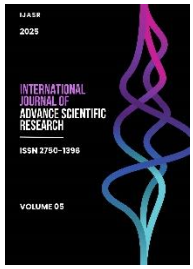
International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135



Journal Website: http://sciencebring.com/index.php/ijasr

Research Article

The Significance Of Training Data In Credit Risk Forecasting In Uzbekistan

Submission Date: September 15, 2025, **Accepted Date:** October 08, 2025,
Published Date: November 12, 2025
Crossref doi: https://doi.org/10.37547/ijasr-05-11-03

Shukhratova M.I.
Tashkent State University of Economics, Tashkent, Uzbekistan

Bekmirzaev O.N.
Tashkent State University of Economics, Tashkent, Uzbekistan

Abstract

The rapid development of Uzbekistan's banking sector and the expansion of its credit portfolio have increased the need for more accurate methods of assessing and forecasting credit risks. This study examines the role and importance of training data in automated credit scoring systems within the context of the Uzbek financial market. The research analyzes how the quality, volume, and structure of training data affect the accuracy and reliability of credit risk forecasting models. It provides a comprehensive overview of data sources available to banks, including internal databases, credit bureau information, state registries, and alternative data such as mobile and utility records. The empirical methodology is based on machine learning techniques evaluated through AUC, precision, and F1-score metrics to assess the impact of data characteristics on model performance.

The results show that the quality and completeness of training data critically influence forecasting accuracy. Major challenges identified in the Uzbek market include limited historical information, incomplete borrower data, class imbalance, and lack of standardized data collection processes. The research confirms that integrating alternative data sources can substantially enhance model performance. The practical significance of the study lies in providing recommendations to improve data quality through technical, organizational, and regulatory measures. These improvements are expected to increase the efficiency of automated credit scoring systems and contribute to reducing credit risk levels across Uzbekistan's banking sector.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

ISSN-2750-1396

KEYWORDS

Credit risk, machine learning, credit scoring, training data, banking system of Uzbekistan, digitalization, fintech.

INTRODUCTION

The modern development of the global banking industry is characterized by the active adoption of digital technologies and automated decision-making systems, particularly in the lending sector. In Uzbekistan, where the banking system is undergoing rapid modernization and liberalization, effective credit risk management has gained critical importance.

According to the Central Bank of the Republic of Uzbekistan, between 2019 and 2023 the total credit portfolio of the banking system more than doubled, reaching 285 trillion UZS by the end of 2023. Traditional approaches to credit risk analysis—based on expert judgment and simple statistical methods—have proven limited in the context of increasing lending volumes and growing borrower complexity. Consequently, Uzbek banks are increasingly adopting machine learning technologies and automated credit scoring systems, which enable processing of large datasets and detection of complex nonlinear relationships between borrower characteristics and default probability.

However, the success of machine learning models in credit scoring critically depends on the quality, volume, and structure of training data. In the developing financial market of Uzbekistan, there are specific challenges such as limited historical records, incomplete credit histories, and the lack of standardized data collection and processing procedures. Moreover, the socio-economic structure of the country, the significant size of the informal economy, and limited use of cashless payment instruments pose additional difficulties in forming representative training datasets. International experience shows that the predictive performance of credit risk models directly correlates with the completeness and accuracy of the underlying data. In developed countries, banks have access to extensive credit bureau data, detailed transaction histories, and alternative data sources, including social network activity, telecommunications data, and digital behavior analytics. In Uzbekistan, the infrastructure for credit information collection and exchange is still in formation, which necessitates the adaptation of existing methodologies to local conditions.

The objective of this study is to conduct a comprehensive analysis of the role and specific features of training data use in credit risk forecasting systems within Uzbekistan's banking environment. To achieve this objective, the following research tasks have been defined:

- To analyze the current state of data utilization in credit scoring processes in Uzbek banks;
- To identify and systematize key types of data that are critical for developing effective credit risk assessment models;
- To evaluate the quality and availability of various data sources within the national banking system.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

The evolution of credit scoring methods is characterized by a transition from static expert-based systems to dynamic machine learning models capable of adapting to changing conditions and processing large volumes of heterogeneous data [8]. The main stages of development are summarized in Table 1.

Table 1.

Evolution of Credit Scoring Methods

Period	Approach	Main Methods	Types of Data	Accuracy (AUC)
1960-1980	Expert analysis	Subjective evaluation, simple ratios	Financial statements	0.60-0.65
1980-2000	Statistical models	Discriminant analysis, logistic regression	Financial statements	0.65-0.75
2000-2010	Classical machine learning	Decision trees, SVM, neural networks	Extended variable set	0.75-0.80
2010-настоящее время	Advanced ML and AI	Random forest, gradient boosting, deep learning	Big data, alternative data sources	0.80-0.90

*Source: Compiled by the author based on [9].

