

Research Article

Big Data Mining and Analysis of Agricultural Products Based on e-Commerce Platform

Hu Meiling 

School of Economics and Trade, Shandong Management University, Jinan 250357, China

Correspondence should be addressed to Hu Meiling; 14438120090181@sdmu.edu.cn

Received 14 June 2022; Revised 9 August 2022; Accepted 18 August 2022; Published 13 September 2022

Academic Editor: Kapil Sharma

Copyright © 2022 Hu Meiling. 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.

Solve the problem of agricultural product big data mining based on e-commerce platform, meet the needs of e-commerce development to agricultural products, meet the diversified needs of e-commerce platforms, and improve people's living standards and convenience. According to 1000 online questionnaires, 866 people believe that e-commerce can bring them convenience, and 134 people believe that the convenience is insufficient. Even agricultural products, as a traditional primary industry, have begun to be "involved" in the sales mode of e-commerce platforms. In the face of the increasingly huge online consumer demand market, the agricultural product economy has redisplayed a strong market vitality. Of course, the huge market base also makes the e-commerce model of agricultural products pay attention to big data mining and analysis. This paper focuses on how to carry out big data mining and analysis of agricultural products more efficiently from the technical level. Therefore, the agricultural product user data mining technology of e-commerce platform based on Hadoop is proposed. Through the intervention of association rule analysis and algorithm, the improvement of relevant algorithms and agricultural products user behavior analysis system under e-commerce platform based on Hadoop is proposed. The results show that the system can realize the analysis of commodity association degree under various agricultural products user behavior modes and can better help the e-commerce platform of agricultural products realize precision marketing.

1. Introduction

In recent years, with the development of urban and agricultural policies in our country, the entire agricultural sector in our country has developed rapidly. According to data from iiMedia research as of the end of October 2021, China's total agricultural output value has grown for the tenth consecutive year since 2010. By the end of 2020, China's total manufacturing output value will reach 10.7 trillion yuan. At the same time, with the deepening of the Internet, the Internet has had a great impact on all walks of life, network marketing plays a very powerful role, bringing great business opportunities for many enterprises, the Internet enables enterprise managers to realize the cross-regional input of management information, and realize the global management, and agricultural products have been pushed into the

development channel of online shopping [1]. Major e-commerce platforms have launched a series of agricultural products. At the same time, in-depth research in the field of e-commerce, the integration of e-commerce organizations in various fields of agricultural products, even if entrepreneurs see the future development direction, even if they have good strategic ideas, they may eventually achieve their strategic goals due to the organizational inertia, and because the organizational ability cannot keep up and become an important participant in the development of e-commerce. Of course, the process of agricultural product e-commerce is similar to the operation of other products. It needs to subdivide the customer group and mine the user information of agricultural products e-commerce platform with the help of big data analysis. Therefore, how to correctly use big data technology and deepen the extraction, analysis, and

utilization of big data in agricultural information is an important link in the development of agricultural e-commerce. Therefore, this paper takes the agricultural product e-commerce platform as an example and proposes the user data mining of the e-commerce platform based on Hadoop, hoping to develop more research and benefits for agricultural e-commerce data mining and provide more technical support and benefits for the development of agricultural e-commerce data mining. Products in this field [2] in the Hadoop ecosystem are shown in Figure 1.

2. Literature Review

Izzati and others said that at present, in the field of e-commerce, data generated by users in the shopping process can be used as a tool for data analysis, and the association rule algorithm in data mining technology is often used [3]. Li and Huang pointed out that the commodities that meet the association rules can be analyzed through the association rule algorithm, and on this basis, the relevant commodity information that users may be interested in can be pushed uniformly on the e-commerce platform. It is analyzed from two aspects: users and e-commerce platform [4]. Zuo believes that the wide application of data mining algorithms maximizes the benefits of both [5]. Lv and Li also proposed a new way for users to push related advertisements when browsing goods on e-commerce platforms [6]. Janijjevi et al. proposed in the work of text data mining that first collect the scattered or fragmented data generated by users and convert it into structured data and then analyze it in combination with users' behavior, so as to realize the functions of finding common friends and launching related advertisements [7]. In finance and trade-related fields, data mining algorithms gradually play an important role. Chen believes that experts in the financial field can analyze customers' capital status and consumption ability by analyzing customers' deposits, loans, and daily consumption bills, so that they can recommend corresponding financial products for customers [8].

Zhao's research on big data mining, its mining technology, and methods is mostly reflected in the research on clustering methods, feature selection methods, and granular computing, such as using text semantic processing, clustering, and other mining methods and technologies to process social network big data [9]. Zhou et al. verify the effectiveness of the proposed method through specific cases. For high-dimensional data mining, a new feature selection method is proposed [10]. At the same time, existing processes and procedures are compared and evaluated to further determine the performance of the given framework.

