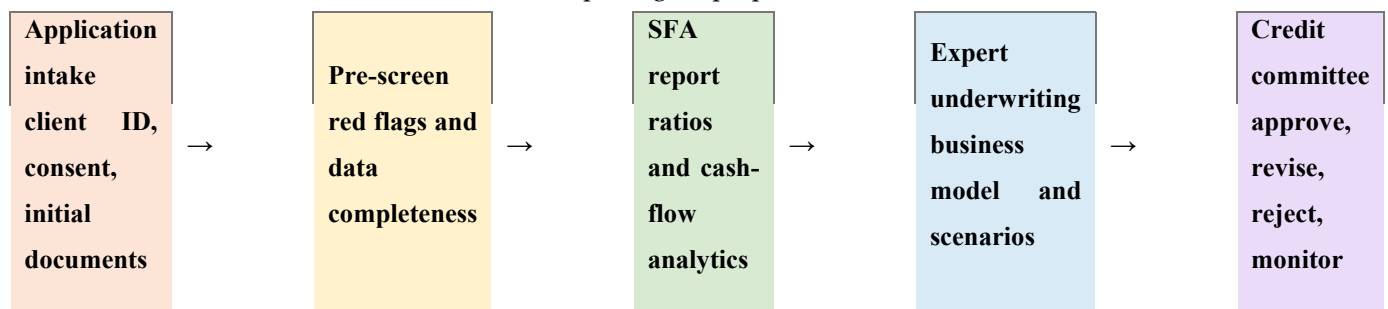


Figure 2. Native editable DRSEP decision flow

3.4. Empirical design and statistical procedures

The empirical evaluation uses a pre/post quasi-experimental design. It compares credit-process indicators before and after the implementation of the FT-CCAF, SFA-Engine and DRSEP instruments. The available data are implementation records rather than a randomized experiment, so causal interpretation should remain cautious. Still, the same bank process, the same type of borrowers and the same underwriting environment make the before/after comparison useful for assessing practical feasibility.

Table 3. Empirical design

Evaluation block	Before implementation	After implementation	Test or comparison
Early identification of elevated risk	41 high-risk applications out of 560	214 high-risk applications out of 684	Two-proportion comparison; Pearson chi-square
Financial calculation errors	39 errors out of 300 reviewed cases	7 errors out of 320 reviewed cases	Two-proportion comparison; Pearson chi-square
Analytical preparation time	6.7±1.2 hours	2.0±0.8 hours	Welch two-sample t-test and standardized effect size
Cycle time of application review	5-7 days	2-3 days	Operational comparison using reported time bands
Unjustified forwarding to committee	Baseline level before DRSEP	41% lower after DRSEP	Relative operational reduction

4. Results

4.1. Completeness of the borrower risk profile

The first result is methodological rather than purely statistical. The integrated model changes the unit of analysis. In the traditional approach, a bank often evaluates the documents supplied by the borrower and the turnover visible in the lender's own accounts. In FT-CCAF, the unit becomes a consolidated risk profile built from several independent sources. This reduces dependence on one self-reported data stream and gives the underwriter a more stable basis for judgement [11].

The same logic also improves internal discipline. When ratios are computed from reconciled data, differences between financial statements, tax records and account turnover become visible earlier. This is important in corporate lending because a weak borrower may still look acceptable if the analysis sees only part of its cash flow or ignores obligations outside the lending bank [12].

4.2. Early identification of elevated risk

The pre/post comparison indicates a large increase in early identification of elevated-risk applications. Before implementation, 41 of 560 applications were flagged at an early stage, equal to 7.3%. After implementation, 214 of 684 applications were flagged, equal to

31.3%. Recalculation from these counts gives a statistically significant difference, chi-square = 108.51, $p < 0.001$, and Cramer's $V = 0.295$ [13].

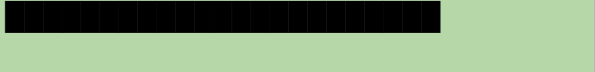
Indicator	Value	Native editable bar
Before implementation	7.3%	
After implementation	31.3%	

Figure 3. Native editable bar chart: early risk detection before and after FT-CCAF/DRSEP

4.3. Reduction of calculation errors and analytical time

The SFA-Engine had its strongest effect on repetitive analytical operations. Calculation errors declined from 13.0% to 2.2%, based on 39 errors among 300 reviewed cases before implementation and 7 among 320 cases after implementation. The chi-square statistic is 26.35 with $p < 0.001$ and Cramer's $V = 0.206$ [14].

