

## Research Article

# Enterprise Decision-Making and Analysis Based on E-Commerce Data Mining

Xiaozhi Wang and Jieqiong Huang 

Department of Management Engineering, Hebei Petroleum University of Technology, Chengde, 067000 Hebei, China

Correspondence should be addressed to Jieqiong Huang; [huangjieqiong@cdpc.edu.cn](mailto:huangjieqiong@cdpc.edu.cn)

Received 29 November 2021; Revised 23 December 2021; Accepted 16 February 2022; Published 9 March 2022

Academic Editor: Deepak Kumar Jain

Copyright © 2022 Xiaozhi Wang and Jieqiong Huang. 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.

Enterprises have established a number of business processing systems and business websites according to their own characteristics and business needs, such as e-commerce websites and shopping websites. Over time, a large number of sales transaction data and customer purchase information have been generated, but no useful information has been generated stored in the database. Therefore, the management and decision-making levels of enterprises try to get useful information from huge and complex data. With the development of network technology and database, data mining technology has emerged. Nowadays, data mining technology has become one of the most concerned technologies of e-commerce. It can select appropriate data mining methods according to the characteristics of commodities, carry out effective statistical analysis and decision support for data, predict and analyze future market trends, greatly improve the business intelligence analysis of e-commerce enterprises, and make enterprises have greater advantages in market competition. This paper explains some important methods related to data mining and designs a data mining system. One of the systems can help e-commerce decision-makers analyze and predict data. Through experiments, it is concluded that the average absolute percentage errors of prediction data are 8.6% and 5.3%, respectively, with small errors and high accuracy. Second, make better recommendations for users. After investigation and analysis, the highest satisfaction has increased by 25% after using the system.

## 1. Introduction

*1.1. Background.* Data mining can analyze a large number of data stored in enterprises to find different customers or market divisions and analyze consumers' preferences and behaviors. It is of great significance for e-commerce enterprises engaged in online transactions. In recent years, with the rapid development of e-commerce in China, it is more and more favored by people for its convenience, interactivity, and efficiency. Therefore, many enterprises use various resources to establish and improve their own e-commerce platform to meet the needs of the market and maximize profits [1]. Data mining came into being. The term knowledge discovery first appeared at the 11th International Conference on artificial intelligence. Now, this technology is well known by more and more people. Its development has also been recognized by the society. It has been applied to many fields of society and achieved certain results. In order to

meet the needs of society, many companies have developed new data mining products. The emergence of these products makes the use of data mining technology more and more widely. It can be seen in many fields, and its application has also achieved good results.

*1.2. Significance.* The use value of data mining technology is very high and extensive. It is mainly used in the fields that are easy to produce a large amount of data and have the necessity of deep mining, such as banking, finance, communication, aviation, and government; these fields also have a common feature that it is necessary to predict the future development trend and market prospect of enterprises. Today, we will explain the application of e-commerce enterprises. Today's transactions are greatly transferred from offline to online, so there are many e-commerce companies. Now these e-commerce enterprises have many problems, such as the hidden customer value mining is not enough.

Data mining technology can extract a large amount of data for analysis, provide users with customer preferences and needs, and help users mine customer value [2]. Now many enterprises have adopted big data mining to more accurately grasp and meet the needs of customers by analyzing data information. Therefore, in recent years, many e-commerce companies have adopted data mining technology to deal with these huge amounts of data with high-value information. All these bring convenience and comfort to people's life.

