Developing A Risk Control System for Fintech Products Using Big Data and Real-Time Analysis Technology

Yingsong Gao *

Department of International Business School, Tianjin University of Finance and Economics, Tianjin, China

* Corresponding Author Email: 1811030413@stu.hrbust.edu.cn

Abstract. Big data technology has revolutionized the financial sector, significantly impacting the design and operation of future financial products. This paper explores the use of big data technology in risk control in financial products, examining its potential benefits and challenges in product design. By briefly introducing the operation logic of big data, this paper analyzes the application of big data in financial products for risk control and the challenges it faces. The study reveals that big data technology enhances the efficiency of financial technology companies in internal management and decision-making, while significantly reducing systemic risk in their financial product design. Big data technology has exacerbated the conflict between traditional financial institutions and fintech companies, as well as the interest conflict between fintech companies, users, and regulatory authorities. Therefore, Big data financial products not only need to improve their information processing methods and information analysis methods internally, but also need to deal with the conflicts between their users, issuers, and financial regulatory authorities and with the traditional financial sector.

Keywords: Big Data Technology; Financial Technology; Control of Risk Introduction.

1. Introduction

The financial industry has been significantly shaped by big data technology. Through the processing of massive open-source information, financial institutions can understand the investment and financing needs of their potential customers and provide them with suitable financial products. This business model has enabled fintech companies such as Alipay to reap huge user bases and fat profits. In recent years, with the significant increase of uncertainties in the external economic environment, such as the bankruptcy of Silicon Valley Bank, how to effectively manage the risks of financial products and prevent potential systemic risks has become an urgent issue. The combination of big data's powerful information processing and analysis capabilities and financial risk control can theoretically significantly improve the anti-risk ability of financial products and provide guidance for future financial product design.

The research on big data technology and financial risk control has attracted some researchers. Yu W et al. found that big data technology can effectively control the uncertainty in the supply chain. Jennifer et al. found from the case study that big data can effectively manage risks and control budgets. Spatt et al. found that machine learning can detect relationships between different economic variables and build more complex models to predict future trends more accurately. However, the research on big data's application in risk control of financial products is still limited, with few cases available. This paper explores the use of big data technology in financial product risk control through case studies, aiming to enhance industry understanding and address challenges in its development.

This paper is separated into three parts. In the first part, this paper introduces the characteristics of big data technology, its data processing and analysis capabilities, aiming to enhance understanding of this technology. In the second part, this paper will present a case study on the application of big data technology in risk control of financial products. In the final part, this paper explores the potential risks associated with big data in financial risk control and envisions its future application in technology.

2. Overview of Big Data

Data is becoming huge in today's digital economy. Over the past 30 years, thanks to the falling cost of data storage and computing, the storage and application of data can be used by increased people, so massive amounts of data have been created and stored on electronic devices [1]. How to select, organize, analyze these data, and predict the future situation has been an increasingly difficult problem for traditional data processing methods timelessly and accurately. That is why people need a new application which could cope with a huge amount of data precisely and timelessly. And big data technology is designed of such needs.

2.1. Big Data characteristics

For some time, there was no accepted definition of the characteristics of big data until 2014, when data in big data was defined as three characteristics, namely the "3V" that “Volume” “Variety” and “Velocity” [2]. The term "volume" describes the enormous amount of data that big data analysis must manage. “Variety” refers to the variety of data sources and its storage applications that big data must select, such as libraries database where data is usually structured or social platform where has massive unstructured data. “Velocity” refers to the processing speed of big data, big data must obtain time-sensitive data from its efficient data asaprocessing applications, which often have a higher value.

Then big data was added new definition along with subsequent development, “Veracity” was added in 2016, It usually refers to the interference information encountered in the process of big data analysis, such as noise and abnormal information that increase the uncertainty of data accuracy and effectiveness [2]. “Veracity” also is used to determine the credibility of the information [3]. And “Value” was added in 2018, which refers to the awareness and analysis strategy of the correlative degree of things in data processing [4]. These definitions make the concept of big data clearer.

2.2. Big data framework

These characteristics make big data technology must ensure its own processing of information speed, timeliness, and intelligent. It must avoid its own various subroutines appear task overlap this inefficient situation. As a result, creating a framework is required to allocate the duties that the various subsystems are responsible for and coordinate their operations. To understand how the framework works, this paper introduces a popular big data framework: Hadoop.

