American Research Journal of Computer Science and Information Technology

Volume 8, Issue 1, 12-17 Pages
Research Article | Open Access
ISSN (Online)- 2572-2921
DOI : 10.21694/2572-2921.25003

Application of Machine Learning for Risk Prediction in Credit Institutions

Di Feng

Director of Cyber Data Risk & Resilience Division, China/Canada (Montreal, Quebec).

ABSTRACT

The article examines modern financial systems, which are a multi-layered structure where data processing requires the use of analytical methods. The relevance of the topic is due to the need to develop algorithms that take into account the complexity of information and the volume of data. The main focus of the article is on the use of machine learning algorithms for managing credit risks and assessing their adaptation to financial activities.

The article discusses approaches aimed at predicting risks in credit institutions. The emphasis is on models adapted to the specifics of financial processes, taking into account atypical situations. This allows you to create tools that meet the requirements of the industry.

The use of algorithms increases the accuracy of forecasting borrowers' creditworthiness, detects fraud, and improves loan portfolio management processes. Methods related to the identification of relationships between data characteristics contribute to the development of more effective models. Such approaches demonstrate resilience to changing financial market conditions.

The conclusions of the article are addressed to specialists involved in risk management, developers of analytical solutions, and researchers in the field of economic applications of machine learning. The use of these tools helps to optimize the internal processes of credit institutions and increase their efficiency in working with data.

The use of machine learning algorithms opens up prospects for accurate analysis of credit risks. These technologies improve the quality of forecasting, strengthen the sustainability of organizations, and create conditions for successful adaptation to changes in the financial environment.

KEYWORDS: *machine learning, credit risks, financial sector, risk management, forecasting, data analysis.*

INTRODUCTION

In the context of economic instability and the globalization of financial markets, credit risk management remains a critical task for financial organizations. Traditional evaluation methods based on statistical analysis or expert approaches often demonstrate limited effectiveness when dealing with large volumes of complex data. This highlights the need for more accurate solutions capable of ensuring reliable forecasts.

Machine learning, a field within artificial intelligence, offers extensive possibilities for data analysis. Its algorithms enable the detection of hidden patterns, the analysis of time series, client base segmentation, and borrower behavior prediction. These technologies are employed to address tasks such as credit scoring, fraud detection, and credit portfolio management.

The growing volume of data in the financial sector necessitates the implementation of automated approaches that minimize the influence of subjective factors. The use of machine learning algorithms enhances organizations' adaptability to market changes and improves the quality of decision-making. Integrating these technologies into analytical processes strengthens the competitive position of credit institutions.

This article explores machine learning algorithms and their application in credit risk management, evaluating their practical significance. It examines the prospects for integrating these technologies into financial analytics to improve the efficiency of credit organizations.

MATERIALS AND METHODS

Contemporary approaches to credit risk assessment based on machine learning algorithms encompass the study of existing methods, the development of new tools, and the adaptation of models for use in financial institutions.

The article by Noriega J. P., Rivera L. A., and Herrera J. A. [1] examines risk forecasting methods utilizing machine learning. Decision trees, neural networks, and ensemble approaches are analyzed in terms of their characteristics and potential applications in real-world tasks. A detailed review forms the basis for further research in this field.

The study of algorithms highlights their specificity. Kanaparthi V. [2] emphasizes ensemble methods such as random forests and gradient boosting, noting their accuracy. Tumuluru P. et al. [4] describe logistic regression and support vector methods, evaluating their application in credit analysis. The research by Alonso A., and Carbó J. M. [10] underscores the importance of initial data and its quality processing, which significantly influences model outcomes.

The article by Wang Y. et al. [3] proposes combining traditional financial tools with machine learning algorithms. The study by Liu Q. et al. [9] focuses on cross-features that account for parameter interdependence, improving prediction accuracy.

Practical adaptation of algorithms necessitates consideration of data specifics and processing methods. Ortiz R. H. et al. [8] examine the use of algorithms in banks for risk management. The study by Suhadolnik N., Ueyama J., and Da Silva S. [7] confirms the need to adapt models to the characteristics of specific organizational data.

The analysis of risks for small enterprises with limited information volume is addressed in the article by Mi H. [6]. Machine learning approaches provide tools for working with clients lacking sufficient credit history, reducing uncertainty levels.

The study by Trivedi S. K. [5] explores factors influencing forecasting. The use of methods like LASSO enables the elimination of redundant parameters, thereby improving prediction quality.