In China, some researchers believe that data mining is the process of finding useful information from various incomplete data using different techniques such as archives and professional knowledge. Based on the challenges faced by big data mining, granular computing is regarded as a kind of the new big data mining method clearly explained some existing problems; based on K -means and FP, the Spark platform was developed as a method to extract large amounts of thermal energy data for parallel computing.

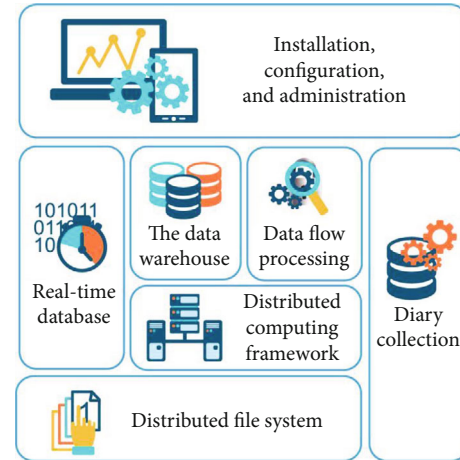


FIGURE 1: Hadoop ecosystem components.

Once analyzed, this method can improve the efficiency of large-scale thermal energy data mining; use the Canopy algorithm to improve K-shell, and develop new algorithms in parallel on the Hadoop platform; apply for a case of traffic data collection equipment based on big data mining technology, and define the scheme through the case. A computational model based on the output of big data mining technology is developed to analyze various time series and data analysis methods of big data. Predictive models are based on big data environments and can effectively make accurate predictions.

3. Method

3.1. Hadoop Cluster and Related Technologies

3.1.1. Hadoop Platform. When using Hadoop cluster for development, users can conduct independent research and development on the completed underlying platform without in-depth research and analysis of the underlying architecture. At the same time, users can perform the task of data analysis by changing the corresponding parameters in Hadoop cluster. All these provide convenience for R&D work and save a lot of time [11]. When configuring files, several copies of data backups are usually kept, which effectively prevents data loss. Please note that the command you typed is what you want before executing the backup command. Performing a backup command may take quite a very short period of time. In the process of data processing and analysis, if the data block fails and cannot continue, the work of the corresponding data block is transferred to other nodes, which will not directly lead to data loss or job failure. Although the Hadoop platform is widely used and has many of the above advantages, Hadoop itself cannot directly perform computing processing, and it needs the help of other components in its ecosystem [12].

3.1.2. Distributed File System HDFS. The architecture of HDFS consists of the master node namenode and several

Datanode nodes. The work of the master node includes monitoring the metadata in the HDFS directory and the status of the Datanode to see if there will be problems.

3.1.3. Distributed Database HBase. HBase is a distributed storage system based on HDFS. HBase is different from MySQL, which is a common database. MySQL often makes relevant queries through index. HBase can complete millisecond fast query through line key or realize multidimensional query by combining line key with cell value. Therefore, the design of row keys in HBase Table is particularly important. Reasonable row keys cannot only improve the query speed of HBase but also ensure the query efficiency when the rows and columns of HBase Table change [13]. The architecture is shown in Figure 2.

3.2. Association Rule Analysis Method. The association strategy algorithm usually needs to be analyzed in combination with the scene in the process of data analysis and is driven by data. As we all know, events are created by users, and at the same time, events generate a lot of information. Therefore, there is a relationship between the above three, as shown in Figure 3.

The main idea of the ensemble rule algorithm is to find the true relationship between products through seemingly unrelated purchases. For set L , it is a collection of multiple transaction items. That is $\{L1, L2, L3, \dots, Ln\}$. For sets X and Y composed of multiple transaction items in set L , they must belong to a subset of set L . The degree of correlation between X and Y is determined by the relationship between X and y . When it comes to the association rule algorithm, we must mention two professional terms, namely, support and confidence. In the above set X , the number of times an item is called the number of supports for those items. The formula is shown in

$$\text{Support_Count}(X). \quad (1)$$

Suppose T is used to represent the total number of occurrences of all kinds. At this time, the support of item set X is shown in

$$S(X) = \frac{\text{Support_Count}(X)}{|T|}. \quad (2)$$

If two or more transactions occur at the same time, for example, the number of times X and Y occur at the same time is called the absolute support of $X \Rightarrow Y$. The method of calculating the absolute support is shown in

$$S(X \Rightarrow Y) = \frac{\text{Support_Count}(X \Rightarrow Y)}{|T|}. \quad (3)$$

