Application of Data Mining Technology in Financial Data Analysis Methods under the Background of Big Data

1st Zhengyin Wang
Shandong Jiaotong University
Jinan, Shandong, China
wt165198764@163.com

2nd Yan Wang
Shandong Jiaotong University
Jinan, Shandong, China
wangyan19860218@163.com

3rd Aruna T M
Department of AI and ML
Nitte Meenakshi Institute of Technology
Bengaluru, India
aruna.tm@nmit.ac.in

Abstract—With the development of the times, the financial industry has gradually become the focus of people's attention. Financial data analysis can help banks analyze various information of customers and decide which business to engage in with them. However, the analytical capabilities of current financial data analysis methods are still insufficient, so new technologies need to be used to improve. The application of data mining technology in financial data analysis methods under the background of big data studied in this article aimed to improve the analytical ability of financial data analysis using data mining technology. This article tested the repayment rate of customers after using data mining technology through experiments. The data showed that the highest repayment rate was 92%, indicating that financial data analysis using data mining technology is very accurate in analyzing customers' repayment ability. The experimental test showed that the use of data mining technology can increase bank revenue by up to 19%, indicating that data mining technology can bring significant benefits to banks. Through these two experiments, it can be proven that data mining technology and financial data analysis methods have a high degree of compatibility.

Keywords—financial data, big data background, data mining, data analysis

I. INTRODUCTION

The financial industry often uses financial data analysis methods to analyze various aspects of customer data, which is beneficial for banks to evaluate customers and decide whether business can be processed for them. This is very helpful for banks to reduce unnecessary losses. How to optimize financial analysis methods is currently the focus of research, and improving the analytical ability of financial analysis methods can bring greater benefits to banks.

Many scholars have conducted extensive research on the financial industry. Zetzsche D A believed that finance is in a paradigm shift, shifting from a framework based on neoclassicism to a framework based on psychology. Behavioral finance is the application of psychology in financial decision-making and financial markets. Behavioral finance is the process of replacing neoclassical assumptions with corresponding behavioral assumptions [1]. Aziz H M believed that the development of the financial industry has led to an increase in its ability to diversify risks, as well as an increase in the economy's risk tolerance and actual risk bearing capacity, resulting in a series of financial transactions that have so far been impossible and creating greater financing channels for businesses and households [2]. Goldstein I explored whether there exists a threshold beyond which financial depth no longer has a positive impact on economic growth [3]. Although there is currently a lot of research in the financial industry, there is not much research on financial analysis.

Financial industry related units can use data mining technology to discover the inherent meaning of financial information at a deeper level, which provides a strong basis for unit development decision-making and enables relevant units to occupy a leading position in the fierce financial market competition. Therefore, this article studied the application of data mining technology in financial data analysis methods under the background of large data.This article tested the improvement of customer repayment rate and bank income after using data mining technology through experiments, and found that the data was good, indicating that the use of data mining technology in financial data analysis is good.

II. USE OF DATA MINING IN FINANCIAL DATA ANALYSIS METHODS

A. Overview of Data Mining Technology

Data mining refers to the process of searching hidden information from a large number of data by algorithms. Data mining is often associated with computer science and is accomplished through a variety of methods, including statistics, online analytical processing, intelligence retrieval, machine learning, expert systems (relying on past rules of thumb), and pattern recognition. Data mining is a technique to find the law from a large number of data by analyzing each data, which mainly includes three steps: data preparation, law finding and law representation. Data preparation is to select the required data from the relevant data sources and integrate it into the data set for data mining; Law finding is to find out the law contained in the data set by some method; The rule representation is to represent the found rule as much as possible in a way that the user can understand (such as visualization). The tasks of data mining include association analysis, cluster analysis, classification analysis, anomaly analysis, specific group analysis and evolution analysis.

The essence of data mining is to identify potential knowledge from databases. This is also an important research focus in disciplines such as artificial intelligence and data processing.

Data mining refers to the systematic and automated analysis of data from various industries based on big data technology, statistics, automatic learning, artificial intelligence, and recognition technology. In short, it is necessary to operate a network, identify possible patterns and potentials, guide market trends, reduce risks, and enable decision-makers to make accurate predictions. From the technical point of view, data extraction requires extracting

implicit and useful information from massive Uncertain data, and ensuring that the extracted information and knowledge are unknown in advance, so as to ensure its potential value. Problem data includes defective data, incomplete data, and arbitrary data.