The average preparation time for a credit analysis declined from 6.7 hours to 2.0 hours. This is not only a speed gain. It also changes the role of the underwriter. Less time is spent on mechanical calculation and repeated document reconciliation, while more time remains for judgement about business model, repayment schedule and transaction structure [15].

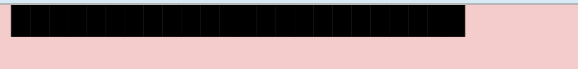
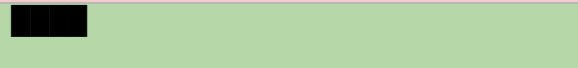
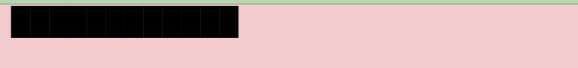
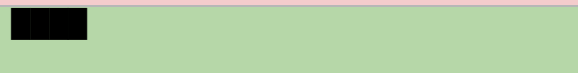
Indicator	Value	Native editable bar
Calculation errors before	13	
Calculation errors after	2.2	
Time before, hours	6.7	
Time after, hours	2	

Figure 4. Native editable bar chart: calculation errors and analytical preparation time

Table 4. Main empirical effects of FT-CCAF, SFA-Engine and DRSEP

Indicator	Before	After	Change	Interpretation
Early identification of elevated-risk applications	41/560 = 7.3%	214/684 = 31.3%	+24.0 percentage points	Earlier identification of weak or risky applications
Financial calculation errors	39/300 = 13.0%	7/320 = 2.2%	-10.8 percentage points	Lower manual-calculation and reconciliation risk

Indicator	Before	After	Change	Interpretation
Preparation time for analytical report	6.7±1.2 hours	2.0±0.8 hours	-70.1%	Faster preparation of standardized credit package
Application review cycle	5-7 days	2-3 days	About 58% reduction by midpoint	Faster movement from intake to decision
Unjustified applications forwarded to committee	Baseline	41% lower	Operational reduction	Lower burden on credit committee

4.4. International transferability

The model is not tied to the names of Uzbek institutions. KATM, the Tax Committee and the pledge registry are local expressions of more general data domains: credit information, fiscal data and collateral rights. This is why the framework can be transplanted to other countries [16]. The essential condition is not the same institutional label, but the presence of reliable data access, legal permission, data matching rules and governance around model outputs.

	High-income banking system	Emerging banking system	Low-data banking system
Credit bureau	Direct API / registry	Usually available	May require registry upgrade
Tax/fiscal data	E-invoicing or tax API	Phased access possible	Pilot with consent-based records
Pledge registry	Digital collateral registry	Often partially available	Priority reform area
Bank transactions	Open banking / internal data	Account aggregation pilot	Start with lender data
Sector benchmarks	Commercial databases	Statistical agency / industry data	Manual benchmark library

Figure 5. Native editable transferability matrix

5. Discussion

The empirical results are consistent with the theory of information asymmetry. A lender that sees more independent data sources can identify risk earlier. Still, the model should not be interpreted as a claim that more data automatically means better lending. More data may also create noise, privacy concerns and false confidence. The useful part of FT-CCAF is not the volume of information alone. It is the structured conversion of information into a transparent risk profile [17].

The framework also addresses a practical tension in banking. Credit departments often want speed, while risk departments want depth. A single-stage manual process struggles with both. DRSEP separates speed from depth. The pre-screen is designed to be fast and rules-based; the second stage remains analytical and expert-led. This design helps banks avoid the two extremes of purely bureaucratic underwriting and fully automated black-box lending [18].

From a model-risk perspective, the most sensitive issue is AI governance. Gradient boosting, random forests or neural models may increase predictive accuracy, but their use in corporate credit decisions requires validation, documentation, bias monitoring and explainability. For this reason, the proposed architecture allows banks to start with an explainable scorecard and add machine-learning components gradually. A bank with limited historical default data can still use the framework for data integration and standardized ratio calculation before moving to a full predictive model [19].

The Uzbekistan pilot offers a useful lesson for other emerging markets. A bank does not need a complete national open-banking system to begin modernization. It can start with a practical data map, controlled access to official registries, standardized templates and a two-stage workflow. The first gains may come from eliminating manual errors and repeated document requests rather than from sophisticated AI. This is sometimes less impressive than a fully automated scoring platform, but in real corporate lending it may be more important.

The framework is also relevant to financial inclusion. Small and medium-sized firms often have weak collateral and uneven accounting practices. Global development institutions treat SME finance as a major constraint for enterprise growth. Better data integration can help sound borrowers become more visible to banks. At the same time, safeguards are needed so that official data are used for proportional credit-risk analysis rather than excessive surveillance.

