Proceedings of the 3rd International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/178/2025.22775

Big Data Technology Empowers Credit Risk Management in Commercial Banks

Jiayang Tang

Law and Business School, Wuhan Institute of Technology, Wuhan, China
2308010205@stu.wit.edu.cn

Abstract: The big data era, driven by internet and AI advancements, has revolutionized finance, particularly challenging commercial banks post-pandemic with rising credit fraud risks and surging funding demands from SMEs, necessitating innovation in traditional credit risk management. This study examines how big data technology enhances banks' risk management through risk identification, dynamic prevention, and response to emerging risks. By analyzing pre-loan, in-loan, and post-loan processes, it highlights the role of multi-source data fusion, generative AI, and decentralized credit scoring in improving risk assessment speed and accuracy. However, challenges persist: poor data quality, "black-box" AI models, and cross-institutional collaboration barriers in data governance and privacy protection. Findings show big data significantly boosts risk control efficiency via intelligent early-warning systems and real-time monitoring, yet its full potential requires standardized data protocols, modular IT upgrades, and privacy-preserving computation frameworks. The research offers actionable strategies for banks' digital transformation while proposing innovative approaches for the financial sector to support SMEs, mitigate systemic risks, and align with real-economy needs.

Keywords: Big Data Technology, Commercial Banks, Credit Risk Management, Digital Transformation, Privacy Protection.

1. Introduction

With the rapid advancement of internet, Internet of Things, and artificial intelligence technologies, human society has officially entered the era of big data. The ability to collect, store, and analyze massive amounts of data has not only reshaped the operational logic of traditional industries but also injected new vitality into the financial sector.

In the "post-pandemic era," the uncertainty surrounding global economic recovery has intensified volatility in financial markets. As the core entities responsible for credit resource allocation, commercial banks are facing unprecedented risk management challenges. According to statistics from the China Banking Association, by the end of 2020, the loan balance of Chinese commercial banks had exceeded 160 trillion yuan [1]. The surge in financing demands from small and micro enterprises and the rise in credit fraud risks following the pandemic have further highlighted the limitations of traditional credit risk management models. Against this backdrop, big data technology, through multi-dimensional data integration and intelligent analysis, offers a new paradigm for risk management in commercial banks.

DOI: 10.54254/2754-1169/178/2025.22775

Furthermore, the "Five Major Themes" proposed at the Central Financial Work Conference position "digital finance" as the core foundation supporting areas such as technology finance and inclusive finance, explicitly requiring financial institutions to embrace digital transformation to meet the demands of high-quality economic development [2]. Simultaneously, the pandemic has accelerated the migration of financial services online, leading to exponential growth in customer behavior data and credit assessment needs. Traditional credit models reliant on manual reviews and static financial statements can no longer adapt to the requirements of dynamic risk monitoring. The introduction of big data technology not only enables the construction of customer credit profiles through data mining but also allows real-time tracking of transaction anomalies, significantly enhancing the timeliness and accuracy of risk warnings.

Given this context, this paper primarily explores the impact of big data technology on credit risk management in commercial banks. It provides a detailed introduction to the optimization of risk identification mechanisms enabled by technology, the efficiency improvements in dynamic risk prevention and control through big data technology, and the application of big data technology in addressing emerging risks.

The application of big data technology can effectively address the shortcomings of traditional credit risk management. It improves upon issues such as reliance on manual experience and static indicators, difficulty in covering long-tail customers, and "adverse selection" caused by information asymmetry in traditional models. In the post-pandemic era, against the urgent backdrop of economic recovery, some financially struggling enterprises have resorted to falsifying their operational conditions to more easily obtain bank credit support, leading to an increase in commercial fraud. Big data technology can provide reliable early warnings in such scenarios [3]. Additionally, to enhance service efficiency and resource allocation rationality, big data technology, through intelligent decision-making models, can reduce the issuance of ineffective loans and optimize capital allocation.

In summary, this study aims to systematically analyze the empowerment pathways and potential challenges of big data technology in credit risk management, providing theoretical support and practical references for the strategic transformation of commercial banks in the digital economy era.

