The role of big data in the secondary market

Xiaozhuo Pan

Vanke Meisha Academy, Shenzhen, 518085, China

xiaozhuop@gmail.com

Abstract. With the rapid development of information technology, Big Data has emerged as one of the important characteristics of the contemporary era. In the secondary market, the application of Big Data is increasingly widespread, which has a profound impact on market structure, trading patterns, and investment decisions. This paper aims to explore the utilization of Big Data in the secondary market, including data analysis techniques, enhancement of decision efficiency, augmentation of market transparency, and innovation in risk management practices while discussing the drawbacks of using Big Data. Then, the study focuses on how to improve and enhance the application of big data in the secondary market in the future. The findings underscore that Big Data plays an instrumental role in shaping the dynamics of the secondary market, this is predominantly manifested in the prediction of stock market trends, the assessment of investment risks, and so on. However, certain challenges persist. Continual development and improvement of Big Data can facilitate more informed investment decision-making.

Keywords: Secondary market, big data, investment, data analysis.

1. Introduction

In an era when information is as valuable as money, the stock market has proven the transformative power of Big Data. Financial markets are driven by the need to mine vast amounts of data due to the vast number of decisions being made every second. Big Data participation in the secondary market can help companies identify market opportunities and provide customers with high-quality products and services. Financial institutions, notably, have extensively integrated big data analytics to facilitate more informed investment decisions and secure sustainable returns. Algorithmic trading, in synergy with Big Data, leverages extensive historical datasets and intricate mathematical models to enhance portfolio profitability. The persistent utilization of Big Data is poised to transform the paradigm of financial services [1]. Nowadays, accelerating the construction of Big Data infrastructure and taking Big Data construction as a new economic growth point for enterprises has become an important choice for enterprises [2]. Nevertheless, amidst the distinct advantages brought by Big Data, the capacity to harness vast datasets encounters substantial impediments. This article further studies the problem of big data. Moreover, it discusses the development trend of Big Data in the future and how Big Data can be used more perfectly in the secondary market.

2. An overview of secondary market and big data

2.1. Secondary market

The secondary market is where investors buy and sell securities from other investors. Compared to the primary market, which is the initial offering market where a government or company sells securities for the first time, the secondary market is the place where bonds and stocks are traded after their initial issuance, also known as the circulation market or secondary market. The primary market is private, whereas the secondary market is public-facing.

2.2. Big data

Big Data refers to extremely large and diverse collections of structured, unstructured, and semi-structured data that have grown exponentially over time. In the year 2001, Gartner potentially set off a cascade of cumulative confusion through their publication of an article that prophesied the trends within the industry. They meticulously compiled and collated these predictions under the encompassing umbrellas of data volume, data velocity, and the various kinds of data. As time unfolded, the relentless rise of inflation was paralleled by the evolution of the buzzword 'Big Data,' which itself graduated from the '4Vs' to the '7Vs' and eventually to the '10Vs,' reflecting the exponential growth and complexity in the field. Until 2017, Big Data had 42 Vs characteristics [3]. Although each is closely related to big data, this paper primarily focuses on the basic '4Vs' of Big Data, which are Volume, Velocity, Variety, and Veracity. First, volume. The high volume of Big Data covers the continuous collection and generation of large amounts of data from various means, such as devices and sources. Velocity refers to the rate at which data is generated. Today's data is often generated in real-time or near real-time, which also requires that data must be processed, accessed, and analyzed at the same speed to produce the best effects. As for variety, data is heterogeneous, which means that it can come from many different sources and can be structured, unstructured, or semi-structured. Last, veracity. Massive datasets are often characterized by their Messiness, Noise, and Propensity for Error, which collectively hinder the manipulation and verification of data integrity. Enormous datasets can be ponderous and perplexing, whereas minuscule datasets may offer an inadequate representation. The authenticity of the data is inversely proportional to its reliability.

Big Data refers to the large amounts of structured and unstructured information produced at an unprecedented scale. In the context of market analytics, Big Data includes a variety of data sources, including social media, customer feedback, web analytics, and sales data. This wealth of information enables businesses to gain valuable insights into consumer behavior, market trends, and competitor analysis.