The rule of confidence is to calculate the probability of $X \Rightarrow Y$ support number on the premise that X exists, as

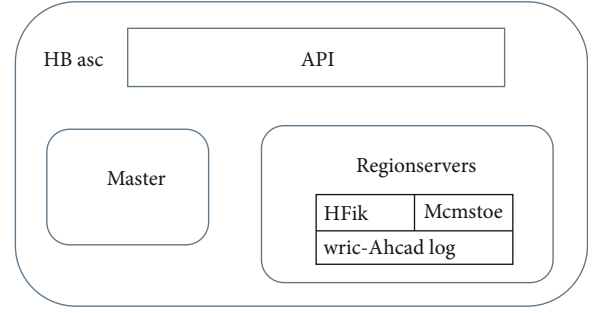


FIGURE 2: Architecture of HBase.

shown in

$$\text{Con}(X \Rightarrow Y) = \frac{\text{Support_Count}(X \Rightarrow Y)}{\text{Support_Count}(X)}. \quad (4)$$

3.3. Big Data Processing and Mining

3.3.1. Big Data Processing. Big data has rich resource types and various application processing methods, but the basic process of information processing is roughly similar, including four stages: data collection, processing and integration, analysis, and data interpretation. The functions are as follows: obtain the required data from the data source, process the data of different models in a unique way, assemble the models into different files, and then process and analyze the data from the necessary analysis processes and tools. Finally, use visual aids to get user recognition results [14].

3.3.2. Big Data Mining. Data mining methods are generally used to describe the characteristics of the target data set, or summarize and summarize the current information to further predict the future situation. According to the different functions, they are divided into descriptive and predictive types, as shown in Figure 4.

3.3.3. Cluster Analysis. Population analysis is an unsupervised study and an important activity of data analysis. Its idea is to aggregate the features of unlabeled samples by similarity. Cluster analysis is a kind of unsupervised learning, which only requires data without marking the results. It can make a large number of observations into several classes according to a certain rule. The observation values within each class are similar, and the difference between each class is large. In order to achieve the automatic division of sample categories, that is, the process of dividing data objects into subsets through static classification method. The data with similar properties are divided into a subset, so that each subset is a cluster, and the data in the cluster has some similar attributes, and the data attributes between clusters have great feature differences [15]. Cluster analysis methods rely only on the distance of the data. The application of clustering analysis in the field of e-commerce is mainly the clustering of users. From the research of this paper, it is the clustering of agricultural products users on the e-commerce platform,

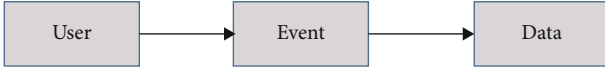


FIGURE 3: Relationship diagram of users, events, and data.

finding out users with common characteristics, and then adopting targeted marketing strategies.

3.3.4. Data Feature Analysis. Take the agricultural apple as an example. The apple sales stores on the e-commerce platform usually sell 5 kg apples as a piece, so the apple sales price data obtained is the total price of 5 kg apples. When drawing the frequency histogram, the total price data of 5 kg is used. Draw statistical histogram in order to more intuitively describe the data characteristics and distribution form, so as to find the function curve in line with the data change. Because the apple price data is positive, combined with the distribution form of the drawn histogram, it is obvious that it does not conform to the normal distribution, but through the corresponding data transformation, it is found that the logarithm of its random variable conforms to the normal distribution [16], as shown in Figure 5.

In lognormal distribution, let X be a continuous random variable with a positive value, as shown in

$$\ln X \sim N(\mu, \sigma^2). \quad (5)$$

Then, the probability density of X is shown in

$$f(x, \mu, \sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right], & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (6)$$

Then, it is said that the random variable x obeys the lognormal distribution and is recorded as $\ln X \sim N(\mu, \sigma^2)$.

Let x obey lognormal distribution, and its density function is shown in

$$p(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{(\ln x - \mu)^2 / 2\sigma^2}. \quad (7)$$

The mathematical expectation and variance are shown in

$$E(X) = e^{(\mu + \sigma^2)/2}, \quad (8)$$

$$D(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}. \quad (9)$$

In formula (9), the abscissa represents the apple sales price, the ordinate represents the probability density, and the red curve represents the lognormal distribution curve of price fitting. It can be seen from the formula that the lognormal distribution can approximately describe the distribution and trend characteristics of the data. Through further calculation, the parameter values of the fitted distribution curve are obtained: the value of log likelihood function is

-4163.35, the mean value is 53.9495, and the variance is 1082.58 [17].

