Proceedings of ICMRED 2024 Workshop: Identifying the Explanatory Variables of Public Debt and Its Importance on The Economy
DOI: 10.54254/2754-1169/96/2024MUR0076

Prospects of the Application of Data Science in the Financial Field

Hongrui Ling[1,a,*]
[1]*School of Mathematics, Shanghai University of Finance and Economics, Shanghai, 200433, China*
a. 2023111233@stu.sufe.edu.cn
**corresponding author*

Abstract: The application of data science in the financial field has become an important trend in the financial industry. This article reviews the current application status of data science in the financial field, including application scenarios in risk management, credit assessment, market forecasting, etc. Through the analysis of existing research, this article summarizes the benefits brought by data science in the financial field, such as improving decision-making efficiency, reducing risks, optimizing customer experience, etc. In addition, this article also looks forward to the future development trends of data science in the financial field, including the application prospects of new technologies such as machine learning and deep neural networks in the financial field. Finally, this article calls on financial institutions to strengthen the training and technical investment in data science talents to adapt to the new trends in the development of financial technology and achieve digital transformation and innovative development in the financial industry. This article provides a relatively comprehensive introduction to the application of data science in multiple financial fields, reflecting the integration of mathematics, finance, computer science, and other disciplines. Most of the literature on academic websites starts from a specific aspect and introduces a certain field in depth. Although it is quite specific, it lacks an overall perspective.

Keywords: Finance, data science, quantitative investment.

1. Introduction

With the widespread popularization and maturity of big data technology, the application of data science in finance has become a hot trend in the industry. Take fund investment as an example. At the beginning of this century, Wall Street fund managers realized that data science would be a practical and efficient tool in the future of investment. James Simons, the famous quantitative fund manager and legend of quantitative investment, used a series of data to prove his success after moving to the "second battlefield" in the investment world for twenty years: between 1989 and 2009, he traded The return rate of the Medallion Fund is as high as 35%, which is more than 20 percentage points higher than the average annual return rate of the S&P 500 Index, and 10% higher than the trading performance of "financial giant" Soros and "stock god" Buffett. remaining percentage points. Even in 2007, when the subprime mortgage crisis broke out, the fund's return rate was still as high as 85%. Whether it was the Russian bond crisis in 1998 or the Internet bubble at the beginning of this century, the Medallion Fund has survived several financial crises and has always stood firm, eclipsing the

DOI: 10.54254/2754-1169/96/2024MUR0076

efficient market hypothesis. This example fully demonstrates the huge advantages of data science in the financial field, especially in the investment field [1].

Because quantitative investment is a high-frequency application field of financial data, this article first introduces the development overview of quantitative investment and lists the early and foundational literature on quantitative investment. Discusses the rise and development of global quantitative investment since the issuance of the world's first passive quantitative fund in 1971. It is then clarified that the development of machine learning has played an important role in promoting quantitative investing, especially with the advancement and application of deep neural network technology. Furthermore, the application of financial data science in specific industries is introduced, including both macro and micro levels. Finally, the prospects of this industry are prospected.

