Big Data in Financial Field: a Brief Review

Jiansheng Jiang

Jinhua Jinyi New District New Town Construction Management Center, Jinhua, 321035, China

84033233@qq.com

Abstract. With the advent of the digital age, the development and application of big data is profoundly changing the landscape and operating modes of the financial field. As a new type of information resource, big data characterized by its massive, high-speed, diverse and low value density solves the problem of information explosion in the financial field and provides unprecedented opportunities and challenges for financial institutions. This paper aims to discuss the applications of big data in the financial field and analyzes its important roles in risk management, trading decisions, customer relationship management and other areas, and look forward to its future development.

Keywords: Big data; Financial field; Information resource.

1 Introduction

As the core and pillar of the modern economy, the good operation of the financial field not only concerns the stability and development of the economic system, but also directly affects people's livelihood and social stability [1, 2]. However, with the rapid development of and the accelerated advancement of global digitalization, the financial field is also facing unprecedented challenges and opportunities. The traditional operating models and data processing methods of financial services have become difficult to adapt to the needs of the information explosion era. In this background, the rapid rise of digital has brought a change in the financial field and given birth to digital finance [3].

Digital finance, as a product of the deep integration of digital and financial business, is becoming an important engine to promote the forward development of the financial field with its advantages of low cost, high efficiency, wide coverage and sustainability [4, 5]. With the popularization of the Internet and the development of mobile Internet technology, people's financial needs have become increasingly diversified and personalized. The singularity of the traditional financial business model can no longer meet people's needs. Through digital finance, individuals and enterprises can more conveniently obtain financial products and services such as financing services, payment and settlement, investment and wealth management, insurance protec-

L. Liu et al. (eds.), *Proceedings of the 3rd International Conference on Financial Innovation, FinTech and Information Technology (FFIT 2024)*, Advances in Computer Science Research 118,
https://doi.org/10.2991/978-94-6463-572-0_5

tion and so on, thus enhancing the accessibility and inclusiveness of financial business.

In the era of digital finance, data has become the core resource and competitive force of financial business. The operation of the financial field can not be separated from the efficient management and utilization of data such as market trends, customer demand, and risk conditions. However, traditional data processing is often unable to meet the demand for massive, diverse and high-speed data processing [6]. This requires the intervention of big data technology. The development of big data has brought new data processing methods and analysis means for financial institutions. Through the real-time collection, storage, processing and analysis of massive data, financial institutions can more accurately understand market dynamics, customer behavior and risk conditions, thus providing strong support for business making decision [7].

As an important support for digital finance, big data will play an increasingly important role in the intelligent, personalized and efficient development of financial business. In this paper, we will discuss the current situation, advantages and challenges of the application of big data in the financial field, and look forward to the direction and prospects of its future development. Through depth research and discussion in this field, it aims to provide theoretical support and practical guidance for the development of the financial field, and to promote the financial field to take more solid steps towards digitalization and intelligent transformation.

2 Overview of Big Data

Big data is a collection of data that is huge in size, diverse in type, fast in processing speed and low in value density. It is a new type of data processing and analysis, which can help enterprises and organizations to understand and use data better, in order to improve the decision-making efficiency and business value [8, 9]. In the financial field, big data can help banks and insurance companies understand customer needs and risks better, and improve risk control and customer service.

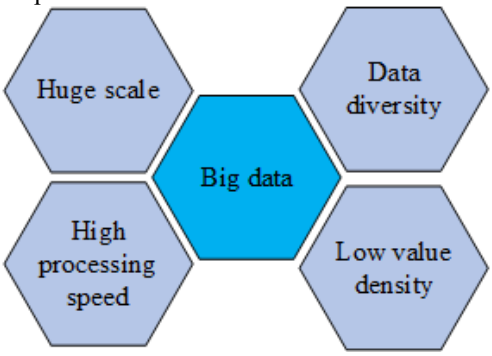

Fig. 1. Characteristics of Big Data.

