

Retraction

Retracted: Big Data-Based E-Commerce Transaction Information Collection Method

Complexity

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] B. Yan, C. Wu, R. Yu et al., "Big Data-Based E-Commerce Transaction Information Collection Method," *Complexity*, vol. 2021, Article ID 8665621, 11 pages, 2021.

Research Article

Big Data-Based E-Commerce Transaction Information Collection Method

Bingwen Yan ^{1,2}, Chunqiong Wu ^{1,2}, Rongrui Yu ^{1,2}, Baoqin Yu,^{1,2} Nafang Shi,^{2,3} Xiukao Zhou,^{1,2} and Yanliang Yu^{1,2}

¹Business College, Yango University, Fuzhou 350015, Fujian, China

²Big Data Business Intelligence Engineering Research Center of Fujian University, Fuzhou 350015, Fujian, China

³Zhejiang Yuying College of Vocational Technology, Hangzhou 310018, Zhejiang, China

Correspondence should be addressed to Bingwen Yan; bwyan@ygu.edu.cn

Received 29 April 2021; Revised 5 June 2021; Accepted 11 June 2021; Published 17 June 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Bingwen Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of e-commerce industry, online shopping has become a craze. With the rapid growth of transaction volume on e-commerce platforms, a large amount of transaction data has been accumulated. From the transaction information of these users, a lot of very valuable information can be mined, such as the defects of products and the actual needs of users. In view of the existing e-commerce transaction information collection method is not mature, in this paper, the electric business platform system architecture planning and design increases the function management module. In this paper, a new Naive Bayes model is established by using HBase distributed database instead of traditional database. Based on the optimization and extraction of the important transaction information in the product, the dataset of e-commerce transaction information is updated. Through the efficiency test of the collection method, the information scalability ability test, and the accuracy test, the important context was sorted out after integration, the sources of trading information were sorted out, and the data analysis of the collected information was conducted to optimize the information collection method and verify the feasibility of the method.

1. Introduction

With the further development of network technology into the era of Web2.0, e-commerce platforms are blossomed, social networks are getting bigger and bigger, shopping is becoming more and more convenient, and communication is becoming more and more convenient [1]. The amount of network data is also growing exponentially. In recent years, the total amount of global data has accumulated to more than 35 ZB. This is also a symbol of the big data era. Although the development of network technology has also expanded the way to obtain information from the extremely large amount of data, the reality is that as the volume of data continues to accumulate, it is still very difficult and will become more difficult to obtain truly useful and comprehensive data [2].

With the increase of commodity volume and the growth of users on the platform, transaction information converges

into a sea. It can be difficult to sift through transaction information to determine shopping needs [3]. Therefore, mining useful information from extremely large information sets through technical means has become a hot application of big data mining technology in e-commerce platforms, which also makes this technology a new research hotspot. With the development of economic globalization, e-commerce has attracted more and more people's attention. This business model is changing people's life and changing people's shopping ways and habits. The rapid popularization of the Internet in the world and the establishment of transaction information collection methods, for the people of the whole world to establish a unified Internet, can carry out barrier-free information exchange [4]. With the continuous improvement of consumers' economic level, the rapid development of booming e-commerce, the gradual rise of e-commerce penetration, and the continuous expansion of the scale of the online shopping market, the online

shopping market is still a powerful engine for the growth of Chinese consumption [5]. The development trend of e-commerce will be directly related to the collection of transaction information. By enhancing the refined collection and processing of data, it will enhance its own advantages.

Xie et al. [6] under the market mechanism, reputation mechanism, and search ranking mechanism can effectively alleviate the commodity. There is no information asymmetry of standard information in the description of physical properties, and it cannot effectively alleviate the information asymmetry of commodity price information in physical standards. Solving the physical attributes of standard commodity price information asymmetry requires enterprise mechanisms. Replacing the market mechanism partly with the enterprise mechanism on the e-commerce platform can reduce the transaction cost. Malhotra and Rishi [7] e-commerce transactions are conducted based on computer networks, bringing great convenience to the majority of traders, confirming the identity information of both parties in the transaction and ensuring the security of the transaction. Bin Li built an electronic commerce transaction system based on PKI, analyzed its transaction process, and proposed information security thought based on Java2 system for data security and antifraud in online transaction. A concrete implementation based on Java is given. Abbasi et al. [8] analyzed the specific manifestations and effects of information asymmetry in C2C e-commerce transactions and then put forward strategies to eliminate information asymmetry in C2C e-commerce transactions. Nagesh and Prabhu [9] put forward feasible policy suggestions for the further construction and improvement of e-commerce trading platforms in view of the credibility problem and summarized the rectification suggestions from three aspects: the supervision system, the enterprise itself, and the legal environment. Suguna et al. [10] designed a web data extraction algorithm based on Jericho HTML Parser and taking HTML data tags as landmarks, aiming at the store sales data. Aiming at the user information of two websites, a web data extraction algorithm based on regular expression is designed. A web extraction system is designed and implemented, which can extract data from different sites according to different extraction rules. Finally, the effectiveness of the design scheme is verified by the actual data extraction on the above two platforms. The experiment proves that the designed prototype system has a high recall rate and accuracy. Wang et al. [11] invented an e-commerce trading platform and e-commerce trading information collection method based on big data, which can overcome the shortcomings of existing technologies.

2. Electronic Commerce Platform System

Electronic commerce is the use of computer technology, network technology and remote communication technology, the realization of the whole business process of electrification, digitalization, and networking. No longer are people face to face, looking at tangible goods, buying and selling on paper documents (including cash). The transaction was carried out through a comprehensive online

product information, a complete logistics distribution system, and a convenient and safe fund settlement system. The essence of e-commerce should be a set of complete network business management and management information system. More specifically, it is the use of existing computer hardware equipment, software, and network infrastructure, through a certain protocol connected with the electronic network environment to carry out a variety of business activities [12].

