Exploration of Commercial Banks' Credit Business in the Context of Big Data

Chenran Li

Abstract:

With the development of big data technology, there are many problems in the credit system of commercial banks, commercial banks continue to improve the credit system to better ensure the good development of loan business. Micro, Small, and Medium Enterprises (Msmes) are the foundation of national economy, it is very important to study the financing of msmes, commercial banks should make correct credit decision for different msmes. In this paper, we solve the problem of commercial bank making loan decision to loan enterprise from the source, and study the efficient decision-making method. Based on RFM model, we propose a new YNA model to determine the main value of the studied credit customers, k-means clustering algorithm is used to determine the stratification of certain type of enterprises, KNN model and logistic regression model is used to help banks to identify the risk of msmes. It is found that the use of big data technology to facilitate bank credit decision-making improves the efficiency and accuracy of commercial bank credit business review of different enterprises. Based on the essence of commercial bank credit business model, this paper combines the characteristics of commercial bank credit business and big data through the financing of small and medium-sized enterprises.

Keywords: commercial banks; credit service; Financing for micro, small and medium-sized enterprises; big data

1.Introduction

Big Data, as a product of the new internet age, has played an integral role in the field of data mining, contributing significantly to the exploration and development of data within society. With the evolution of technology, commercial banks are actively investing and hastening their integration and application of Big Data. By conducting comprehensive research, analysis, excavation, and organization of Big Data, they provide customers with all-encompassing financial services, thereby enhancing the customer experience, reducing client attrition, and bolstering customer loyalty. This, in turn, aids in the development of the enterprise, delivering greater value to customers.

Against the backdrop of Big Data, financing for small and micro-enterprises is confronted with numerous challenges and opportunities. In terms of traditional financing, these enterprises could opt for conventional methods such as bank loans, bond issuance, or equity financing from investors. However, these channels may involve higher costs, longer review periods, and greater risks. In the aspect of Big Data risk assessment, small and micro-enterprises can appraise their risks and credit ratings more accurately with the aid of Big Data technology, facilitating more effective financing acquisition. In China, as significant providers of corporate credit services, commercial banks play an essential role in financing for small and micro-enterprises. Banks should actively respond to national policies, strengthen the integration of credit services with Big Data technology, enhance the efficiency of credit service processing, and offer superior financing channels to small and micro-enterprises, aiming for a win-win situation and fostering positive economic development.

The advent of Big Data has provided commercial banks with new credit pathways. However, it has also diverted customer resources from banks, impacting traditional credit services. Under these circumstances, commercial banks have witnessed a decline in market share and profits, necessitating an enhancement in credit efficiency and the planning of innovative directions, with an urgent need for business transformation. Thus, the study of credit service models of commercial banks is beneficial in helping these institutions understand the multifaceted impacts of Big Data, accelerate their data strategies, and offer valuable references and decision-making bases for the banks' transformation and development.

2. Literature Review

2.1 Research on the Impact of Big Data on

Commercial Banks' Credit Operations

With the advent of the Big Data era, competition in the financial industry has intensified. This period is marked by the adjustment of credit structures and the maturity of loans, as well as the narrowing of the interest spread between deposits and loans. Commercial banking credit operations, closely linked to customer information data, are facing innovation and challenges. Enhancing the ability to organize and analyze information data and control credit risks is a crucial task for commercial banks at present.

In the traditional credit model, commercial banks spend substantial costs collecting customers' credit data, which suffer from untimely updates and weak processing capabilities, leading to severe information asymmetry between banks and borrowers, low operational efficiency of banks, and poor credit risk control. In the era of Big Data, Xue Kezhen (2023) believes that Big Data will greatly assist commercial banks in seizing market resources and is the key to enhancing industry competitiveness. Only by possessing the ability to mine and manage massive amounts of data can banks improve their core competitiveness and sustainable development capabilities. Long Jing (2023) suggests that commercial banks should establish specialized data-related departments, build data warehouses, and set up screening mechanisms to improve the efficiency of data mining; strengthen data sharing between departments to address the problem of data fragmentation and independence; and accelerate the integration of Big Data technology with commercial banking operations by utilizing these measures.