The graphic analysis of apple sales cannot only more intuitively see the changes of apple sales but also provide data analysis basis for the realization of mining task. Data analysis, as an important basis and means for finding problems, adjusting strategies and optimizing directions in the current enterprise operation, has gradually been paid attention to in various industry departments. Assignment is always more important than classification, and classification has positive effects in many fields. An environmental distribution is defined as follows: if the difference in difference x follows a probability, its potential density is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (10)$$

Then, x is called a canonical random deviation, and the division obeyed by a canonical random variable is called a canonical distribution.

3.4. Apriori Algorithm. The strategy algorithm of the institute includes a variety of algorithms, such as Apriori algorithm and FP growth algorithm. Association rules are a big data mining task, initially motivated by the shopping basket analysis (Market Basket Analysis) problem. With the advent of the prior algorithm, the algorithm has been widely used in the data mining industry due to its high efficiency in corporate policy analysis [18].

Table 1 shows the purchase records of agricultural product users in the e-commerce platform database. Here, we specify that nonempty itemsets with a support number of no less than 2 are frequent itemsets. The minimum support number is not fixed. For practical problems, the minimum support number can be changed flexibly. Start with choosing a minimum right edge; after each step, always choose the least right edge of the unselected edge, and make it to not form a circle with the selected edge. Let the four records in Table 1 be all the records of the database, and it is easy to find that transactions T1-T4 are $\{A, B, C\}$, $\{A, D, E\}$, $\{A, E\}$, and $\{A, B, D, E\}$. Then, the number and size of the database are 4, and the number of transactions is 4.

- (1) Scan the transaction data of the e-commerce platform, and count the support of each product in all the data, including the number of products purchased, product price, and product type. If you buy a lot of things, there are a greater number of supports of the commodity in the process of calculating the number of supports

Calculate the transaction candidate set through step (1) to obtain the support number of each item set [19]. The number of supported sets is shown in Table 2.

- (2) After completing the above steps, according to the minimum support 2, the collection components that do not meet the conditions are eliminated, and only the specified items are enabled. As an initial hold

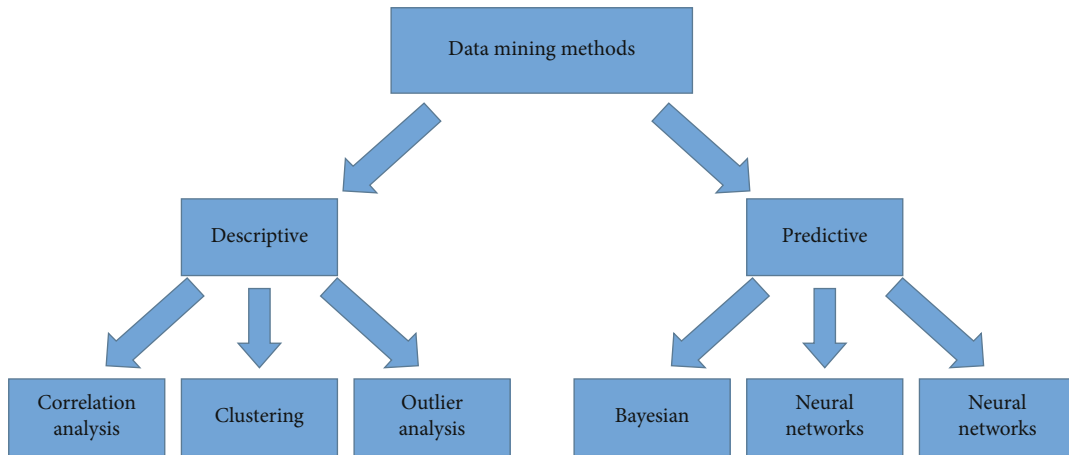


FIGURE 4: Data mining method.

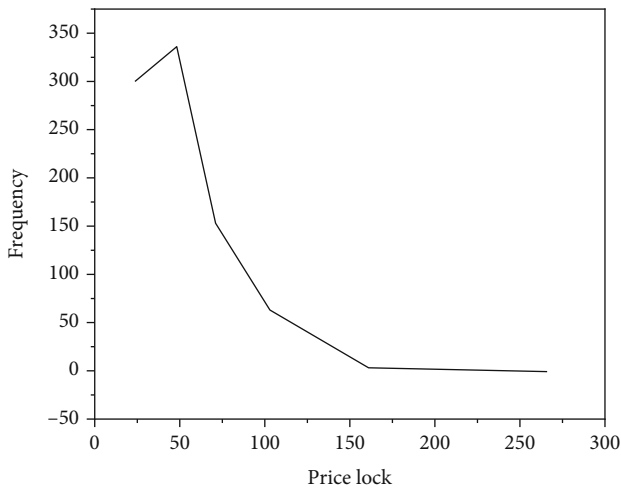


FIGURE 5: Apple sales price chart (broken line).