The characteristics of big data mainly include the following aspects in Figure 1:

- Huge size: the amount of data is usually calculated at the level of TB, PB, EB, etc., which is far beyond the scope of traditional data processing capability.
- Data diversity: data sources are diverse, containing many types of data such as structured data, semi-structured data and unstructured data, such as text, images, audio, and video.
- Fast processing speed: big data can improve the speed of information processing, not only in the collection of information, but also in the comprehensive analysis of information and storage, which can significantly improve the efficiency of processing and analysis of various types of data, and obtain a more ideal effect of the use of information.
- Low value density: big data contains a large amount of redundant and useless information, which needs to be extracted through data mining and analysis techniques to extract valuable information.

The main of big data include data acquisition, data storage, data processing, and data visualization [10], as shown in Figure 2. When the financial field applies big data technology, it should choose the technical means suitable for its own characteristics, and it should also focus on the organic combination of technical methods and business models.

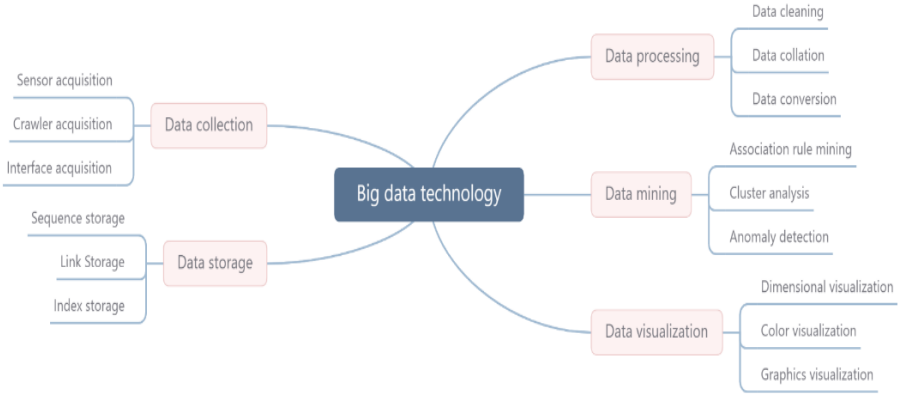

Fig. 2. Key of Big Data.

Big data analyses and applications are no longer limited to the information industry. It is now across all industries and disciplines, including business and financial management [11, 12]. Big data analytics play an important role in supporting business management, financial decision-making, global manufacturing and supply chain management by creating data transparency to improve decision-making and facilitate innovative business and financial models. Organizations are discovering new potentials and possibilities, discovering hidden knowledge, improving decision-making processes and supporting their strategic planning from big data. Big data is the next frontier for financial innovation, competition and productivity. The following discussion will explore the application of big data in the financial field.

3 Application on Big Data in Financial Field

After a long time of information data collection, big data has helped the financial field to grasp the high value of information data in the development, and big data finance has become the main trend of the future development of the financial field. In order tohelp the financial field to deal with the fierce competition in the future, the refinement of the processing based on the big data environment has become the development needs of the financial field in the new period [13]. In the future development, the big data products under the big data environment will develop more mature, and the can be applied more abundantly and diversely, providing more powerful support for the rapid development of the financial field under the big data environment.

3.1 Risk Management

The financial risks included cover all aspects of the financial field and their major forms are shown in Figure 3.

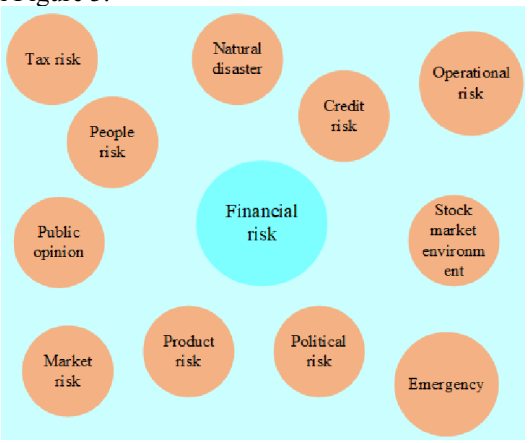

Fig. 3. Financial risk management.