2.2 Studies related to microfinance for small and medium-sized enterprises

Numerous scholars have studied the financing difficulties faced by small and micro-enterprises under the traditional commercial banking model. Huang Xia (2021) noted that the existing financing methods for small and micro-enterprises are predominantly based on mortgage loans, with limited financing channels; the cost of financing is excessively high; financing relies mainly on short-term funds, with long-term funds being difficult to obtain; and the cash flow is unstable, with poor sustainability of capital inflow. Against the backdrop of Big Data, many scholars have proposed new solutions for financing small and micro-enterprises in commercial banks. Zhang Junjun (2022) contributed to inclusive finance by building models, proposing a digital customer acquisition model — the small and micro-merchant financing access model, and a digital risk control model — the small and micro-enterprise loan risk monitoring model.

2.3 Research on Commercial Banks' Credit

Business for Micro, Small and Medium Enterprises (MSMEs)

In the context of national economic significance, micro and small enterprises (MSEs) are increasingly prioritized by commercial banks, particularly in their pursuit to minimize non-performing loans and maximize profitability. Gao Lubing (2021) utilized the entropy-weighted TOPSIS method for a quantitative assessment of credit risks and ratings, followed by the application of logistic regression to ascertain the probability of default. This scientific approach aids banks in the formulation of credit policy strategies, facilitating the optimization of resources. Liu Xinying (2021) observed that companies with high credit ratings often correlate with lower profitability and potential default risks, suggesting the necessity of models that concurrently consider reputation and financial strength in credit outcome predictions, aligning with the requirements for a literature review as per SCI journal standards.

3. Method

3.1 RFM model

The present article delineates the YNA model, a derivative of the RFM paradigm, tailored to appraise customer creditworthiness within loan operations. The acronym RFM stands for Recency, Frequency, and Monetary value—metrics instrumental in grasping customer requisites and categorizing them per their value to the enterprise. The YNA model transmutes these metrics into the year of company establishment, client count, and annual average yield, thereby equipping credit institutions with pivotal decision-making tools. This methodology enables assessments grounded in customer behavior, offering customizable and visually interpretable insights that are conducive to the formulation of potent marketing strategies.

| FRM	Rencency	Frequency	Monetary
YNA	Year	Number	Average Return

Embracing the ethos of the RFM framework, the YNA approach ascertains a company's age through the temporal span between the earliest and latest billing periods on input and output tax invoices; it gauges customer quantity via the number of valid invoices on record; and it infers the annual average interest rate from a comprehensive analysis of active, sale, and voided invoices, facilitating a stratification of customer value.

This classification bifurcates clients into five distinct segments. For the high-risk category 1 and category 5 businesses characterized by a scant customer base and low profits, banks are advised to exercise caution by offering lower interest rates and promotional initiatives to mitigate risk. Category 2 enterprises, with average RFM scores,

are habitual clients who may benefit from enhanced services and incentives designed to foster their latent growth potential. Categories 3 and 4 businesses, despite their considerable revenue, are short-lived and represent core clientele. They necessitate personalized services and substantial banking investments to sustain their high-value status and forestall attrition.