The work of Yu. Y. Agarkov [12] analyzes the applied use of spiking neural networks in various economic sectors. The author describes the functional features of such networks, including their energy efficiency and ability to process temporal data sequences. The article provides examples of implementing these technologies in forecasting tasks and optimizing production processes. Attention is focused on the potential use of these solutions in industrial system automation and technological process management.

Sources [13] and [14] explore the implementation of machine learning methods in the banking sector. The article "Machine learning in banking: 8 use cases and implementation guidelines," published on www.itransition.com [13], examines eight areas of algorithm application, including creditworthiness analysis, fraud detection, investment portfolio management, and cash flow forecasting. The text describes approaches to configuring systems for specific tasks. The material on the portal habr.com [14] highlights changes in credit risk analysis using machine learning technologies. Examples in the publication demonstrate process automation, the handling of large data volumes, and increased calculation accuracy.

Despite advancements in data analysis, certain aspects remain unresolved. Model interpretation requires expanded approaches, and processing limited datasets demands updated solutions. Issues related to bias continue to provoke discussions within the scientific community, emphasizing the importance of addressing ethical considerations.

Machine learning is actively being integrated into credit risk management. Researchers focus on developing solutions capable of adapting flexibly to changes in the market environment.

Scientific studies were based on a systematic review of publications that examined machine learning algorithms, including decision trees, neural networks, regression approaches, ensemble methods, and support vector algorithms.

To evaluate the models, financial industry data were used, including both real and synthetically generated datasets. Emphasis was placed on prediction accuracy, result interpretability, algorithm adaptability to changes in input data, and application in the specific conditions of financial organizations.

The research helped identify the functional features of machine learning algorithms, determine their potential for solving applied problems, and clarify their practical significance for credit risk management.

RESULTS AND DISCUSSION

Financial risks encompass credit threats, market fluctuations, and operational challenges. Their analysis relies on diverse data, including clients' financial indicators and macroeconomic variables. Managing risks is complicated by data variability, which reflects the behavioral characteristics of clients or market processes. Relationships often emerge under the influence of multiple factors. Historical data may prove insufficient in new customer segments or product lines. Machine learning algorithms address these challenges effectively by processing heterogeneous data and enhancing the adaptability of risk management systems [3].

Algorithms such as XGBoost, LightGBM, and CatBoost are successfully applied in credit scoring and default probability estimation tasks. These technologies identify complex relationships between variables, allowing consideration of multiple factors. This enables the analysis of borrower payment risks by incorporating data such as transactional activity and demographic characteristics.

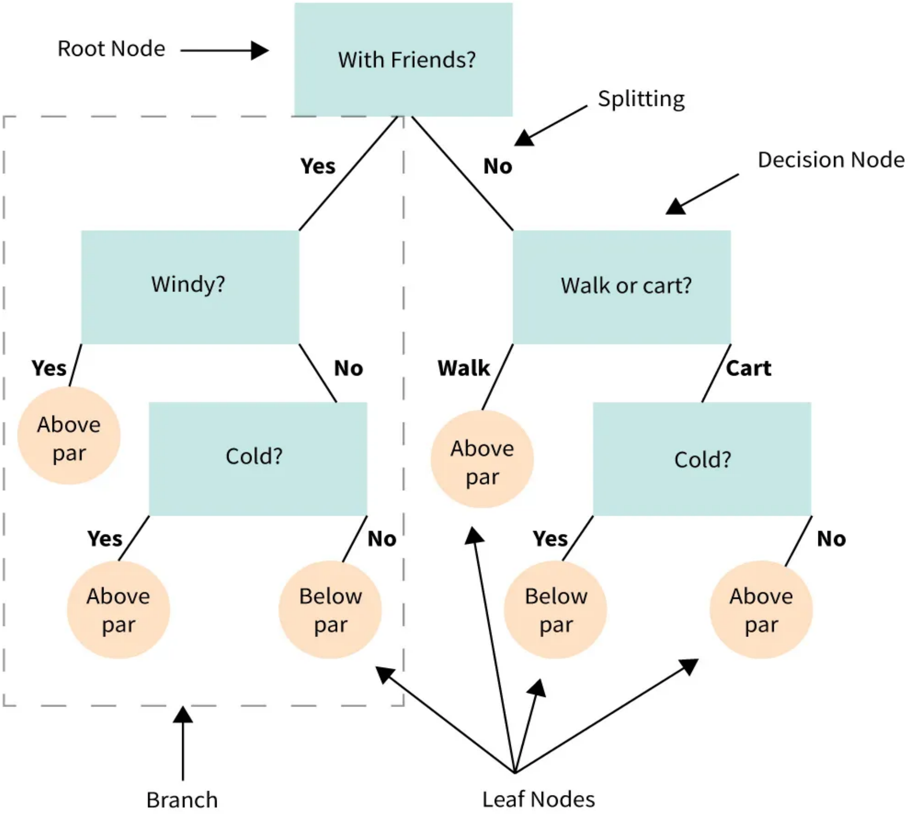

Figure 1. Gradient boosting [14]