2. Early Quantitative Finance Literature

With the rapid development of data science, many academics have published many works in the fields of financial mathematics, financial engineering, and quantitative investment. The most classic of these is undoubtedly the option pricing model published by Black in 1973. It is also an important milestone in the field of financial engineering and can be called a pioneering work. It also laid the foundation for subsequent quantitative finance research and is widely used in option pricing and risk management [2]. A key conclusion of the Black-Scholes model is triangular hedging, which involves adjusting the underlying portfolio to offset changes in option value. This hedging strategy can help investors manage risk and potentially increase returns. The model emphasizes the importance of implied volatility, which is the market’s expectation of future volatility based on option prices. Implied volatility plays a crucial role in option pricing and can influence trading decisions. The paper written by Cox in 1985 proposed an interest rate term structure model based on risk-neutral pricing theory. By considering the risk preferences and uncertain attitudes of market participants, it explains the formation and changes in the term structure of interest rates. Unlike some books that focus on financial derivatives pricing and risk management, it provides an important theoretical basis for bond pricing and interest rate risk management and has had an important impact on empirical research and risk management [3]. Duffie's 1992 book also played an important role in the development of the field of financial data science [4]. It mainly discusses the dynamic asset pricing theory, which aims to explain the dynamic changes in asset prices and the dynamic investment behavior of investors. The book introduces the theoretical framework of dynamic portfolio selection, which considers investors' allocation decisions for different assets at different points in time. The dynamic evolution of asset prices is modeled as a state price process, that is, the random changes of asset prices in different states. Through the state price process, asset price fluctuations and the formation of risk premiums can be better understood. The literature published by Shreve introduces the application and principles of the binomial asset pricing model in the financial field. That is, by constructing a binary tree model to simulate changes in asset prices, thereby calculating the probability of the price and risk of options. At the same time, it involves random calculus methods commonly used in the financial sector, including Brown Movement, ITO, and option pricing, which can be called one of the classics in the field of finance and mathematics [5]. The papers published by Gray further introduced the quantitative method of value investment and discussed how to use data and algorithms to identify and underestimate stocks and invest [6]. These well-known books provide important information about theory, methods, and practical applications of financial engineering and other fields, and have important reference value.

DOI: 10.54254/2754-1169/96/2024MUR0076

3. Algorithm in Financial Data Science

3.1. Application of Machine Learning

Machine learning algorithms, as an important algorithm in financial data science, have naturally received widespread attention from the scientific community. Machine learning is a branch of artificial intelligence, which aims to make computer systems self-improvement by learning data and modes rather than clear programming instructions. In short, machine learning enables computer systems to learn and automatically improve from data without human intervention. The core idea of machine learning is to allow the computer system to learn and discover the model through data learning and discover, to make predictions, make decisions, or implement tasks. By analyzing massive data, machine learning can help financial institutions more accurately identify risks, predict market trends, optimize investment portfolios, and improve transaction execution efficiency. At the same time, machine learning algorithms can also be applied to anti-fraud, customer credit evaluation, personalized recommendation, and other fields to improve the quality and efficiency of financial services. Through the application of machine learning algorithms, financial institutions can better meet customer needs, reduce costs, improve profitability, and promote digital transformation and innovation development in the financial industry. Peter's thesis explores how to use machine learning algorithms to build quantitative trading strategies and optimize investment portfolios, highlighting the importance of machine learning in the field of financial trading [7]. David et al. The literature review of people and others focused on the application of artificial intelligence methods and machine learning algorithms in the financial service industry, providing a comprehensive overview of research in this field [8]. Debidutta's literature expounds on the contribution of machine learning algorithms to the science of economics and financial data and emphasizes the important role of machine learning in these fields [9]. Finally, the papers of Lin et al. Based on the Taiwan epidemic, the application of machine learning algorithms in the analysis of the financial time sequence analysis was displayed, and it provided a useful perspective of financial challenges in the special period [10]. Table 1 summarizes the above documents.

Table 1: Some specific applications

Literature	[7]	[8]	[9]	[10]
Application	Building quantitative trading strategies	Financial service industry	measuring economics fields	Time series analysis

3.2. Deep Neural Network -the Forefront of Machine Learning

Deep neural network algorithms use a multi-layer neural network structure to learn complex modes and characteristics of data. Deep neural networks train neural network models through a large amount of data so that the model can automatically learn the laws and characteristics in the data, thereby achieving tasks such as forecasting, classification, and identification. Deep neural networks can learn a lot of data, identify complex models and trends.For instance,Liu's article provides intelligent and efficient solutions for financial data processing [11]. Hilal introduced a new type of algorithm based on artificial intelligence-based financial data science optimal functions, which provides innovative methods for financial data analysis [12]. The core idea of this algorithm is to link multiple functions together to form an optimal function chain to achieve better modeling and prediction of financial data. Its advantage is that it can make full use of the non-linear modeling capabilities of neural networks. At the same time, the model is more flexible and efficient through the method of linking the optimal function. This algorithm can not only be applied to financial data analysis but also is widely used in other fields such as medicine and transportation. Valaitisv explained in detail the machine learning