*1.3. Related Work.* Data mining technology is the product of the information age. Many scholars have carried out relevant research on it to promote the development of industry technology. Zhou et al. proposed a big data mining method based on particle swarm optimization (PSO) back propagation (BP) neural network for financial risk management of commercial banks deploying the Internet of Things. This method uses Apache spark and Hadoop HDFS technology to build a nonlinear parallel optimization model on the data sets of on balance sheet items and off balance sheet items. The experimental results show that the parallel risk management model has fast convergence speed, strong prediction ability, and good performance in screening default behavior [3]. Yan et al. proposed a new multimedia big data mining system based on MapReduce framework to find the negative correlation between semantic concept mining and retrieval. This multimedia big data mining system consists of a big data processing platform with Mesos for efficient resource management and Cassandra for processing data across multiple data centers, but this system is difficult to use and lacks practicability [4]. In order to protect sensitive information in mining data, Xu et al. propose a method to organize various ongoing work, classify protection methods with rampart framework, and encourage interdisciplinary solutions to solve more and more privacy problems found from data. This method is very humanized, but it is lack of practicality [5]. Hu and Tan proposed a big data processing model from the perspective of data mining. This data-driven model involves demand driven aggregation, mining and analysis of information sources, user interest modeling, and security and privacy considerations, but the research lacks some detailed design and introduction [6]. Batran et al. propose a method based on mature technology to extract the start end (OD) journey from the original CDR data of mobile phone users and process the data to capture the mobility of these users. Then, the travel map, scenic spot map, and OD matrix of the study area are generated, which can be used for urban and transportation planning. The proposed method and research prove that mining big data sources can infer the temporal and spatial flow of people in cities and understand their flow patterns, which is valuable information for urban planning. This method analyzes people's travel big data through data mining, which is a reference for the design of data mining system in this paper. The stiffness method is complex and difficult to operate [7]. Atta-ur-Rahman proposed the research on data mining technology of teacher evaluation data organization survey, including data drilling, data cleaning, data fusion, data dimension,

analysis, visualization, and prediction. This technology will help administrators take the necessary steps to guide/consult staff and thank those who have performed well. The data has been entered into a database file with an appropriate schema. The system will help managers find out various results immediately without any problems [8]. Jha et al. introduced the literature review, including the importance, challenges, and application of big data in various fields, as well as different methods of big data analysis using data mining technology. The results of this review provide researchers with relevant information about the main trends of big data research and analysis using different analysis fields [9]. Cong applies data mining and optimized neural network to big data inventory forecasting. Data mining allows extract useful information from huge data sets, while neural network can predict the future trend of large databases [10]. Therefore, the combination of these two technologies can achieve very reliable and powerful prediction. Das et al. try to make better prediction results for complex stock markets, but the operability of this method is not strong [11].

*1.4. Innovation.* The innovations of this paper are as follows: (1) this paper introduces the related technologies of data mining and the application methods in e-commerce system in detail, including data mining technology, data warehouse technology, e-commerce recommendation system technology, and association rules. (2) Combined with personalized e-commerce recommendation and sales data analysis and prediction, a system is designed to bring better user experience to consumers and help managers make decisions.

## 2. Application Methods of Big Data Mining in E-Commerce Enterprises

### 2.1. Data Mining-Related Technologies

*2.1.1. Data Mining Technology.* Data mining means the process of extracting useful information and knowledge from huge, incomplete, and messy information and then transforming the extracted knowledge into definitions, models, laws, and other forms. No matter in which field, the application of data mining is generally divided into three stages, as shown in Figure 1 [12].

The first stage is data preparation stage, which includes data selection and data preprocessing. The second stage is the stage of selecting algorithm and implementation of data mining and selecting algorithm according to the type of data function and data characteristics. The third stage is the performance and interpretation stage of the results, which interprets and evaluates the results of data mining and transforms them into knowledge that can be understood by users [13].

*2.1.2. Data Warehouse Technology.* With the wide application of information systems and the fierce competition in the market, people are no longer satisfied with the simple transaction processing of computers but need valuable information to guide managers to make correct decisions.