2.1. Electronic Commerce Platform System Architecture Design. The mature e-commerce platform adopts the SSH architecture as a whole, and the application architecture adopts the B/S model. On the basis of the use of MVC III, the system is designed with full consideration of the high availability, usability, maintainability, advancement, scalability, and security of the system, and the entire platform system is divided into the following: (1) presentation layer: the presentation layer goes up to serve the application layer and down to receive services from the session layer. The presentation layer provides a service for representing the information that is transmitted between application processes. The architecture design diagram in Figure 1 is composed of user and application system interaction pages. (2) Business layer: the core business is concentrated on the intermediate components, which can guarantee the maximum flexibility of the business. In Figure 1, the architecture design diagram consists of infrastructure service, resource integration service, data acquisition, format conversion, data exchange, file management, metadata management, and other functional components. (3) Data layer: the standard relational database is used to store data, which effectively guarantees the consistency of data and the security of data, laying a good foundation for the future data analysis. The security of database data is shown in the architecture design drawing in Figure 1, which lays a good foundation for future data analysis. In Figure 1, the architecture design diagram consists of the detection database, the authentication service database, the user information database, the GIS database, and the data sharing centre. Adopting this structure is beneficial to the flexible development and stability of the system, as well as the maintenance of the system.

2.2. Overall Functional Design. The overall function of the platform system is mainly divided into user management, organization management, service management, online tracking, online inquiry, and other major functions of the e-commerce platform. The user rights system consists of user management, rights management, organization management, single sign-on, and other functions. The portal content management system is composed of functions such as template maintenance, custom style and information release, as well as the data maintenance centre for maintaining basic data and the report management system for detecting data report management. Specific functional planning is shown in Figure 2:

Portal and content management subsystem: to achieve unified management content within the platform: including

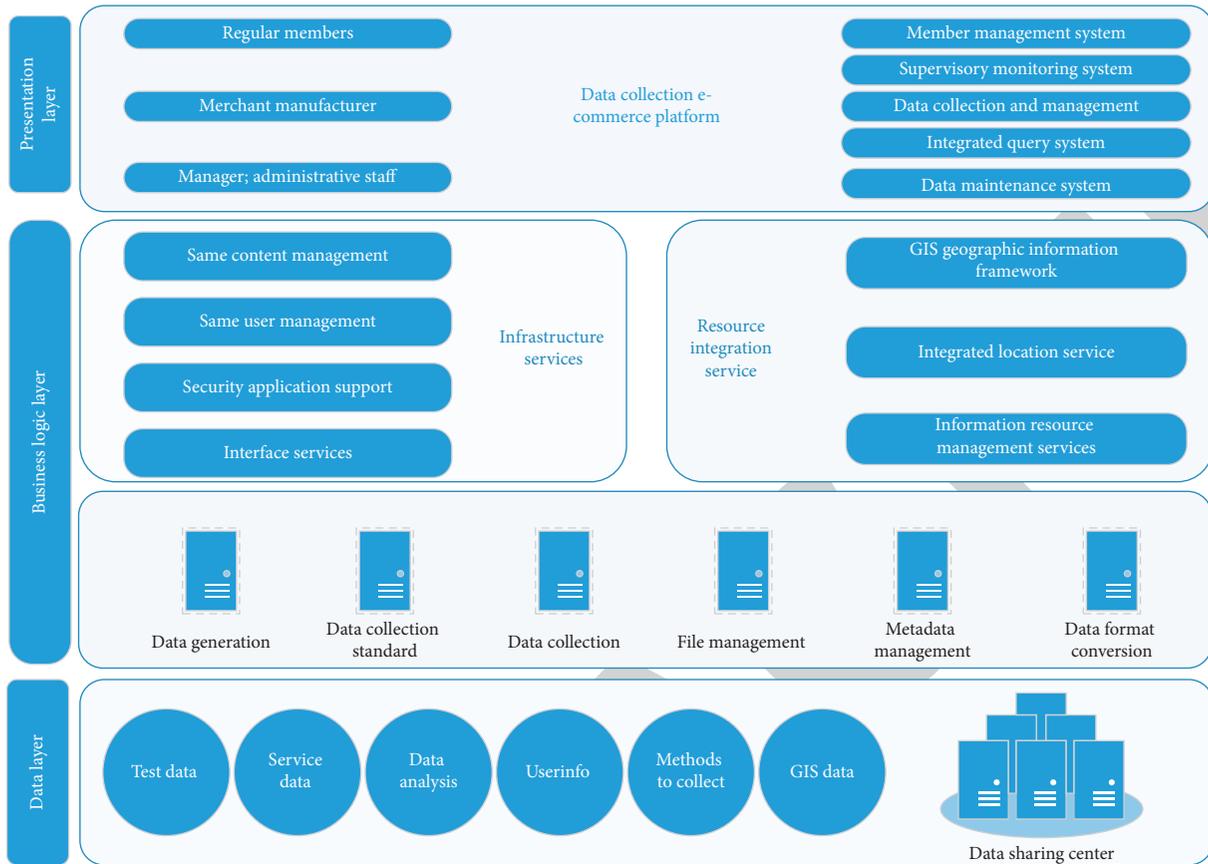


FIGURE 1: Architectural design drawing.