TABLE 1: User purchase record.

Affair	Project
T1	A, B, C
T2	A, D, E
T3	A, E
T4	A, B, D, E

TABLE 2: Number of items supported by each itemset.

Itemset	Support number
{A}	4
{B}	2
{C}	1
{D}	2
{E}	3

minimum support, normally 1 item, set ok, as shown in Table 3

- (3) According to the result obtained in step (2), connect the above five itemsets in pairs, and calculate the support number of the itemset formed by the combination of the two after the connection, as shown in Table 4
- (4) The connected 2-itemsets are obtained from the above steps, and the qualified itemsets are retained according to the minimum threshold, as shown in Table 5
- (5) Connect according to the frequent 2-itemsets obtained in step (4). According to the characteristics of frequent 2-itemsets, its nonempty subsets must also be frequent. Therefore, after connecting the frequent 2-itemsets, you can judge whether they are frequent itemsets according to the first two items. If the conditions of frequent itemsets are not met, prune them. It is not difficult to find the candidate 3-itemset, as shown in Table 6

- (6) The candidate 3-itemsets are obtained from the above steps, and the qualified itemsets are retained according to the minimum threshold, as shown in Table 7

After multiple scanning and discrimination, when the two itemsets are related, the Apriori algorithm calculates that {AB}, {AD}, {AE}, and {DE} have a high degree of correlation. Finally, when the three itemsets are related, it can be seen that the itemsets {A, D, E} are highly related.

TABLE 3: Frequent 1-itemsets.

Itemset	Support number
{A}	4
{B}	2
{D}	2
{E}	3

TABLE 4: Connected 2-itemsets.

Itemset	Support number
{A, B}	2
{A, C}	1
{A, D}	2
{A, E}	2
{B, C}	1
{B, D}	1
{B, E}	1
{C, D}	0
{C, E}	0
{D, E}	2

TABLE 5: Frequent 2-itemsets.

Itemset	Support number
{A, B}	2
{A, D}	2
{A, E}	2
{D, E}	2

TABLE 6: Candidate 3-itemset.

Itemset	Support number
{A, B, D}	1
{A, B, E}	1
{A, D, E}	2

TABLE 7: Frequent 3-itemsets.

Itemset	Support number
{A, D, E}	2

The purpose of association analysis is to find interesting associations or interrelationships between the sets of items from a large number of data, and the most classical Apriori algorithm has a great influence in the field of association rule analysis. Through the support number, we can determine

whether there is an association relationship between itemsets; that is, for the purchase data of all users, how many goods they buy have an association relationship [20]. However, if we want to determine the strength of the correlation degree, we need to calculate it in combination with the confidence degree, as shown in

$$\text{Confidence}(X \Rightarrow Y) = P\left(\frac{X}{Y}\right) = \frac{\text{Support_Count}(Y \Rightarrow X)}{\text{Support_Count}(X)}. \quad (11)$$

The degree of association is judged by setting the value of confidence. For example, if the confidence is 50%, then less than 50% is a weak correlation; otherwise, greater than or equal to 50% is a strong correlation. In this way, we can get the itemset with high degree of correlation that meets the conditions [21]. Although the Apriori algorithm can accurately find the itemsets with strong correlation, the Apriori algorithm also has some disadvantages that cannot be ignored: in the process of Apriori algorithm mining, it is necessary to scan all the contents of the database for several times. Every time a frequent itemset is found, the database needs to be completely scanned. Moreover, the candidate item set generated in this way is also very large. When the database structure is relatively simple, the Apriori algorithm can work better, but when the database is relatively large, the I/O times of Apriori algorithm will be very high. Therefore, the Apriori algorithm needs to be improved [22].

4. Results and Analysis

Combining the MapReduce framework in Hadoop ecology with the improved Apriori algorithm, the traditional single computer operation mode is transformed into parallel operation and then processed [23]. The improved algorithm is divided into two stages.

In the first stage of the algorithm, the specific process is as follows:

- (1) Input transaction database s , divide the data set in database s into N data blocks, and allocate these n data blocks to each computer node
- (2) Convert each data block node into a data structure $\langle T, k \rangle$ key value pair that meets the requirements of MapReduce, where T is the transaction name in the dataset and k is the transaction item corresponding to each transaction
- (3) Mapper function will be executed to scan the data blocks in each computer node according to the corresponding key value and output the key value pairs of $\langle k, v \rangle$, where k is the transaction item corresponding to each transaction and v is the number of transaction items supported)
- (4) Integrate the key value pairs generated in step (3) through the combine function, and take the integrated key value pairs as the input of the reducer

TABLE 8: Running time of improved Apriori algorithm under two operating environments (s).