Hadoop, which is regarded as a distributed storage system with a specific fault tolerance rate. Hadoop comprises two essential parts. The further part is "Hadoop Distributed File System (HDFS)," which is intended for data storage, and “MapReduce,” which was developed by Google, is a data processing framework in distributed environment. The chart that how HDFS manages data is shown on the Figure 1 and Figure 2.

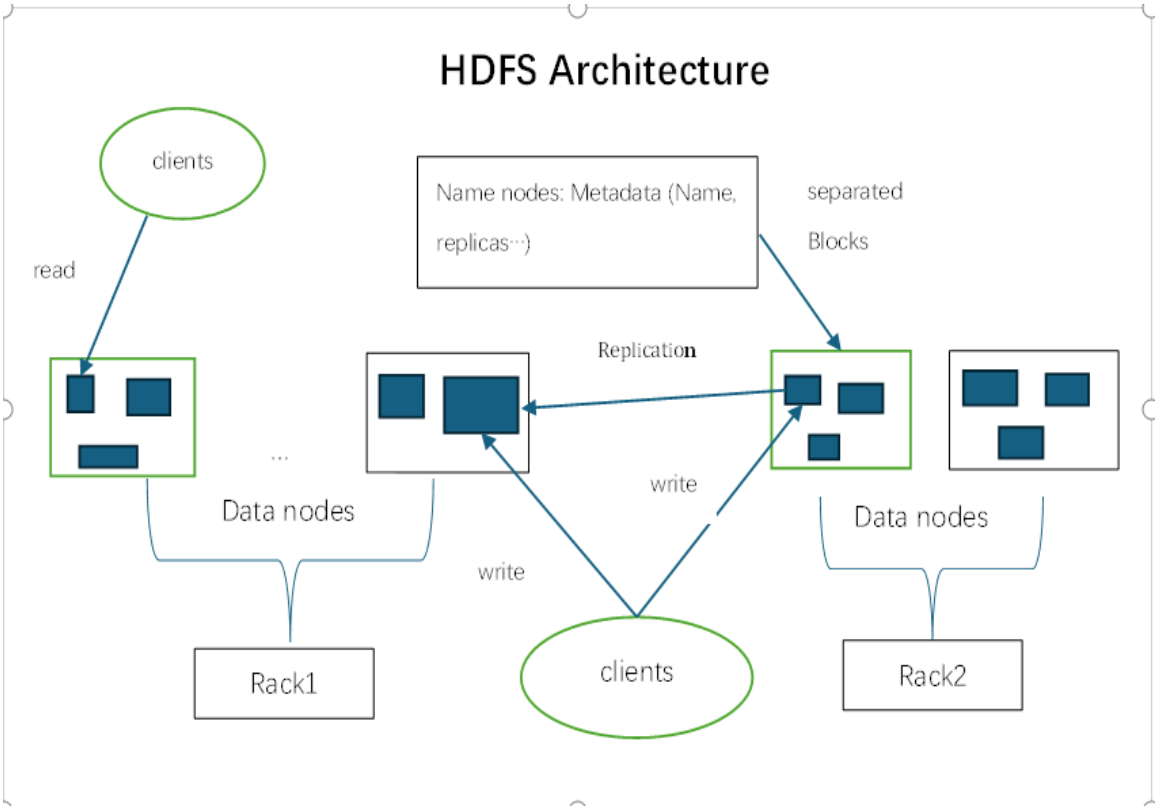

Figure. 1 HDFS Architecture [2]