This method requires real and large-scale data sources, the extracted information that can be cared for by the target user, and can be applied in practice. The source data is a working database at various levels, which can be structured data from relational databases or semi structured data, such as partial textual data, image data, etc. It's more like data scattered throughout the entire network without any connection.

Among them, deductive reasoning, mathematical model method, non-mathematical model method and other methods can be used to extract useful information and knowledge. This information can be used for information retrieval, rapid retrieval, process control, etc., and can also be used to save data. So, exploring big data through data fusion technology, from a single individual to complex representations, from simple appearances to hidden knowledge, has become the core technology for decision-making and guidance in today's society.

B. Use of Data Mining in the Financial Industry

The establishment of data warehouses and diversified data analysis: Data mining has outstanding performance in financial transactions and specific applications, because the collection and processing of financial transactions often require a large amount of data analysis, and most banks and financial institutions often provide a lot of services to society, which leads to the frequent and complex transmission of

information. Due to the asymmetry of information transmission, it is difficult for financial institutions to find the correct corresponding information from the vast amount of data. Financial data mining technology can quickly and accurately discover useful information from massive amounts of data [4-5].

Risk management and credit assessment: Applying data mining analysis methods to the financial system, in addition to establishing an information base and conducting comprehensive analysis, is actually more about mining and analyzing deep-seated data. An important application of data mining technology in the financial field is credit risk assessment [6-7]. Due to the risks involved in the trading process, for securities firms and banks, it is necessary to ensure the safety of customers' funds as much as possible. By analyzing the potential risks of products through data mining, a comprehensive product credit risk assessment can be made and corresponding solutions can be provided [8-9].

C. Evaluation Methods of Data Mining in the Financial Industry

To apply data mining technology to banks, an important prerequisite is to build a unified central customer database, thereby improving the ability to analyze customer information. At the beginning of the analysis, all customer related information, transaction records, model building, analysis data, and prediction of customer future behavior were collected from the database. The specific application of data mining in the banking industry can be divided into three stages, and the specific application of data mining in the banking industry is shown in Figure 1.
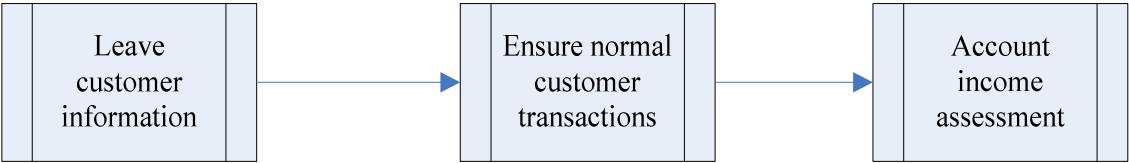

Fig.1. The concrete application of data mining in banking industry

The first stage is mainly to leave customer information and lay the foundation for integration into the central customer information database. At the same time, it is necessary to eliminate the inconsistency between the existing customer information and the information in the database and ensure the correctness of the data. Its accuracy is very helpful for various industries related to finance [10-11].

The second stage is to ensure the normal transaction of financial customer information, which includes the economic transformation of the securities industry, the distribution stage of banks, counters, third-party storage platforms, remittances, transfers, purchases of financial products, etc. This can better mobilize the connection between customers and enterprises, making future operations more efficient [12-13].

The third stage is to evaluate the returns of each account. Based on the revenue situation of each account, an evaluation is conducted to assess future revenue and input the data from the system into the central database. After this process is completed, banks can analyze profit contribution from three perspectives: organization, users, and products. For example, banks can arrange appropriate distribution channels based on customer profit contribution, simulate and predict the profit contribution of new products to the bank [14-15].

D. Impact of Data Mining on Banking and Finance

In recent years, with the rapid development of data mining technology, traditional banking business has been greatly affected and facing enormous challenges. The challenges faced by traditional banks are shown in Figure 2.