Table 1. Type Styles.

Specific applications	Definition
Data collection & integration	Collecting information on an individual's asset status, credit history, repayment history.
Feature selection & modeling	Select important features related to credit risk from massive data and construct models
Risk scoring & classification	Scoring and classifying borrowers using established credit risk models
Warning and monitoring	Monitoring the borrower's credit situation in real time and providing an early warning mechanism
Anti-fraud analysis	Identify potential patterns of fraudulent behavior by analyzing transaction data

In the past, the financial risk management and control model had to deal with a variety of risk management factors, such as the most basic market factors, credit factors

and the more common operational, legal and managerial factors. Financial institutions need to constantly assess and monitor market risk, credit risk and other types of risk to ensure the stable operation of the financial system. Big data plays an important role in risk management and prediction. Using big data analysis technology, risks can be identified and quantified more accurately, and potential risk events can be warned and prevented in time. For example, big data technology can help banks enhance the accuracy of their credit assessment of customers. As shown in Table 1, through analyzing multi-dimensional information such as customers' historical transaction data and social network behaviors, a more accurate credits model is constructed, thus improving the efficiency and accuracy of lending decisions. Among them, the credit score model uses a linear regression model to predict credit scores, as shown in formula (1).Another example is in terms of tax risk, constructing tax big data as shown in Figure 4. Through comprehensive analysis, rational allocation and active management to reduce the loss of national taxes and use big data to reduce the risk of tax management.

$$F = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{1}$$

Where, F is the credit score, which is the variable that affects the credit score, such as historical transaction data, social behavior, etc. β is the model parameter.

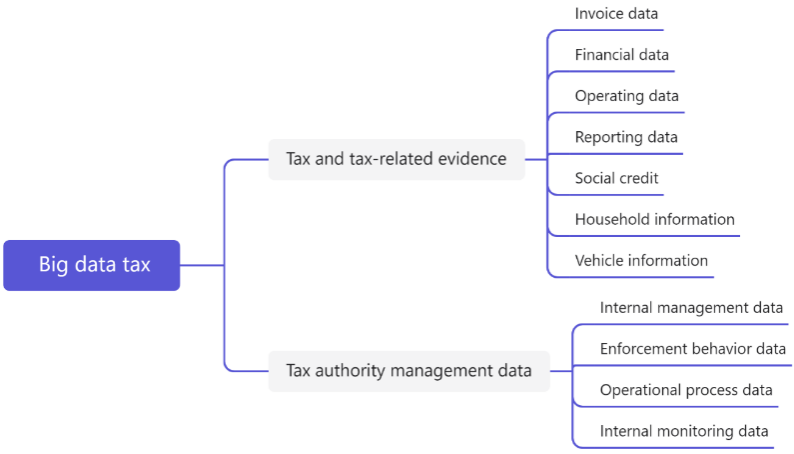

Fig. 4. Big data tax.

The support of big data has greatly improved financial services and expanded the financial of the business [14]. In addition, with the increasing diversification of current financial products, the market risks they face are also increasing. There are all kinds of risks hidden in these financial products, and financial risk management in the big data environment can bring out the bigger business risks.In addition, the risk can be minimized by introducing the concept of value at risk to measure the maximum loss that financial assets may suffer in a certain period of time at a certain level of confidence, as shown in formula (2).

$$VaR(\alpha, T, \sigma, z) = z \cdot \sigma \cdot \sqrt{T} \tag{2}$$

Where, α is the confidence level, T is the time range, σ is the standard deviation of asset return, z is the quantile of the normal distribution.