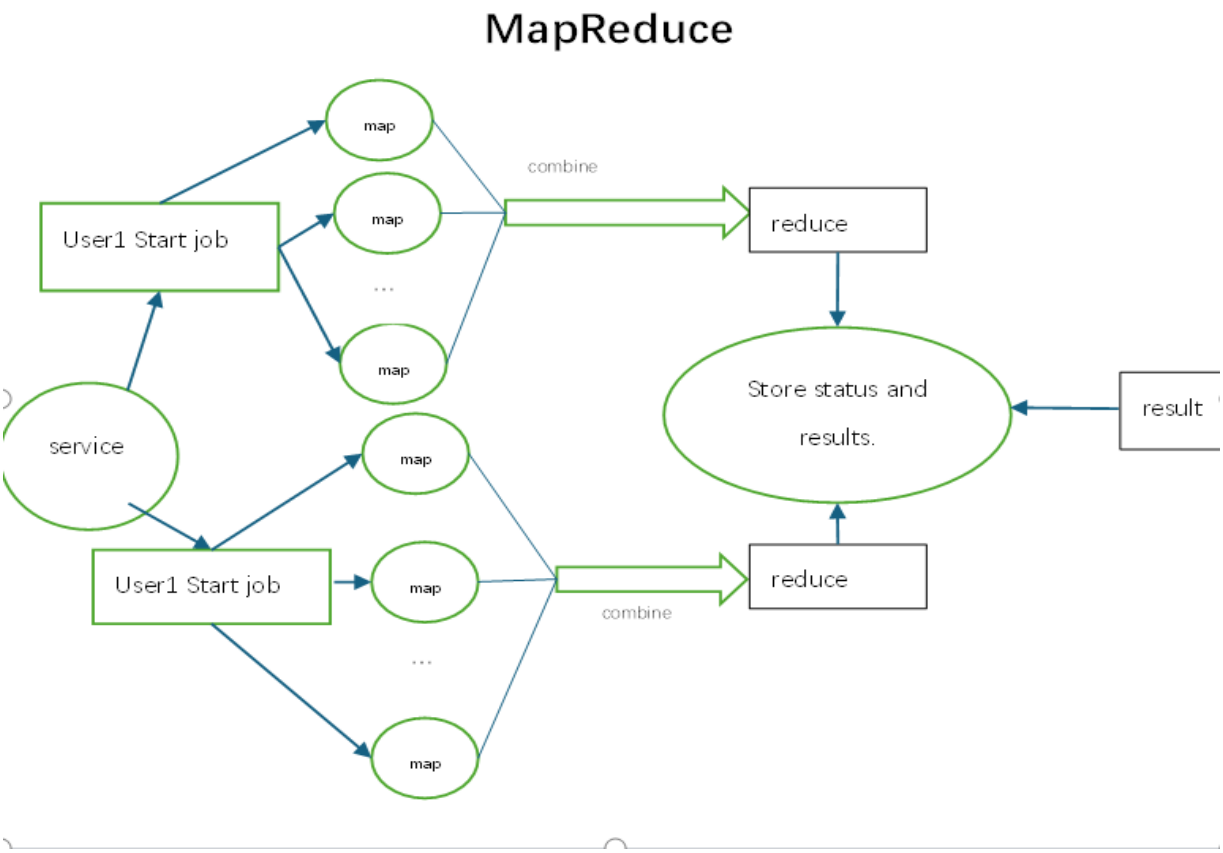

Figure. 2 HDFS Architecture and MapReduce [2]

In figure 1, the name node is the primary node responsible for HDFS running and the data node is the slave node responsible for tracking HDFS running trends. In fact, only data node could store data,

they are responsible for storing the data in the personal terminal, so that individual users can log in to the client on their personal terminal to read and rewrite the data.

Name node will separate the Meta data file into small blocks and storage them in a server. The server could create three backups of blocks to ensure the safety of the data. In addition, the server shares this data with all Hadoop servers for storage and backup.

Compared to a traditional single server, the information stored by Hadoop has stronger security because its data can be stored in thousands of servers and backed up. In this way, once several servers occurred data loss or tampering, these faulty servers can compare the data with other normal servers and complete the data to modify the faulty data.

Figure 2 demonstrates how MapReduce works. MapReduce could separate the data into several subsets and analyses them, then the analysis results will be combined and contribute to the result. Compared with traditional processing method where the progress needs to be finished with special email application such as ETL, in MapReduce, the progress could be written as important level of luggage such as pig to make the operation easier.

In a conclusion, the framework of big data contain needs to contain three steps. Step one, managing data where the framework is capable of separating huge amounts of data of different categories to make data useful for analysis, the categories could be structural or unistructural, Step two, analysis and modeling where the framework needs to be written as higher lavage to make the process easier, and the third step is prediction output where the framework needs to integrate the analysis results of previously categorized data.