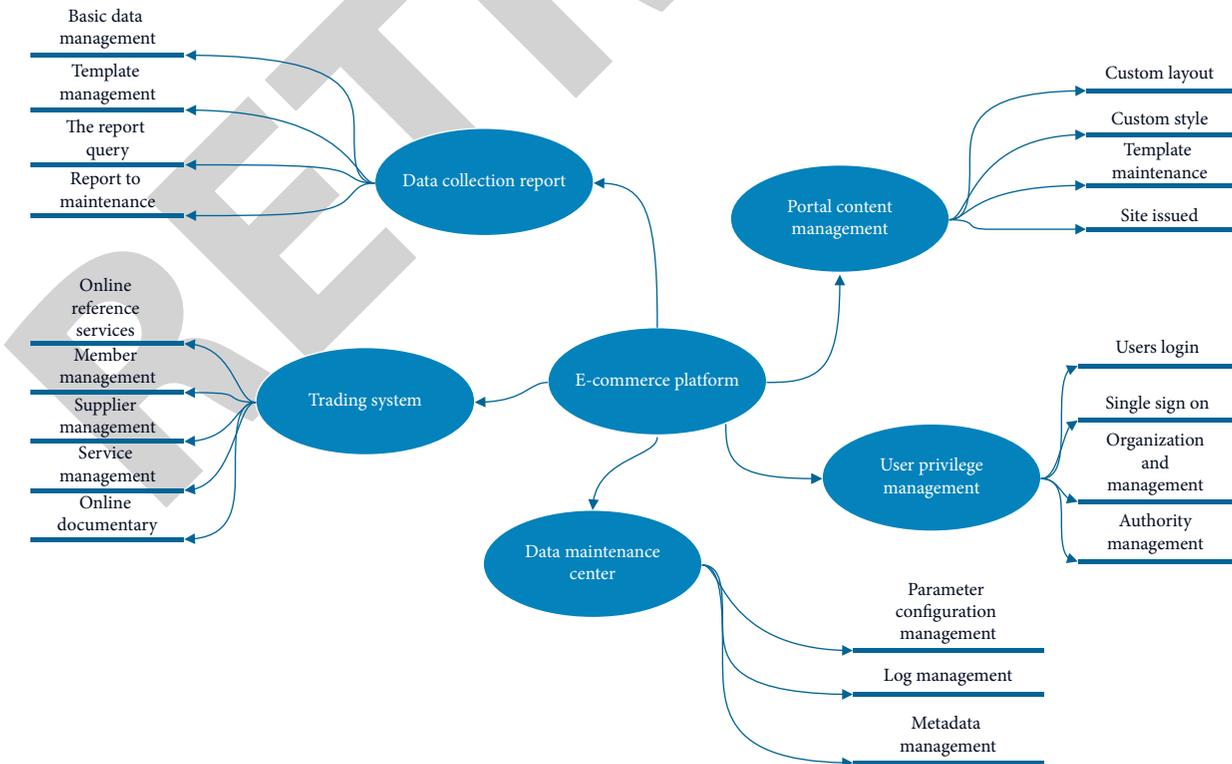


FIGURE 2: Function overall module.

page layout, page style, site maintenance, template maintenance, and information publishing five main functions; assist users to operate a series of content management functions such as template customization, column management, channel management, and interaction management in the subsites and subportals of the platform; and manage users' data such as static pictures such as files uploaded and information released during the use process.

User and authority management subsystem: implement the role to manage menu access authority and data operation authority. The system realizes privilege combination management by assigning different roles, so that different organizations and users have different role permissions and operation permissions. It is also responsible for managing the user's token and implementing custom token permissions for future integrated third-party applications.

Data maintenance centre subsystem mainly provides basic data maintenance, including data dictionary, authority maintenance, menu maintenance, role maintenance, and other data table operations.

E-commerce platform subsystem mainly realizes the functions of service management, online merchandiser, and online communication through the member management and supplier management function module and through online inquiry, online communication, order goods, online order, order review, order payment, order confirmation, and a series of operations to complete the whole business process of inspection and testing e-commerce [13].

Inspection and testing report subsystem mainly provides the management of report template, query, and maintenance of test report. From here, system administrators can create and edit corresponding report templates and formulate a unified report format. Detection users can browse the standard report information, export, and download the report information through the report query.

2.3. E-Commerce Process. The whole e-commerce process can be divided into the following three stages. Through the organic combination of the three stages of commodity release, information review, and commodity purchase, a complete business process can be formed: (1) after members log in the platform, they can carry out commodity management operations such as commodity off-shelf maintenance and price maintenance. (2) The website administrator verifies and passes the application for commodity release operation of members of the testing institution, so that ordinary members can browse and view specific commodity information. (3) Ordinary members browse the products through the platform, freely choose the products they need, and communicate and consult with the customer service of the merchant by using the online communication function. The specific process is shown in Figure 3.

2.4. Data Collection of E-Commerce Transactions. Electronic commerce database generally adopts the relational database, through the database management system to manage because the database management system is built on the operating system, so as to ensure the security of the

database, not only to ensure the security of the database management system itself but also to ensure the security of the operating system. A database management system is a group of software that defines, manages, maintains, and retrieves a database [14].

- (1) Data mining is a comprehensive interdisciplinary subject. Data mining refers to extracting or "mining" knowledge from large amounts of data. It integrates machine learning, statistical analysis, and database technology. It elevates people's application of data from low-level simple query to mining knowledge from data and providing decision support. Under the traction of this demand, researchers from different fields, especially scholars and engineers in database technology, artificial intelligence technology, mathematical statistics, visualization technology, parallel computing, and so on, have joined the emerging research field of data mining and formed new technical hot spots.

The results of data mining and knowledge discovery research should be practical. Knowledge discovery consists of an iterative sequence of the following steps: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge representation. Data mining is a step in the knowledge discovery process. The broad view of data mining function: data mining is the process of discovering interesting knowledge from a large amount of data stored in database, data warehouse, or other information base [15]. Based on this view, a typical data mining system has the main components shown in Figure 4.

- (2) HBase is a distributed database, so it can easily deal with the storage of massive data and provide corresponding data services through the management mechanism. HBase relies on other file systems to run and is similar in functionality to big table databases. As a database system for storing and processing large data, HBase can either use the local file system directly or use the HDFS file storage system in Haddon [16–18].

HBase is a sparse, multidimensional, sorted mapping table. The index of the table is the row key, column key, and timestamp. HBase, like any other NOSQL database, stores string data without any type. The server architecture of HBase follows the simple master and slave server architecture, as shown in Figure 5. Hamster is responsible for managing all things, and Zookeeper is responsible for coordination. HRegionServer does not store data itself, but provides storage management function, and HRegion serves as storage carrier.

2.5. Data Analysis. Naive Bayes method is a classification method based on Bayes' theorem and the assumption of conditional independence between features [19–22]. The method uses the given training data, first assumes the