3.2 K-Means model

In the study of enterprise clustering, the YNA model delineates three pivotal indicators, leading to the establishment of four rating clusters labeled A, B, C, and D. Under the K-Means model, it is posited that data points in close proximity are aggregated into the same category, which analogously suggests that enterprises within each cluster exhibit similar repayment capabilities. Building on this premise, the K-Means algorithm from the Sklearn library is employed to compute the mean values for the four principal categories A, B, C, and D, with assigned values ranging from 0 to 3. This clustering process not only enhances the precision of creditworthiness evaluation but also streamlines the categorization of enterprises based on their repayment aptitude.
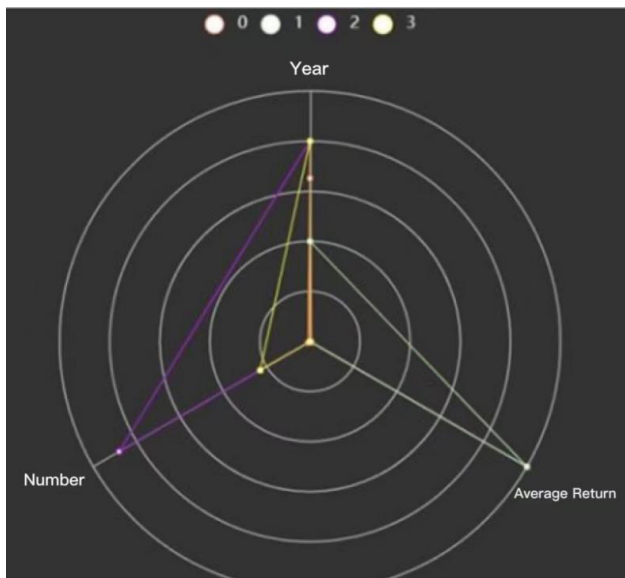
In scenarios with fixed credit limits, 'A' rated enterprises, known for their excellent credit history and long-established presence, receive preferential loan terms. Their stable financial performance, strong balance sheets, and high profitability make them attractive to lenders. 'B' rated companies, with moderate credit, are assessed on specific metrics like profitability and business longevity, and they usually secure reasonable financial support, reflecting their stable yet less prominent market position and financial health. 'C' rated entities, marked by average credit, are evaluated based on customer numbers and operational duration. Their financial support from banks is more conservative, mirroring their lower profitability and weaker market standing. Finally, 'D' rated firms, with the lowest creditworthiness, face stringent scrutiny, focusing on their smaller customer base and shorter business history. They typically encounter limited financing options with high rates, owing to their unstable operations and frail financial positions.

3.3 KNN model

In the refined implementation of the YNA model, the processed dataset is utilized as a training set. This setup enables the application of the K-Nearest Neighbors (KNN) algorithm to analyze the accuracy of credit ratings across various enterprises. It was observed that when the number of neighbors, denoted as 'n', is set to 2, the model attains a credit rating accuracy of 72% for the enterprises, which is considered to be an acceptable level of precision. Building on this insight, the algorithm is then applied to predict the credit ratings of enterprises in a separate sample dataset.

Survival Years	Number of Clients	Profit	Return Rate	Credit Rating
2	52	27119655	0.009367	0
3	63	-1.7E+07	-0.00362	0
3	143	-5.2E+08	-0.19848	0
3	25	-7.2E+08	-0.24935	0
3	7	-2.5E+08	-0.31975	1
3	796	-3.2E+08	-0.18826	0
3	14	-4E+07	-0.07992	1

Table 2 presents the majority of enterprises, particularly those with negative profits and lower return rates, are classified with a credit rating of 0, indicating a lower creditworthiness. In addition, there's no apparent direct correlation between survival years, number of clients, and credit rating. For example, enterprises with a higher number of clients (e.g., 796 clients) still received a lower credit rating.

3.4 Logistic model

Following the implementation of the YNA model for value stratification and the K-Means method for credit rating clustering, the KNN model was subsequently applied to forecast the credit ratings of corporations in a secondary dataset. Subsequent to this, the logistic regression model, initially developed using the first database, was employed to ascertain the probability of default occurrences. In this initial application, the logistic regression model, leveraging other variables, achieved a prediction accuracy of 97.5% for default occurrences. The model was then further utilized, in conjunction with the sklearn library, to predict the occurrence of defaults in the second database

Survival Years	Number of Clients	Profit	Return Rate	Credit Rating	Non-compliance
2	52	27119655	0.009367	0	否
3	63	-1.7E+07	-0.00362	0	否
3	143	-5.2E+08	-0.19848	0	否