3. The close analysis of the application of big data in the secondary market

3.1. The current use of big data in the secondary market

The use of Big Data requires high-quality data analysis to achieve the user's purpose and achieve the best results. Research has shown that the ability of enterprises to develop strong Big Data analytics can promote superior market performance [4]. "Big data analytics requires a combination of quantitative and qualitative data," the founder of FINQ, an online secondary market trading platform for people to invest and trade, Eldad Tamir said [5]. This dual approach ensures that hard facts such as earnings reports and market trends are considered as well as qualitative factors such as company culture and competitive advantage.

The selection of appropriate data analysis tools and algorithms is crucial for delving into the intrinsic value of datasets. Popular Big Data analysis tools and open-source software solutions include Python, the R language, Excel, and SQL, among others. When opting for these tools, beyond the fundamental technical prerequisites, it is imperative to take into account the real-world requirements and resources available to the enterprise or individual, encompassing aspects such as cost-effectiveness, maintainability, and performance capabilities. Furthermore, utilizing appropriate Big Data analysis tools

enables comprehensive and meticulous investigation as well as model development. In-depth analysis encompasses techniques such as data visualization and statistical examination, among others, aimed at identifying patterns, trends, and discrepancies within the dataset.

Model construction is a pivotal aspect of data analysis, encompassing the selection of an appropriate model, its training, evaluation, and subsequent optimization [6]. Initially, the model is established by carefully selecting appropriate algorithms and methodologies. Subsequently, utilizing historical data to construct a predictive model facilitates the accomplishment of the model's training objectives. The assessment of the model's performance necessitates the application of test data for evaluation purposes. Finally, in the iterative optimization process, the model's accuracy is enhanced through persistent parameter tuning.

Table 1. Testing multiple linear regression analysis between big data analytics techniques[7].

Dependent Variables	Multiple Correlation on Coefficient R	Correlation Coefficient Square R²	F Value	Statistical Indication	Independent Variables	Non-Standard Transactions (B)	Standard Coefficients (β)	T Value	Statistical Indication
PB	0.624	0.513	1321.27	**0.000	Constant	0.028		1.523	0.176
					Volume	0.536	0.751	36.154	**0.000
					Velocity	0.898	0.381	16.214	**0.000
					Variety	0.827	0.574	23.223	**0.000
					Veracity	1.537	0.491	22.285	**0.000

** Statistically significant at the significance level (α=0.05)

Table 1 illustrates the correlation coefficient between the methodologies of Big Data analysis and the projected stock prices of Jordanian financial intermediaries, employing the linear regression approach in the analytical process. The data presented in this chart reveals that the magnitude of the multiple correlation coefficient is 0.624, with the coefficient of determination equaling 0.513. Such findings indicate that the four variables under consideration account for 51.3% of the variance in the predicted changes in stock prices. The F value is 1321.271, and the significance level is $\alpha \leq 0.05$, thus indicating a statistically significant impact of the four V's—Volume, Velocity, Variety, and Veracity— on the prediction of stock prices for Jordanian financial intermediaries. This implies that, from the perspective of Jordanian financial intermediaries, the deployment of Big Data analytics techniques in conjunction with these 4V's has affected the forecasting of stock prices.

It can be concluded that all four independent variables exert an impact on the prediction of stock market prices, with the regression coefficient, B, varying among these four variables, suggesting a differing degree of influence among them. Among these variables, Veracity exhibits the highest regression coefficient, indicating a preponderant influence. The Volume variable is positioned at the bottom of the ranking, indicating that, in comparison to other informational attributes, the organization assigns a higher priority to the veracity of the collected data over the quantity of datasets.

3.2. The current impact of big data on the secondary market

3.2.1. Advantages of big data participation in the secondary market
In the secondary market, most financial-related information and data suffer from low transparency, which often leads to unequal status between both parties in financial transactions. However, after the widespread application of Big Data technology, the transparency of financial-related data will be greatly improved. There will be no significant difference in information acquisition before and after the transaction, which can effectively avoid many risks caused by uneven information acquisition. Therefore, with the application of Big Data technology, the secondary market should use it to complete overall risk avoidance and improve the security of the secondary market [8]. Besides, some businesses leverage Big Data for predictive analytics, which uses historical and real-time data to predict future trends, risks, and opportunities. Financial analysts can use Big Data to predict market trends and make investment decisions. By analyzing historical market data, economic indicators, and current trends, predictive

models can provide valuable insights into asset price movements and market direction, helping investors make informed decisions.