3. Applications and examples of big data technology in financial product risk control

Big data technology has plenty of applications in the field of financial risk control. From operating perspective, by collecting wide range of information which including supply chain, regulation standard, firms could make improve its strategy to decrease its operating risk. For example, a company can monitor its own business situation by collecting the information of its internal procurement, logistics, operation and other departments, as well as collect the information of upstream and downstream firms in its own industrial chain and make decisions to improve the relationship with them to ensure the safety of its own industrial chain [5]. From investment perspective, big data helps investors avoid high risk investment by monitoring mass firm’s fundamental statistics, credit rank or events that could make significant influence on firms’ management risk. Furthermore, portfolio managers can use big data analysis technology to find relationships among hundreds of investment assets and create a portfolio that lowers the systemic risk the portfolio faces. From lenders perspectives, big data. Big data can help them select suitable lenders by collecting their economic conditions. Banks are the most representative example. They can analyze the credit ratings of existing customers and potential customers to decide their repayment probability, and then manage the risk of bad debts [6]. In addition, the bank can also use its data as an asset as collateral, which can improve the bank's liquidity and reduce the bank's liquidity risks.

To help readers better comprehend the use of big data technologies in financial risk management, this article will provide two application instances.

3.1. Case 1: "E-finance" series products of Bank of Jiangsu

In 2015, Bank of Jiangsu found that it was more difficult for small and micro firms to obtain financing compared with mature firms and large firms, mainly because in traditional financial credit products, borrowers need lenders to provide collateral assets, credit investigation records of the central bank and existing financial statements. Due to their small scale, low credit investigation record and lack of formal financial reports, it is difficult for small and micro firms to obtain the favor of borrowers [7].

To solve this problem, Bank of Jiangsu incorporated the big data credit investigation system into its new product "E-finance". Before the loan, "E-finance" collected internal and external data, semantics and public opinion of small and micro firms on the credit investigation platform, determined the authenticity of the lender's identity through face recognition, and matched the "E-finance" series loan products according to their loan demand. In this way, through precise push and big data drainage, "e-finance" has attracted several small and micro firms who come to borrow money.

Secondly, in loans, "E-finance" has set up a credit audit and early warning mechanism. By collecting tax data of small and micro firms on the credit investigation platform and comparing them with their external data, it has conducted credit ratings on thousands of small and micro firms and set different loan interest rates for small and micro firms in different credit rating ranges according to different credit ratings. Therefore, "E-finance" can not only meet the diversified financing needs of small and micro firms, but also greatly simplify the loan approval procedures, with certain intelligence.

In traditional lending, due to information asymmetry, it is difficult for borrowers to track the cash flow of lenders, and lenders may invest loans in high-risk projects, thus increasing the risk of bad debts. To avoid the problem, Bank of Jiangsu has developed an automatic risk early warning system to analyze the situation of small and micro firms around the world after lending. With the use of big data technology, the automatic risk early warning system not only gathers these businesses' operational conditions but also follows and monitors the business's capital flow in several dimensions to upon identification of risks and hidden threats, issues forward-looking early warnings. Finally, “E-finance" finally incorporates machine learning to modify its own credit evaluation criteria and algorithms in a timely manner for the growth of financial markets to maintain the effectiveness and dependability of its system operation.

3.2. Case2: "Future Book" system in Bridgewater.

Bridgewater is the world’s largest hedge fund. which collects data such as interest rates, and sales to build algorithms that measure the intrinsic value of financial assets. In 2016, Bridgewater proposed the automation of decision-making on top of the automation of trading processes. To this end, Bridgewater Economics developed an ai system called "Future Book", which collected the strengths and weaknesses of employees within the company and ranked them to select excellent employees. furthermore the "Future book" will replace human management to automatically adapt employees' work efficiency to their tasks of the day and supervise the completion of tasks by itself [8].

As the functions of the company's management are gradually replaced by big data algorithms, the company's decision-making process is simplified, and the efficiency of decision-making will be improved.

3.3. Challenges facing the application of big data

The challenges associated with applying big data technology are briefly examined in this work, starting with the four stages of big data information collection, processing, analysis, and prediction.