4. Conclusion

This study primarily focuses on extracting the underlying information from existing data, aiming to maximize the revelation of deep insights. By constructing models and employing machine learning algorithms, the research facilitates value stratification of target enterprises and enables precise predictions for missing data, thereby enhancing the efficiency of commercial bank lending. Utilizing the RFM value stratification model based on current data, a novel enterprise value model was established, delving deeper into data stratification for businesses. Subsequently, the K-Means algorithm was applied for secondary clustering to refine the accuracy of these classifications. Finally, on the foundation of the existing model, deep learning predictions for missing data were conducted using both the KNN model algorithm and the logistic regression model, illustrating the potential of these methodologies in advanced data analysis and prediction. To mitigate credit risk, this study undertook a comprehensive analysis, processing, modeling, and visualization of existing data. The data mining process for lending enterprises was categorized into three key aspects: years of establishment, number of clients, and average annual return rate, facilitating stratification of businesses. The information gleaned was used to develop model algorithms for evaluating and predicting enterprises with missing data. This approach not only ensured completeness of information for businesses with incomplete data but also facilitated effective analysis of real-world lending operations. Consequently, it addressed the issue of unsuccessful loan applications due to missing data parameters, thereby accelerating the loan processing speed for small and medium-sized enterprises (SMEs) and resolving issues related to missing data in loan applications. This approach has significantly mitigated the challenges faced by SMEs in securing loans, thereby advancing the practical efficiency of lending institutions.

Reference

[1] Fang Wei, Zheng Yu, Xu Jiang. A Comprehensive Review on Big Data: Concepts, Technologies, and Applications. Journal of Nanjing University of Information Science & Technology, 2014, Volume 6(5): 405-419.

[2] Chen Chunzhao, Xie Rui, Zha Jingyi, Zhu Jiaming. Research on Credit Risk Assessment and Lending Strategies for SMEs Based on Big Data. Journal of Harbin Normal University, 2021, Volume 37(4): 20-30.

[3] Lv Xiumei. Credit Assessment of SMEs in the Era of Big Data Finance. Financial Monthly, 2019, (13): 22-27.

[4] Sun Tingting, Yang Ting, Xu Junfan, Huang Shaolang. Investigation and Research on the Financing Status of SMEs under the Background of Digital Finance. Small and Medium Enterprises Management and Technology, 2022, (23): 168-170.

[5] Guo Tongdan. Analysis of the Impact of Digital Finance on the Financing Environment of SMEs. Science and Technology Information, 2022, Volume 20(21): 135-138.

[6] Gao Bojin. Analysis of Internet Financial Financing Models — Solutions to the SME Loan Problem. Modern Market, 2018, (1): 121-122.

[7] Meng Xiaofeng. Meng Xiaofeng: Challenges of Opening Big Data. China Education Network, 2014, (4): 23.

[8] Huang Xia. Discussion on Financing Channels for SMEs in the Era of Big Data. Mass Investment Guide, 2021, (27).

[9] Zhang Junjun. Exploration and Practice of Big Data and Artificial Intelligence in the Financing Field of Small and Micro Enterprises. China Financial Computer, 2022, (7): 50-52.

[10] Zhang Huixia. Discussion on the Financing Problems of SMEs under the Background of Big Data and Internet Finance. Knowledge Economy, 2017, (1): 32-33.

[11] Ge, M. (2021). Construction of Intelligent Risk Control System for Commercial Banks in the Era of Big Data. International Journal of Frontiers in Sociology, 3(11).

[12] Bhuvana, M., Thirumagal, G. P., & Vasantha, S. (2016). Big Data Analytics - A Leveraging Technology for Indian Commercial Banks. Indian Journal of Science and Technology, 9(32).

[13] Zhou, J., & Jie, Z. (2020). Analysis on the Innovation Hierarchy of Commercial Bank's Financial based on the Big Data. Journal of Physics: Conference Series, 1648(2), 022108.

[14] Kang Xiaodi. The Impact of Internet Finance on Commercial Bank Credit Business and Bank Countermeasures. China Market, 2021, (29): 34-35.

[15] Lu Zhiyuan. Exploring the Financing Relief Path for SMEs — From the Perspective of Digital Finance. Northern Economy and Trade, 2023, (1): 132-134.