Implementation of Bigdata Analysis in Banking: Debt Default, Anti-money Laundering, Investment Behavior Forecasting

Qimo Hu[1, *], Haoting Li[2], and Yangning Liu[3]

[1]Department of Computer Science, China Three Gorges University, Yichang, China

[2]Jinan xinhang experimental foreign language school, Jinan, China

[3]Zhengzhou forty-one Senior Middle School, Zhengzhou, China

*Corresponding author: 202110110106@ctgu.edu.cn

Abstract. With the development of computer science and technology, people live become more and more convenience. The bank's account security and investment system have also become more and more perfect due to the technology. In our article, we will focus on analyzing debt default system, anti-money laundering system, and investment prediction system. In order to know about the structure and operation of the banking system deeply. After studying and research, we understood the law and trend of global economic development in recent years, including the economic turning down because of COVID-19 from 2021, and the recovery from 2023. Besides, we learned several program systems such as ARIMA, AML and KYC solutions and so on, which can be also used in other business fields. In summary, the meaning of writing this article is not only develop our interests, but also learn more about the technology that touches out lives, So, it will be easier for us to manage our finance and understand the operation rule of society in future.

Keywords: ARIMA; AML; debt default.

1. Introduction

The banking industry has a long and prosperous history, dating back to ancient civilizations. In history, various forms of financial institutions developed into the modern banking system that we observe in today's society. Ancient civilizations such as Babylon, Assyria, and Egypt provided important financial services like loans, savings accounts, and currency exchange. Similarly, Ancient Rome and Greece also developed similar financial institutions, often closely related to religious organizations or governments. The pioneers of our modern banking system can be traced back to the Renaissance period in Venice, Italy. Venetian merchants played a crucial role by providing loans and other financial services in various ways. This led to the establishment of specific regulations and institutions that gradually evolved into the modern banks we recognize today. For example, Venetian trade companies, anonymous partnerships, and bills of exchange. The 17th century marked an important milestone in the practice of modern banking. Amsterdam, in particular, became the headquarters of the Dutch East India Company. It was not only the first publicly traded company but also a true modern bank. The success of the company paved the way for business models that attracted significant investments and promoted financial innovation. The 18th century further advanced the banking industry. In particular, in 1694, the Bank of England was established as a unique central bank responsible for managing currency issuance and regulating economic activities [1-3].

With the progress of the 19th and 20th centuries' industrial revolution and globalization trends, the banking industry experienced rapid growth on a global scale. Countries began establishing their own central banks and witnessed the emergence of various types of financial institutions, including commercial banks, investment banks, credit cooperatives, and more. In today's era, with advancements in technology and financial innovations, new opportunities have emerged in the banking industry. Online banking, mobile payments, and blockchain, among other advanced technologies, have changed the way people interact with finance to meet the evolving market demands.

In order to improve their consumer loan business, banks can utilize big data analytics to gain a better understanding of their customers' preferences and needs. This allows them to customize their

services and products to meet specific requirements. Additionally, analyzing customer data enables banks to classify customers based on their risk profile, identifying those who are more likely to default on their loans and those who are less risky.

Furthermore, the use of big data analytics in consumer loans can help banks predict future trends in consumer borrowing. By analyzing historical data, banks can identify patterns and indicators that signal when customers are more likely to seek a loan. This information allows banks to proactively offer loan products to customers based on their past behaviors and financial needs.

It is important to note that effective utilization of big data relies on the bank's ability to make informed decisions based on insights derived from data analysis. The data should be used to develop strategies and policies that can prevent default accounts and mitigate risk in consumer loans. This includes setting parameters and implementing effective risk management practices [4].

Overall, integrating big data analytics in consumer loans has the potential to significantly benefit banks by providing deeper customer insights, enhancing risk management capabilities, and enabling the development of targeted loan products and offers. However, it is important for banks to handle customer data responsibly and in accordance with privacy regulations to maintain customer trust and comply with legal requirements [5, 6].

The banking industry, with the development of digitization and financial technology, is increasingly relying on big data analytics. Banks can utilize big data analytics to analyze and optimize various aspects such as customer risk, fraud detection, marketing recommendations, innovative product development, and operational efficiency to improve service quality and efficiency, and achieve sustainable development. However, in the process of applying big data analytics, banks need to pay attention to the protection of customer privacy and data security, and comply with relevant laws and regulations to ensure the legality and transparency of data usage. In summary, the banking industry can achieve better business development and customer service goals through the application of big data, while actively addressing the challenges and risks brought by big data applications.