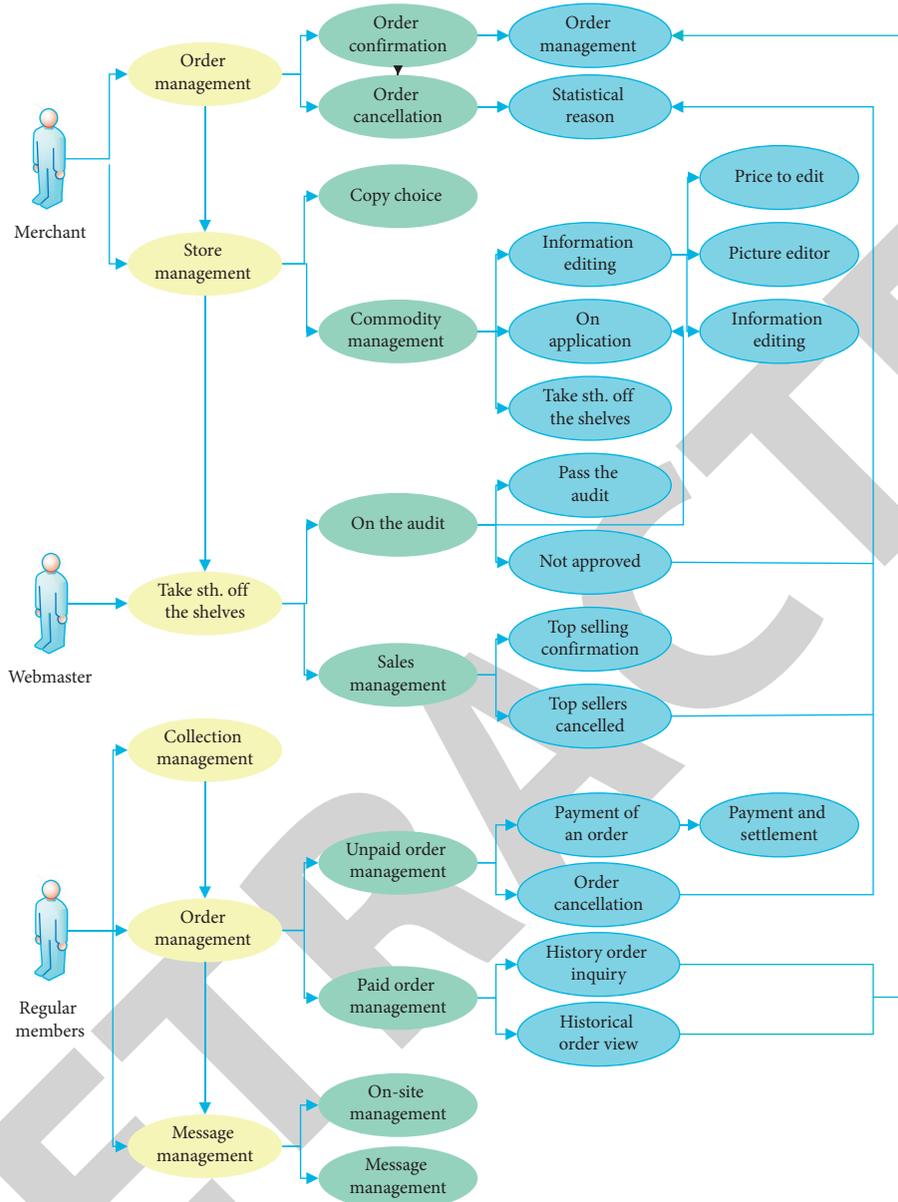


FIGURE 3: E-commerce flow chart.

conditional independence between the features, and learns the input and output based on this assumption. In the model input quantitative, Bayes' theorem is used to find the maximum posterior probability big output [23–26].

2.5.1. Naive Bayes Model. Let the input eigenvector x be a B-dimensional vector, the output set of class tags. Let x be a random variable defined in the input space, y is a random variable defined in the output space, and P is the joint probability distribution of x and y . The formula can be obtained from the following Bayesian formula:

$$P(Y = c_k | X = x) = \frac{P(X = x | Y = c_k)}{P(X = x | Y = c_1) + \dots + P(X = x | Y = c_k)}. \quad (1)$$

Naive Bayes makes the conditional independence assumption for conditional probability as follows:

$$P(Y = c_k | X = x) = \prod_{j=1}^b P(X^{(j)} = x^{(j)} | Y = c_k). \quad (2)$$

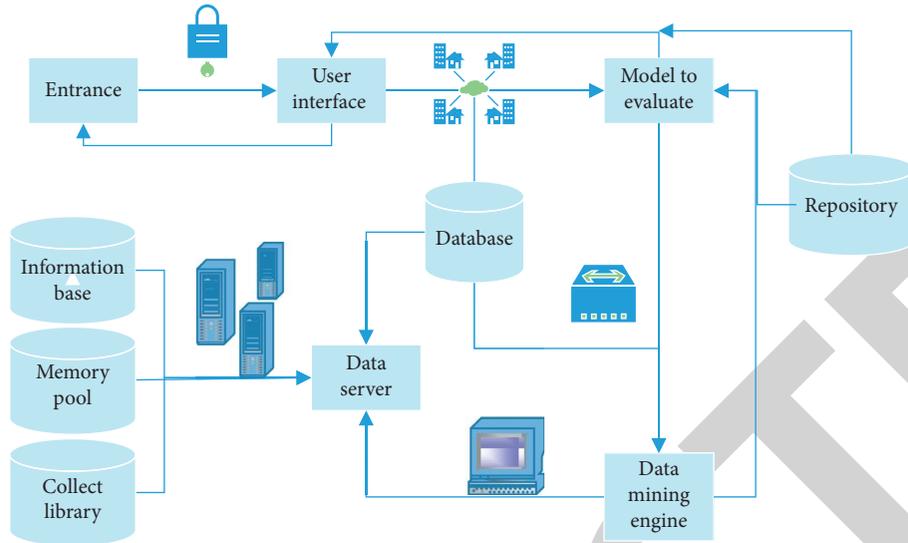


FIGURE 4: Typical data mining system structure diagram.

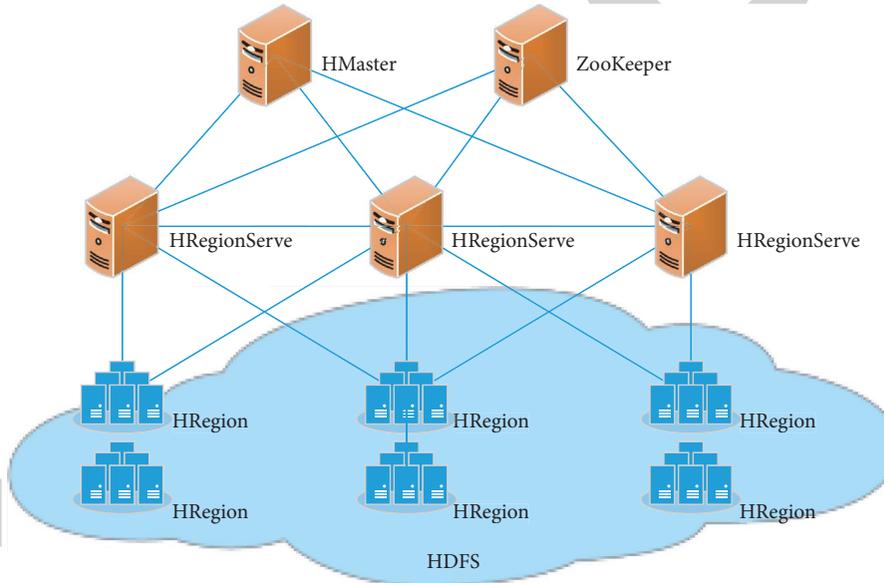


FIGURE 5: HBase structure.