Big Data also enhances decision-making by providing insights into market trends and customer behavior while enhancing risk management through predictive analytics. The ability to analyze individual customer data makes it possible to provide personalized services and investment advice. By collecting information such as customers' income level, assets, and personal expectations, big data finally gives customers a conclusion about investment risk assessment. For customers with different characteristics, they will have different risk-bearing capacities, and Big Data will provide customers with references for their own investment products. This greatly helps customers find the right product for them and thus manage their assets more effectively in the secondary market.

Furthermore, the deployment of computer algorithms and mathematical models for the execution of trading activities within financial markets is employed. This methodology entails the utilization of sophisticated technology to autonomously engage in the purchase and sale of assets in accordance with predefined rules and strategies, thereby aiming to enhance investment returns and optimize risk management.

Here is a computational example of an algorithmic trading strategy. The following figures are derived using a model of algorithmic trading called the Long Short-Term Memory Model (LSTM). The LSTM model was introduced by Hochreiter and Schmidhuber in 1997 [9]. As a variant of recurrent neural networks, it not only has the memory of recurrent neural networks but also introduces the gating mechanism to solve the problems of gradient disappearance and gradient explosion in the previous fitting process [10].

Table 2. RMSE, MAE, and R^2 of ICBC in different time steps[11].

Time Step	RMSE		MAE		R^2	
	Training Set	**Testing Set**	**Training Set**	**Testing set**	**Training Set**	**Testing Set**
3	0.01773	0.1635	0.01057	0.01422	0.95922	0.79635
5	0.01610	0.01174	0.00832	0.00834	0.96176	0.87408
10	0.01568	0.01098	0.00772	0.00685	0.96388	0.88555
20	0.01821	0.01644	0.01044	0.01310	0.95581	0.77146
50	0.01570	0.01132	0.00784	0.00760	0.96289	0.87658

Table 2 shows the estimated value of the evaluation index of ICBC stock. When the time step is 10, the RMSE on the training set and the test set are 0.01568 and 0.01098, and the MAE on the test set are 0.00772 and 0.00685, respectively, which are smaller than the RMSE and MAE when the time step is 3, 5, 20 and 50. At the same time, the value of R^2 is greater than the value of other time steps.

Combined with the forecast curves of ICBC under different time steps in Figure 1, it can be seen that when the time step is 10, the predicted value is basically consistent with the real value. Therefore, the optimal interstep of ICBC in the model forecast is 10.

Lastly, through a diverse array of methodologies, including real-time monitoring, proactive alerts, risk evaluation, and user behavior analytics, supplemented by network sentiment monitoring and the application of Big Data analytics, fraud detection in the investment sector can be significantly enhanced in terms of both accuracy and efficiency. This, in turn, safeguards investor interests and maintains market stability.
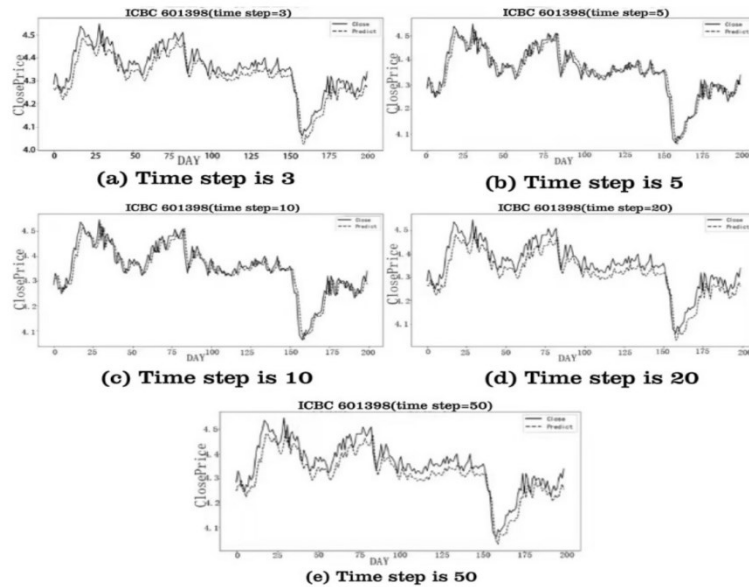
Figure 1. Forecast curves of ICBC under different time steps [11].

3.2.2. Problems of big data participation in the secondary market

The advent of Big Data has precipitated significant privacy and security issues, particularly in the context of amassing extensive financial data from individuals and businesses. Should such information be compromised, it would inevitably exert a profound and detrimental effect on both individuals and corporations.