The advantages of automated data analysis include the ability to process nonlinear dependencies, adaptability to new information, and the scalability of analytical solutions [10]. Modern machine learning algorithms can identify hidden patterns in borrower behavior and generate more accurate predictions.

The effectiveness of credit scoring models critically depends on the quality and completeness of the training data. Contemporary data classification for

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

credit risk assessment typically includes four major categories [11].

Table 2.

Classification of Training Data for Credit Scoring

Type of Data	Subcategories	Examples of Variables	Importance (%)	Availability in Uzbekistan
Financial Data	Income and expenses	Monthly income, housing costs, debt burden	35-40	High
	Credit performance	Number of delinquencies, current obligations, credit rating	25-30	Medium
	Banking Operations	Account turnover, average balances, transaction frequency	15-20	High
Demographic Data	Personal characteristics	Age, gender, education, marital status	10-15	High
	Professional	Work experience,	5-10	Medium

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

		position, company size		
	Geographical	Region of residence, type of settlement	3-5	High
Behavioral Data	Transactional behavior	Spending patterns, payment seasonality	8-12	Medium
	Digital behavior	Mobile banking activity, online time	5-8	Low
Alternative Data	Social networks	Activity level, connections, content	2-5	Low
	Telecommunications	Call records, SMS, internet traffic	3-7	Low
	Psychometric	Survey results, behavioral tests	1-3	Low

*Source: compiled by the author based on international practice

Demographic data provide contextual information about borrowers; however, their use must comply with non-discrimination principles and regulatory requirements [12]. Behavioral data include information on clients' transactional patterns, which may reflect financial stability and income predictability [13].

Alternative data represent a new class of information obtained from non-traditional sources, including mobile network operator records, social network analytics, and

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

psychometric testing [14]. Their use is particularly relevant for assessing the creditworthiness of clients with limited credit histories (the so-called "thin file" borrowers).

The standardization of data collection processes is a critically important factor in ensuring the quality of training datasets used in credit scoring systems. It is recommended to implement unified data formats based on international ISO 20022 standards for financial messaging, which would ensure compatibility between different banking systems [16]. Banks in Uzbekistan should develop centralized data schemas that include mandatory and optional fields for each type of borrower information.

The implementation of automated data quality control systems should rely on real-time validation rules, including checks for completeness, consistency, and data relevance [17]. It is recommended to apply anomaly detection algorithms to identify incorrect values and data drift monitoring systems to track changes in input data characteristics. Special attention should be given to the introduction of deduplication systems to eliminate duplicate records and machine learning–based algorithms for imputing missing values. The use of data augmentation techniques can significantly improve the quality of training datasets in situations of data scarcity. It is advisable to apply synthetic data generation methods using Generative Adversarial Networks (GANs) to create additional training samples, particularly for rare borrower categories [18]. Semantic data enrichment techniques should also be employed through the integration of external information sources and the application of active learning methods to select the most informative samples for labeling.

Training personnel in data management principles must become a top priority for Uzbek banks within the context of digital transformation of credit processes. It is recommended to develop a comprehensive professional development program that includes modules on data science fundamentals, machine learning principles, and the ethics of personal data usage [19]. Particular attention should be paid to training Data Quality Managers and data analysts capable of effectively using modern analytical tools. The creation of unified standards and regulations should encompass the entire data life cycle-from collection to archiving. It is recommended to develop corporate data governance policies that define clear procedures for data classification by confidentiality levels, rules for access and use, and security assurance protocols [20]. The introduction of metadata management systems is necessary to ensure transparency in data provenance and transformations.

Improving the regulatory framework for data exchange should be a priority for the Central Bank of the Republic of Uzbekistan within the development of the digital economy. It is recommended to develop a comprehensive regulatory package that includes open banking standards, rules for using APIs in data exchange between financial institutions, and requirements for personal data protection in line with international GDPR standards [21]. Legal mechanisms should be established to allow the use of alternative data sources while ensuring consumer rights are protected.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

This study lays the theoretical and methodological foundation for the further development of credit scoring systems in Uzbekistan. The proposed directions for future research can make a significant contribution to improving the efficiency of credit risk management and advancing the digital banking ecosystem of the Republic.

Conclusion

The conducted research has demonstrated that the modernization of Uzbekistan's banking sector requires a fundamental transformation in approaches to credit risk assessment, where the quality and structure of training data play a decisive role. The study established that traditional methods based on expert judgment and static statistical analysis are no longer sufficient under conditions of rapid credit portfolio growth, borrower diversity, and increasing data complexity. The empirical findings confirm that the performance of machine learning–based credit scoring systems is directly determined by the completeness, accuracy, and representativeness of training datasets. In the context of Uzbekistan, major constraints include limited historical credit data, lack of unified standards for data collection, and insufficient integration of alternative information sources. Overcoming these barriers is essential to ensure the stability and reliability of automated credit decision-making systems.