3.2 Investment Trading Decisions

Big data support for trading and investment decisions is also one of its key applications. In the financial market, the acquisition and analysis of information is crucial to the formulation of trading strategies and optimization of investment portfolios. Through the in-depth mining of big data, financial institutions can achieve real-time monitoring and analysis of market data, capturing minute changes and pulse of the market, in order to make more accurate trading decisions. At the same time, big data can also provide support for optimizing investment portfolios, constructing effective asset allocation models through analysis of historical data and market trends, reducing investment risks and improving investment returns. At the same time, big data can also provide support for portfolio optimization. By analyzing historical data and market trends, effective asset allocation models are constructed to reduce investment risks and improve investment returns.

In Figure 5, it shows a model for risk investment decision-making based on big data. This model is divided into two parts: data collection and integration, and analyzing data using data mining and machine learning. They can help investors assess investment opportunities and control investment risks more accurately. By collecting and analyzing a large amount of investment case data, the model can predict the possible investment return from a certain investment opportunity and help investors make more informed investment decisions [15].

Portfolio optimization generally aims at improving investment returns or reducing investment risks, and recombines products with different targets. The classic capital asset pricing model defines the relationship between the risk premium that tells us about investment and the market risk premium:

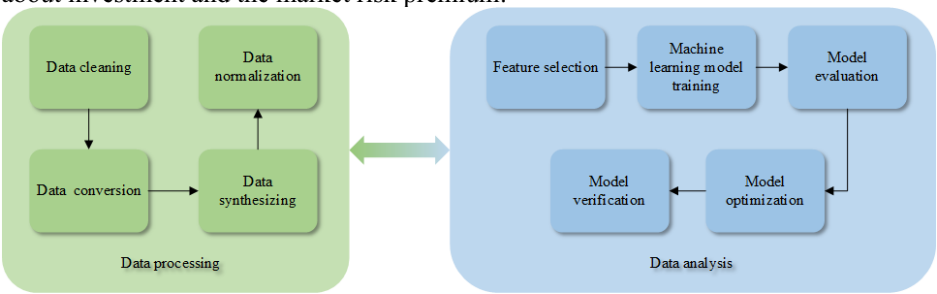

Fig. 5. Venture Capital Decision Models

$$E[R_I] - R_f = \alpha_I^M \left(E[R_M] - R_f \right) \tag{3}$$

Where, R_I and R_M represent the return of the investment*I* and the market *M*respectively, and R_f is the risk-free rate. We could calculate the α value of the investment*I* in relation to the market by the following formula:

$$\alpha_I^M = \frac{\mathrm{cov}(R_I, R_M)}{\mathrm{var}(R_M)} \tag{4}$$

The above formula shows that the risk premium of the investment is proportional to the risk premium of the market, and the proportionality constant is the risk of the investment relative to the market.

3.3 Customer Relationship Management

Big data has also brought new ideas and approaches to customer relationship management for financial institutions. In the highly competitive financial market, the provision of personalized financial products and customized services has become a competitive advantage for major financial institutions. With the help of big data, enterprises and institutions can better understand the needs and preferences of customers, and by analyzing data such as customers' consumption behavior and historical transaction records, they can accurately recommend products that meet the needs of customers and enhance customer satisfaction and loyalty.

Basic telecoms operators have a unique advantage in big data, as the large amount of data streams carried on their networks record all of the user's communication actions. Compared to Internet companies that record user data behaviors through mobile applications, the user data held by basic telecoms operators is unique due to the uniqueness of their mobile numbers. Compulsory real name makes the data more accurate. Meanwhile, due to its national comprehensive network, it ensures that user data can be collected in real time and uninterruptedly through the communication carrier network [16, 17]. In Figure 6, it shows the application of big data in telecom customer relationship management, based on data mining to study the customers. The overall process of data mining is to carry out the process of targeted extraction of data. Effective information is presented and its regularity is analyzed for valuable expression.

Another typical example is the application of personalized recommendation systems. Many enterprises and institutions have begun to use big data to build personalized recommendation systems, which analyze customers' transaction behaviors, preferences and historical data to recommend financial products that best meet their needs. Whether it is a financial product, a loan scheme or an insurance plan, they can be accurately matched through big data to improve the sales conversion rate. This kind of personalized recommendation not only meets the needs of customers, but also effectively improves the sales performance of enterprises and organizations. This case is widely used in e-commerce platforms.In addition, cluster analysis can be introduced to divide customers into different market segments, so as to provide personalized products.