In terms of data collection, the balance between velocity and value arises when big data technology is applied to financial risk management. “Velocity” requires big data technology needs to collect real-time data quickly; this means that it conflicts with the filtering and filtering of information required by "value. The pursuit of high speed might sacrifice the time of filtering and processing information and lead to the loss of critical information. For example, in risk management, analysts need to be aware of "black swan" events, which includes rare and unpredictable variables. However, big data technologies may ignore to ensure the speed of information collection. For stock market, big data technology has the potential to quickly process low-value data from hundreds of companies while ignoring notable events such as insider transactions and share buybacks, thereby exacerbating errors in analysis and decision-making [9].

Big data has information security challenges throughout the data collection stage in addition to the velocity vs. value conundrum. As mentioned above, due to the reduction of information storage and

computing consumption, the types and number of application devices of data collection technology are increasing, which increases the difficulty and complex of big data supervision and makes big data users have relative anonymity, which makes them collect other users' private information on the network without being awarded. Although some information technology firms have developed some applications to protect customers' privacy security by tracking the terminals used by "data pickpockets" and revoke their access to search engines. The effect of this initiative is limited, companies also need to strengthen internal employee supervision to prevent the misuse of big data technology. The information security involves a new problem, which will be shown on Figure 3.

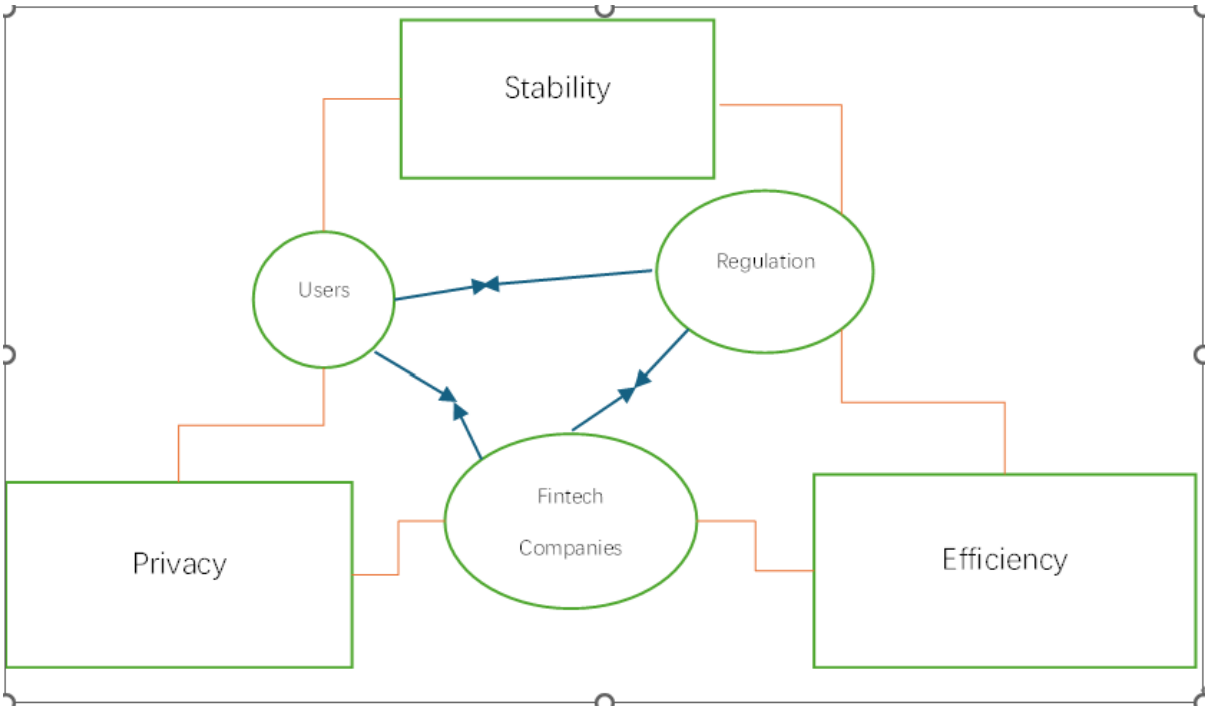

Figure 3: Policy trade-offs from digital transformation in finance [10].