Minimum support	Improved Apriori algorithm for stand-alone version	Hadoop cluster improved Apriori algorithm
0.2	76	41
0.4	55	33
0.6	38	26
0.8	29	22

TABLE 9: Running time of improved Apriori algorithm under two operating environments (s).

Dataset size	Improved Apriori algorithm for stand-alone version	Hadoop cluster improved Apriori algorithm
67544	22	24
88424	36	26
102256	48	29

function. Then merge the local candidate 1-itemset, that is, add the key value pairs with the same K in the key value pair set to obtain the global candidate 1-itemset. Then, a two-dimensional array is constructed with row values as transaction sets and columns as transaction item sets, and the column values that do not meet the minimum support number are deleted. Reduce the size of the two-dimensional array and the number of transaction items in the transaction database s , so as to reduce the I/O times when scanning the database during iteration, so as to save time [24].

- (5) Step (4) is completed. The nonempty transaction itemset $Q1$ that meets the conditions is the frequent 1-itemset

In the second stage of the algorithm, the implementation steps are as follows:

- (1) Max_l and Min_l are determined from the frequent 1-itemset $Q1$ obtained by the optimization in the first stage, in which the maximum value of the length of the highest frequent set is $Max_l = N$ and the minimum value is $Min_l = 1$
- (2) Based on the idea of halving, the length L of the frequent itemset set to be iterated in the next MapReduce stage is obtained according to

$$L = \frac{Min_l + Max_l}{2} \quad (12)$$

- (3) After iterating the set length L in the next stage obtained in (2), take out the combination of frequent itemsets with length L in the frequent 1-itemset with set length N . The number of "connected" sets obtained in this way is C_N^L

- (4) The process of the mapper and reducer in the second stage is the same as that in the first stage, except that the value of K in the key value pair $\langle k, v \rangle$ is changed from 1 to L ; that is, the process of Mapper and Reducer for one transaction item is changed to 1 transaction items. After scanning transaction database S , frequent L -itemsets are obtained

Aiming at the problems of repeated scanning and large I/O overhead in the iterative operation of traditional Apriori algorithm, a feasible improvement strategy is proposed. Firstly, by constructing a two-dimensional transaction array, the initial input data set is simplified, and the items that are certainly not frequent item sets are removed to reduce the time of scanning the transaction database; then, the strategy of half thought is used to solve the leapfrog frequent itemset, which no longer uses the traditional step-by-step iterative solution, and reduces the number of iterations; Finally, the two improved methods are combined and deployed on Hadoop platform. Using MapReduce framework on Hadoop platform can greatly improve the processing efficiency. Resource Manager is a global resource manager with two components: scheduler (Scheduler)+Application Manager (Application Manager). Then, the specific implementation process and some pseudocodes of the algorithm are described. Through the analysis of the algorithm, the results obtained are consistent with those obtained by the traditional Apriori algorithm, which can prove the correctness of the algorithm [25].

- (1) Experiment 1: first, 98566 data are selected in the agricultural product data behavior of the e-commerce platform and set 4 groups of minimum support degrees, which are 0.2, 0.4, 0.6, and 0.8, respectively. The Apriori algorithm is improved on the stand-alone version, and the Apriori algorithm is improved on the Hadoop cluster version for collaborative purposes. In the case of unified data processing, the running time of the improved Apriori algorithm in the two processes is compared [26]. The unit is seconds, and the result is a number. The experimental results are shown in Table 8

According to the data of the experimental results obtained in Table 8, the running time of the improved Apriori algorithm of the stand-alone version and the improved Apriori algorithm of the Hadoop cluster version.

In (1), the performance of the two algorithms in different operating environments is compared by setting different

minimum support degrees. Now, the minimum support is set to 0.5. Just change the file size, and use the default behavior set to select 67544, 88424, and 102256 files in the file. Compare the running time of the improved stand-alone version of Apriori algorithm and the improved Hadoop cluster version of Apriori algorithm in different files [27]. The experimental results are shown in Table 9.

The experimental data obtained from Table 9 shows that the runtime is used to improve the standalone version of the Apriori algorithm and the improved Hadoop cluster version of the Apriori algorithm. As the first set of experimental results of Operation 1, the upgrade completion of the Apriori algorithm of the Hadoop cluster version is greater than that of the stand-alone version of the Apriori algorithm, and the time required for parallelization may be negligible. Combined with the results of the above two experiments, it has been proved that the algorithm improvement effect in the Hadoop cluster is better, and the performance of the algorithm will be better when the amount of data collection is also large.