In customer relationship management, cluster analysis is a common technique used to divide customers into different market segments based on their characteristics and behaviors. The cost function is usually used to evaluate the effect of clustering, and its purpose is to minimize the variance within the cluster, that is, the members within the cluster are expected to be as similar as possible, which is obtained by formula (5). Ideally, as the iteration progresses, the value of the cost function should gradually decrease, indicating that the data points within the cluster are getting closer and closer. Then the characteristics of each cluster are analyzed, and each cluster is labeled, such as "high value customers", "growth customers", etc., in order to facilitate targeted customer relationship management.

$$J(w) = \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{5}$$

Where, $J(w)$ is the cost function of the cluster, $k$is the number of clusters, S_i is the i-th cluster, $x$is the data point, $\mu_i$is the cluster center.

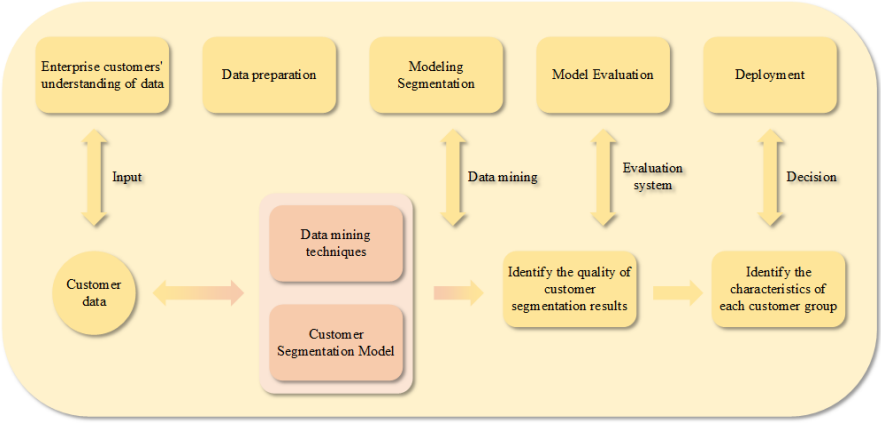

Fig. 6. Overall scheme of data mining.

4 Challenges and Perspectives

The application of big data in the financial field presents many opportunities. It also faces some challenges such as data privacy and security, infrastructure development and data quality and consistency. The challenges need to be continuously explored and solved by financial institutions.

4.1 Data Privacy and Security

In the era of big data, data and information are characterized by a very significant value. By giving guidance to high-value data, financial institutions can move in the

right direction. When acquiring information, analyzing and collecting information, it is done with the help of network which highlights its security and privacy even more. Financial data often contains extremely sensitive data such as personal identity and financial information. Such data, once leaked, can lead to serious consequences such as identity theft and fraud. It brings huge losses to customers and financial institutions. In addition, with the continuous development of hacking technology, financial institutions face the threat of hacking attacks from outside. They may attack the databases or systems of financial institutions to steal customers' account information, transaction records and other sensitive data.

To address these privacy and security challenges, financial institutions can take a number of steps to protect the security and privacy of customer data. The following are some of the recommendations given:

1.Strengthen Data Security Protection Measures and Establish Access Control Mechanisms.

The strong encryption technology could be used to protect data transmission and storage security (such as formula (6)). High-strength encryption is used to protect the security of data transmission and storage and a sound access control mechanism is established to restrict access to data. In addition, data anonymization is used to hide personally identifiable information when sharing and analyzing data to protect customer privacy.

$$c = \text{Encrypt}(k, p) \tag{6}$$

Where, c is ciphertext, $\text{Encrypt}(\cdot)$ encryption function, k is key, p is plaintext.