Substitute formula (2) into formula (1) to get the following formula:

$$P(Y = c_k | X = x) = \frac{\prod_{j=1}^b P(X^{(j)} = x^{(j)} | Y = c_k)}{P(Y = c_1) + \dots + P(Y = c_k)}. \quad (3)$$

Formula (3) is the basic formula for Naive Bayes and represents the probability that the output is marked as c_k with a given feature x . When classifying, we choose the one with the highest probability in (2) as the final classification, which can formalize (3):

$$y f(x) = \arg \max \frac{\prod_{j=1}^b P(X^{(j)} = x^{(j)} | Y = c_k)}{P(Y = c_1) + \dots + P(Y = c_k)}. \quad (4)$$

Note that, in equation (4), the denominator is constant for all categories, so the formula can be simplified to the following:

$$y f(x) = \arg \max \prod_{j=1}^b P(X^{(j)} = x^{(j)} | Y = c_k). \quad (5)$$

2.5.2. Parameter Estimation of Naive Bayes Method. According to equation (5), learning a Naive Bayesian model means estimating. We can use the maximum likelihood estimation method to estimate the corresponding probe.

The maximum likelihood estimate of P is expressed as follows:

$$P(X^{(j)} = a_{jk}|Y = c_k) = \frac{I(x_1^{(j)}, y_1 = c_k) + \dots + I(x_k^{(j)}, y_k = c_k) - I^2}{I(y_1 = c_k) + \dots + I(y_i = c_k) - E_j}, \quad i = 1, 2, \dots, b; l = 1, 2, \dots, S; k = 1, 2, \dots, K. \quad (7)$$

Using Naive Bayes method to model user consumption behaviour, our definition of user consumption behaviour is to establish hypothesis. The browsing behaviour on the e-commerce platform is generated because users have consumer demand and the final purchase of users depends on whether users browse the goods that meet the consumer demand or not [27–30]. Therefore, we can model the user's consumption behaviour from the satisfaction degree of the characteristics of each product to the user's consumption demand.

Firstly, we can formalize the user's consumption behaviour with the Naive Bayes method as follows.

Assume that a product has B features and is represented by x , and the class is marked as y , where 0 represents no purchase and 1 represents purchase. Then, according to the Bayesian formula,

$$P(Y = y|X = x) = \frac{P(Y = y|X = x)P(Y = y)}{P(X = x)}. \quad (8)$$

According to the conditional independence assumption of conditional probability in Naive Bayes, the solvable formula (9) is as follows:

$$UB = \frac{P(Y = 2) \sum_{j=0}^o P(X^{(j)} = x^{(j)}|Y = 2)}{\sum_{k=1}^2 P(Y = c_k) \prod_{j=1}^0 P(X^{(j)} = x^{(j)}|Y = c_k)}. \quad (9)$$

The maximum likelihood estimation of P is obtained as follows:

$$P(Y = 2) = \frac{\sum_{i=2}^B I(Y_i = 2)}{\sum_{k=1}^2 P(Y = c_k) \prod_{j=1}^0 P(X^{(j)} = x^{(j)}|Y = c_k)}. \quad (10)$$

The maximum likelihood estimation can be obtained as follows:

$$P(X^{(j)} = a_{jl}|Y = c_k) = \frac{\sum_{i=2}^B I(x_i^{(j)} = a_{jl}, Y_i = c_k)}{\sum_{i=2}^B I(y_i = c_k)}. \quad (11)$$

UB can be obtained by combining equations (9)–(11).

In this paper, two-thirds of the data of each user is taken as the training data and one-third of the data is taken as the test data to test the classification effect of the Naive Bayesian model. Due to the large number of users, we randomly selected 10

$$P(Y = c_k) = \frac{I(y_i = c_k)}{I(y_1 = c_k) + \dots + I(y_k = c_k)}, \quad k = 1, 2, \dots \quad (6)$$

Assume the set of possible values of the JTH feature. Then, the maximum likelihood estimate of the conditional probability is as follows:

users for the test and obtained the classification readiness rate as shown in Figure 6. It can be found from the figure that the accuracy rate of modelling user consumption behaviour using Bayesian method can basically reach more than 70%.

3. Test Results and Analysis

3.1. Configuration of Experimental Environment. All performance collection tests were carried out on the big data platform. Since big data has the best performance on the Linux system, the big data cluster was directly deployed on the virtual machine in this experiment. A total of 8 PCs were used in the experiment, and one of them was assigned as the Name Node to act as the task manager. The other seven as data nodes act as workers. The software and hardware configuration of these eight PCs are the same. The details are shown in Table 1:

3.2. Experimental Dataset. The experimental data comes from Netflix dataset, which contains 1,000,000,000 pieces of rating data and is the rating record of 17,770 movies by 480,189 anonymous users. The rating data is the discrete integer value of the interval [1, 5]. The data in the dataset is quite sparse as follows:

$$Spa = \frac{N_s^2}{N_u \times N_i}. \quad (12)$$

Among them, N_s represents all the ratings of the project by users, N_u represents the number of users, and N_i represents the number of projects. If the Spa value is closer to 1, it means the dataset is more sparse. According to equation, the data sparsely of this dataset can be calculated as 0.9882, which indicates that the data of Netflix dataset is very sparse, which is very suitable.

3.3. Experimental Results and Analysis

3.3.1. Efficiency Test of Collection Method. Collect method efficiency, i.e., running speed to test. Taking the running time as the evaluation index, the shorter the running time is, the stronger the computing power is and the higher the efficiency is; on the contrary, the longer the running time is, the weaker the computing power is and the lower the efficiency is.

