World Journal of Management Science
ISSN: 2959-9628
DOI: 10.61784/wms3037

APPLICATION OF BIG DATA ANALYTICS IN COMMERCIAL CREDIT RISK PREDICTION: CHALLENGES AND OPPORTUNITIES

Jun Dai
School of Business Administration, Baise University, Baise 533000, Guangxi, China.
Corresponding Email: 598656371@qq.com

Abstract: This study aims to explore the application of big data technology in commercial credit risk prediction, analyzing the opportunities and challenges it presents. The research reveals that big data technology offers significant advantages in credit risk prediction, notably enhancing the accuracy and real-time capabilities of such predictions. However, it also faces issues related to data quality, privacy protection, and model transparency. The study further suggests that future technological advancements will continue to propel the intelligentization of credit risk management, while emphasizing the need to strengthen attention to privacy and compliance.
Keywords: Big data; Commercial credit; Credit risk prediction; Challenges and opportunities

1 INTRODUCTION

In today's globalized economy, commercial credit risk management has become an essential component for businesses and financial institutions. Effective credit risk management not only impacts a company's cash flow and operational efficiency but also directly influences financial institutions' loan decisions and fund security. Traditional credit assessment methods primarily rely on limited financial data and historical credit records, which exhibit significant shortcomings in terms of accuracy and comprehensiveness. These methods typically capture only static, single-dimensional information, failing to reflect the dynamic credit performance of businesses or individuals across different time points. Moreover, with the continuous changes in the market and increasing economic uncertainties, assessing corporate credit risk has become more complex, rendering traditional credit evaluation models increasingly inadequate to meet current challenges.

The rise of big data technology offers a novel solution for credit risk management. Unlike traditional methods, big data analytics can integrate vast amounts of information from diverse sources, including social media behavior, transaction records, internet search habits, publicly available corporate financial data, industry reports, and more. This multi-dimensional data sourcing not only broadens the scope of data but also brings higher precision to credit risk prediction. Through deep mining and analysis of massive datasets, companies can promptly identify potential risk signals, thereby effectively enhancing the scientific and timely nature of decision-making. Similar to the real-time analysis of student behavior data through MOOCs (Massive Open Online Courses) in the education sector, commercial credit risk prediction can also monitor the real-time performance of credit entities using big data technology [1]. With big data platforms, companies can swiftly adjust their risk management strategies based on the dynamic behaviors of credit entities, improving their responsiveness to potential risks.

Furthermore, the introduction of big data technology breaks the boundaries of traditional credit assessment. Utilizing techniques such as machine learning and data mining, businesses and financial institutions can process large volumes of data in real-time and make precise predictions about future credit risks through predictive models. For example, machine learning algorithms can learn from historical data to identify features highly correlated with credit defaults, thereby providing essential bases for credit assessments. Particularly in the context of the booming internet finance and sharing economy, traditional credit scoring systems face unprecedented challenges. Leveraging big data technology to build more flexible and dynamic credit evaluation systems has become a topic of mutual interest in academia and practice.

In summary, as data resources continue to expand and technological means advance, the application of big data analytics in commercial credit risk prediction demonstrates immense potential. However, issues such as data quality, privacy protection, and model interpretability and transparency still require further exploration and resolution in practice. This offers ample opportunities for future research and provides new possibilities for innovation in credit risk management.

2 LITERATURE REVIEW

In recent years, with the rapid development of big data technology, both academia and industry have shown strong interest in its application in commercial credit risk management. The introduction of big data analytics offers new perspectives for credit risk prediction, overcoming the limitations of traditional methods and enabling businesses and financial institutions to assess credit risks more accurately [2]. However, the complexity of big data applications and the challenges of data management also introduce new problems.