In Figure 7, it is a possible anonymous authentication scheme. The authentication function is better achieved by using a flexible and security enhanced key for the end device. At the same time the anonymity of the terminal equipment (TE) is guaranteed through an invisible ID. The scheme consists of three phases. The first is TE registration. A registration request is submitted to the Registration Centre (RC) with a unique identification number ID. After receiving the registration request, the TE management first checks whether it is registered or not. Then, the RC computes the TE's hash value and private key and returns these two values to the TE. Then it is the Mobile Edge Computing Server (MECS) registration. It first sends a registration request to the RC and provides the ID. After receiving the request, the RC first checks whether the ID is previously registered. Then the RC computes the private key of the MECS and sends it to the MECS. Finally, there is a mutual authentication key agreement phase. In this phase, TE and MECS authenticate each other and negotiate a common key for further communication.

2.Blockchain.

The application of blockchain in the financial field can effectively improve data security and privacy protection. Through encryption, decentralization, smart contracts and other technical means. It can effectively address the privacy and security chal-

lenges facing financial data. Provide more secure, transparent and efficient data management solutions for the financial industry.

a) Blockchain-based encrypted storage: Financial data is stored on the blockchain. The tamper-proof characteristic is used to strengthen encryption algorithms, in order to ensure that data is encrypted and protected from unauthorized access during transmission and storage.

b) Distributed storage: Blockchain enables decentralized storage of financial data. Data is stored in multiple nodes of the network in a decentralized manner. It is no longer dependent on a single data center or server. This will not lead to the leakage of all data even if a node is attacked.

c) Secure data sharing: Financial institutions can use blockchain to establish a secure data sharing platform. On this platform, data ownership and access rights can be managed and enforced through smart contracts ensuring that data is only shared with authorization.

d) Data traceability: Tamper-proof characteristics of blockchain ensure that all changes to data are recorded and can be traced back to its initial state. This is important for auditing and regulation of financial data.

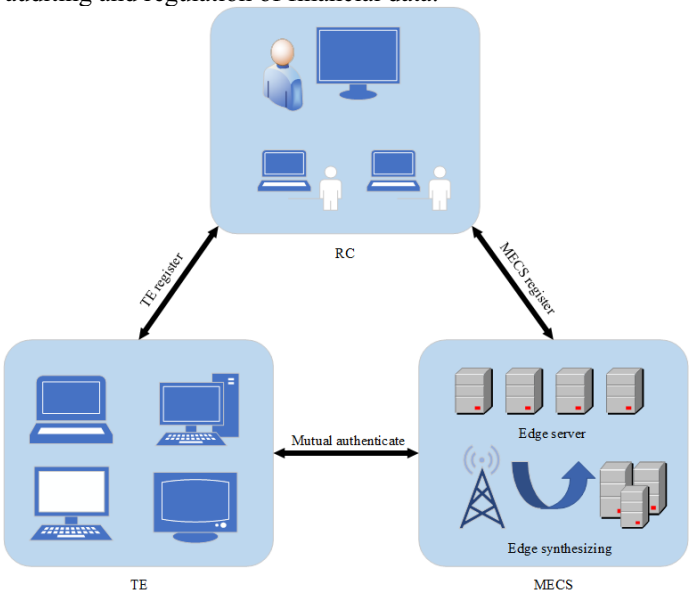

Fig. 7. Anonymous authentication scheme.

4.2 Infrastructure Development

The application of big data in the financial field faces a number of infrastructure challenges. Firstly, financial institutions typically deal with large amounts of data, including transaction records, customer information, market data, and so on. This requires powerful data storage and processing capabilities. As a result, financial institutions

need to invest heavily in building and maintaining high-performance data centers to ensure that data storage, processing and analysis can meet demand. Secondly, data collection and integration are also a challenge. Financial data comes from a wide range of sources and may come from different trading platforms, banking systems, third-party data providers, etc. These data often have different formats and standards, and require effective integration and cleansing so that they can be analyzed and utilized in a uniform manner. As a result, financial institutions need to invest a lot of manpower and resources to develop and maintain tools and processes for data integration.

