International Journal of Information Technologies and Systems Approach
Volume 16 • Issue 3

Financial Data Collection Based on Big Data Intelligent Processing

Fan Zhang, Shanghai Urban Construction Vocational College, China*
Ye Ding, Shanghai Urban Construction Vocational College, China
Yuhao Liao, Business School, University of Shanghai for Science and Technology, China

ABSTRACT

With the rapid growth of big data technology, its field plays a crucial role in financial data processing. No matter in the past or in the future, the financial industry has always been an important part of leading the development of the world economy, and the premise of this is to stabilize the financial environment. The current turmoil in the world economy also means that the financial environment is volatile. Therefore, the collection and analysis of financial data is an indispensable step. According to the collection and analysis of financial data intelligently processed by big data, this article studied the necessity of financial data collection and analysis, and used python crawler and k-means algorithm to process financial data. By crawling the stock trading information of a website, part of the data was extracted and visualized. According to the trend chart of the stock price and trading volume, the trend of the stock could be clearly and intuitively understood.

KEYWORDS

Big Data, Data Collection and Analysis, Financial Data, K-means Algorithm, Python Crawler

INTRODUCTION

Financial Data Collection Based on Big Data Intelligent Processing

With the rapid growth of the Internet, almost all human and commercial behaviors can be digitized, resulting in the accumulation and deposit of a large amount of personal and business process data, which also include financial data. The financial industry has always been the climax of the development of the world industry, playing a huge role in the growth of the world economy. The application of financial big data has become a hot trend in the industry. It has been widely used in specific businesses involving banking, securities, insurance, payment, and clearing, and Internet finance in many fields, such as transaction fraud identification, consumer credit, credit risk assessment, stock market prediction, stock price prediction, and intellectual property risk pricing. For example, the international financial market can gather a large amount of idle funds on an international scale, thus meeting the needs of international economic and trade development. In recent years, advanced

DOI: 10.4018/IJITSA.320514 *Corresponding Author

information technology has developed at high speed. Emerging technologies such as the Internet, big data cloud computing, and artificial intelligence have been more widely used in the financial industry. Financial data processing is particularly important, in this context, because data can truly reflect the attributes of objective things. Finance is an important part of the national economy and has a close relationship with all walks of life. Financial data are an objective description of the operation of financial enterprises and can also reflect the operation of the national economy. Only through financial data can the characteristics, laws, and operating conditions of financial activities be understood. Financial data analysis can help financial enterprises make economic decisions and play an essential role in improving economic efficiency.

The scale of financial data is growing due to the rapid growth of the financial industry. Increasingly, more kinds of data are being generated, and the real-time performance is stronger. However, it has become more difficult to obtain data, which has also attracted the attention of researchers in related fields. Several studies have focused on this issue. For example, Liang, Das & Kostyuk (2018) proposed a framework for understanding the national monitoring infrastructure by discussing how various government agencies cooperate to establish such a centralized data infrastructure to score credit and examining the different but interrelated processes of data collection, aggregation, and analysis. Müller, Fay & Brocke (2018) used a unique panel dataset that contained details of big data and analytics (BDA) solutions that 814 companies had between 2008 and 2014 and found that real-time BDA assets were related to an average increase of 3%–7% in enterprise productivity. The research results provided strong empirical evidence for BDA's commercial value and highlighted essential boundary conditions. Goertzen (2017) discussed the key features of quantitative data and various research questions they could answer. Then, Goertzen listed various performance measurement standards and indicators that could be used in the information management environment to support the conclusions and provide evidence for the development decisions of e-book collections. Goertzen also provided a research framework that could be used to plan and define the set analysis project. By focusing on potential processes and strategic activities, Chanias, Myers & Hess (2019) showed that digital strategy formulation not only represented a breakthrough in the previous practice of strategic information system planning, but also revealed the new extreme of emergency strategy formulation. Chanias et al. also concluded that a digital transformation strategy was sustainable and had no predictable outcome. Many researchers have provided theoretical methods for financial data collection and analysis and have made certain achievements. However, with the development of a more intelligent life, financial data collection and analysis requirements have become more precise.

Financial applications are becoming increasingly common with the continued popularization of big data. Big data processing and analysis have greatly contributed to financial data collection and analysis. Lee (2017) presented an integrated view of big data by tracing the evolution of big data over the past 20 years. Lee discussed the data analytics necessary to handle various structured and unstructured data and merchant review data for data analysis applications and evaluated the influence of big data on key operational performance. Choi, Wallace & Wang (2018) discussed the existing big data-related analysis technologies and identified their advantages, disadvantages, and main functions. They discussed various big data analysis strategies to overcome their respective computing and data challenges, reviewed the literature, and revealed how different types of big data methods (i.e., technology, strategy, and architecture) were applied to different subject areas. Finally, Choi et al. investigated the practical use of big data analysis in top brand enterprises through case studies. These studies have certain reference values, but most focus on the theoretical level, not combined with their actual application status for analysis.