Firstly, traditional credit assessment methods heavily rely on historical financial data, bank transaction records, credit scores, and other conventional financial data. These data mainly originate from credit reports, company financial statements, and loan records. However, the single nature and static characteristics of traditional data make them

inadequate for responding to rapidly changing market environments. Traditional models often employ classical algorithms such as linear regression and decision trees, which typically assume that data distributions meet certain specific conditions. In practice, however, data are often nonlinear and complex [3]. Consequently, these methods have limitations in handling the complexity of credit default predictions. To compensate for the shortcomings of traditional methods, the application of big data analytics in credit risk prediction has gradually emerged in recent years. Compared to traditional models, big data technology integrates multi-source heterogeneous data, providing more dynamic and comprehensive credit evaluation methods [4-5]. For instance, unstructured data generated from social media, e-commerce platforms, and mobile applications offer new dimensions for credit assessments. These data encompass not only financial transaction behaviors of individuals or enterprises but also non-traditional factors such as consumption habits, social interactions, and online evaluations, making credit risk prediction models more comprehensive and diverse. Similarly, the concept of multi-dimensional information integration reflects the importance of data fusion. In commercial credit risk prediction, integrating multi-source data allows credit assessments to more comprehensively capture customers' credit performance, enhancing the accuracy of credit risk predictions [6].

In this context, the widespread application of machine learning and artificial intelligence technologies is particularly noteworthy. Literature indicates that machine learning algorithms, such as random forests, support vector machines, and deep neural networks, have been widely used in the field of credit risk prediction. By learning from extensive historical data, machine learning models can identify potential credit default risk features and predict future risks based on these features. For example, some studies have found that by analyzing individuals' payment behaviors, social network activities, and e-commerce purchase records, machine learning models can significantly improve the accuracy of credit scores, especially in cases lacking traditional credit records. Related to this is data mining technology, which can deeply analyze potential patterns in large-scale data to help uncover hidden factors influencing credit risk.

Secondly, in the era of open banking, data integration and sharing have become key to enhancing the efficiency of credit risk assessments [7]. Open banking policies allow financial institutions to share customers' financial data with third parties, providing a broader data source for credit risk prediction. By integrating customers' transaction records across different financial institutions, companies can gain a more comprehensive understanding of customers' financial status and credit performance. For instance, some literature points out that open banking data can provide more comprehensive income and expenditure information than single-bank transaction data, thereby aiding in the construction of more refined credit risk assessment models. Additionally, open banking policies promote real-time data analysis, enabling businesses to identify credit risks and take countermeasures at the earliest signs of market changes.

Moreover, despite the numerous advantages that big data technology demonstrates in credit risk management, it also faces several challenges [8]. On one hand, data quality issues are the primary challenge in big data applications. Due to the wide range of big data sources, inconsistencies, data omissions, and noise may occur during data collection and processing, directly affecting the accuracy of credit assessments. Additionally, handling unstructured data (such as text and images) in big data is more difficult, making the effective extraction of useful information a critical research direction in academia. On the other hand, data privacy and security issues are also of significant concern. With the large -scale use of personal data, finding ways to reasonably utilize data while safeguarding user privacy has become a major dilemma. Relevant literature indicates that although big data analytics enhances the precision of credit risk predictions, the inability to protect data privacy can severely undermine public trust in big data applications.

Lastly, regarding the application of big data technology, attention must also be paid to the issues of model transparency and interpretability [9]. Complex machine learning models, especially deep learning models, while significantly improving prediction accuracy, have "black box" characteristics that result in poor interpretability of their results. This is particularly challenging for highly sensitive fields like credit risk assessment. Regulatory bodies and businesses need to clearly understand how models derive credit scores to review and manage them effectively. Therefore, ensuring model transparency while enhancing performance remains a hot topic in current academic research.

Overall, the application of big data technology in credit risk management offers new possibilities and challenges for credit assessments. However, businesses can improve the accuracy and real-time capabilities of credit risk predictions by integrating multi-dimensional data and introducing machine learning algorithms. Nevertheless, bottlenecks in technological applications, such as data quality, privacy protection, and model interpretability issues, require further research and exploration. In the future, as big data technology continues to mature, its application prospects in credit risk management are broad, but it also necessitates corresponding balances and innovations in regulation and technology.