6. Conclusion

This article proposed FT-CCAF as a next-generation integrated framework for corporate creditworthiness assessment. The model combines official data, internal bank records, transaction analytics, automated financial ratios, explainable AI and human credit governance. The empirical implementation in an emerging banking market indicates meaningful operational and analytical improvements: earlier risk identification, fewer calculation errors and shorter analytical preparation time.

The main scientific value of the study lies in its transferable design. Uzbekistan is used as an empirical base, while the model itself is formulated in general data domains that can be adapted across countries. A bank in another jurisdiction can replace KATM with its credit bureau, the Uzbek tax data source with its own fiscal or e-invoicing records, and the local pledge registry with its collateral registry. The structure remains the same.

Future research should test the framework on a larger multi-bank dataset, connect it with observed default or restructuring outcomes, and compare transparent scorecards with machine-learning models under the same data conditions. Another useful direction is fairness testing. Corporate borrowers differ by size, sector, region and accounting capacity; a data-driven model should improve risk discipline without excluding viable firms merely because their data traces are thinner.

REFERENCES

- [1] A. N. Berger и G. F. Udell, «A More Complete Conceptual Framework for SME Finance», *Journal of Banking & Finance*, т. 30, вып. 11, сс. 2945–2966, 2006.

- [2] S. M. Lundberg и S.-I. Lee, «A Unified Approach to Interpreting Model Predictions», в Proc. 31st Conf. on Neural Information Processing Systems, 2017.
- [3] S. Lessmann, B. Baesens, H.-V. Seow, и L. C. Thomas, «Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring», *European Journal of Operational Research*, т. 247, вып. 1, сс. 124–136, 2015.
- [4] J. E. Stiglitz и A. Weiss, «Credit Rationing in Markets with Imperfect Information», *American Economic Review*, т. 71, вып. 3, сс. 393–410, 1981.
- [5] World Bank, «Credit Reporting Knowledge Guide». Washington, DC, 2019 г.
- [6] L. Gambacorta, Y. Huang, H. Qiu, и J. Wang, «How Do Machine Learning and Non-Traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm», Bank for International Settlements, BIS Working Papers 834, 2019.
- [7] B. Y. Yusupov, «Implementation Materials on FT-CCAF, SFA-Engine and DRSEP in Corporate Credit Assessment». 2026 г.
- [8] M. Brown, T. Jappelli, и M. Pagano, «Information Sharing and Credit: Firm-Level Evidence from Transition Countries», *Journal of Financial Intermediation*, т. 18, вып. 2, сс. 151–172, 2009.
- [9] J. M. Liberti и M. A. Petersen, «Information: Hard and Soft», *Review of Corporate Finance Studies*, т. 8, вып. 1, сс. 1–41, 2019.
- [10] T. Berg, V. Burg, A. Gombovic, и M. Puri, «On the Rise of FinTechs: Credit Scoring Using Digital Footprints», *Review of Financial Studies*, т. 33, вып. 7, сс. 2845–2897, 2020.
- [11] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, и A. Walther, «Predictably Unequal? The Effects of Machine Learning on Credit Markets», *Journal of Finance*, т. 77, вып. 1, сс. 5–47, 2022.
- [12] Basel Committee on Banking Supervision, «Principles for Effective Risk Data Aggregation and Risk Reporting». Basel, 2013 г.
- [13] S. Djankov, C. McLiesh, и A. Shleifer, «Private Credit in 129 Countries», *Journal of Financial Economics*, т. 84, вып. 2, сс. 299–329, 2007.
- [14] International Monetary Fund, «Republic of Uzbekistan: Financial System Stability Assessment». Washington, DC, 2025 г.
- [15] World Bank, «Small and Medium Enterprises (SMEs) Finance». Washington, DC, 2024 г.
- [16] Basel Committee on Banking Supervision, «Sound Practices: Implications of Fintech Developments for Banks and Bank Supervisors». Basel, 2018 г.
- [17] M. A. Petersen и R. G. Rajan, «The Benefits of Lending Relationships: Evidence from Small Business Data», *Journal of Finance*, т. 49, вып. 1, сс. 3–37, 1994.
- [18] J. Jagtiani и C. Lemieux, «The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform», *Financial Management*, т. 48, вып. 4, сс. 1009–1029, 2019.
- [19] T. Chen и C. Guestrin, «XGBoost: A Scalable Tree Boosting System», в Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016, сс. 785–794.