Recurrent neural networks (RNNs), including LSTM and GRU, are utilized for processing time series and modeling client revenues. Convolutional architectures are employed for image analysis, including document authenticity verification. Spiking Neural Networks (SNN) represent another subtype of artificial neural networks with a distinct architecture. They mimic the functioning of biological neurons and their interactions, resembling processes in the brain. Artificial neurons transmit information via spikes generated in response to incoming spikes (Figure 2. a).

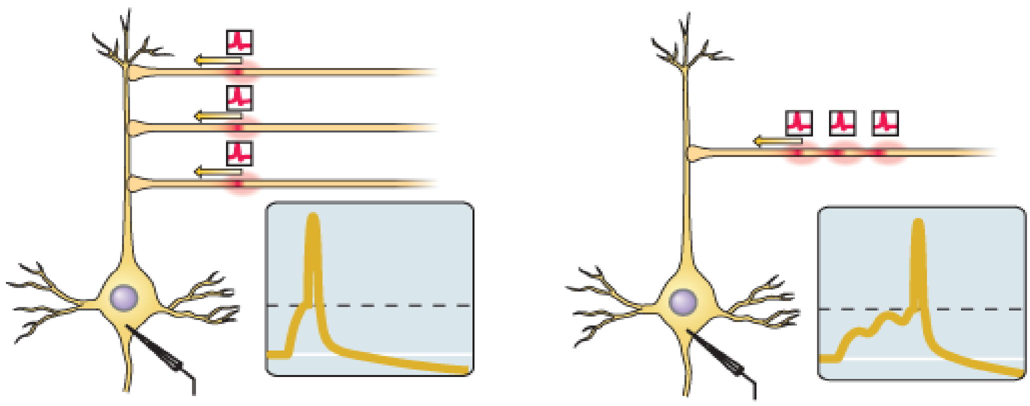

Figure 2. a: Parallel spike transfer of information between neurons; b: Sequential spike transfer of information between neurons [12]

Spiking Neural Networks (SNN) operate based on the principles of biological neurons, each accumulating an electrical charge known as membrane potential. Upon reaching a threshold level, the neuron generates an electrical pulse—a spike—that is transmitted to other neurons. Signals are conveyed as sequences of pulses with specific time intervals, forming a unique interaction structure (Figure 2. b).

Within SNNs, neurons enable sparse asynchronous computations, where only certain elements of the network remain active. This approach emulates the operation of biological neural systems, characterized by parallel information processing and random activity distribution. Asynchronous processing reduces the load on computational resources and increases data processing speed.

In information security, Spiking Neural Networks (SNN) are applied to address the following tasks:

- Detection of network anomalies and attack prevention. SNNs analyze temporal signals in network traffic, identifying deviations from typical patterns. This approach enables the detection of suspicious activities, including malicious operations or unusual behavior.

- Malware classification. Training models on datasets containing malware signatures allows for accurate threat identification and recognition of new types of programs. These methods enhance detection capabilities and improve classification quality.

- User behavior monitoring. SNN models analyze user behavior data to identify deviations from standard activity patterns. This approach helps detect unauthorized actions and prevent internal threats.

- IoT device security. SNNs process data from IoT devices, analyze interconnections and identify potential risks. These methods minimize attack probabilities, considering the inherent vulnerabilities of such devices.

- Real-time risk analysis. SNNs process data on events, vulnerabilities, and threats, enabling rapid decision-making and incident response. This approach ensures the stability of information systems [12].

Clustering algorithms, such as k-means and Gaussian Mixture Models, segment clients by risk categories. Transaction activity analysis helps identify groups of clients without credit histories, enabling the creation of optimal service conditions.

Methods like Isolation Forest and autoencoders detect anomalous transactions, minimizing fraud risks. These algorithms analyze transactional time-series data, identifying deviations characteristic of suspicious operations.