The following recommendations are given to address the above challenges:

1.Mobile Edge Computing.

Typically, all the data of a financial institution is transmitted to a central server for aggregation and analysis.Due to resource and arithmetic constraints, it is not possible to cope with the huge amount of data that needs to be processed. The use of mobile edge computing can solve this problem by deploying edge nodes close to the data terminals to offload data processing tasks from the central server. The edge nodes aggregate and process the data before uploading it to the central server. By using a distributed architecture, data processing and storage can be dispersed across multiple nodes and servers, improving the fault tolerance and reliability of the overall system.

2.Machine Learning and Artificial Intelligence.

Financial institutions can use machine learning and artificial intelligence to automate the identification and correction of data quality issues. These techniques can be applied to all aspects of data cleansing, such as identifying and dealing with missing values, outliers, and duplicate data. By building machine learning models, the system can learn and understand patterns and regularities in the data and automatically identify anomalies in the data. For example, machine learning algorithms can identify outliers in a data field and provide automated processing solutions such as replacement, deletion, or interpolation. Meanwhile, AI techniques can be used for data standardization and normalization to unify data of different formats and structures into a standard format for more effective integration and analysis. The most commonly used method is maximum-minimum normalization, such as formula (7).These methods not only improve the efficiency of data processing, but also reduce human error and improve data accuracy providing financial institutions with a more reliable and high-quality data base.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{7}$$

Where, x is the original data, x_{norm} is the normalized data, $\max(x)$ and $\min(x)$ are respectively the minimum and maximum values in the data set.

4.3 Data Quality

Financial institutions typically obtain information from multiple data sources, which may come from different systems, departments, or third-party vendors. This integration of multiple sources of data can lead to data quality challenges. Data quality challenges include aspects such as data completeness and accuracy.

1.Completeness.

Financial data often contains a large number of fields and information where there may be missing values or incomplete data records. This may lead to bias in data analysis and modelling which in turn affects the financial institution's accurate understanding of customer behavior, market trends, etc.

2.Accuracy.

The accuracy of financial data is critical for risk assessment, investment decisions, etc. However, data may have entry errors and duplicate recording problems, resulting in the accuracy of the data being compromised. For example, an incorrect transaction amount or account balance may lead to misleading analytical results, affecting the reliability of financial decisions.

To address the above challenges, the following points are proposed:

a) Establish a data quality management framework: establish data entry process standards, including setting required fields, data format specifications, data validation rules, etc., to ensure the standardization and accuracy of data entry. Establish a data quality monitoring mechanism to detect and deal with data quality problems in a timely manner through real-time monitoring of data quality indicators such as data missing rate and outliers.

b) Establish data dictionary and metadata management: Develop a data dictionary to record information such as the definition, source, format and specification of data. Establish a metadata management system, including the creation of data models, metadata standards and documented data structures to help understand and manage data. This can be achieved through the use of metadata management tools and databases to ensure data consistency and maintainability.

3.Improve Security Specifications for Big Financial Data.

Network security firewalls, authentication and data encryption must be used across the board. Actively using multi-party security design, federated learning, differential privacy, alliance chain and other means, and trying to build a statistical information platform for cross-subject data and information sharing security to open up data and information silos and effectively release data and information resources.

5 Conclusions

With the rapidly development of information technology, big data is widely used in the financial field, which solves the processing and analysis problems brought by the rapid growth of data in this field. The paper aims to systematically discuss the application of big data in the financial field, and analyze its role in risk management, investment and transaction decision-making, and customer relationship management. By analyzing and investigating relevant cases, the challenges brought by the application of big data in the financial field are summarized and corresponding solutions are given. The research on big data and the financial field was aimed at promoting the continued innovation and development of the financial field in the era of big data.