Fig.2. Challenges facing traditional banks

1) The impact and challenge on the corporate culture of the banking industry: This is largely because banks have

shifted from liking complete and pure data to liking unstructured data, allowing for inaccurate and incomplete

data; expanding from local shocks to global shocks, in the era of big data, banks need to use massive amounts of data to shock corporate culture, thus colliding with a brand new corporate culture that is suitable for the new era [16-17]. (2) The impact of digitization on the banking industry: In the big data environment, the challenges faced by the banking industry are: on the one hand, the development of big data technology has promoted the research and development of new products and services in the banking industry. On the other hand, it is to address the structural deficiencies of traditional data [18-19]. (3) New issues in the transformation of the banking industry: In the era of big data, the development of banks must undergo rational transformation, and the key to this transformation is the ability to analyze and utilize bank data rationally [20].

E. Application of Big Data in the Financial Field

Personalized customization and precise real-time marketing: Commercial banks are gradually transitioning from "product oriented" to "customer oriented", and customer needs and preferences are the source of bank development. Personalized customization refers to commercial banks recommending financial products and services to customers based on their preferences, and accurately positioning them based on their economic scale, financial preferences, and age region, in order to carry out targeted advertising. Precision real-time marketing refers to the promotion and marketing of customers based on their current actual situation, and a comprehensive examination of their location, recent financial behavior, and consumption records, in order to achieve real-time and effective tracking of customers and achieve accurate services.

1) Customer management: Banks can use the method of building a customer relationship warehouse to manage their own customers. On this basis, data mining tools are used to finely divide customer groups and analyze customer behavior patterns. Based on different customer groups, banks provide personalized and accurate financial and loan services to them. In addition, using big data warehouse technology can predict the customer flow and loss of commercial banks, providing decision-making basis for banks to explore new customers and maintain old customers. Banks use clustering to naturally classify, analyze, predict, and optimize customers, and achieve profitability by analyzing customer returns, service fees, risks, and other factors.

F. Data Mining Technology Exploration Methods

1) Cluster: Cluster analysis refers to the data found in data to illustrate the relationships between objects in the data. The target has similarity (correlation) between targets in the group, and differences (irrelevance) between targets in the group. As the similarity within the group increases, the differences between groups increase, and the differences between groups increase, resulting in better clustering results for the group.

2) Classification and prediction: Classification mainly refers to predicting classification labels (based on discrete attributes), while prediction refers to building a continuous value function model to predict the values of the dependent variable corresponding to a given independent variable. The advantage of this method is that it can reflect the requirements

of data items through certain data characteristics. This method can be applied to analyze consumer purchasing trends. This is a functional model, and when data accumulates to a certain extent, it becomes more valuable and continuously stored.

3) Feature analysis method: Using feature analysis method, feature data is extracted from the database based on customer categories, and the overall characteristics of the data are clarified. After identifying the reasons for customer churn, corresponding measures are taken to retain customers.

4) Association rule method: Association rule method refers to the method of inferring one or more data items from one data item based on the relationship between them. The use of association rule method for financial data mining can be divided into two stages. One is to mine all high-frequency project groups in the vast financial data information center, and the other is to propose other project categories based on the relevance of high-frequency project groups themselves. At present, the association rule method has been widely applied in consumer demand prediction. For example, banks would package the information they care about into ATMs based on customer expenses to provide the information they need. This not only greatly improves customer satisfaction, but also improves their marketing management.

G. Specific Applications of Financial Data Analysis Methods in Data Mining

1) Bank loan repayment prediction: As a financial institution, a bank must use data mining technology to conduct deeper analysis of the data. The scope of bank lending is very extensive, and in order to ensure timely disbursement, it is necessary to prepare for disbursement in advance. Predicting loan repayment based on customer information can also better identify customer credit data, reduce bank loan losses, and achieve maximum economic benefits. Both lending and credit card transactions are indispensable for banks. However, the repayment of loans and the evaluation of customer credit ratings are influenced by many factors, such as loan interest rates, loan terms, income levels of lenders, and the education level of lenders. The application of financial data can effectively eliminate these irrelevant factors. By using data mining, it can be seen that the cultural level, loan term, and repayment speed of the lender are not closely related. However, the income level of the borrower is a relatively important factor for the lender. In this way, banks can formulate corresponding lending plans based on the borrower's income and repayment level, and start lending.