Machine learning technologies are utilized in risk management due to their ability to process large volumes of data, uncover patterns, and eliminate subjective errors. Such methods automate analytical processes that traditionally require significant time investments. In banking, algorithms are employed for document processing, financial information analysis, and workflow simplification. For instance, automated systems extract key insights from text data, optimizing costs and improving efficiency.

Quontic Bank, a community bank in New York, implemented machine learning technologies to enhance its platform, improve digital services, and expand access to banking products for underserved groups. Over a year, the bank increased its customer base, retail deposits, and credit financing volumes.

This initiative became a foundation for strategic growth. Expanding its headquarters, forming a mortgage product team, and preparing to launch a branch for the Chinese-American community strengthened the bank's market position. This example illustrates how leveraging modern technologies can drive business transformation and enhance customer engagement.

In document processing, machine learning can significantly reduce the time required for labeling, classifying, and organizing financial documents. When combined with optical character recognition (OCR), machine learning models also support the digitization of paper-based records.

Datamatics, an Indian data management company, assisted a major American bank in classifying over 1.8 million unstructured mortgage documents. Following the merger of six banks, the organization needed to integrate these documents into a unified document management system for further retrieval.

To automate the indexing and classification of more than 35 million pages of mortgage documents across 275 categories, Datamatics implemented a workflow equipped with OCR, RPA, NLP, and machine learning. As a result, the bank achieved an 87% increase in classification accuracy, a 50% reduction in operational costs, and saved over 150 person-hours per month [13].

The application of machine learning technologies is associated with certain challenges. Data quality control requires particular attention due to the susceptibility of models to errors, omissions, and noise. Complex algorithms are often difficult to interpret, raising concerns among regulatory authorities. Regulatory restrictions limit the use of tools that interact with personal information. Issues of confidentiality and compliance with ethical standards necessitate detailed development of appropriate approaches.

In risk management, machine learning algorithms are used to optimize processes and create robust systems that adapt to changes in external conditions.

Machine learning methods introduce new approaches to data analysis, transforming financial process management. BBVA develops algorithms to assess default probabilities based on transactional information and macroeconomic indicators. Santander applies techniques for anomaly detection, enabling efficient identification of fraudulent activities. Goldman Sachs develops models that integrate neural networks and statistical methods, achieving high prediction accuracy.

Algorithm transparency remains a critical aspect of implementing machine learning. Solutions incorporating blockchain systems enhance data protection. Fully autonomous analytical platforms provide the ability to perform complex tasks without human involvement. Using such technologies requires precise consideration of data characteristics, technical infrastructure parameters, and applicable legal regulations.

Financial institutions in various countries are adopting artificial intelligence technologies to optimize financial management. In China, banks and fintech companies utilize algorithms to process large datasets, assess economic processes, and analyze credit risks. The case of Evergrande highlighted the necessity of applying technologies to evaluate liquidity and redistribute resources [8,13].

Figure 3 below illustrates a four-step approach to implementing machine learning in the banking sector, developed by machine learning consultants at Itransition. It provides insight into the efforts and risks to be addressed at each stage.