With the advent of the big data era in recent years, more and more financial industries are engaged in big data application practices. Big data are used to mine, collect, classify, integrate, process financial data, and obtain reliable information from them to analyze big data, discover economic operation rules, and formulate business strategies. Big data are a breakthrough technology with huge potential. In this context, using big data to collect and analyze financial data is the choice of most enterprises.

Therefore, in this paper the authors combined financial data collection and analysis based on big data intelligence to study its effects on related fields in the financial industry. This study is a simulation experiment of financial data collection and analysis based on a Python crawler and k-means algorithm. Through the crawler financial Web site data, combined with the data visualization library in Python, the trend of stock turnover, price trend, and trend price and turnover can be expressed through the line, scatter, and other intuitive charts. The above data show this method has strong timeliness, which can be accurate to the data of every hour and minute. K-means, a Python-based crawler and clustering algorithm, has a powerful processing function for financial data. Hence, this case can be used as a reference for enterprises to conduct financial analysis and can be judged according to the extracted valuable data for data analysis.

FINANCIAL DATA COLLECTION

Financial Data

A large amount of data is generated every minute and every second in the world. According to the International Data Center Company report, by 2020, an individual would generate 1.7 megabytes in just one second. Internet users generate about 250 million gigabytes of data every day. Since the rise of the Internet industry, data have gradually covered almost all industries, representing an overall trend in society. It is a huge challenge to collect, store, and analyze such massive data. How these data can be collected, stored, analyzed, and valuable information extracted from them are all issues that need to be studied by various industries, in the future. Financial data is an umbrella term for a range of data involved in the financial industry sector, including market data, company data, industry indices, and pricing data. Financial data represent the performance of enterprise operations, which can be divided into four categories, namely, user data, market transaction data, analysis data, and other data:

1. **User Data:** They can also be called basic data, which include personal customer information, such as name, gender, age, identity, and contact information. They also includes enterprise customer information data, name, affiliated enterprises, industry, sales amount, and registered capital. Product development and decision support are carried out based on other transaction data. For example, the consumer group can be understood according to the customer's age, occupation, income, and other information. For this part of the group, corresponding financial products are promoted, and customer consumption measures are formulated to provide customers with different financial services according to their needs. The actual sales of the product can be determined after data analysis. Marketing expenses are reduced by promoting specific customers.
2. **Market Transaction Data:** They refer to all trading activities in the exchange or trading platform. Market data are more standardized and frequent. The amount of data is larger and processing is more difficult, but the value is greater. Enterprises can use these data to analyze the consumption behaviors of individual consumers to develop products and provide financial services. At the same time, these data can also be used to provide supply chain financial services and establish its commodity trading ecosystem. Integrated market transaction data can be used for risk management to intervene in risk events in advance, reducing the risk of debt default. For example, an enterprise can predict the sales trend of a product in the next quarter based on its market transaction data and formulate a sales plan based on the sales trend to obtain more profits. When an enterprise collects and uses market transaction data, it is challenging for the big data industry to determine which data should be used and how they should be presented.
3. **Analysis Data:** They represent the results of the analysis of the original data information, also called the derived data of the original data, such as the in-depth analysis of the company's business model, activities, competition, and prospects. An important aspect in analyzing data is the choice of their processing method, which can be extracted directly from the source for the

target customers. However, data analysis also involves high cost, and the entire methodology may be opaque or biased.

4. **Other Data:** They include other types of data that need to be used in addition to the above three kinds of financial data. They also play an important role in data analysis and decision-making, and include behavioral, location, business, and business process data.