In Figure 3, it is challenging for big data financial products to ensure user information security, the stability of financial environment and big data efficiency at the same time. Tech companies aim to protect user information in addition to using big data technology to gather as much data as possible to guarantee its operating efficiency and effective analytical capacity. Such measures allow these tech companies to earn higher profits than traditional financial industries such as banks, which makes it easy for them to develop into a single dominant situation, undermining the stability of the financial environment. Customers want to attain their information security and the stability of the financial environment, which will limit the information gathering of big data technology and impair its efficiency. For the regulatory authorities, they intend to build big data technology based on ensuring the stability of the financial environment to better supervise the financial industry. Their arrangement will require regulatory agencies to monitor more users' information, reducing user data security.

In terms of data processing, big Data technology is confronted with the challenging task of handling complicated data. In the financial sector, the selecting of data is particularly important, analysts neither excluding rare but critical data nor including unreliable data. In the financial sector, Data screening is especially crucial, as is not excluding rare but critical data or including inaccurate data. Human intervention, as well as the website itself, can have an impact on the information value. Existing big data technology can build its own logical chain through mechanical learning, but it still faces the problem that it can only use homogeneous data [11]. In addition, due to various sources of information, information from various sources may be different or even conflict in content, which brings latest problems to big data technology [12].

In process in data analysis, the massive amount of data presents a greater challenge to existing methodologies. For example, the variables in the traditional autoregressive model are less than 10,

but the variables in the big data are likely to reach hundreds of variables. If the traditional autoregressive model is still copied, the problem of overfitting may occur, and the credibility of the model will be reduced. In addition, in asset management, too many stocks will increase the error of covariance calculation, which will affect the credibility of analysis results [13]. Therefore, it is urgent to design a new analysis method for big data frameworks.

In terms of prediction, big data is faced with Survivor bias problem, for open market, analysts are more likely to make investment decision as they are easier to obtain information from the market. However, in alternative financial market, because of its opacity, analysts have difficulty obtaining valuable information, so they may overlook financial products of alternative investments and miss investment opportunities.

Not only the problem of big data itself, but also the problem of conflict of interest in the use of big data technologies in financial risk control. For example, when big data makes low-risk decisions for investors and tries to avoid investing in high-risk companies, big data technology also reduces the opportunities for some capital demanders to obtain investment [14]. In addition, big data technology will potentially widen the gap between companies. Because large companies or existing firms have more data for big data analysis, they can make more efficient decisions and reduce more investment risks compared with startups and small companies, and thus have better operating profits [15].

4. Future development

As traditional analysis methods are difficult to use under the condition of a vast quantity of data, big data analysts need to build a new analysis method to apply to the analysis of massive data. Here, machine learning and artificial intelligence play an indispensable technology. These models usually have a better fit than traditional analytical models, and AI decisions based on analysis results can help to avoid the problem of human bias. Thus, research on machine learning in large data analysis methods and artificial intelligence in decision-making systems will continue.

The emergence of big data technologies has exacerbated the conflicts of interest between traditional financial institutions and technological corporations. On the one hand, fintech companies have developed more intelligent, convenient, and rapid financial products using big data technology, which are becoming increasingly popular. However, these profit-making fintech companies are also affecting the market share of traditional financial entities such as banks. The promotion of big data technology in financial risk control may further intensify the game between traditional financial sectors and new fintech companies. The two sides may reach an agreement and achieve win-win results in the future, or one side may merge with the other.

5. Conclusion

By studying the application of big data technology in the risk control of financial products, this paper makes the following findings:

Big data technologies can process and analyze large volumes of data. HDFS and MapReduce components built on the Hadoop architecture can effectively classify and store data. Financial products supported by big data can make accurate risk assessment by collecting many relevant information of existing customers and potential customers in real time and reduce the possibility of financial losses. Taking the "E-finance" of Bank of Jiangsu as an example, "E-finance" has introduced a big data credit investigation system to collect and compare many internal and external information of small and micro companies for credit rating, this efficiently measures the credit risks of small and micro businesses.

Big data technology can strengthen the timeliness and intelligence of decision-making ability. Machine learning techniques can discover relationships between enormous amounts of data in real time and establish complex risk analysis and predictive modeling. In addition, the development of artificial intelligence technology makes it possible for financial products to skip manual decision-

making. Take Bridgewater's "Future Book" AI system, for example, which improves the company's operational efficiency by collecting and analyzing employees' data to formulate daily tasks on its own.