DOI: 10.54254/2754-1169/96/2024MUR0076

projection method based on the macro-financial model, providing a new perspective for financial data modeling and forecasting [13]. The machine learning projection method is a data processing method based on machine learning technology. It aims to achieve the dimension of data reduction and feature extraction by projecting high-dimensional data into low-dimensional spaces. This method can help researchers better understand the structure and characteristics of the data while reducing the complexity of data, and improving the efficiency of data processing and analysis. Machine learning projection methods usually include technologies such as the main component analysis (PCA), linear judgment analysis (LDA), and T-distributed neighborhood embedded (T-SNE). These methods can help researchers to discover the potential models and associations in the data, to achieve visualization and understanding of data. In the financial field, machine learning projection methods can help analysts and decision-makers better understand the characteristics and laws of financial data, and find the information hidden behind the data to improve the accuracy of financial data modeling and prediction. By projecting high-dimensional financial data into low-dimensional space, researchers can see the relationship between data more clearly, to better carry out data analysis and decision-making formulation. Zhiwei and others introduced a high-frequency prediction method based on model fusion and feature reconstruction of neural networks, providing a more accurate volatility prediction tool for stock market participants [14]. In addition, Paolo's research concentrates on the unified tensor-based clustering coefficient approach in the financial multiple networks, providing more in-depth research and understanding of financial network analysis [15]. These studies jointly show the potential of using neural networks, machine learning, and artificial intelligence algorithms in the financial field, bringing innovation and development opportunities to financial science and practice.

4. Applications of Financial Data Science in Specific Industries

4.1. Macro Level

Financial big data plays a wide range of roles in various industries. At the macro level, the digital transformation of financial data has provided important support for government supervision and intelligent decision-making. The research of Cheng et al. discusses the practical significance of financial big data in financial services and rural revitalization [16]. The papers of Huang have explained the application of financial big data from the perspective of tax policy formulation [17]. The literature of Galena et al. introduced the importance of financial data science from the overall macro perspective [18]. During the pandemic, the application of big data has become particularly important. The research of Wen and other research fully demonstrates the indispensable effect of financial big data in dealing with emergencies [19].

4.2. Micro Level

Financial data science is widely used at the micro level, and the most effective one is undoubtedly the field of quantitative investment. This field is also one of the focuses of this research review. While quantitative investment is booming, many quantitative investment technologies have emerged. They cover almost the entire investment process, including quantitative stock selection, stock index futures arbitrage, risk control, etc. If a rational investor owned $ 1,000 in 1988, then he invested in the S&P 500, Buffett's Berkshire Hathaway, and the Medallion Fund, it would be very different, which is shown in Table 2 [1].

DOI: 10.54254/2754-1169/96/2024MUR0076

Table 2: Comparison of benefits of investing at 1,000 US dollars

Invest in	1988	2016
S & P 500	1000$	6600$
Berkshire Hathaway	1000$	52000$
Medallion Fund	1000$	13830000$

At the micro level, financial big data also has a profound impact on other various fields. Yang's research fully demonstrates the key role of financial big data in investment decisions [20]. In addition, there are studies on the management of supply chain financial risk management based on big data [21] and articles that use data science methods to predict stock price prediction [22]. These studies show the importance and extensive application of financial big data in various fields. Table 3 summarizes the application of macro and micro-level financial data in the above documents.

Table 3: Examples of applications for financial data science

Macro applications	Rural economic development	Tax policy formulation	Economic development during epidemic periods
Micro applications	Investment decision	Corporate risk management	Stock price forecast

5. Evaluation and Outlook

Through the analysis of this article, it can be inferred that data science is a cross-disciplinary discipline that integrates statistics, computer science, and field knowledge. Through the analysis and mining of massive data, it provides more accurate and efficient decision-making support for financial institutions and promotes finance. The innovation and development of the industry will undoubtedly empower the financial industry and have good prospects in the future.