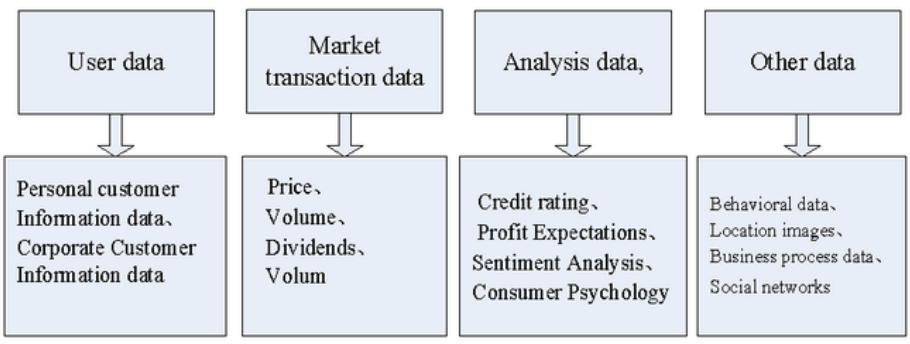
Figure 1 shows the classification of financial data. In addition to the general characteristics of data, financial data also have specific characteristics, which include extensiveness, comprehensiveness, reliability, and continuity:

1. **Extensiveness:** Financial institutions have a special position in the national economy and are related closely to each economic cell and the micromain body of the whole society. Therefore, data must be widely available to society, broadening financial data coverage.
2. **Comprehensiveness:** As a comprehensive department of the national economy, financial data directly involve all sectors of the national economy and provide financial services for all social groups. Through these services, especially financial services, comprehensive data reflecting the operation of the national economy can be collected. Therefore, financial data are highly comprehensive.
3. **Reliability:** Financial enterprises provide various financial services, and have not only a service relationship, but also a contractual one. In particular, the financial services provided by financial enterprises reflect the relationship of credit, custody, and agency. Hence, financial enterprises should not make mistakes in their services, and the data fed back into their operations must be true and reliable.
4. **Continuity:** Financial data, whether on financial business or national economic activities, are a dynamic reflection of the entire economic activity. With the continuous development of economic activities, financial data are constantly generated and systematically reflect the development and changes of economic activities. The core financial data and relevant information are organically organized, and users can obtain all information about a financial transaction through a report. The report provides considerable information but not all information is useful for data analysis of different users. Only the data about which users care are effective. Hence, financial information analysis should be carried out to collect useful and reliable information.

Data Collection

Although big data technology has developed rapidly in recent years, it has also been applied to the financial market. However, predicting the current situation and future of the financial market remains a challenge. Therefore, the collection and analysis of financial data play a decisive role.

Figure 1. Classification of financial data



Data Acquisition

Data acquisition involves the extraction, processing, transformation, and organization of data, with the final aim of unearthing its potential value. Data acquisition is an important part of the data analysis life cycle. Due to the complexity of the types of financial data, data in complex formats must be extracted through extraction technology to obtain the required data from the original data format. Finally, the data in the data warehouse are analyzed and processed.

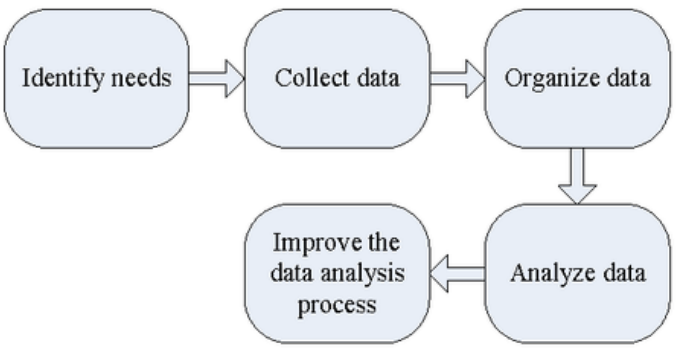
Information has become increasingly available, generating large amounts of individual-level data for ranking and analyzing categories of risk or value (Fourcade & Healy, 2017). However, the use of financial data faces challenges, such as large data volume, strong real-time, high data quality requirements, data security, and reliability. Data collection is the process of obtaining useful information from massive and messy data with the help of collection tools and methods and integrating them for management and use. Collecting financial data is an important step in the operation of the financial industry, which generally includes demand research, positioning, data preparation, collection, and sorting. Financial data collection can also be divided into enterprise internal and external data collection. Enterprise internal data are generally related to the production of enterprises, involving the confidentiality of user information and trade secrets. The company's internal data are stored in the database, meaning that the company's business, channel, cost, income, and other production processes are digitized and stored in machines. Generally, most enterprises develop their own data storage systems and establish a secure database system for internal data storage. The data analyst can extract the desired data table and collect data through database language. The data collected from the external data of the enterprise are more macroscopic and open. Most of these data are not specific to the operation and production of a company but more specific to the external environment of society and the industry's economic situation.

Data Analysis

After financial data are collected and processed, valuable data are extracted and stored on the data platform, providing a convenient, fast, and real-time environment for financial data analysis. Financial data analysis is an indispensable step for enterprise development. Data acquisition and data analysis are inseparable. Data collection is the premise of data analysis, and is an essential process after data collection. Data collection can be classified into the following: Identifying needs, collecting data, sorting data, analyzing data, and improving the data analysis process (Figure 2).