However, there are several problems in developing big data financial products. From the viewpoint of big data technologies itself. Firstly, traditional large data analysis methods have poor results, which directly affects the effectiveness of financial product analysis and decision-making. Second, big data technology may be limited by its own computing power to ignore rare but vital information in the financial market, thus increasing the risks in the operation of financial products. In terms of the operation of financial products, big data financial products bring about conflicts of interest among users, fintech companies and financial regulators, which may directly affect the operation of big data financial products. From the financial sector, big data technology brings about a conflict of interest between the traditional financial sector and fintech companies. How to balance the interests of users, fintech companies and financial regulators, and make the traditional financial sector and emerging fintech companies reach a consensus, will be the most pressing challenge in the development of big data financial solutions.

By investigating the use of big data technology in the risk control of financial goods, this paper provides the industry with a more thorough understanding of the use of big data technology and emphasizes the importance of big data technology to the financial sector. Furthermore, this paper discusses the risks faced by big data financial products in terms of conflicts of interest among users, financial companies, and financial regulatory authorities, as well as conflicts between traditional industries and fintech companies as big data technology develops, which will help the industry gain a more comprehensive understanding of the challenges faced by big data financial products.

However, there are many areas that need to be improved. In terms of information sources, this paper does not use primary sources but many secondary or even third-hand sources, which reduces the timeliness of this paper. In terms of analytical approaches, this work lacks a more extensive analysis of the application of big data in financial products. It just briefly covers two scenarios, which lacks general applicability. In the future, primary and secondary information sources will be collected as far as possible through investigation to improve the timeliness of information sources. Furthermore, in the future, research methodologies will be merged with big data technology and logic to have a better knowledge of how big data financial products are designed.

References

[1] The Economist, Nov 2011, "Drowning in numbers – Digital data will flood the planet and help us understand it better", http://www.economist.com/blogs/dailychart/2011/11/big-data-0

[2] Bhosale H S., Gadekar D P. A review paper on big data and hadoop[J]. International Journal of Scientific and Research Publications, 2014, 4(10): 1-7.

[3] Hannan S A. An overview on big data and hadoop[J]. International Journal of Computer Applications, 2016, 154(10).

[4] Alfityani A., Al-Lozi E., Abadalmajeed Alsmadi A., et al. The role of big data in financial sector: A review paper[J]. International Journal of Data and Network Science, 2022, 6(4): 2-13.

[5] Yu W., Wong C Y., Chavez R., et al. Integrating big data analytics into supply chain finance: The roles of information processing and data-driven culture[J]. International journal of production economics, 2021, 236: 108135.

[6] Peric A., Polic N., Kozarevic E. Application of Data Science in Financial and Other Industries/Primjena Znanosti Podataka U Financijskom I Drugim Sektorima[C]//International Scientific Symposium Economy of Eastern Croatia–Vision And Growth. Sveuĉilište Josipa Jurja Strossmayera u Osijeku, 2016.

[7] Zhou, L., Yan, X., Zhu, Y., et al. Case study on big data credit investigation service for small and micro-Companies financing: Taking "E-rong" series products of Bank of Jiangsu as an example [J]. Southwest Finance,2022(10):54-64.

[8] Solon O. World's largest hedge fund to replace managers with artificial intelligence[J]. The Guardian, 2016, 22: 2016.

[9] Trelewicz J Q. Big data and big money: The role of data in the financial sector[J]. IT professional, 2017, 19(3): 8-10.

[10] Doerr S., Frost J., Gambacorta L., et al. big techs in finance[M]. Bank for International Settlements, Monetary and Economic Department, 2023.

[11] Yaseen S G., El Qirem I A., Intention to use e-banking services in the Jordanian commercial banks[J]. International Journal of Bank Marketing, 2018, 36(3): 557-571.

[12] Fan J., Han F., Liu H. Challenges of big data analysis[J]. National science review, 2014, 1(2): 293-314.

[13] Huttunen J., Jauhiainen J., Lehti L., et al. Big data, cloud computing and data science applications in finance and accounting[J]. ACRN Journal of Finance and Risk Perspectives, 2019, 8: 16-30.

[14] Song S., Bickel P J. Large vector auto regressions[J]. arXiv preprint arXiv:1106.3915, 2011.

[15] Begenau J., Farboodi M., Veldkamp L. Big data in finance and the growth of large firms[J]. Journal of Monetary Economics, 2018, 97: 71-87.