2. The evolution of big data technology

2.1. The emergence of big data tools and their impact on the financial sector

The birth and evolution of big data tools have profoundly reshaped the operational models and service systems of the financial industry. From a technological development perspective, the origins of big data tools can be traced back to the early 21st century when Google introduced distributed computing frameworks, laying the foundation for financial institutions to process massive datasets [4]. With the growing demand for real-time computation, tools such as Apache Spark and Kafka further enhanced the efficiency of financial transaction analysis. The proliferation of cloud computing platforms lowered barriers to data storage and processing, driving financial institutions to transition from traditional data warehouses to data lake architectures. At the application level, big data tools have revolutionized risk control through multi-dimensional data analysis. For instance, dynamic credit assessment models built on unstructured data have significantly improved the ability to identify non-performing loans.

2.2. Big data technologies applied to the financial sector and credit risk management

2.2.1. Enhancing traditional credit risk models

With the advent of big data, credit institutions can now integrate vast amounts of structured and unstructured data from diverse sources, enabling a more comprehensive and dynamic risk assessment

DOI: 10.54254/2754-1169/178/2025.22775

process [5]. By leveraging big data analytics to enhance credit risk models, financial institutions can incorporate a broader range of factors influencing creditworthiness, thereby developing more accurate and predictive credit risk models. For example, big data analytics can analyze customer behavior data, such as transaction patterns and social media activity, to generate valuable insights into borrowers' financial habits and latent risks [6]. Additionally, by analyzing real-time market conditions and macroeconomic indicators, these models can identify external factors that may impact borrowers' repayment capabilities, enabling more granular risk ratings for individuals and businesses.

Furthermore, with borrower consent or within legally permissible boundaries, big data models can track borrowers' transaction histories, online activities, and even social media interactions. For instance, analyzing spending patterns and social media posts can reveal insights into financial stability and potential risk factors. Similarly, evaluating borrowers' stock market engagements and housing statuses allows for a more holistic assessment of their economic context and prospects.

2.2.2. Machine learning algorithms

Currently, machine learning algorithms have become a powerful tool in credit risk management, enabling financial institutions to process and analyze large datasets with speed and accuracy unattainable through traditional manual methods. These algorithmic models are designed to identify patterns and correlations within vast amounts of structured and unstructured data, thereby facilitating more precise risk assessments [7]. For example, supervised learning algorithms such as decision trees are commonly used to categorize borrowers into distinct risk classes based on historical data, while unsupervised learning techniques like clustering can detect potentially risky behaviors hidden in borrower activities. Moreover, the flexibility and adaptability of these algorithms are recognized as critical development priorities, driving continuous advancements in the field.

2.3. Recent innovations in big data technology

Since 2023, newly developed big data tools have largely followed upgrades of existing frameworks while also introducing entirely novel solutions. These tools heavily leverage AI technologies that have undergone rapid development and breakthroughs in recent years, representing both innovation and adaptive responses to increasingly complex economic conditions. Notably, emerging big data technologies are actively addressing or improving upon limitations observed in earlier tools. This section highlights cutting-edge technologies already in practical deployment to provide actionable insights for readers.

2.3.1. Generative AI-driven credit risk assessment

This type of credit risk assessment model offers novel approaches to resolving challenges of data imbalance and class overlap in credit datasets. Traditional methods like SMOTE generate minority-class samples through linear interpolation but risk distorting original data distributions and struggle with overlapping class boundaries. Generative AI techniques—such as Generative Adversarial Networks GANs, VAEs, and diffusion models—learn the true data distribution to synthesize high-quality minority-class samples while optimizing feature spaces to mitigate class overlap [8]. For example, GANs utilize adversarial training between generators and discriminators to produce synthetic default samples that align with real-world distributions [8].

Furthermore, generative AI supports dynamic data augmentation, enabling the generation of new samples in response to shifting economic conditions or changes in borrower behavior, thereby enhancing the adaptability and robustness of models. These strengths allow generative AI to excel in

DOI: 10.54254/2754-1169/178/2025.22775

improving default sample detection capabilities, mitigating class overlap challenges, and strengthening model generalizability.