On the one hand, data analysis can reduce enterprise costs and improve the efficiency of the company. In the face of a large amount of financial data, considerable human time is needed to collate and find valuable data for classification. Therefore, during data analysis, analysis tools are used to find the relationship between data and some reports or charts. On the other hand, financial data analysis can

Figure 2. Data acquisition and analysis steps



increase the accuracy of analysis results in predicting financial risks, thereby reducing financial risks of enterprises. Enterprises can make judgments, predict risks, and make business decisions based on the analysis results of financial data. For example, companies often take into account the future expansion of production capacity when considering adding new investments to increase production capacity and operating income. Enterprises can judge the popularity of products and popular groups according to the consumption volume and consumer groups in a quarter to prepare for product sales in the next quarter.

Financial Data Processing of Big Data

The term "big data" has been a hot word in recent years. "Big data" refers to a large amount of information that can create miracles. In the past two decades, it has become a topic of special interest because of huge potential. Various public and private sectors generate, store, and analyze big data, with the aim of improving the services they provide (Dash, Shakyawar & Sharma 2019). Big data is a new engine for sustained and rapid growth of the financial industry. New technologies, new products, new services, and new forms of business facing the financial market are constantly emerging, and big data technology brings huge prospects to the financial industry. Big data mining technology based on back-propagation neural network can optimize input variables and effectively extract hidden information from data (Song & Wu, 2022). Big data technology refers to the use of relevant technologies to develop these meaningful data after mastering huge amounts of data and information. Big data technology includes massive parallel processing database, data mining power grid, distributed file system, distributed database, cloud computing platform, Internet, scalable storage system, and many other technologies. The use of big data technology to process data can be said to be the core of big data. Therefore, it is an unquestionable choice to use big data technology in collecting and analyzing financial data. Data collection and analysis methods based on big data intelligence include database collection, log file collection, Web crawler collection, and sensor data collection. Because of the features of financial data, the use of traditional methods, such as research, collection, text processing, and other traditional manual methods, would incur huge costs to human and financial resources. This paper introduces the data and analysis based on Python and solves the problem of collecting and analyzing financial data through Python data crawler and data mining.

Python Crawler

Python is an interpreted and object-oriented scripting language with concise language and no complex data types, which greatly reduces the development difficulty. In recent years, Python software development has progressed positively. Python has many built-in standard libraries, including crawler class urllib library, sketch library, Python library, computing class numpy library, pandas library, scipy library, image processing class pillow library, matchlotlib library, and other powerful third-party libraries. The crucial technologies of Web crawlers are reflected in three main areas: Web page acquisition, Web page analysis, and data storage (Raguseo, 2018). Python Web crawler refers to the automatic crawling and collection of network information through Python tools, and the collection of data from the Internet is usually completed using a Web crawler (Gunawan, Amalia & Najwan, 2017). The crawling tool has high performance requirements because of the huge information contained in financial data. The authors selected Python crawler to collect financial data. The basic process is as follows. First, according to the research requirements or the part of data that needs to be crawled, the Uniform Resource Locator (URL), that is, the target page to be accessed, is obtained. The second step is to build a request and send a request to the server to obtain the corresponding content. Because crawling data require the urllib library of Python, the urllib library should be imported first. The syntax is as follows: import urllib. The third step is to parse and extract data after obtaining corresponding data. Regular expressions can be used for this step. Regular expression is a special character sequence that can help to conveniently check whether a string matches a certain pattern. In this study the authors used the re.match function to match. The syntax is as follows:re. match(pattern, string, flags=0). The re.match function is a regular expression processing function commonly used in Python, and

re.match tries to match a pattern from the starting position of the string. If the starting position is not matched successfully, match() returns none. The re. match() method returns a matching object, not the matching content. The location of the matching result can be obtained by calling span(). The fourth step is to output the data after extracting the required content. The data can be saved to the dictionary, returned to the local, and can be saved to the local plain text file (CSV) format using the data frame format. The steps are as follows:

$$\text{item} = \{\text{“financial data”}\}$$

$$\text{return item}$$

Item represents the acquired data content and financial data represents the acquired data. After extracting the data content, the data are returned to the local. The syntax is as follows:

$$\text{df} = \text{pd.DataFrame}(\text{items})$$

$$\text{df.to\_csv}(\text{"flie name"})$$

Crawling financial data involves repeating the analysis and extraction process and randomly entering the next capture workflow. Figure 3 shows the Python crawler workflow.