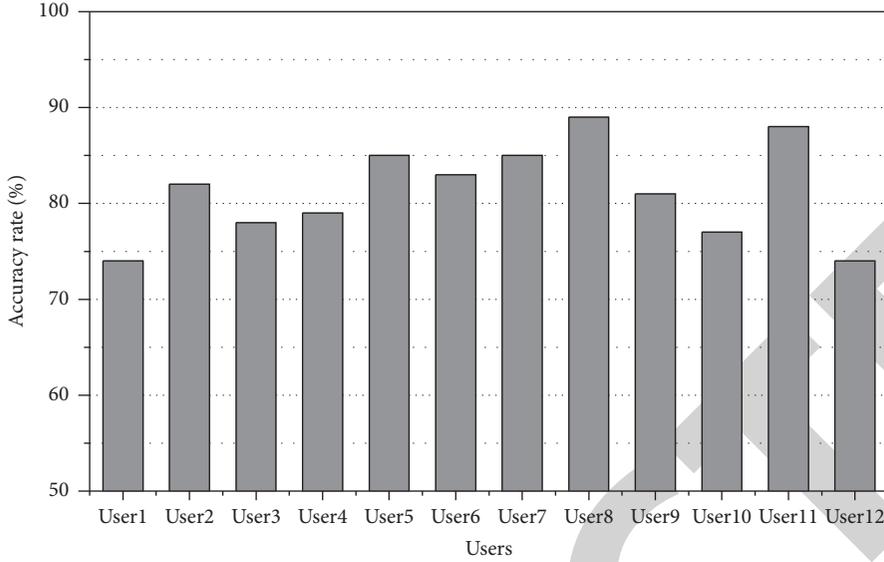


FIGURE 6: Classification accuracy of Bayesian method model.

TABLE 1: Configuration information of experimental environment.

Designation	Detailed information
Virtual machine operating system	OS/2
Host operating system	Window
Hardware configuration	AMD Athlon64 X2 4450e
Random access memory	8G
Read only memory	20G
Big data versions	1. 21. 2
JDK versions	JDK-6b48-OS/2-i598.Bin

The dataset input for the experiment contains data information sets of different sizes, respectively, 100,000 200,000 and 300,000; 400,000; and 500,000. The algorithm is run under different Data Node nodes (1 nodes, 2 nodes, 3 nodes, 4 nodes, 5 nodes, 6 nodes, and 7 nodes, respectively), and the running time of the algorithm under different nodes is compared. The experimental results are shown in Figure 7:

In Figure 8, the vertical axis represents the running time. For convenience of observation, the running time is calculated in seconds. The horizontal axis represents the size of the input dataset in thousands. In the test process, the overall running time of different nodes grows slowly and continuously when the amount of input data information continues to increase. The test results show that the larger the amount of information, the slower the processing speed. However, the overall running time under 7 nodes is better than the running time under 5 nodes. The running time under 5 nodes is better than that under 3 nodes, and the running time under 3 nodes is better than that under 2 nodes, indicating that the more nodes in the distributed platform, the stronger the computing power and the higher the execution speed. Therefore, it is proposed that the improved operation speed is fast and the collection method is efficient.

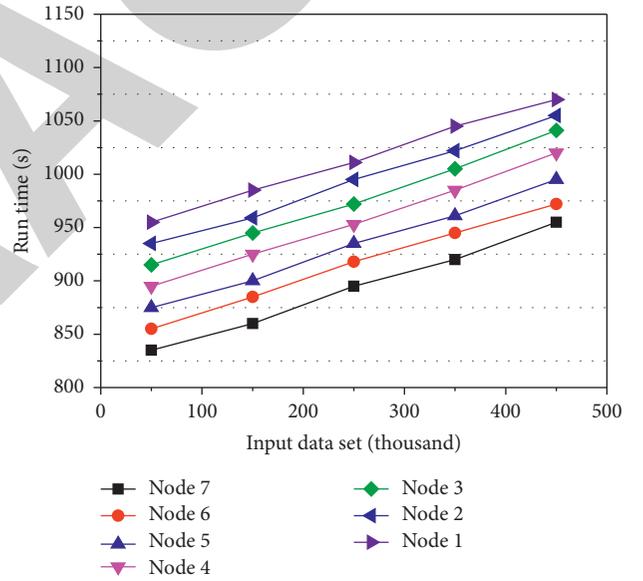


FIGURE 7: The run-time test results.

3.3.2. *Information Extensibility Ability Test.* Taking the acceleration ratio as the evaluation index, the larger the acceleration ratio is, the better the information scalability is; otherwise, the smaller the acceleration ratio is, the worse the information scalability is. Here, we first introduce the definition of speedup ratio. Speedup is the ratio of the running time of the same event task in a single processor computer system to the running time of the same event task in a parallel computer system with multiple processors. The calculation formula is as follows:

$$Sp = \frac{t_1 + 3}{t_n - 1}, \quad (13)$$

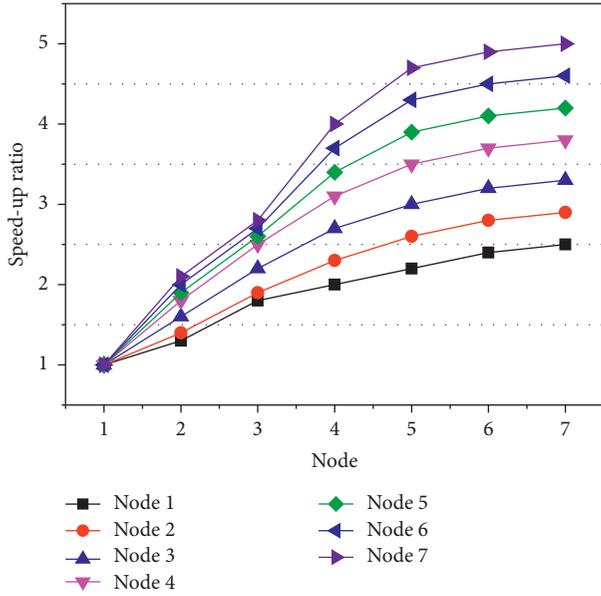


FIGURE 8: Acceleration ratio test 1.

where the acceleration ratio is represented by Sp , $T1$ is the time taken by the event in the single machine environment, that is, when the number of nodes on the distributed platform is 1, and T is the running time of the same event in the distributed environment, that is, when the number of nodes on the distributed platform is more than 2.

The input dataset of the experiment is 1 million pieces of user evaluation information about the goods. The methods are, respectively, run under different Data Node nodes (1 nodes, 2 nodes, 3 nodes, 4 nodes, 5 nodes, 6 nodes, and 7 nodes, respectively) to compare the acceleration ratio of the methods under different nodes. The experimental results are shown in Figure 8.