2.3.2. Decentralized credit scoring systems

Decentralized credit scoring system is a credit evaluation method based on privacy-preserving technologies, aiming to achieve cross-institutional credit scoring modeling through distributed data sharing and encryption techniques while ensuring user data privacy. The core of the system lies in the privacy-preserving collaborative modeling process, including three main steps: data processing, data encryption, and model result decryption. Each data provider holds a portion of user data and preprocesses it before collaborative modeling. To protect data privacy, data providers use encryption methods based on reversible random matrix transformations to encrypt data, generating encrypted data matrix enc_x. On the basis of encrypted data, parties collaboratively train credit scoring models, with model parameters βi and β computed in encrypted states, and the final results are decrypted through a decryption algorithm to restore interpretable credit scores [9].

The security of the decentralized credit scoring system relies on the strength of encryption technologies. For example, the privacy-preserving method based on reversible random matrix transformations ensures that financial institutions cannot deduce raw data during collaborative modeling [10]. Financial institutions only hold encrypted data matrix enc_x, model parameters βi and β, as well as variable relationships from the collaborative modeling process. Due to the use of random matrix transformations in the encryption process, financial institutions cannot obtain raw data through reverse computation. Additionally, encrypted data from each provider is independently stored and computed, preventing financial institutions from directly accessing other providers' raw data, thereby ensuring data isolation and privacy.

3. How big data technology empowers credit risk management in commercial banks

Currently, some commercial banks still employ traditional credit risk management methods, which have exhibited significant issues across pre-lending, post-lending, and pre-lending stages up to now. This chapter primarily analyzes the problems inherent in traditional technical approaches, exploring how big data technologies can address these challenges and the ultimate outcomes they can achieve.

3.1. Pre-loan management

Traditional credit risk management primarily relies on manual review, yet this approach proves inefficient. Manual review not only operates slowly, leading to prolonged approval cycles that hinder applicants' project progress, but is also susceptible to subjective influences, increasing risks of erroneous judgments. Additionally, labor-intensive manual collection of customer information yields limited data dimensions, single-source data streams, and lengthy model optimization cycles, rendering it inadequate to adapt to rapidly changing socio-economic environments. Consequently, risk analysis remains fragmented.

Furthermore, while due diligence plays a critical role in traditional credit risk management, its execution demands substantial resource investment and carries operational risks. Insufficient professional expertise or industry-specific knowledge among due diligence personnel may result in distorted reports, compromising risk assessment accuracy. In extreme cases, personnel accepting improper benefits for personal gain may engage in fraudulent activities, severely undermining the foundation of credit risk management.

Meanwhile, traditional credit risk management overemphasizes collateral while neglecting multidimensional factors such as information and technology. This creates barriers for groups like small and micro enterprises and low-income populations who struggle to provide guarantees or collateral, hindering the realization of inclusive finance objectives. For businesses that appear high-risk yet possess growth potential, traditional strategies often fail to meet demands, restricting development opportunities for many enterprises [11].

Regarding the inefficiency of manual reviews, big data tools offer effective solutions through integrated processing and data mining. By establishing proprietary database systems and integrating external big data sources such as government disclosures, consumption patterns, social security, and medical records, these tools provide a theoretical foundation for streamlined review and approval processes.

3.2. During-loan management

During the lending period, borrowers'operational conditions may deviate significantly due to external factors such as macroeconomic fluctuations, industrial policy adjustments, and shifts in market competition. For instance, a manufacturing enterprise might face an industry-wide overcapacity crisis after securing credit, or a foreign trade company could experience a sharp decline in orders due to international trade disputes. Under traditional models, banks rely on low-frequency, fragmented information channels—such as quarterly financial statements and manual on-site due diligence—to detect risk signals. This approach often leads to delayed information acquisition and misjudgments, causing banks to initiate countermeasures only after risks materialize, thereby missing the optimal window for risk mitigation.

A deeper issue lies in the "Silo Effect of Information" [12]. In traditional frameworks, banks' operational systems—including core systems, credit management systems, and anti-fraud systems—feature disjointed data architectures. External unstructured data, such as industry databases, supply chain data, and regional economic statistics, struggle to integrate effectively. This lack of multidimensional data creates "blind spots" in risk assessment models, hindering accurate predictions of systemic risks arising from industry cyclical shifts.

To address these challenges, big data technologies fundamentally reconstruct dynamic credit risk monitoring systems. By establishing multi-source heterogeneous data integration platforms, commercial banks can now access over 12 real-time data streams, including but not limited to: real-time operational data from enterprise ERP systems, tax invoice data, customs import/export records, social media sentiment analytics, and third-party payment transaction records.