K-Means Algorithm

Python clustering algorithms usually merge the input data according to the central point or hierarchical method. All clustering algorithms attempt to find the internal structure of the data to classify them according to the largest common point. The k-means algorithm is the most representative clustering algorithm in Python. This paper uses the k-means algorithm to analyze financial data. The k-means algorithm is a relatively basic clustering algorithm. It divides a given dataset into k clusters and gives the corresponding center point of each sample data. The algorithm steps are as follows:

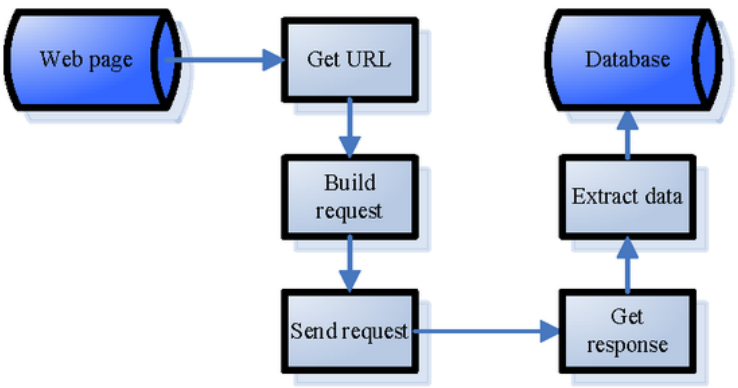
1. Data normalization and standardization (data preprocessing).
2. K centers are randomly selected, which is recorded as: λ_i , i=1,2,3, k.

The distance from each sample point to the “cluster center” is calculated as follows:

$$d\left(\mu_{j,}\lambda_i\right), i = 1,2,3..,k., j = 1,2,3,...,k \tag{1}$$

According to d, the new cluster is divided. If

Figure 3. Python crawler workflow



$$d\left(\mu_j, \lambda_i\right) = \min\left\{d\left(\mu_j, \lambda_{i'}\right), i' = 1,2,..,k\right\}, \tag{2}$$

then the cluster center is updated as follows:

$$\lambda_i = \sum_{n}^{1} \mu_j \epsilon M_i \tag{3}$$

The above steps are repeated until the clustering function $E=\sum_{j=1}^{n}\sum_{i=1}^{k}\left\|\mu_{j-\lambda_i}\right\|^2$ and the clustering center no longer changes significantly, that is, it converges as shown in Figure 4.

FINANCIAL DATA COLLECTION EXPERIMENT SIMULATION TEST

Data Crawler

Financial data collection and analysis are crucial to the financial industry's growth. In this paper, the authors used a financial Web site as an example to test the collection and analysis object. Data visualization is carried out in combination with matablib library in Python through the Python crawler and k-means algorithm. First, the URL of the financial Web site is obtained, and the encoding format is changed to "UTF-8." Then, it waits for the server to respond and starts crawling the securities information. Next, the information is saved, and the data are saved locally in CSV. Table 1 shows the acquisition of securities information from a financial Web site.

Table 1 lists a security's stock price, average price, and trading volume every other minute. Only a part is shown here because of the large amount of data involved. Python crawlers are used to obtain data every other minute. The next few are data every 10 minutes. The crawler's good data are saved locally. Table 1 shows that the stock price is different every minute, and the data have strong timeliness. Using Python crawlers allows to obtain real-time data, which are powerful.

Financial Data

After collecting the financial data of the Web site, it is necessary to analyze them. The regularity contained in financial data is often highly time-sensitive. The rules are constantly changing with the changes in the time and economic environment. Data visualization transforms abstract data into understandable forms, such as charts and reports, for understanding. Data visualization represents each data item in the database as a single primitive element, with many datasets forming a data image. At the same time, the various attribute values of the data are represented in the form of multidimensional