Through Figure 8, you can see that, with the increasing number of cluster nodes, the speed ratio is more and more big, the early changes quickly, the late changes slowly, seven node under acceleration ratio on the whole is greater than 5 node ratio, five node acceleration ratio on the whole is greater than three nodes ratio, and speed ratio under three nodes on the whole is greater than 1 node under the speed ratio. It shows that the more nodes in the distributed platform, the higher the information scalability. However, it can also be found from the experiment that when the number of nodes is less than 5, the acceleration ratio basically shows a linear change. When the number of nodes is greater than 5, the acceleration ratio changes slowly with the continuous increase of the number of nodes. This shows that simply increasing the number of nodes in a cluster does not improve performance indefinitely.

The optimized acceleration ratio is shown in Figure 9, where the horizontal axis represents the number of cluster nodes and the vertical axis represents the acceleration ratio. It can be seen from the experiment that the acceleration ratio becomes larger and larger with the increasing number of cluster nodes, indicating that the collection method has good

scalability under the Haddon platform. Observing the change curve of the acceleration ratio, you can find the speed-up ratio of the Haddon cluster nodes.

3.3.3. Collection Accuracy Test. The average absolute error is taken as the evaluation index for the collection accuracy test. The smaller the value of the average absolute error is, the better the accuracy of the collection results will be. On the contrary, the more accurate the collection results are, the worse the collection results will be. Mean absolute difference (hereinafter referred to as MAD) is the average of the absolute value of the difference between the predicted score of all single events and the arithmetic mean of all events as a whole. Transferring this concept here, this evaluation index can reflect the degree of deviation between the predicted score and the actual score. To achieve the accuracy of the recommended results test effect,

$$MAD = \frac{\lim_{n \rightarrow \infty} \sum_{i=1}^n P_{ij} = R_i}{N}, \quad (14)$$

where I represents the user, j represents the project, P_{ij} represents the predicted score, R_i represents the actual score, and N represents the number of projects. The smaller the MAD value is, the smaller the error is and the better the accuracy of recommendation is. The larger the MAD value is, the larger the error is and the lower the accuracy of recommendation is.

The dataset input in the experiment contains data information sets of different sizes, respectively, 100,000; 200,000; 300,000; 400,000; and 500,000. As can be seen from Experiment 4.3.2, the algorithm has reached its peak when the node of the Haddon cluster is 5. In order not to waste resources, this experiment is carried out in an environment where the Haddon cluster node is 5. The experimental results are shown in Figure 10:

In Figure 10, the vertical axis represents the mean absolute error value. The horizontal axis represents the input dataset in tens of thousands. It can be seen from the experiment that when the amount of input data information keeps increasing, the MAD value is basically constant and changes little, and their MAD value is all less than 1, indicating that the recommendation accuracy of the algorithm is high. Observing the change curve of MAD value, the collection accuracy is high.

3.3.4. Experimental Summary. The experimental analysis of the collection method was carried out through the evaluation indexes of running time, acceleration ratio, average absolute error, etc. The experimental results can be drawn as follows: the collection method has high efficiency, and the collection speed has been significantly improved, which can greatly improve the speed of e-commerce transaction information collection; high collection accuracy; and high information scalability. Collecting a large amount of extensible information facilitates the integrity of trading information.

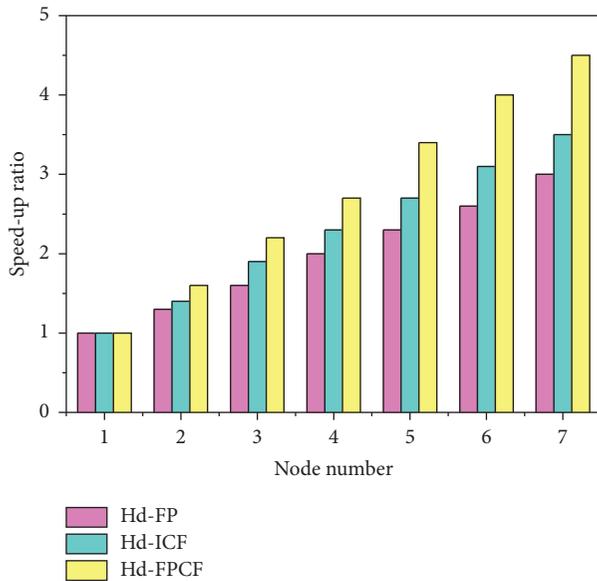


FIGURE 9: Acceleration ratio test 2.

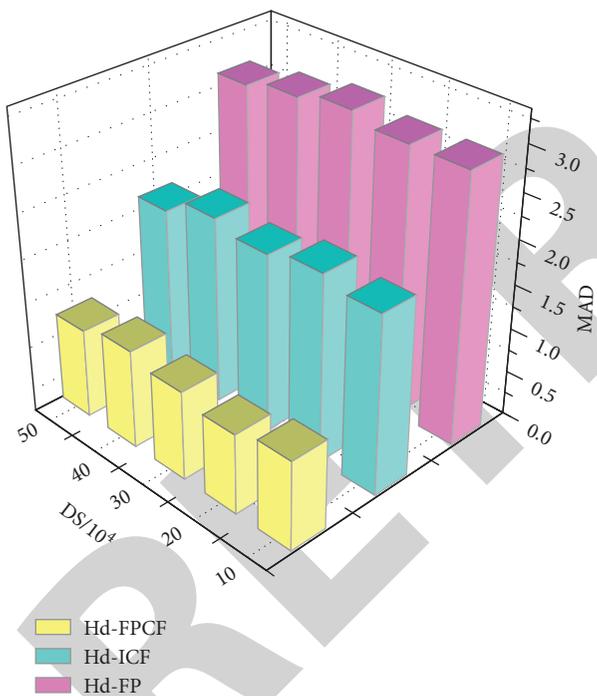


FIGURE 10: Mean absolute error test.

4. Conclusion

This paper focuses on the acquisition technology and cluster analysis of transaction information in the era of big data, which provides a strong support for the rapid, accurate, and mass collection of e-commerce data. In this paper, by consulting a large number of relevant literature, the development of e-commerce, platform design, trading information collection, and a series of data have done a detailed investigation and research. By combining the current situation of e-commerce transaction information collection at

home and abroad, this paper puts forward the method of e-commerce transaction information collection and realizes the collection of e-commerce transaction data systematically by platform design, data collection, data calculation, and experimental verification.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known conflicts of interest.