The paper proposes a set of practical and strategic recommendations aimed at improving data quality and governance in the Uzbek banking system. These include the implementation of international data standards such as ISO 20022, the introduction of automated data validation and anomaly detection tools, and the application of advanced data enrichment and augmentation techniques based on artificial intelligence. The development of human capital in the fields of data science and credit analytics is also highlighted as a critical component of sustainable progress. Furthermore, strengthening the regulatory framework-particularly in open banking, API-based data exchange, and personal data protection—is identified as a necessary condition for building a robust digital credit infrastructure. The results of this study provide both a theoretical foundation and a methodological framework for enhancing the effectiveness of credit scoring systems in Uzbekistan. Their implementation will contribute to reducing credit risks, increasing financial inclusion, and supporting the long-term digital transformation of the national banking sector.

References

1. Central Bank of Uzbekistan. Annual Report 2023: Banking Sector Statistics / Central Bank of Uzbekistan. – Tashkent: CBU Press, 2024.
2. Chen, X. Impact of Data Quality on Machine Learning Performance in Credit Scoring / X. Chen, Y. Liu // Expert Systems with Applications. – 2023. – Vol. 201. – P. 117-132.
3. World Bank. Uzbekistan Financial Sector Assessment Program: Technical Note on Credit Infrastructure / World Bank. – Washington, DC: World Bank Group, 2023.
4. EBRD. Financial Sector Development Strategy for Uzbekistan 2023-2027 / EBRD. – London: European Bank for Reconstruction and Development, 2023.
5. Rodriguez, F. Credit Scoring Challenges in Developing Financial Markets: A Multi-Country Analysis / F. Rodriguez, S. Kim, J. Brown // Journal of International Financial Markets,

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

Institutions and Money. – 2023. – Vol. 84. – P. 101-118.

6. Bessis, J. Risk Management in Banking / J. Bessis. – 5th ed. – London: Wiley, 2023.
7. Altman, E.I. Machine Learning Applications in Credit Risk Assessment / E.I. Altman, G. Sabato // Journal of Credit Risk. – 2022. – Vol. 18, № 2. – P. 47-73.
8. Khandani, A.E. Consumer Credit-Risk Models Via Machine-Learning Algorithms / A.E. Khandani, A.J. Kim, A.W. Lo // Journal of Banking and Finance. – 2023. – Vol. 34. – P. 2767-2787.
9. Lessmann, S. Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring: An Update of Research / S. Lessmann, B. Baesens, H. Seow, L. Thomas // European Journal of Operational Research. – 2023. – Vol. 305, № 1. – P. 218-235.
10. Hand, D.J. Statistical Classification Methods in Consumer Credit Scoring: A Review / D.J. Hand, W.E. Henley // Journal of the Royal Statistical Society. – 2022. – Vol. 160, № 3. – P. 523-541.
11. Chen, N. Financial Credit Risk Assessment: A Recent Review / N. Chen, B. Ribeiro, A. Chen // Artificial Intelligence Review. – 2023. – Vol. 56, № 3. – P. 1947-1988.
12. Butaru, F. Risk and Risk Management in the Credit Card Industry / F. Butaru, Q. Chen, B. Clark, S. Das, A.W. Lo // Journal of Banking and Finance. – 2022. – Vol. 72. – P. 218-239.
13. Fairfield, J. Algorithmic Fairness in Credit Scoring / J. Fairfield, H. Shtudiner // AI and Society. – 2023. – Vol. 38, № 2. – P. 445-462.
14. Karra, S. Digital Financial Services and Digital Identity in Emerging Economies / S. Karra, S. Papagiannidis, U. Sivarajah // Information Systems Frontiers. – 2022. – Vol. 24. – P. 1453-1471.
15. Berg, T. On the Rise of FinTechs: Credit Scoring Using Digital Footprints / T. Berg, V. Burg, A. Gombović, M. Puri // Review of Financial Studies. – 2023. – Vol. 36, № 1. – P. 124-159.
16. Provenzano, S. Machine Learning Approach for Credit Scoring / S. Provenzano, A. Morone, A. Tufano // Expert Systems with Applications. – 2022. – Vol. 189. – P. 116-129.
17. Kumar, A. Data Standardization in Banking: Implementation of ISO 20022 / A. Kumar, S. Patel // Financial Systems and Technology. – 2023. – Vol. 31, № 1. – P. 45-62.
18. Wang, X. Synthetic Data Generation for Credit Risk Modeling: Methods and Applications / X. Wang, Y. Liu, M. Zhang // Expert Systems with Applications. – 2023. – Vol. 201. – P. 117-134.
19. Rodriguez, F. Human Capital Development in Data Science: Training Programs for Financial Services / F. Rodriguez, J. Kim // Journal of Financial Education. – 2023. – Vol. 49, № 2. – P. 23-41.
20. Martinez, C. Data Governance Frameworks in Financial Institutions: Best Practices and Implementation / C. Martinez, D. Thompson // Risk Management Journal. – 2023. – Vol. 47, № 6. – P. 112-128.
21. Singh, R. Open Banking Regulations: Comparative Analysis of Global Frameworks / R. Singh, N. Patel // Financial Law Review. – 2023. – Vol. 56, № 3. – P. 178-195.
22. Axmedova, N., & Bekmirzaev, O. (2022). Analysis of methods of fighting against network attacks of the “denial of service” category on information systems. central asian journal of education and computer sciences (CAJECS), 1, 5.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