Furthermore, by establishing a three-tier response mechanism of "signal acquisition–model computation–hierarchical alert," artificial intelligence enables intelligent escalation in risk analysis [13]. The system automatically triggers a primary alert when negative social media sentiment about an enterprise exceeds predefined thresholds. The alert level escalates to intermediate when multidimensional indicators—such as abnormal supply chain payment delays and consecutive declines in industry PMI—are concurrently observed. Upon deep analysis using cash flow rupture probability models, if the system identifies high-risk signals of capital chain fractures, a top-level alert is activated. This dynamically progressive warning system employs machine learning to conduct penetrative analysis of multi-source heterogeneous data, evolving risk identification from single-dimensional assessments to panoramic evaluations, significantly enhancing the timeliness and precision of risk warnings.

This hierarchical alert mechanism allows banks to dynamically adjust post-lending management strategies based on risk levels. For example, automated financial data verification is initiated for "blue alert" enterprises, while immediate risk mitigation protocols are activated for "red alert" cases.

DOI: 10.54254/2754-1169/178/2025.22775

3.3. Post-loan management

Through an in-depth analysis of the credit structures and credit asset distribution of regional commercial banks, it becomes evident that these institutions commonly face challenges stemming from the uneven distribution of financial market resources, leading to significant imbalances in their credit structures. In particular, regional commercial banks maintain close partnerships with traditional industries, rendering them more susceptible to cyclical fluctuations within these sectors and thereby increasing the volatility of asset quality [14]. During the credit approval process, rudimentary credit management systems result in lax risk assessments by review personnel for local government projects. Evaluations of corporate debt burdens tied to collaborations with local governments or large enterprises are often delayed, and credit standards may even be relaxed due to policy biases—all of which substantially amplify credit risks. More critically, most regional commercial banks have yet to establish robust mechanisms for assessing asset quality risks, overlooking vulnerabilities in their internal asset risk management frameworks. This further exacerbates the challenges of credit risk management [14].

Big data technology enables real-time monitoring of corporate operational status by establishing a customer behavior profiling system that integrates dynamic data streams across multiple dimensions, including business registration, tax payments, utility payments, and supply chain transactions [15]. Meanwhile, a market environment monitoring system, through crawling publicly available data across the web, can construct an intelligent analytical matrix encompassing hundreds of economic indicators, thereby facilitating prosperity index predictions for key industries [16].

At the risk mitigation level, big data technology drives the formation of a three-tier "prediction–early warning–contingency planning" response mechanism [17]. When the system detects contractual non-compliance by a client, it automatically triggers a risk signal transmission process: sending tailored financial optimization recommendations to the client, simultaneously activating backup credit enhancement measures, and forwarding the resolution plan to relationship managers' decision terminals. Crucially, by constructing industry risk heatmaps and customer linkage mapping, commercial banks can dynamically optimize credit resource allocation.

Under traditional models, risk assessment is constrained by the costs of information acquisition and limited data processing capabilities, creating a vicious cycle of "data silos → cognitive biases → delayed decision-making". In contrast, big data technology transcends temporal and spatial constraints, enabling granular monitoring of daily corporate operations and real-time analysis of supply chain dynamics across industries. This shifts risk management from reactive problem-solving to proactive prevention [18]. Such transformation not only enhances individual institutions' risk resilience but also establishes dynamic risk buffer mechanisms for the entire financial system, carrying profound strategic significance in the digital economy era.

4. Current challenges and improvement recommendations for big data technologies

4.1. Data quality and model design issues

Deficiencies in data quality and integrity remain primary challenges for big data tools [19]. The diversity of data sources and inconsistent formats, coupled with extensive unstructured data, amplify difficulties in cleaning and integration. Insufficient or incomplete historical data—particularly the lack of credit histories for emerging industries and micro enterprises—constrains the comprehensiveness of risk assessments [19]. Furthermore, delayed data updates, cross-departmental data silos, and the difficulty in quantifying critical risk indicators further undermine evaluation accuracy and depth.