References

- [1] G. Ilieva, T. Yankova, and S. Klisarova, "Big data based system model of electronic commerce," *Trakia Journal of Sciences*, vol. 13, no. 1, pp. 407–413, 2015.
- [2] K. Zheng, Z. Zhang, and B. Song, "E-commerce logistics distribution mode in big-data context: a case analysis of JD.COM," *Industrial Marketing Management*, vol. 86, pp. 154–162, 2020.
- [3] L. Li, T. Chi, and T. Hao, "Customer demand analysis of the electronic commerce supply chain using big data," *Annals of Operations Research*, vol. 268, no. 1, pp. 113–128, 2018.
- [4] S. Akter and S. F. Wamba, "Big data analytics in e-commerce: a systematic review and agenda for future research," *Electronic Markets*, vol. 26, no. 2, pp. 173–194, 2016.
- [5] E. W. K. See-To and E. W. T. Ngai, "Customer reviews for demand distribution and sales nowcasting: a big data approach," *Annals of Operations Research*, vol. 270, no. 1, pp. 415–431, 2018.
- [6] K. Xie, Y. Wu, J. Xiao, and Q. Hu, "Value co-creation between firms and customers: the role of big data-based cooperative assets," *Information & Management*, vol. 53, no. 8, pp. 1034–1048, 2016.
- [7] D. Malhotra and O. P. Rishi, "An intelligent approach to design of e-Commerce metasearch and ranking system using next-generation big data analytics," *Journal of King Saud University-Computer and Information Sciences*, vol. 3, no. 2, pp. 183–194, 2018.
- [8] A. Abbasi, S. Sarker, and R. H. L. Chiang, "Big data research in information systems: toward an inclusive research agenda," *Journal of the Association for Information Systems*, vol. 17, no. 2, p. 3, 2016.
- [9] H. R. Nagesh and S. Prabhu, "High performance computation of big data: performance optimization approach towards a parallel frequent item set mining algorithm for transaction data based on hadoop MapReduce framework," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 1, p. 75, 2017.
- [10] S. Suguna, M. Vithya, and J. I. C. Eunaicy, "Big data analysis in e-commerce system using hadoop map reduce, 2016 international conference on inventive computation technologies (ICICT)," *IEEE*, vol. 2, pp. 1–6, 2016.
- [11] J. Wang, S. Yang, and Y. Wang, "The crawling and analysis of agricultural products big data based on jsoup," in *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 1197–1202, IEEE, Jinan, Shandong, China, October 2015.

- [12] Y. Zhao, Y. Zhou, and W. Deng, "Innovation mode and optimization strategy of B2C e-commerce logistics distribution under big data," *Sustainability*, vol. 12, no. 8, p. 3381, 2020.
- [13] Y. Tao and W. Zhang, "Establishment of cross-border e-commerce credit evaluation system based on big data," *Management & Engineering*, vol. 24, p. 3, 2016.
- [14] X. Du, B. Liu, and J. Zhang, "Application of business intelligence based on big data in e-commerce data analysis," *Journal of Physics: Conference Series*, vol. 13995, no. 1, Article ID 012011, 2019.
- [15] B. Binbin Wang and X. Xiaoyan Li, "Big data, platform economy and market competition: a preliminary construction of plan-oriented market economy system in the Information Era," *World Review of Political Economy*, vol. 8, no. 2, pp. 138–161, 2017.
- [16] X. Zhao, "A study on e-commerce recommender system based on big data," in *Proceedings of the 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pp. 222–226, IEEE, Chengdu, China, April 2019.
- [17] S. Ahmad, P. Singh, and A. K. Sagar, "A survey on big data analytics," in *Proceedings of the International Conference on Advances in Computing, Communication Control and Networking-ICACCCN*, pp. 256–260, IEEE, Noida, India, October 2018.
- [18] Z. Zhou, C. Gao, and C. Xu, "Social big-data-based content dissemination in internet of vehicles," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 768–777, 2017.
- [19] K. Vassakis and E. Petrakis, *Big Data Analytics: Applications, Prospects and Challenges, Mobile Big Data*, Springer, Cham, Switzerland, 2018.
- [20] X. Guan, L. Zhang, and H. Zhao, "Research on e-commerce supplier selection based on big data analysis technology," *Journal of Physics: Conference Series*, vol. 1757, no. 1, Article ID 012135, 2021.
- [21] H. Xu, K. Li, and G. Fan, "Novel model of e-commerce marketing based on big data analysis and processing," in *Proceedings of the 2017 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 80–84, IEEE, Xi'an, China, September 2017.
- [22] C. Zhao and J. Li, "Equilibrium selection under the bayes-based strategy updating rules," *Symmetry*, vol. 12, no. 5, p. 739, 2020.
- [23] L. Ouyang, S. Zhu, and K. Ye, "Robust Bayesian hierarchical modeling and inference using scale mixtures of normal distributions," *IISE Transactions*, vol. 14, pp. 1–21, 2021.
- [24] M. C. Chen, S. Q. Lu, and Q. L. Liu, "Uniqueness of weak solutions to a keller–segel–navier–stokes system," *Applied Mathematics Letters*, vol. 121, Article ID 107417, 2021.
- [25] G. Dong, W. Wei, and X. Xia, "Safety risk assessment of a pb-zn mine based on fuzzy-grey correlation analysis," *Electronics*, vol. 9, no. 1, p. 130, 2020.
- [26] Y. Li and J. Yang, "Few-shot cotton pest recognition and terminal realization," *Computers and Electronics in Agriculture*, vol. 169, Article ID 105240, 2020.
- [27] B. Yang, X. Li, and Y. Hou, "Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-a review," *Energy and Buildings*, vol. 224, Article ID 110261, 2020.
- [28] W. Wang, Z. Gong, and J. Ren, "Venue topic model-enhanced joint graph modelling for citation recommendation in scholarly big data," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 20, no. 1, pp. 1–15, 2020.
- [29] M. J. Akhtar, Z. Ahmad, and R. Amin, "An efficient mechanism for product data extraction from e-commerce websites," *Computers, Materials & Continua*, vol. 65, no. 3, pp. 2639–2663, 2020.
- [30] S. Liu and W. Zhang, "Application of the fuzzy neural network algorithm in the exploration of the agricultural products e-commerce path," *Intelligent Automation & Soft Computing*, vol. 26, no. 3, pp. 569–575, 2020.