H. Use of Algorithms in Data Mining Technology

Deep neural network model is often used in data mining. The deep neural network uses unsupervised learning pre training to generate initial weights. In a deep neural network, the feature observation vector of timing a is n, and the calculation method for calculating the probability of the optimal training method in the deep neural network is shown in Formula (1):

$$P(\mathrm{d}\,|\,\mathrm{n}) = \frac{\exp\{z(d)\}}{\sum_d \exp\{z(d)\}} \tag{1}$$

In Formula (1), z(d) is the state of the output layer, and d is the output value. The probability calculation method for the optimal output layer is shown in Formula (2):

$$P = P(d \mid n) - \log P(d) \qquad (2)$$

In Formula (2), P(d) represents the probability of state d appearing in the training data.

For deep neural networks, cross entropy is generally used as the objective function, and this process is usually realized by the random gradient descent algorithm. For multi state classification, the negative value of the logarithmic probability should be used as the objective function, as shown in Formula (3):

$$y = \sum_{a=1} \log(d) \qquad (3)$$

Formula (3) is the cross entropy between the functional state and the predicted state distribution.

III. SIMULATION EXPERIMENTS ON THE USE OF DATA MINING IN FINANCIAL DATA ANALYSIS

A. Experiment

Data mining can better identify customer credit data and determine whether to lend to customers. This article tested the customer repayment rates of five banks using data mining technology for financial data analysis. The repayment rate of customers after financial data analysis using data mining technology is shown in Figure 3.
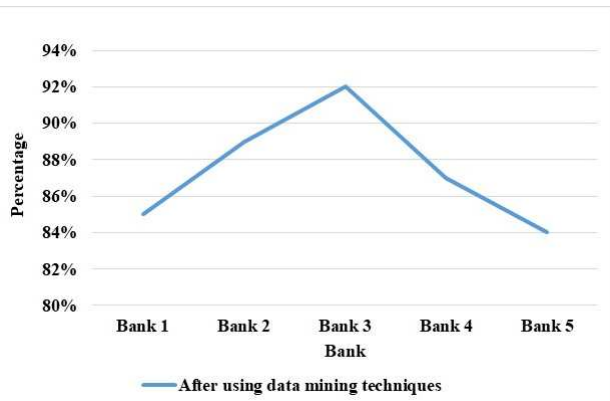

Fig.3. Using data mining technology to analyze financial data after customer repayment rate

From the experimental results in Figure 3, it can be seen that the highest and lowest repayment rates of financial data analysis customers using data mining techniques can reach 92% and 84%, respectively. This indicated that with the support of data mining technology, banks can have a clearer understanding of customers' repayment ability, so they can achieve such a high repayment rate, which also proved that data mining technology has a good effect in financial data analysis.

This article also tested the revenue improvement of banks after using data mining technology to analyze financial data. This article still selected five banks to test their use of data mining technology for financial analysis of bank income. The test results are shown in Figure 4.

From the experimental results in Figure 4, it can be seen that the revenue of banks increased by at least 12% and at most 19% after using data mining technology for financial data analysis. From this experimental data, it can be seen that banks have a comprehensive understanding of customer data when using data mining for financial analysis, which can reasonably handle business for customers and avoid situations similar to loans that customers cannot afford.

B. Discussion

Data mining is a process of extracting useful information and patterns from massive data. It includes the collection, extraction, analysis and statistics of data, also known as the process of knowledge discovery, that is, knowledge mining from data or data pattern analysis. This is a logical process of looking for useful information to find useful data.

One of the most important tasks in data mining is to choose the right data mining technology. Data mining techniques must be selected according to the type of business and the problems the business faces. Therefore, a general approach must be used to improve the accuracy and cost effectiveness of using data mining techniques. This article focuses on seven data mining techniques that are considered to be more commonly used by business people.
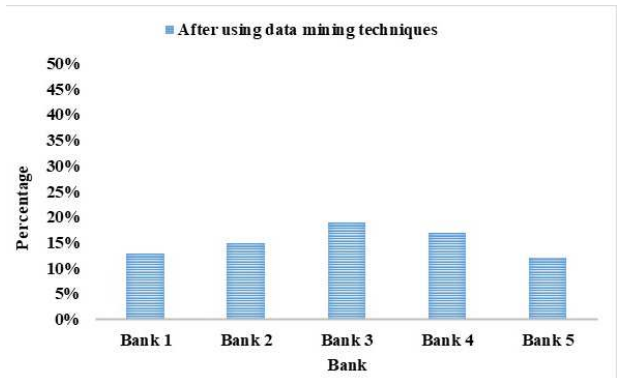