2. Big Data Analysis

Big data analysis can be a valuable tool in understanding and predicting debt defaults. By analyzing large volumes of data from various sources, such as financial statements, credit reports, economic indicators, and social media sentiments, organizations can gain insights into the likelihood of debt default and take proactive measures to mitigate risks. The fintech development can reduce the likelihood of debt defaults by alleviating financing constraints and providing better investment opportunities. Additionally, industrial policy support plays a moderating role in the influence of fintech on debt default risk. This research provides valuable insights for the fintech industry and policymakers, enabling them to effectively manage and address the risk of corporate debt default, thereby promoting stable economic development. Unhealthy consumer loans will become a debt burden and create instability financial conditions, these instability conditions will expose the costumer to others disadvantages such as social exclusion and physiological and health issues . In Germany, the growth of consumer loan relatively stable, the total amount of consumer loan to an individual in the half-year of 2019 is approximately 1.254 million Euro. Compare to only 871.4 million Euro 20 years ago in 1999 [4-6].

3. Debt default

In the context of predicting debt default, a neural network can be trained using historical data that includes various financial and non-financial factors as input features. These factors may include credit scores, income levels, employment status, loan amount, payment history, and economic indicators. Neural networks aim to mimic the human brain but have notable distinctions. While the human brain is capable of independent learning and forms unique connections based on individual experiences, an

artificial neural network typically begins with predefined relationships. Both the human brain and artificial neural networks learn to identify patterns and uncover relationships in data [7, 8].

Patterns and relationships in artificial neural networks are represented by neurons, which collectively produce an output. If the predicted results do not align with actual outcomes, adjustments must be made to these neurons. One advantage of artificial neural networks over the human brain is their ability to analyze and recognize patterns within large datasets. Artificial neural networks excel at processing vast amounts of data that would overwhelm a human brain. The Fig. 1 and Fig. 2 illustrates how an artificial neural network system operates. The inputs represent independent variables, while the hidden layer consists of neurons responsible for processing this input data. Finally, the output layer represents the dependent variable – in this study's case: allowance for bad debt.

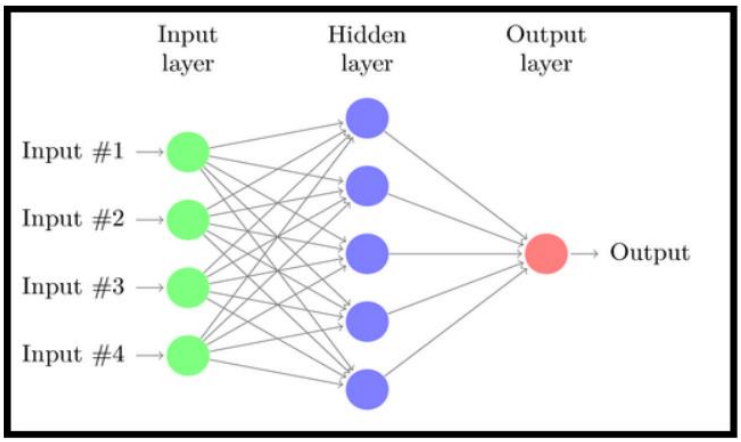


Fig. 1 A sketch of neural network.

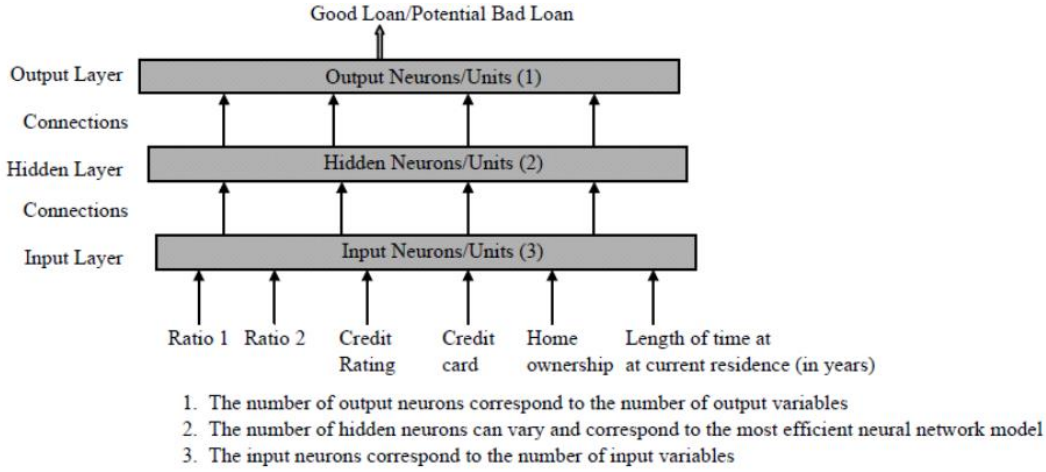


Fig. 2 Judgement of the loan.