DOI: 10.54254/2754-1169/178/2025.22775

In model design, traditional statistical models struggle to handle high-dimensional, nonlinear big data characteristics, while the lack of interpretability in machine learning models reduces transparency in risk decision-making. Additionally, models overly reliant on historical data often neglect emerging risk factors, and single-model frameworks fail to adapt to diverse client needs, resulting in inadequate flexibility and precision in risk assessments. These issues are not inherently tied to big data technologies themselves but stem from long-term data accumulation challenges and manual input dependencies, posing significant unresolved difficulties under current conditions.

4.2. Technical implementation and privacy protection challenges

Technical implementation and integration pose significant challenges for commercial banks adopting big data technologies. Existing IT infrastructure struggles to meet the high demands of big data systems, with costly and complex system upgrades. Compatibility issues between big data platforms and legacy core banking systems disrupt data flows and compromise real-time processing capabilities. Moreover, severe internal data silos hinder the construction of holistic risk profiles, while the tension between rapid technological iterations and banking system stability complicates continuous optimization [12].

Simultaneously, privacy protection and compliance concerns are escalating. Big data analytics involves vast sensitive information, creating dilemmas in balancing data value extraction with customer privacy preservation. Cross-domain data integration risks regulatory violations, and cross-border data flows face intricate compliance hurdles. Furthermore, the "black-box" nature of algorithms may lead to discriminatory treatment of specific groups, violating fair lending principles and complicating regulatory oversight, thereby undermining the fairness and transparency of decision-making processes [20].

Currently, banks and enterprises can address these challenges by refining big data models to incorporate tailored categorization and computation for special populations. For sensitive information avoidance, external regulatory intervention is required to conduct audits and enforce compliance frameworks.

5. Conclusion

The paper systematically reviews the evolutionary trajectory of big data technology, provides an in-depth analysis of its empowerment pathways in credit risk management at commercial banks, and examines the current limitations in technological applications from multiple dimensions including data governance, technical implementation, and privacy protection. Subsequently, targeted improvement strategies are proposed. By integrating technological development, practical application, and critical reflection, the study constructs a comprehensive analytical framework of "technology-driven innovation - risk prevention and control - system optimization and upgrading". This research offers a reference solution that integrates theoretical depth with practical value for the digital transformation of commercial banks, establishing a multidimensional analytical perspective that connects technological advancement with financial risk management practices.

The evolution of big data technology has driven the refinement and dynamic enhancement of credit assessment models, with its core value lying in the optimization of risk identification mechanisms, improved efficiency in dynamic risk prevention and control, and strengthened capabilities in addressing emerging risks. However, significant challenges remain in data governance, technical implementation, compliance, and ecosystem development. Looking ahead, these issues can be effectively addressed through the following measures: establishing unified data standards and cross-departmental sharing mechanisms; upgrading banking IT infrastructure with a "modularization + microservice" architecture; collaborating with regulatory bodies to formulate technical specifications for privacy-preserving computation; promoting regional credit information sharing platforms; and incentivizing financial institutions through policy measures to participate in market-oriented allocation pilots for data elements. These integrated strategies aim to systematically resolve existing constraints while advancing the maturity of big data-enabled financial risk management systems.

The deep integration of big data technology not only represents a technological advancement in commercial banks' risk management, but also serves as a critical enabler for the financial system to better serve the real economy and advance inclusive finance. It provides replicable solutions for systemic risk prevention and control in the digital economy era, establishing a robust technical foundation that bridges financial innovation with macro-prudential supervision requirements.

This study provides multidimensional practical guidance for the digital transformation of financial institutions through systematic analysis of the implementation pathways and potential challenges of big data technology in commercial banks' credit risk management. At the operational level, the proposed "technology-driven innovation - risk prevention and control - system optimization and upgrading" framework enables commercial banks to accurately identify and dynamically manage risks throughout the pre-loan, in-loan, and post-loan phases. For instance, real-time risk monitoring can be achieved through multi-source heterogeneous data integration platforms, while default sample identification capabilities can be enhanced via generative AI technologies, thereby improving credit asset quality and resource allocation efficiency. For regulatory authorities, the discussion on privacy-preserving collaborative modeling and decentralized credit scoring systems offers technical references for balancing data sharing with compliance requirements, facilitating the standardized development of cross-institutional credit information platforms. Furthermore, the proposed data governance strategies and technical implementation recommendations provide theoretical foundations for formulating fintech policies, promoting the healthy development of market-oriented data element allocation. In academic research, this paper bridges the knowledge gap in deep integration of big data technology with credit risk management by synthesizing technological evolution, application scenarios, and critical reflections. It lays groundwork for future exploration of dynamic risk model optimization and algorithmic fairness.