Figure 4. Schematic diagram of K-Means algorithm

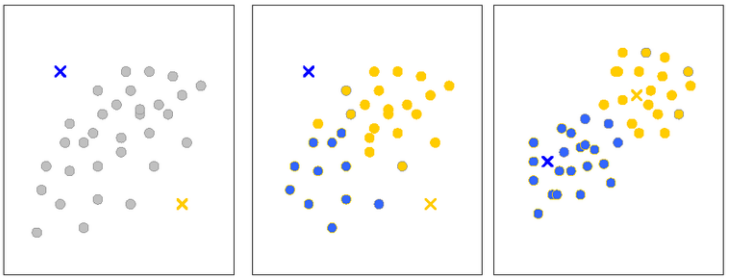


Table 1 Data of some financial web sites crawled by Python

Time	Price	Average price	Volume
10:00	19.98	19.98	10034
10:01	19.96	19.97	6780
10:02	19.97	19.97	1580
10:03	19.96	19.96	3567
10:04	19.98	19.96	4790
10:05	19.97	19.96	8854
10:10	20.04	20.02	4759
10:20	20.17	20.05	29339
10:30	20.34	20.11	9987
10:40	20.45	20.22	23345
10:50	20.59	20.29	6675
11:00	20.68	20.34	3348
11:10	20.79	20.39	7990

data. The data can be observed from different dimensions to conduct more in-depth observation and analysis. The increase in available data has attracted attention to clustering methods to integrate them coherently and determine the pattern of big data (Jahwar & Abdulazeez, 2020). In this study, the authors used the k-means clustering algorithm to visualize financial data in combination with matplotlib library, numpy library, padans library, and other libraries in Python. The development of big data analysis offers new paradigms and solutions for big data sources, storage, and advanced analysis (Saggi & Jain, 2018). Data analysis is generally reflected in three aspects:

1. **Data Preprocessing:** The original data obtained may have irregular problems, such as data loss or duplication. Therefore, data preprocessing is required before data analysis.
2. **Descriptive Analysis:** Some data rules can be displayed intuitively through graphs and tables. In this study, the authors used Matplotlib and Pandas for the analysis. Matplotlib can generate data visualization modules to help users draw, and Pandas data analysis package can quickly provide Python with structured data resources. Matplotlib is the most widely used component in Python drawing. It can draw a variety of graphics, including scatter plots, line plots, bar plots, contour plots, grayscale plots, pie plots, volume field plots, and polar axis plots.
3. **Prediction Analysis:** Some rules found through descriptive analysis are processed through feature engineering, mathematical statistics algorithm, machine learning model, and deep learning model, and are selected to predict the development of things (Zhang,Huang & Zhang, 2022). It is more accurate and real-time than other traditional data visualization algorithms. Figure 5 shows the analysis results of the k-means algorithm.

Figure 5 shows that the trend of stock price and the average price is rising, and the stock price is higher than the average. The stock price is generally on the rise between 10:00 and 11:30 AM, and the average price is also on the same trend. Figure 6 shows the trading volume of the stock during this period.

Figure 6 shows that the trend of stock trading volume between 10:00 and 11:30 AM fluctuates greatly, with two trading peaks, and the data are updated quickly. This algorithm can visualize data well in the case of large amounts of data in the time block of data update. Figure 7 shows the relationship between stock price and trading volume with the k-means algorithm.

Note. Figure 7a shows broken line chart of stock trading volume; Figure 7b shows scatter chart distribution of stock price and trading volume.

Figure 7a shows the daily turnover of stocks in recent days. The daily peak of stock turnover in recent years and the trend of turnover in recent days can be seen by using this line chart. Figure 7b