First of all, through data science and technology, financial institutions can more accurately evaluate risks, predict market trends, and optimize investment portfolios. For example, using machine learning algorithms to analyze customer data can help banks better understand customer needs and provide personalized financial products and services; using big data technology to dig market data can help investors better grasp market dynamics to optimize their investment portfolio. The application of data science not only improves the decision-making efficiency of financial institutions but also provides customers with a better financial experience and promotes the digital transformation of the financial industry.

Secondly, with the continuous development of emerging technologies such as artificial intelligence, big data, and blockchain, the application of data science in the financial field will become wider and more diversified. For example, the application of artificial intelligence technology can realize the functions of intelligent risk management, intelligent customer service, etc., and enhance the service level and competitiveness of financial institutions; the application of blockchain technology can achieve decentralization and security of financial transactions, change the operating method of the traditional financial system. In the future, data science will be more closely combined with the financial field, bringing more innovation and changes to the financial industry.

In addition, the application of data science in the financial field also faces some challenges and problems. The first one is the problem of data privacy and security. Financial institutions need to protect customer privacy information when using big data to prevent data leakage and abuse; the second is the problem of data standardization and integration. Financial institutions need to integrate data from different sources to ensure the accuracy and consistency of the data. Finally, there are problems of talent shortage and technical barriers. Financial institutions need to increase the

DOI: 10.54254/2754-1169/96/2024MUR0076

cultivation and introduction of data science talents and improve the level of application of data science and technology in the financial field.

In summary, the wide application prospects of data science in the financial field will bring more opportunities and challenges to the financial industry. Financial institutions should increase investment and application of data science and technology, continuously improve the application level of data science in the financial field, and achieve digital transformation and innovative development of the financial industry. The application of data science will bring more efficient, smarter, and safer services to the financial industry, and promote the financial industry to move towards more prosperous and sustainable development.

Based on the discussion of this article, it could be concluded that in the context of the current digital transformation, financial institutions need to have more talent with data science skills. They can be proficient in data science tools and technologies, deeply dig data value, and provide more accurate and personalized services for financial business. At the same time, financial institutions also need to increase investment in data science and technology, build a strong data science team and technology infrastructure, promote the application and innovation of data science in the financial field, and realize the digital transformation and upgrade of the financial industry.

Only by strengthening the cultivation and technological investment of data science talents can financial institutions better adapt to the new trend of financial technology development and achieve digital transformation and innovative development of the financial industry. Financial institutions should promote the application of data science in the financial field, bring more opportunities and challenges to the financial industry, realize the intelligence, efficiency, and security of financial business, provide customers with better financial services, and promote the financial industry Towards more prosperous and sustainable development.

6. Conclusion

The application of data science in the financial field has achieved significant achievements. By analyzing and excavating a large amount of financial data, data science, and technology provides more accurate and more efficient solutions for financial institutions, which helps reduce risks, improve efficiency, and create value.

In terms of risk management, data science, and technology can help financial institutions better evaluate and manage risks and reduce potential losses. Regarding investment decision-making, data science methods can help investors better understand market dynamics, discover investment opportunities, and increase investment return. In terms of market forecasting, data science, and technology can predict market trends through modeling and analysis of financial data to provide investors with decision support. In terms of financial product innovation, data science, and technology can help financial institutions to design more personalized and efficient financial products and improve customer experience.

With the continuous development of fintech and data science and technology, the application prospects of data science in the financial field will be very broad. It is expected that data science will play an increasingly important role in financial risk management, intelligent investment, and personalized financial services. At the same time, the continuous integration and innovation of new technologies will bring more opportunities and challenges to financial data science, and bring more innovation and development opportunities to the financial industry.

In summary, the application of data science in the financial field has achieved remarkable achievements and shows broad development prospects. With the continuous progress of technology and the continuous expansion of application scenarios, data science will continue to play an important role in the financial field, bringing more opportunities for innovation and development to the financial industry.