Overall, this research not only delivers implementable solutions for commercial banks' risk management practices but also contributes innovative yet feasible pathways for the financial system to better serve the real economy and mitigate systemic risks. The findings establish an analytical paradigm that connects technological advancement with financial stability maintenance, while providing methodological references for addressing emerging challenges in data-driven financial regulation.

References

[1] *Xv, B. (2025) The foundation for the steady development of the banking industry has been continuously consolidated. Retrieved from: https://www.financialnews.com.cn/. Accessed Feb.24,2025.*

[2] *Wang, X. (2024) Do a good job in the 'five big articles' of finance. Retrieved from: https://www.cpcnews.cn/. Accessed Sept.25,2024.*

[3] *Zhou, W. (2025) Exploration of Credit Risk Management Issues in Commercial Banks under the New Economic Normal. China Circulation Economy, Jan,2025.*

[4] *Singh, J., & Singla, V. (2015) Big data: tools and technologies in big data. International Journal of Computer Applications, 112, 15.*

[5] *Bi, W., & Liang, Y. (2022) Risk assessment of operator's big data Internet of Things credit financial management based on machine learning. Mobile Information Systems, 2022(1), 5346995.*

[6] *Wang, H. (2021). Credit risk management of consumer finance based on big data. Mobile Information Systems, 2021(1), 8189255.*

[7] *Wei, Y. (2022). [Retracted] A Machine Learning Algorithm for Supplier Credit Risk Assessment Based on Supply Chain Management. International Transactions on Electrical Energy Systems, 2022(1), 4766597.*

DOI: 10.54254/2754-1169/178/2025.22775

[8] *Gu, Z. & Huang, B. (2024) Improved Active Generative Oversampling for Personal Credit Risk Assessment: A Machine Learning Approach. Software Guide.*
[9] *Kuang, X., Ma, C., & Ren, Y. S. (2024) Credit risk: A new privacy-preserving decentralized credit assessment model. Finance Research Letters, 67, 105937.*
[10] *He, H., Wang, Z., Jain, H., Jiang, C., & Yang, S. (2023). A privacy-preserving decentralized credit scoring method based on multi-party information. Decision Support Systems, 166, 113910.*
[11] *Yv, S. (2023) Research on Credit Risk Management of H Rural Commercial Bank under the Background of Big Data. Nov.30,2023.*
[12] *Abedalla, R. W. (2014). The silo effect of information technology on the organizational productivity: A qualitative research study. International Journal of Computer, 15(1), 42-50.*
[13] *Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial Intelligence Trust, risk and security management (AI trism): Frameworks, applications, challenges and future research directions. Expert Systems with Applications, 240, 122442.*
[14] *Fang, H. (2024) Research on Credit Risk Management of Local Commercial Banks under the New Economic Situation. Chinese Agricultural Accounting, Feb,2024.*
[15] *Tang, C., Tang, Y., Zeng, Z., Zhang, L., & Xiang, S. (2023). Customer characteristics analysis method based on the selection of electricity consumption characteristics and behavioral portraits of different groups of people. Journal of Intelligent & Fuzzy Systems, 44(3), 4273-4283.*
[16] *Environmental Monitoring Market 2018 Global Industry Drivers, Opportunities, Trends, Segmentation and Forecast to 2022. M2 Presswire, 2018.*
[17] *Lu, S. (2023) Research on the Application of Big Data Analytics in Financial Risk Assessment. Shanxi Science and Technology News, Nov.28,2023.*
[18] *Dr. Pragyansmita, N. (2024) Breaking data silos to achieve AI readiness. Nextgov.com (Online), Dec.27,2024.*
[19] *Wang, W. (2024) Big Data-Driven Credit Risk Assessment and Mitigation Strategies in Commercial Banks: A Comprehensive Study. China Conference & Exhibition, 2024.*
[20] *Fang, M. and Zhang, H. (2024). Regulatory Framework for Algorithmic Black Boxes on Super Algorithm Platforms. Journal of Chinese Legal Studies (JCLS), 1(3), 52-61.*