There are two paradigms for classifying information systems. The first paradigm focuses on explaining human or organizational behavior through theories. On the other hand, the second paradigm adopts a design science approach that aims to enhance and support human and organizational capabilities by inventing new and innovative computer-based solutions known as "artifacts". These artifacts, according to Hevner et al. aim to improve user experience, creativity, intuition, and problem-solving abilities.

4. Anti-money laundering

The application of big data analysis in anti money laundering scenarios in the banking industry can cover multiple aspects. Firstly, banks can identify suspicious transaction patterns by monitoring customer transaction data. Secondly, banks can use big data analysis technology to analyze customer behavior patterns in order to detect abnormal activities in a timely manner. In addition, big data

analysis can also be used to construct customer profiles, helping banks better understand customer risk characteristics [9].

In the context of anti money laundering in the banking industry, commonly used big data analysis models include association rule mining, clustering analysis, and anomaly detection. Association rule mining can help banks discover the association relationships between different accounts, thereby identifying potential money laundering behaviors. Cluster analysis can divide customers into different groups to better identify abnormal behavior. Anomaly detection can detect abnormal activities in a timely manner by monitoring customers' transaction patterns.The principles of big data analysis mainly include data collection, data cleaning, data storage, data analysis, and result display. Firstly, banks need to collect a large amount of transaction data and customer information. Then, clean the data to remove noise and outliers. Next, store the cleaned data in an appropriate database for subsequent analysis. In the data analysis stage, banks can use various big data analysis algorithms and models, such as association rule mining, clustering analysis, and anomaly detection, to conduct in-depth analysis of the data. Finally, the analysis results will be presented to help banks better understand the risk characteristics of customers. One of the most commonly used technologies related to large-scale processing of big data is Hadoop, and the processing (actual processing) of data when it is created is called Apache Storm. The Hadoop ecosystem consists of two key units:

HDFS(Hadoop Distributed File System): the distributed file system is responsible for data storage. HDFS splits data into blocks, copies them and stores them in Hadoop cluster. It aims at networking more efficient computers, thus forming a computer system called. In HDFS system, there are many independent components and applications.Use as needed, for example, HBase is used as a repository for data storage, Hive is used as a data warehouse, and Pig is used as a platform for creating MapRaduce programs (advanced data streams).

MapRaduce: a software framework for distributed data processing, which allows large-scale scalability on thousands of servers in Hadoop cluster. Function Map reads raw data, filters it and creates results in the form of key values. After grouping all the values associated with the key, reduce the number of processes. All values are associated with the same key and produce a smaller set of outputs.

Through the implementation of big data analysis in anti money laundering scenarios in the banking industry, the following results can be achieved:

Identify potential money laundering behaviors. Through association rule mining and clustering analysis, banks can discover the association relationships and abnormal behaviors between accounts, thereby identifying potential money laundering behaviors

Improve the accuracy and efficiency of anti money laundering. Big data analysis can help banks identify suspicious transactions and abnormal activities more quickly and accurately, improving the accuracy and efficiency of anti money laundering.

Building customer profiles. Through big data analysis, banks can segment and classify customers, construct customer profiles, better understand customer risk characteristics, and take corresponding risk control measures.

Improving risk management capabilities. Big data analysis can help banks better monitor and manage risks, timely detect, and respond to potential risk events, and protect the interests of banks and customers.

The application of big data in the anti-money laundering work of China banks at this stage also shows that the work of protecting the property safety of all parties such as anti-moneylaundering has entered a new stage, and now we have gradually adapted to the new form of anti-money laundering work

5. Investment Behavior Forecasting

The next topic, we will focus on investigating forecast. The stock market has always played an important role in the financial field. Fluctuations in stock prices can reflect changes in a country's

economic cycle to some extent. Investment forecasting is the qualitative and quantitative analysis and measurement of investment benefits, to make scientific judgment. We will talk about one of computer technology: ARIMA model.An ARIMA model for time series forecasting is represented as "AR(p) I(d) MA(q)." The values of "p," "d," and "q" are determined through data analysis and model selection techniques, such as AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). The ARIMA model construction process involves several key steps:

Checking for Stationarity: First, assess whether the original data sequence is stationary. If it isn't, apply methods like differencing to make it stationary.

Determining Model Order: Calculate autocorrelation and partial autocorrelation coefficients using software. The lag orders 'p' and 'q' for the ARIMA model are determined using other methods.