- (2) Experiment 2: when the file size is fixed, the file selection size is 102256, and the minimum support is 0.2-0.8. Experimental comparisons were made by continuously changing the size of the minimum support. The purpose is to identify whether there is an improvement in the performance of the improved Apriori algorithm in a Hadoop cluster

The results of Study 2 show that with the minimum support, the number of iterations of the product solution is gradually reduced, and the performance of the two algorithms is almost the same, but the execution time is the same as that of Apriori. Algorithms are improved, even worse than the improved Apriori algorithm based on compressed matrices. These test teams showed that the improved Apriori algorithm performed better, its application to the big data analysis of agricultural products on the e-commerce platform will be more efficient, and the results will be more accurate.

5. Conclusion

The experiment shows that the agricultural product data mining technology based on Hadoop e-commerce platform is feasible, which can solve the problem of big data mining of agricultural products, meet the needs of e-commerce, and solve the drawbacks of e-commerce at the same time. Agricultural products have improved people's lives and the convenience of shopping. In today's era of big data, the cost of big data starts in the terabyte. Users find products they like or are interested in on a multitude of e-commerce platforms, like looking for a needle in a haystack. These processes will also cost users a lot of time and energy; at the same time, for Internet e-commerce platforms such as Taobao and JD, if the platform cannot correctly identify user data, it will have a negative impact on the fierce competition of e-commerce, resulting in unsuccessful e-commerce business. Therefore, how to accurately and timely identify the

user behavior of e-commerce platform is a research hotspot. For the agricultural products of the e-commerce platform, the main function of the data is to store the internal information such as the sales volume, price, and opening time of the products for sale through the e-commerce platform, through data analysis, agricultural product recommendation, agricultural product inventory site optimization, price analysis, etc. Finally, push value analysis to e-commerce business management to provide intelligent support services for e-commerce enterprises to achieve profitability, quality management, and business success.

Based on e-commerce platform, Hadoop-based data mining technologies are identified, including interfaces such as Hadoop cluster, HDFS data system, MapReduce, and HBase. It also discusses the process of identifying organizations and the processes involved in data mining, usually including the Apriori algorithm and FP to develop algorithms.

The Apriori algorithm always suffers from low performance when dealing with big data, and the improvement strategy has prepared for its benefits. First, the product setting consists of a two-dimensional array, and the rows of the two-dimensional array that do not meet the minimum support are removed to make the product more efficient. Then, using the concept of halving to create live objects, the number of iterations in the work process can be reduced to reduce the running time of the algorithm; finally, two optimization strategies are integrated on the Hadoop platform. Example procedures used for the development of the programs and algorithms are described in detail. The improvement of Apriori algorithm based on Hadoop platform is determined by experiments, and the time spent by the improved algorithm is compared by managing different models. In the experiment, the user behavior data provided by Alibaba Cloud was selected, and the Hadoop platform was designed and tested. Firstly, the improvement of the single-machine version of Apriori algorithm and the algorithm improvement of Hadoop cluster version are compared and analyzed through experiments. Then, the development of the existing Apriori algorithm, the integration of the integration-based Apriori algorithm, and the compression matrix-based Apriori algorithm is carried out on the Hadoop platform. a comparative analysis; finally, a separate collaborative development algorithm is separated, and an experimental comparative analysis of collaborative improvements such as product improvement ideas, iterative improvement ideas, and optimization strategies is carried out. The test results show that the algorithm has a shorter lifespan and a more obvious algorithm improvement; a large number of agricultural products data mining that can be applied to the e-commerce platform can improve the sales volume of agricultural products on the e-commerce platform and the operating environment of the e-commerce platform of agricultural products.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was supported by the National Social Science Fund Project (20BJL086).