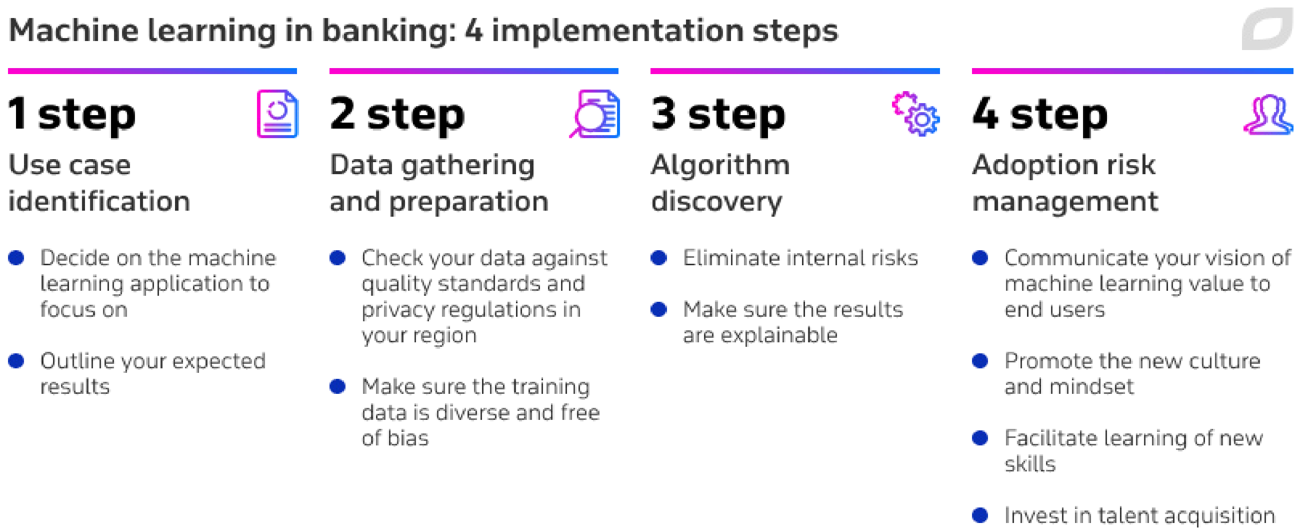

Figure 3. Steps to integrate machine learning [13]

Traditional rule-based approaches are losing relevance as fraudsters change their methods. Machine learning enables the analysis of anomalous transactions, identification of potential threats, and assessment of client actions. Such approaches help minimize financial losses and strengthen organizational security.

CONCLUSION

The article examines methods such as gradient boosting, neural networks, cluster analysis, and anomaly detection algorithms. These approaches effectively address tasks related to credit scoring, default probability assessment, and fraud detection. The application of methods that include feature generation and integration of algorithms with traditional financial models enhances solution accuracy and supports their adaptation to the specific needs of credit organizations.

Identified challenges related to result interpretation, data quality, and adherence to ethical standards require further study. However, machine learning algorithms are already being used to optimize credit risk management processes, reduce losses, and improve organizational resilience to external impacts.

The authors emphasize the importance of integrating algorithms into the operations of credit institutions. This integration creates new opportunities for the financial sector and strengthens market positions. Achieving desired outcomes necessitates the adaptation of technologies to data-specific characteristics and compliance with established regulations.

REFERENCES

1. Noriega J. P., Rivera L. A., Herrera J. A. Machine Learning for Credit Risk Prediction: A Systematic Literature Review //Data, 2023, Vol. 8. No. 11, P. 169.
2. Kanaparthi V. Credit Risk Prediction using Ensemble Machine Learning Algorithms //2023 International Conference on Inventive Computing Technologies (ICICT). – IEEE, 2023, pp. 41-47.
3. Wang Y. et al. The Application Study of Credit Risk Model In Financial Institution via Machine-learning Algorithms //2020 7th International Conference on Information Science and Control Engineering (ICISCE). – IEEE, 2020, pp. 1419-1428.
4. Tumuluru P. et al. Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms //2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS). – IEEE, 2022, pp. 349-353.
5. Trivedi S. K. A study on credit scoring modeling with different feature selection and machine learning approaches //Technology in Society, 2020, Vol. 63, pp. 101413.
6. Mi H. Research on the Application of Machine Learning Algorithms in Credit Risk Assessment of Minor Enterprises //CONVERTER, 2021, PP. 696-706.
7. Suhadolnik N., Ueyama J., Da Silva S. Machine learning for enhanced credit risk assessment: An empirical approach //Journal of Risk and Financial Management, 2023, Vol. 16, No. 12, pp. 496.8.
8. Ortiz R. H. et al. A data analytics method based on data science and machine learning for bank risk prediction in credit applications for financial institutions //2022 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC). – IEEE, 2022, vol. 6, pp. 1-5.
9. Liu Q. et al. Mining cross features for financial credit risk assessment //Proceedings of the 30th ACM international

conference on information & knowledge management, 2021, pp. 1069-1078.

10. Turjo A. A. et al. Cram: a credit risk assessment model by analyzing different machine learning algorithms //2021 4th International Conference on Information and Communications Technology (ICOIACT). – IEEE, 2021, pp. 125-130.
11. Alonso A., Carbó J. M. Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation, 2021.
12. Agarkov Yu.Yu. Applied applications of spike neural networks in the sectors of the national economy // Bulletin of Science, 2023, No.10 (67) volume 3, pp. 499 – 513.
13. Machine learning in banking: 8 use cases and implementation guidelines. [Electronic resource] Access mode: https://www.itransition.com/machine-learning/banking (accessed: 11/19/2024).
14. Machine learning in credit risk assessment: how does ML change the rules of the game? [Electronic resource] Access mode: https://habr.com/ru/companies/data_light/articles/858540 /(accessed: 11/19/2024).

Citation: Di Feng, "Application of Machine Learning for Risk Prediction in Credit Institutions", American Research Journal of Computer Science and Information Technology, Vol 8, no. 1, 2025, pp. 12-17.