Building the Model: It's important to ensure that the model's residual sequence follows a normal distribution. You can analyze autocorrelation and partial autocorrelation of the residuals or test their randomness using correlation tests.

Making Predictions: Once the model is constructed, you can use it to make short-term predictions, such as forecasting stock prices.

For instance, let's predict the daily opening price of Bank of China's stock (601988) from April 10, 2022, to April 10, 2023, using an ARIMA (0,1,1) model with 'p = 0', 'd = 1', and 'q = 1' for the next three days. In summary, constructing an ARIMA model involves checking data stationarity, determining model order, building the model while ensuring normality of residuals, and then using the model for predictions. As an example, we applied this process to predict the daily opening price of Bank of China's stock. After making the time series data stationary, use ACF and PACF diagrams to identify model characteristics and determine lag values 'p' and 'q.' In FIG. 2, most autocorrelation values remain within twice the standard deviation, suggesting a first-order lag. Fitting Test of the model After completing the above parameter estimation steps, the most critical step is to test the normality of the residual sequence of the fitting model. The majority of investors in the stock market are retail investors, which leads to the obvious phenomenon of following the trend of investment in the market, so rational education for investors is essential. Investors should make choices based on their actual situation when buying or selling securities [10].

6. Limitations and Prospects

Nevertheless, there are some shortcomings for implementation of bigdata analysis in banking industry. First of all, there are some limitations in the implementation of debt default. Although big data analysis can help banks identify potential default risks, it still cannot fully predict the occurrence of debt default. Debt default is often affected by many factors, including economic environment, industry development, policy changes, etc. These factors are difficult to capture and predict completely through data analysis. In addition, big data analysis cannot solve the fundamental problems of debt default, such as the borrower's credit status and repayment ability.Secondly, there are some limitations in the implementation of anti-money laundering. Although big data analysis can help banks identify suspicious transactions and money laundering, it still cannot completely eliminate the problems of false positives and false negatives. Big data analysis often depends on the accuracy of models and algorithms, but these models and algorithms may have certain errors and limitations. In addition, money laundering is often hidden and variable, and it is difficult to accurately identify it by traditional data analysis methods.Finally, there are some limitations in the implementation of investment behavior prediction. Although big data analysis can help banks predict market trends and investment opportunities, it still cannot completely eliminate market risks and uncertainties. Market changes are often influenced by many factors, including economic policies, geopolitics, natural disasters, etc. These factors are difficult to accurately predict through data analysis. In addition, investors' risk preference and long-term investment strategy need to be considered in investment decision-making, which is difficult to be comprehensively considered through data analysis.

7. Conclusion

To sum up, the implementation of big data analysis in banking industry in debt default, anti-money laundering and investment behavior prediction is of great significance for risk management and business decision-making in banking industry. Including can help banks improve risk management capabilities, reduce losses, but also can improve the accuracy and efficiency of business decisions. These applications can not only bring economic benefits to the banking industry, but also improve the stability and security of the financial system.

Author Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Goyal K A, Joshi V. Indian banking industry: Challenges and opportunities. International Journal of Business Research and Management, 2012, 3(1): 18-28.

[2] Gupta K, Gauba T, Jain S. Big data in hospitality industry: A survey. International Research Journal of Engineering and Technology, 2017, 4(11): 476-479.

[3] Peng Y. The Chinese banking industry: lessons from history for today's challenges. Routledge, 2007.

[4] Benston G J, Kaufman G G. Risks and failures in banking: overview, history, and evaluation. 1986.

[5] He C, Geng X, Tan C, et al. Fintech and corporate debt default risk: Influencing mechanisms and heterogeneity. Journal of Business Research, 2023, 164: 113923

[6] Putra M P. An analysis of big data analytics, IoT and augmented banking on consumer loan banking business in Germany. Journal of Research on Business and Tourism, 2021, 1(1): 16-36.

[7] Joyner D T. A neural network approach to estimating the allowance for bad debt. Virginia Commonwealth University, 2011.

[8] Chen Y, Wu J, Wu Z. China’s commercial bank stock price prediction using a novel K-means-LSTM hybrid approach. Expert Systems with Applications, 2022, 202: 117370.

[9] Njeguš A, Jovanović V, Veinović M. Izazovi i mogućnosti primene analitike velikih podataka u hotelskom poslovanju. Poslovno udruţenje hotelsko ugostiteljske privrede Srbije, 2015.

[10] Khalil M I, Kim R Y C, Seo C Y. Challenges and opportunities of big data. Journal of Platform Technology, 2020, 8(2): 3-9.