3 THEORETICAL ANALYSIS

The application of big data technology in commercial credit risk prediction provides businesses with new tools and methods to more accurately identify and manage credit risks. Theoretically, big data technology not only enhances the breadth of predictions through multi-dimensional data integration but also improves the depth and precision of predictions through advanced analytical tools and models.

Firstly, the multi-source heterogeneous characteristics of big data significantly expand the range of data acquisition. Traditional credit assessment models rely heavily on single data sources such as financial data and credit reports, whereas big data technology extends data sources to non-traditional areas like social media, e-commerce platforms, and industry analysis reports. By comprehensively processing these multi-dimensional data, businesses can gain a more thorough understanding of an individual’s or enterprise’s credit performance. For example, social media data can reveal

a person's consumption habits, social relationships, and even job stability—all of which are crucial factors in credit risk assessment.

Secondly, in the process of handling big data, the widespread application of machine learning and artificial intelligence algorithms provides robust analytical capabilities for credit risk prediction. Machine learning can identify complex risk patterns through training on historical data, thereby establishing more forward-looking credit risk models. These models can not only predict potential default risks but also offer targeted management strategies for individuals or enterprises with different levels of credit risk. For instance, models like random forests and support vector machines can automatically classify data across various dimensions and refine credit scores based on historical performance data.

Furthermore, the introduction of deep learning algorithms further enhances the precision of credit risk predictions. Unlike traditional linear models, deep learning constructs multi-layer neural networks that can handle vast amounts of complex nonlinear data relationships. Particularly in processing large-scale and unstructured data, deep learning models exhibit high efficiency and accuracy. For example, through deep analysis of massive transaction data and social interaction data, deep learning models can effectively identify credit default risk factors hidden behind surface-level information.

Additionally, techniques such as text mining and network analysis within big data technology provide new perspectives for credit risk prediction. Text mining can extract valuable information from unstructured textual data, such as social media comments and risk indicators in industry reports. These pieces of information are often absent in traditional credit assessment data but play a significant supplementary role in predicting credit risks. Similarly, network analysis can reveal the interconnectedness of individuals or enterprises with other credit entities by analyzing their social network relationships, thereby aiding in the prediction of future credit performance. However, despite the numerous innovations that big data technology brings to credit risk prediction, the theoretical community also points out some challenges in its application. On one hand, data quality remains a core issue in big data analytics. Due to the wide range of data sources, differences in data formats and data integrity exist between different sources, making effective data cleaning and processing essential prerequisites for ensuring the accuracy of prediction models. On the other hand, the complexity of models introduces the "black box" problem—although big data models can provide highly accurate prediction results, their decision-making processes lack transparency and interpretability. This may lead to distrust in model results, especially in highly sensitive fields like credit risk management. Therefore, improving both model performance and its interpretability and transparency has become a joint focus of the theoretical and practical fields.

Lastly, data privacy and legal compliance issues are also significant challenges in the application of big data technology [10]. In the era of big data, credit risk prediction relies on vast amounts of personal and enterprise data. Ensuring reasonable use of this data for prediction while safeguarding data privacy is a major challenge faced by businesses and regulatory bodies. In recent years, with the continuous improvement of data privacy protection laws and regulations, businesses must strictly comply with relevant legal requirements when using big data technology to ensure that data usage does not infringe on personal privacy.

In conclusion, big data technology provides robust theoretical support and practical tools for commercial credit risk prediction. Through multi-dimensional data integration and the application of advanced algorithms, businesses can predict credit risks more accurately. However, issues such as data quality, model interpretability, and privacy protection still require further research and improvement. In the future, as technology continues to advance, the potential for big data technology in credit risk management will be further unleashed.