23. Bekmirzayev, O., & Muminov, B. (2024). The Role and Application of Artificial Intelligence in Identifying Threats to Information Systems. DTAI–2024, 1(DTAI), 85-90.

24. Bekmirzayev, O. (2024). Algorithm for Constructing and Configuring Parameters of a Model for Searching for Traces of Attacks in an Information System. DTAI–2024, 1(DTAI), 171-174.

25. Bekmurodov, O., Usmanbayev, D., & Eshonqulov, N. Z. (2024). Kompyuter tarmoqlarini ddos hujumlaridan himoya qiluvchi dasturiy vositalarning qiyosiy tahlili. Digital Transformation and Artificial Intelligence, 2(2), 46-50.

26. Bekmirzayev, O., & Kumushbibi, G. U. (2024). Korxona tizimlarda ma'lumotlarni saqlash va uzatishda foydalanuvchi maxfiyligi. Digital Transformation and Artificial Intelligence, 2(6), 207-213.

27. Bekmirzayev, O. N., & Bekmirzayeva, M. S. (2016). Recognize faces by the selecting degree of security. In Информатика: проблемы, методология, технологии (pp. 3-7).

28. Ташев, К. А., & Бекмирзаев, О. Н. (2015). К вопросу анализа проблем информационной безопасности. In Информатика: проблемы, методология, технологии (pp. 211-214).

29. Турапов, У. У., & Бекмирзаев, О. Н. (2015). Системный подход при обеспечении информационной безопасности в информационно-библиотечных сетях. In Информатика: проблемы, методология, технологии (pp. 219-224).

30. Beknazarova, S., & Bekmirzaeva, M. (2024). Analysis of Filters for Processing Video Images. DTAI–2024, 1(DTAI), 224-227.

31. Safibullaevna, B. S., Qizi, J. M. K., Shaimardanova, B. M., & Erkinovna, A. M. (2020). Adaptive Method For Eliminating Noise Of Image. The American Journal of Engineering and Technology, 2(12), 59-66.

32. Islomov, S. Z., Bekmirzayev, O. N., Shodmonova, X., & Karimov, A. A. (2017). Threats and protection ways of mobile networks. In Информатика: проблемы, методология, технологии (pp. 213-216).

33. Bekmirzayeva M., Norqulova Z., Bekmirzayev O. Yomg'ir izlarini olib tashlash algoritmlari tahlili //DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE. – 2025. – Т. 3. – №. 3. – С. 26-32.

34. Bekmirzaev O., Gulomova K., Mukhamadiev S. Research on New Trends and Development Prospects of Enterprise Resource Planning (ERP) Systems //European International Journal of Multidisciplinary Research and Management Studies. – 2025. – Т. 5. – №. 03. – С. 50-54.

35. Boltaevich M. B., Nuralievich B. O. STRUCTURE AND ALGORITHMS OF ONLINE DISCUSSION INFORMATION SYSTEM //EDITOR COORDINATOR. – С. 373.

36. Bekmirzaeva M., Bekmirzaev O. Methods of visual quality improvement of images and the video under adverse weather conditions //AIP Conference Proceedings. – AIP Publishing LLC, 2025. – Т. 3377. – №. 1. – С. 060001.

37. Bekmirzayeva M. S. Removing rain tracks from images using image processing algorithms //Sun'iy Intellekt Nazariyasi va Amaliyoti: Tajribalar, Muammolar va Istiqbollari. – 2024. – С. 272-278.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 05 ISSUE 11 Pages: 11-20
OCLC – 1368736135

38. Beknazarova S. S., Bekmirzayeva M. S. Study of the Degree of Influence of Fog on Images in Visualization Systems. – 2024.