DOI: 10.54254/2754-1169/96/2024MUR0076

References

[1] Hongsheng, S., & Hongneng, S. (2017). Simons, the Mathematical Master Who Laughs at Wall Street. Encyclopedia of Knowledge (03), 35-38.

[2] Fischer Black & Myron Scholes. (1973). The Black-Scholes Option Pricing Model. McGraw-Hill Education, 2-10.

[3] John C. Cox, Jonathan E. Ingersoll&Stephen A. Ross. (1985). A Theory of the Term Structure of Interest Rates. Princeton University Press, 1-12.

[4] Darrell, D. (1992). Dynamic Asset Pricing Theory.Pr University Press, 3-10.

[5] Steven, E. Shreve. (2004). Stochastic Calculus for Finance I: The Binomial Asset Pricing Model.Springer, 2-13.

[6] Wesley, R. Gray, & Obias, E. Carlisle. (2012). Quantitative Value: A Practitioner's Guide to Automating Intelligent Investment and Eliminating Behavioral Errors. Wiley, 4-11.

[7] Schwendner, P. (2020). Advances in Financial Machine Learning.Quantitative Finance(2), 189-190.

[8] Pattnaik, D, Ray S, & Raman R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. Heliyon(1), e23492.

[9] Hoepner, Andreas G. F., McMillan, D., Vivian, A., & Wese, S. C. (2021). Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective. The European Journal of Finance (1-2), 1-7.

[10] Hsio, Y. L., & Bin, W. H. (2024). Application of hybrid fuzzy interval-based machine learning models on financial time series — A case study of Taiwan biotech index during the epidemic period. Frontiers in Artificial Intelligence1283741-1283741.

[11] Baifang, L., & Liqiu, Sui. (2023). Intelligent analysis and processing technology of financial big data based on neural network algorithm. Journal of Artificial Intelligence Practice (2).

[12] Anwer, M. H., Hadeel, A., Fahd, N. Al-Wesabi, Mohammed, A. Al-Hagery, Manar A. Hamza, & Mesfer A. D. (2022). Artificial intelligence based optimal functional link neural network for financial data science. computers, Materials & Continua (3), 6289-6304.

[13] Vytautas, V., &Alessandro T. Villa. (2024). A machine learning projection method for macro-finance models. Quantitative Economics (1), 145-173.

[14] Zhiwei S., Zhifeng W., Shuaiwei S., Chengzhi, M., Yingqiao, W., & Laiqi, Z. (2022) .High-frequency forecasting of stock volatility based on model fusion and a feature reconstruction neural network. Electronics(23), 4057-4057.

[15] Bartesaghi, P., Clemente, G. P., & Grassi, R. (2022). A tensor-based unified approach for clustering coefficients in financial multiplex networks. Information Sciences, 268-286.

[16] Kangkang, C. (2024). Development path of rural financial services for rural revitalization based on big data analysis. Applied Mathematics and Nonlinear Sciences (1).

[17] Xiaofang, H., & Jun, Z. (2023). Investigation on tax planning and risk prevention based on financial big data. Financial Engineering and Risk Management (7).

[18] Pisoni, G., Molnár, B., & Tarcsi, Á. (2021). Data Science for finance: best-suited methods and enterprise architectures. Applied System Innovation (3), 69-69.

[19] Long, W., Kwak, W., & Lee, C. F. (2022). Introduction to special issue on financial market and healthcare in COVID-19 pandemic: From the Perspective of Data Science. Annals of Data Science (1), 1-3.

[20] Qifeng, Y., Yingying, W., & Yidong, R. (2019). Research on financial risk management model of internet supply chain based on data science. Cognitive Systems Research, 50-55.

[21] Min, Z., Jiantong, Z., Ruolin, M., & Xiaodong, C. (2019). Quantifying credit risk of supply chain finance: a Chinese automobile supply chain perspective. IEEE Access, 144264-144279.

[22] Budiharto, W. (2021). Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). Journal of Big Data (1), 47-47.