Figure 5. Stock financial data visualized by the k-means algorithm

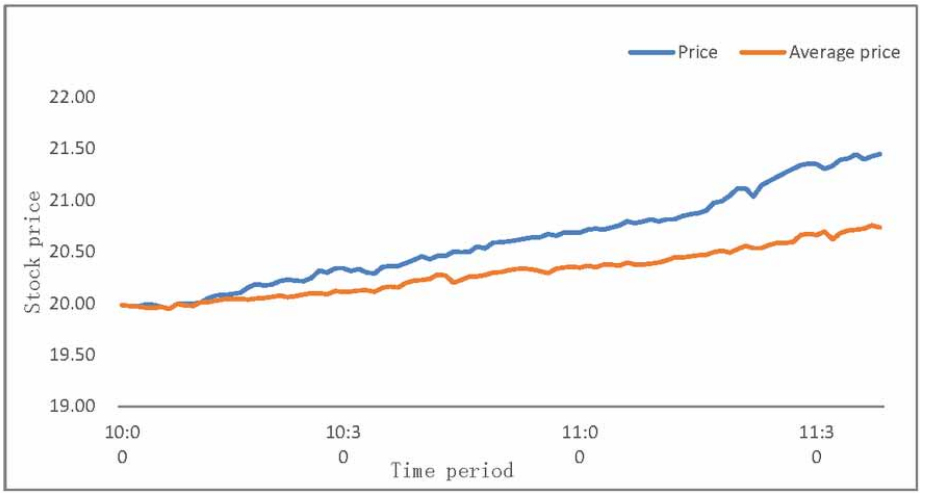
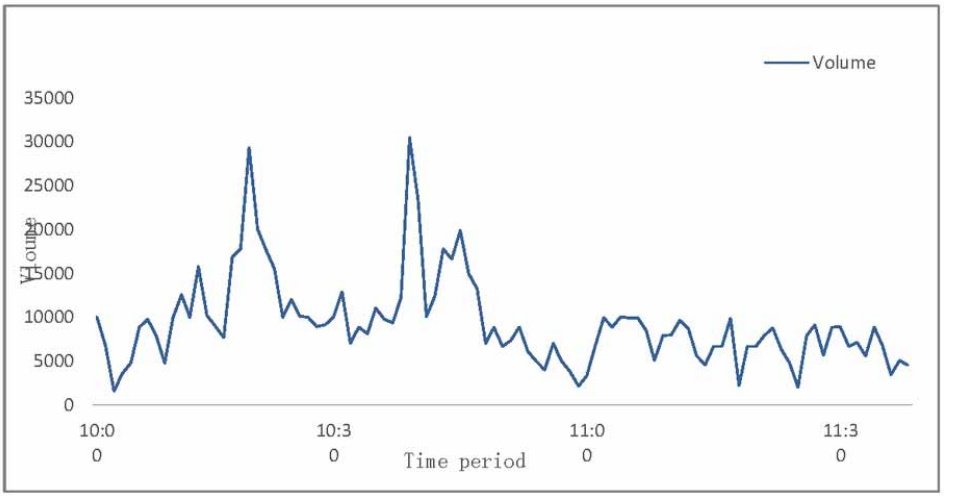


Figure 6. Trading volume of a stock visualized by k-means algorithm data

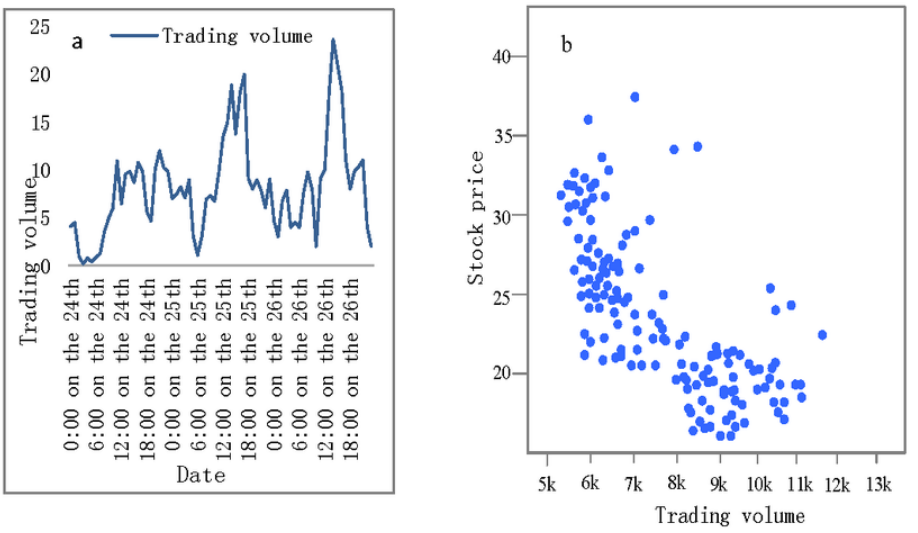


shows the relationship between the stock price and trading volume in recent days using the k-means algorithm and Python data visualization. The scatter plot shows the relationship between the two more intuitively thereby enabling judgement of the relationship between stock price and trading volume. When the stock price is around 20, the trading volume is 8k to 11k. When the stock price is 25 to 35, the trading volume is distributed from 5k to 7k. Therefore, the stock merchant can rely on these stock data to make judgments and analyze it to understand the operating rules and predict future stock trends.

CONCLUSION

Financial data play a very important role for financial enterprises. It is the core of enterprise development and the element of enterprise basic operation. Data value was created through data analysis, thereby improving the service quality of the enterprise and creating more benefits. The

Figure 7. Relationship between stocks and trading volume



ultimate success of any big data and analysis project lies in realizing strategic business value, which brings competitive advantages to enterprises (Grover, Chiang & Liang, 2018; Li & Zhang, 2021). In this paper, the authors introduced the concept of financial data collection and analysis and studied financial data processing based on big data. The authors collected and analyzed financial data through a Python crawler and k-means algorithm (Li, Wang & Yang, 2022; Lv, Lou, Li, Singh & Song, 2021). They extracted certain security of a Web site, and carried out data visualization to analyze the data. The data are updated quickly and in real-time. Financial data collection based on big data can solve the problem of slow updates of traditional algorithms. It can process massive amounts of data in combination with the MySQL database platform, which has strong practicability for processing financial data. However, because of the limitations of the level of simulation analysis, some areas need to be improved. Hence, in future research, the imperfections of this article should be improved to promote big data in financial data collection and analysis (Hou, Li, Liu & Zhang, 2021).