Besides, the integrity and precision of data are paramount when employing Big Data within the secondary market, as these attributes critically determine the veracity of subsequent analytical processes. In the event that the foundational data is erroneous, fragmentary, obsolete, or tainted with bias, it can precipitate flawed analytical outcomes, thereby influencing the investment verdicts of both individual and corporate entities.

4. Future trends of big data in the secondary market

The current use of Big Data in the secondary market is evidently more beneficial than harmful, and it is inevitable that Big Data will continue to develop and progress in the future. The development of financial Big Data intelligence encompasses three crucial factors: data, computing power, and models, all of which are closely interrelated and indispensable. Models serve as the core element for the development of Big Data intelligence; thus, enterprises must have strong computing power and have massive data information in order to build an intelligent data model. Data is the key basis for matching and applying computing power. Computing power is the most important part of Big Data intelligence. Only when enterprises have strong computing power can they obtain more abundant data in time so as to effectively solve complex algorithmic problems [12].

Big Data holds the promise of building more powerful predictive models that can pinpoint complex patterns and correlations in data, leading to higher predictions of stock prices.

5. Conclusion

This essay delineates the deployment of big data in the secondary market, discussing both its merits and demerits. Illustrative examples are provided to elucidate the utilization of big data technology in current applications.

The integration of Big Data into stock market analytics marks a revolutionary shift in how we process, understand, and predict market movements. From defining huge amounts of Big Data to leveraging diverse data sets to employing complex algorithms for analysis, it's clear that the financial world is on

the cusp of a new era. Moreover, the broader impact of Big Data on reshaping the financial sector highlights its integral role in the modern investment landscape. It is imperative to acknowledge the challenges posed by Big Data in the context of secondary market development and to adopt pertinent strategies to address them effectively. By doing so, the potential of Big Data technology can be fully realized, thereby enhancing the decision-making process in investments.

There are some improvements to this essay. It predominantly employs the approaches of literature review and material retrieval to examine the utilization of big data. However, this methodology lacks robust reliability for the numerous conclusions presented within the article. Should circumstances allow, the manuscript could integrate proprietary, manually collected, and computed data to elucidate perspectives, thereby supplementing its reliance entirely on data sourced from literature.

Furthermore, future studies can be increasingly concentrated on the integration of artificial intelligence technology within the domain of big data. Examine the progression of artificial intelligence technology in the market, encompassing the exploration of AI usability and the augmentation of AI's practical applications, among other aspects.

References

[1] Trevir Nath, How Big Data Has Changed Finance, October 24, 2021. https://www.investopedia.com/articles/active-trading/040915/how-big-data-has-changed-finance.asp

[2] Xu, L. (2024). Research on Enterprise financial informatization construction under the background of big data. Enterprise economic, 07: 102-108.

[3] Tom Shafer, The 42 V's of Big Data and Data Science, April 1, 2017, https://www.elderresearch.com/blog/the-42-vs-of-big-data-and-data-science/

[4] Oluwaseun E.O., Nathaniel B., Magnus, H., Constantinos N.L., (2022) Big data analytics capability and market performance: The roles of disruptive business models and competitive intensity, Journal of Business Research, 139:1218-1230.

[5] Eldad Tamir, Data analytics in stock selection: Unlocking market potential, April 21, 2024, https://www.finextra.com/blogposting/25980/data-analytics-in-stock-selection-unlocking-market-potential

[6] Xu X.S., Li L. (2024). Research on corporate financial decision-making Strategy and Portfolio Optimization based on Big Data. Mall Modernization, 12:186-188.

[7] Zahidul I., Maruful H.C., Mohammad M.S., (2023). The Impact of Big Data Analytics on Stock Price Prediction in the Bangladesh Stock Market: A Machine Learning Approach. International Journal of Science and Business, IJSAB International, 28(1): 219-228.

[8] Zhang, J W. (2023), Application and development of big data in financial market. Public Relations World, 09: 134-135.

[9] Zhang A., Zachary C.L., Li M., Alexander J.S. (2023) Dive into Deep Learning. Cambridge University Press.

[10] Zhou S.W. Stock Index prediction based on EEMD-SE-LSTM model [D]. Shandong University of Finance and Economics,2024.

[11] Tao Y.K., Research on Algorithmic Trading Strategy under Deep Learning [D]. North China University of Water Resources and Electric Power,2023.

[12] Zeng W. Z. (2023). Development Status and future trend of financial big data intelligence. Business 2.0, 05: 4-6.