References

1. A. F. Haryadi, J. Hulstijn, A. Wahyudi, H. van der Voort and M. Janssen, "Antecedents of big data quality: An empirical examination in financial service organizations," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2016, pp. 116-121, doi: 10.1109/BigData.2016.7840595.
2. W. Fang, C. Zhu and W. Zhang, "Toward Secure and Lightweight Data Transmission for Cloud–Edge–Terminal Collaboration in Artificial Intelligence of Things," in IEEE Internet of Things Journal, vol. 11, no. 1, pp. 105-113, 1 Jan.1, 2024.
3. Y. Ji, S. Dai, B. Yang, et al. Logical Framework, Developing New Quality Productivity and Promoting High-quality Economic Development. Journal of Industrial Technology and Economy. No. 2 (General, No. 364), 2024, 3-28.
4. X. Wei. Risk and Regulation of Digital Financial Inclusion. Financial Theory & Practice. No. 6 (General, No. 479), 2019, 49-54.
5. W. Fang, C. Zhu, M. Guizani, J. J. P. C. Rodrigues and W. Zhang, "HC-TUS: Human Cognition-based Trust Update Scheme for AI-enabled VANET," in IEEE Network, vol. 37, no. 5, pp. 247-252, Sept. 2023.
6. S. Kurt, J. Heitz, N. Bundi and W. Breymann, "Large-Scale Data-Driven Financial Risk Modeling Using Big Data Technology," 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT), Zurich, Switzerland, 2018, pp. 206-207, doi: 10.1109/BDCAT.2018.00033.
7. Y. Chen and X. Zhao. Analysis on Development of Digital Economy under Background of Big Data. China Circulation Economy. 20, 128-131. https://doi.org/10.16834/j.cnki.issn1009-5292.2023.20.042.
8. H. Mei, X. Du, H. Jin, et al. Big data technologies forward-looking. Big Data Research. Vol. 9, No. 1, 2023, 1-20.
9. W. Fang, C. Zhu, F. R. Yu, K. Wang and W. Zhang, "Towards Energy-Efficient and Secure Data Transmission in AI-Enabled Software Defined Industrial Networks," in IEEE Transactions on Industrial Informatics, vol. 18, no. 6, pp. 4265-4274, June 2022.
10. A. Jaiswal and P. Bagale, "A Survey on Big Data in Financial Sector," 2017 International Conference on Networking and Network Applications (NaNA), Kathmandu, Nepal, 2017, pp. 337-340, doi: 10.1109/NaNA.2017.46.
11. X. Wu, L. Liang, S. Chen. How big data alters value creation: Through the lens of big data competency. Business 2.0, No. 5, 2022, 707-734.

12. W. Fang, N. Cui, W. Chen, W. Zhang and Y. Chen, "A Trust-Based Security System for Data Collection in Smart City," in IEEE Transactions on Industrial Informatics, vol. 17, no. 6, pp. 4131-4140, June 2021, doi: 10.1109/TII.2020.3006137.
13. M. Ahmed, N. Choudhury and S. Uddin, "Anomaly Detection on Big Data in Financial Markets," 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Sydney, NSW, Australia, 2017, pp. 998-1001.
14. S. Lu. Application of big data analysis in financial risk assessment. Shanxi Science and Technology Daily, A06, 2023. https://doi.org/10.28712/n.cnki.nshxk.2023.000508.
15. Y. Zhang. Internet Financial Risk Management under the Big Data Environment. China Journal of Commerce, 035, 2021, 35-37. https://doi.org/10.19699/j.cnki.issn2096-0298.2021.16.035.
16. W. Fang, W. Zhang, W. Yang, Z. Li, W. Gao and Y. Yang, "Trust management-based and energy efficient hierarchical routing protocol in wireless sensor networks," Digital Communications and Networks, vol. 7, no. 4, 2021, pp. 470-478.
17. K. Murugan, V. Selvakumar, P. Venkatesh, M. Manikandan, M. Ramu and K. M, "The Big Data Analytics and its Effectiveness on Bank Financial Risk Management," 2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC), Chennai, India, 2023, pp. 313-316, doi: 10.1109/ICRTAC59277.2023.10480831.