DATA AVAILABILITY

Data sharing is not applicable to this article, as no new data were created or analyzed in this study.

CONFLICT OF INTEREST

The authors state that this article has no conflict of interest.

FUNDING

The authors received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

Chanias, S., Myers, M. D., & Hess, T. (2019). Digital transformation strategy making in pre-digital organizations: The case of a financial services provider. *The Journal of Strategic Information Systems*, *28*(1), 17–33. doi:10.1016/j.jsis.2018.11.003

Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and Operations Management*, *27*(10), 1868–1883. doi:10.1111/poms.12838

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data*, *6*(1), 1–25. doi:10.1186/s40537-019-0217-0

Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-economic Review*, *15*(1), 9–29.

Goertzen, M. J. (2017). Introduction to quantitative research and data. *Library Technology Reports*, *53*(4), 12–18.

Grover, V., Chiang, R. H. L., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, *35*(2), 388–423. doi:10.1080/07421222.2018.1451951

Gunawan, D., Amalia, A., & Najwan, A. (2017). Improving data collection on article clustering by using distributed focused crawler. *Data Science: Journal of Computing and Applied Informatics*, *1*(1), 1–12.

Hou, J., Li, Q., Liu, Y., & Zhang, S. (2021). An enhanced cascading model for e-commerce consumer credit default prediction. *Journal of Organizational and End User Computing*, *33*(6), 1–18. doi:10.4018/JOEUC.20211101.oa13

Jahwar, A. F., & Abdulazeez, A. M. (2020). Meta-heuristic algorithms for k-means clustering: A review. *PalArch's Journal of Archaeology of Egypt/Egyptology, 17*(7), 12002—12020.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, *60*(3), 293–303. doi:10.1016/j.bushor.2017.01.004

Li, L., & Zhang, J. (2021). Research and Analysis of an Enterprise E-Commerce Marketing System under the Big Data Environment. [JOEUC]. *Journal of Organizational and End User Computing*, *33*(6), 1–19. doi:10.4018/JOEUC.20211101.oa15

Li, X. T., Wang, J., & Yang, C. Y. (2022). Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy. *Journal of Manufacturing Processes*, *35*(3), 2045–2058.

Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy and Internet*, *10*(4), 415–453. doi:10.1002/poi3.183

Lv, Z., Lou, R., Li, J., Singh, A. K., & Song, H. H. (2021). Big data analytics for 6G-enabled massive Internet of things. *IEEE Internet of Things Journal*, *8*(7), 5350–5359. doi:10.1109/JIOT.2021.3056128

Müller, O., Fay, M., & vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, *35*(2), 488–509. doi:10.1080/07421222.2018.1451955

Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management*, *38*(1), 187–195. doi:10.1016/j.ijinfomgt.2017.07.008

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, *54*(5), 758–790. doi:10.1016/j.ipm.2018.01.010

Song, Y., & Wu, R. (2022). The impact of financial enterprises' excessive financialization risk assessment for risk control based on data mining and machine learning. *Computational Economics*, *60*(4), 1245–1267. doi:10.1007/s10614-021-10135-4

Zhang, Y., Huang, J., Zhang, J., Liu, S., & Shorman, S. (2022). Analysis and prediction of second-hand house price based on random forest. *Applied Mathematics and Nonlinear Sciences*, *7*(1), 27–42. doi:10.2478/amns.2022.1.00052

Fan Zhang was born in NanChang, JiangXi. P.R. China, in 1975. She received the bachelor's degree from Central South University (formerly Changsha Railway College), the Master degree and the Doctor degree from Nanchang University, P.R. China. Now, She works in Shanghai Urban Construction Vocational College. Her research interests include digital economy, scientific and technological finance.

Ye Ding was born in Shanghai. P.R. China, in 1977. He received the bachelor's degree and the Master degree from Central South University. Now, He works in Shanghai Urban Construction Vocational College, His research fields include block chain, big data and Virtual Simulation Technology.

Yuhao Liao was born in NanChang, sJiangXi. P.R. China, in 2003. Now, he studies in Business School, University of Shanghai for Science and Technology.He is interested in scientific and technological finance.