4 RESEARCH CONCLUSIONS AND PROSPECTS

This study has two main findings: First, the application of big data analytics in commercial credit risk prediction offers unprecedented advantages for businesses, particularly in enhancing the accuracy and real-time capabilities of credit risk predictions. By integrating multi-source heterogeneous data and applying machine learning algorithms, businesses can gain a more comprehensive understanding of credit entities' credit performance and identify potential credit risks early on. This not only optimizes the credit risk management processes of businesses but also enhances the overall stability of financial markets. Second, the application of big data technology in credit risk assessment breaks the limitations of traditional credit scoring models. The use of multi-dimensional data sources makes evaluations more comprehensive, and the introduction of advanced algorithms like machine learning improves the precision and efficiency of predictions. These technologies enable businesses to make more accurate predictions of credit risks based on diverse information such as historical behaviors and unstructured data. Especially in cases lacking traditional credit records, big data technology provides important supplementary means for credit assessment, greatly expanding the applicability of credit evaluations.

However, despite the numerous benefits that big data technology brings to credit risk prediction, there are still some challenges that need to be overcome in practice. Firstly, data quality directly affects the accuracy of prediction models. Issues such as inconsistent data formats across different sources, data omissions, and noise can all lead to biased prediction results. Secondly, the issue of model transparency urgently needs to be addressed. While machine learning and deep learning models excel in prediction accuracy, their complexity results in poor interpretability of results. In highly sensitive fields like credit risk assessment, businesses and regulatory bodies need to clearly understand the basis of model decisions. Additionally, with the continuous improvement of personal privacy protection laws and regulations, finding ways to reasonably utilize big data while ensuring data privacy has become a significant challenge for

businesses. Compliance and privacy protection issues not only involve the ethical responsibilities of businesses but also directly impact their reputation and customer trust.

Looking ahead, as big data technology continues to develop, its application prospects in commercial credit risk management are broad. Technological advancements will provide more intelligent tools for credit risk prediction, particularly achieving further breakthroughs in data integration, model optimization, and real-time risk monitoring. However, while enjoying the dividends brought by big data, businesses also need to continuously improve data management and compliance mechanisms to address challenges related to data privacy protection and model transparency. By reasonably utilizing big data technology, businesses can not only enhance their own credit risk management capabilities but also contribute to the healthy development of financial markets.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Wu PX, Lan ZY, Huang W, et al. Exploration of the application of a MOOC-based blended teaching model in medical undergraduate interns. Journal of Youjiang National Medical College, 2020, (01): 119-122.

[2] Gao L, Xiao J. Big data credit report in credit risk management of consumer finance. Wireless Commun Mob Comput, 2021, 2021: 4811086.

[3] Ilic I, Görgülü B, Cevik M, et al. Explainable boosted linear regression for time series forecasting. Pattern Recognit, 2021, 120: 108144.

[4] Hasan MM, Popp J, Oláh J. Current landscape and influence of big data on finance. J Big Data, 2020, 7(1): 21.

[5] Wang Y, Xiuping S, Zhang Q. Can fintech improve the efficiency of commercial banks?—An analysis based on big data. Res Int Bus Finance, 2021, 55: 101338.

[6] Pang XR, Liu F, Li L, et al. Preliminary exploration of teaching reform for "Introduction to General Medicine" based on the concept of "Curriculum Ideological and Political Education". Journal of Youjiang National Medical College, 2020, (06): 806-809.

[7] Bello OA. Machine learning algorithms for credit risk assessment: an economic and financial analysis. Int J Manag, 2023, 10(1): 109-133.

[8] Luan H, Geczy P, Lai H, et al. Challenges and future directions of big data and artificial intelligence in education. Front Psychol, 2020, 11: 580820.

[9] Xu Y, Hu M, Liu H, et al. A hierarchical deep learning approach with transparency and interpretability based on small samples for glaucoma diagnosis. NPJ Digit Med, 2021, 4(1): 48.

[10] Thapa C, Camtepe S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. Comput Biol Med, 2021, 129: 104130.