References

- [1] N. Wardhani, W. Izaak, and M. Yohanes, "Analysis of ocr's factor and sales promotion on e-commerce," *Journal of Economics Business and Government Challenges*, vol. 3, no. 2, pp. 127–132, 2021.
- [2] N. Maslak, L. Zhang, and X. Lu, "Analysis of agricultural trade in China based on the theory of factor endowment," *Agricultural and Resource Economics*, vol. 6, no. 1, pp. 50–61, 2020.
- [3] R. R. Izzati, N. S. Muntiah, and N. Hidayah, "An analysis of factor that influence the interests in behaviors of using accounting information systems based on e-commerce," *Jurnal AKSI (Akuntansi dan Sistem Informasi)*, vol. 5, no. 1, pp. 1–5, 2020.
- [4] X. Li and D. Huang, "Research on value integration mode of agricultural e-commerce industry chain based on internet of things and blockchain technology," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8889148, 11 pages, 2020.
- [5] J. Zuo, "Analysis of e-commerce characteristics based on edge algorithm and cox model," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6628068, 13 pages, 2021.
- [6] X. Lv and M. Li, "Application and research of the intelligent management system based on internet of things technology in the era of big data," *Mobile Information Systems*, vol. 2021, Article ID 6515792, 6 pages, 2021.
- [7] S. Janijjevi, O. Petrovi, and M. Stefanovi, "Sales prediction on e-commerce platform, by using data mining model," *Serbian Journal of Engineering Management*, vol. 5, no. 2, pp. 60–76, 2020.
- [8] S. Chen, "Analysis of customization strategy for e-commerce operation based on big data," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6626480, 11 pages, 2021.
- [9] Z. Zhao and H. Wang, "The analysis of internet commercial judicial based on big data alliance and mining service process model," *Complexity*, vol. 2021, Article ID 2897823, 17 pages, 2021.
- [10] N. Zhou, Z. Zhang, and J. Li, "Analysis on course scores of learners of online teaching platforms based on data mining," *Ingénierie des Systèmes D Information*, vol. 25, no. 5, pp. 609–617, 2020.
- [11] Y. J. Wan and Z. C. Yu, "Exploring the development of e-commerce for fresh agricultural products under the background of e-commerce law," *The Journal of Internet Electronic Commerce Resarch*, vol. 20, no. 6, pp. 135–163, 2020.
- [12] S. Liu and W. Zhang, "Application of the fuzzy neural network algorithm in the exploration of the agricultural products e-commerce path," *Intelligent Automation and Soft Computing*, vol. 26, no. 3, pp. 569–575, 2020.
- [13] R. R. Ramírez, A. Chamorro, and L. Moreno, "Organic and online attributes for buying and selling agricultural products in the e-marketplace in Spain," *Electronic Commerce Research and Applications*, vol. 42, no. 1, article 100992, 2020.
- [14] S. Teng, "Route planning method for cross-border e-commerce logistics of agricultural products based on recurrent neural network," *Soft Computing*, vol. 25, no. 18, pp. 12107–12116, 2021.
- [15] C. Y. Lin, "An empirical study on decision factors affecting fresh e-commerce purchasing geographical indications agricultural products," *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science*, vol. 71, pp. 541–551, 2020.
- [16] Y. Huo, "Agricultural product E-commerce recommendation system based on neural factorization machine in smart cities," *Internet Technology Letters*, vol. 4, no. 1, 2021.
- [17] J. Wang and Y. Tao, "Reverse integration and optimisation of agricultural products e-commerce omnichannel supply chain under internet technology," *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science*, vol. 71, pp. 604–612, 2021.
- [18] X. Liu, "Empirical study on b2c e-commerce logistics network optimisation of agricultural products," *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science*, vol. 71, no. 5, pp. 346–362, 2021.
- [19] D. Lyu, L. Zhou, and X. Ma, "Income-increasing effect of e-commerce of agricultural products based on IT absorptive theory," *Wireless Networks*, vol. 28, no. 5, pp. 2337–2344, 2022.
- [20] F. Zhang and Y. Zhang, "A big data mining and blockchain-enabled security approach for agricultural based on internet of things," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 6612972, 8 pages, 2020.
- [21] Z. Rao and J. Yuan, "Data mining and statistics issues of precision and intelligent agriculture based on big data analysis," *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science*, vol. 71, no. 9, pp. 870–883, 2021.
- [22] R. Yu, C. Wu, B. Yan, B. Yu, and N. Chen, "Analysis of the impact of big data on e-commerce in cloud computing environment," *Complexity*, vol. 2021, Article ID 5613599, 12 pages, 2021.
- [23] S. Zhang and H. Mao, "Optimization analysis of tennis players' physical fitness index based on data mining and mobile computing," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9838477, 11 pages, 2021.
- [24] B. Yan, C. Wu, R. Yu, B. Yu, and Y. Yu, "Big data-based e-commerce transaction information collection method," *Complexity*, vol. 2021, Article ID 8665621, 11 pages, 2021.
- [25] D. S. Schwing, W. I. Sonntag, and S. Kühl, "Agricultural E-commerce: attitude segmentation of farmers," *Computers and Electronics in Agriculture*, vol. 197, article 106942, 2022.
- [26] S. Liang, "Research on the method and application of mapreduce in mobile track big data mining," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 14, no. 1, pp. 20–28, 2021.
- [27] S. S. Alrumiah and M. Hadwan, "Implementing big data analytics in e-commerce: vendor and customer view," *IEEE Access*, vol. 9, pp. 37281–37286, 2021.