Fig.4. The use of data mining technology after financial analysis of the bank revenue situation

IV. CONCLUSIONS

The development of the financial industry is now becoming increasingly rapid. The business of banks is also gradually increasing, and people are also going to banks to handle business more and more. So now, data mining has begun to be widely used in information processing in the financial industry, which would have great value for the financial industry. This article focused on the application of data mining technology in financial data analysis methods under the background of big data, which aimed to further enhance the analysis of customer behavior through data mining technology and achieve more accurate business management for customers. This article tested the repayment rate of financial data analysis using data mining technology through experiments and found that the repayment rate was very good. This article also tested the income of banks after using data mining technology for financial data analysis, and the income also increased significantly. This indicated that data mining technology has a good effect in financial data analysis. Due to space limitations, the experimental testing conducted in this article is still insufficient and would be improved in the future. Finally, it is hoped that the financial industry would grow better and better.

The corresponding author is Yan Wang.

REFERENCES

[1] Zetzsche D A, Arner D W, Buckley R P. Decentralized finance[J]. Journal of Financial Regulation, 2020, 6(2): 172-203.

[4] Burton B, Kumar S, Pandey N. Twenty-five years of The European Journal of Finance (EJF): a retrospective analysis[J]. The European Journal of Finance, 2020, 26(18): 1817-1841.

[5] De Goede M. Finance/security infrastructures[J]. Review of International Political Economy, 2020, 28(2): 351-368.

[6] Malmendier U. Experience effects in finance: Foundations, applications, and future directions[J]. Review of Finance, 2021, 25(5): 1339-1363.

[7] Ozili P K. Contesting digital finance for the poor[J]. Digital Policy, Regulation and Governance, 2020, 22(2): 135-151.

[8] Ozili P K. Impact of digital finance on financial inclusion and stability[J]. Borsa Istanbul Review, 2018, 18(4): 329-340.

[9] Gomes F, Haliassos M, Ramadorai T. Household finance[J]. Journal of Economic Literature, 2021, 59(3): 919-1000.

[10] Jensen J R, von Wachter V, Ross O. An introduction to decentralized finance (defi)[J]. Complex Systems Informatics and Modeling Quarterly, 2021 (26): 46-54.

[11] Cohn J B, Liu Z, Wardlaw M I. Count (and count-like) data in finance[J]. Journal of Financial Economics, 2022, 146(2): 529-551.

[2] Aziz H M, Sorguli S, Hamza P A. Factors affecting International Finance Corporation[J]. International Journal of Humanities and Education Development (IJHED), 2021, 3(3): 148-157.

[3] Goldstein I, Spatt C S, Ye M. Big data in finance[J]. The Review of Financial Studies, 2021, 34(7): 3213-3225.

[12] Hasan M M, Yajuan L, Khan S. Promoting China's inclusive finance through digital financial services[J]. Global Business Review, 2022, 23(4): 984-1006.

[13] Co Berg T, Reisinger M, Streitz D. Spillover effects in empirical corporate finance[J]. Journal of Financial Economics, 2021, 142(3): 1109-1127.

[14] Cong L W, Li Y, Wang N. Token-based platform finance[J]. Journal of Financial Economics, 2022, 144(3): 972-991.

[15] Sun Y, Shi Y, Zhang Z. Finance big data: management, analysis, and applications[J]. International Journal of Electronic Commerce, 2019, 23(1): 9-11.

[16] An J, Rau R. Finance, technology and disruption[J]. The European Journal of Finance, 2021, 27(4-5): 334-345.

[17] Graham J R. Presidential address: Corporate finance and reality[J]. The Journal of Finance, 2022, 77(4): 1975-2049.

[18] Khan M A, Riaz H, Ahmed M. Does green finance really deliver what is expected? An empirical perspective[J]. Borsa Istanbul Review, 2022, 22(3): 586-593.

[19] He J, Tian X. Finance and corporate innovation: A survey[J]. Asia-Pacific Journal of Financial Studies, 2018, 47(2): 165-212.

[20] Hawkes A G. Hawkes processes and their applications to finance: a review[J]. Quantitative Finance, 2018, 18(2): 193-198.