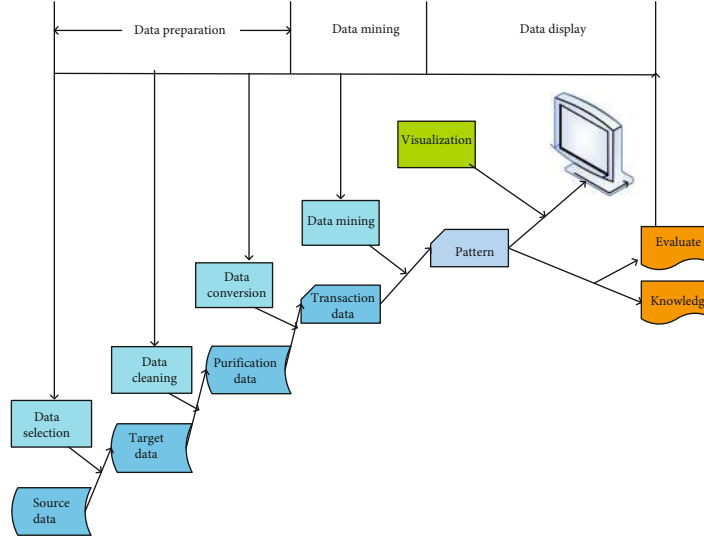


FIGURE 1: Data mining process.

However, the traditional database cannot meet such needs. At this time, a technology that can transform the routine data in the information system into commercially valuable and meaningful information is needed. Then, data warehouse technology is meeting this need. The characteristics of data warehouse are subject oriented, integrated, stable, and time-varying [14].

The physical structure of data warehouse usually uses relational database with star structure. The logical structure is divided into the nearest basic data layer, historical data layer, and general data layer. As shown in Figure 2, the structure is divided into five layers [15].

As can be seen from the figure, the five layers of data warehouse system structure are high comprehensive level, light comprehensive level, data mart, current detail level, and early detail level.

**2.1.3. Time Series Data Mining.** Time series mining is an important research field of data mining. Its role is to analyze and extract useful time-related information from a large number of time series data and make short-term, medium-term, and long-term prediction. It mainly analyzes historical data and records, defines inherent laws, and completes tasks such as predicting future tendencies and actions. Figure 3 is the basic flow of time series data mining [16].

Time series data mining methods can also be divided into many types, including trend prediction method, exponential smoothing method, weighted moving average method, moving average method, and seasonal trend prediction method; here, the seasonal exponential smoothing method is mainly introduced [17].

**(1) Seasonal Exponential Smoothing Method.** The main idea of seasonal exponential smoothing method is as follows: first, establish the tendency equation. Next, investigate the impact of seasons, that is, investigate the impact of seasonal changes on predictors. Finally, combined with seasonal

influence factors and tendency equation, a prediction model describing the development law of total time series is obtained. This method can effectively predict the time series of trend and seasonal tendency. The specific steps are as follows:  $l$  represents the cycle length, and  $x_1, x_2 \dots x_l$  represent the value of the first cycle. The formula for calculating the average value of each cycle is as follows [18]:

$$v_1 = \frac{1}{l} \sum_{i=1}^l x_i, v_2 = \frac{1}{l} \sum_{i=l+1}^{2l} x_i. \quad (1)$$

The seasonal exponential smoothing method calculates the average increment  $C$  and initial exponential smoothing value  $D$  of each cycle in the two cycles as follows:

$$C = \frac{v_2 - v_1}{l}, D = \frac{l-1}{2} C + v_2. \quad (2)$$

Calculate the seasonal factor  $s$  for the first two cycles as follows:

First cycle:

$$S_1 = \frac{2x_1}{2v_1 - l - 1 + 2m} (t = m = 1, 2, 3 \dots). \quad (3)$$

Second cycle:

$$S_2 = x_1 \left( v_2 - \frac{l+1-2m}{2} \right) (t = l+m = l+1, l+2, \dots). \quad (4)$$

Calculate the seasonal factor  $S$  of each period in the first two cycles as follows:

$$S = \frac{S_{n-1} + S_n}{2} (n = l+1, l+2, \dots, 2l). \quad (5)$$

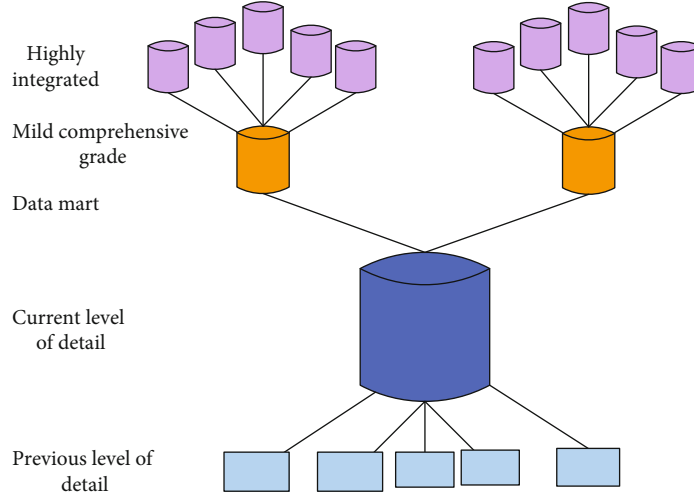


FIGURE 2: Data warehouse system structure.

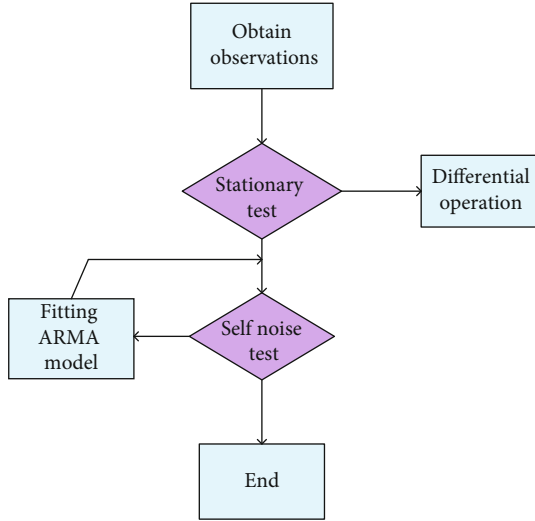


FIGURE 3: Flow chart of time series data mining.

The normalization of seasonal factors is to calculate the sum of  $L$  seasonal factors first, as shown below:

$$l' = \sum_{i=l+1}^{2l} S_i. \quad (6)$$

Then, calculate the normalized seasonal factor  $S_n$ :

$$S_n = \frac{1}{l'} S_n. \quad (7)$$

Predict the third cycle, as shown in the formula [19]:

$$S_{l+m} = (D + mC)S_{l-1+m}. \quad (8)$$

This method is suitable for selling goods with obvious seasonality, such as clothing and fruits, and helps e-commerce enterprises make decisions in different periods.

## 2.2. Key Technologies of E-Commerce Recommendation System Based on Data Mining

**2.2.1. Classical Algorithm of Recommendation System.** Recommendation algorithm is a comprehensive science, including several disciplines: cognitive science, approximation theory, information retrieval, prediction theory, management science, and marketing. The structure of the system is shown in Figure 4 [20].

The specific form of the recommended problem can be described in mathematical language: let  $a$  be all users and  $s$  be the synthesis of items. Many items can be recommended, such as food, books, novels, and games. The number of items in space  $s$  can be very large, ranging from hundreds of thousands to millions. Things such as books and CDS can reach millions. Similarly, in some application scenarios, the number of users can reach millions. Let  $p$  be the scoring function of user  $a$  for item  $s$ , and the mapping relationship is as follows [21]:

$$P : A \times S \longrightarrow R, \quad (9)$$

where  $R$  is a fully ordered set (a nonnegative integer or real number within a certain range of values). Then, according to the function  $u$ , we select the highest rated item  $s \in S$  for each user  $a \in A$  and recommend it to the user.

In the recommendation system, scoring is usually used to describe the user's favorite degree. For example, after watching the film "Changjin Lake," Pratama et al. gave the film 9 points, with a total score of 10. According to the above definition,  $u$  function can be any function. In the normal application environment,  $u$  can generally be a user-defined scoring function [22].

**2.2.2. Content-Based Recommendation Algorithm.** In the prediction function  $U(C, s)$  of content-based recommendation method, when estimating the score of user  $C$  on item  $s$ , it is mainly based on the user's previous score in similar items of  $S$ . For example, in food recommendation, the

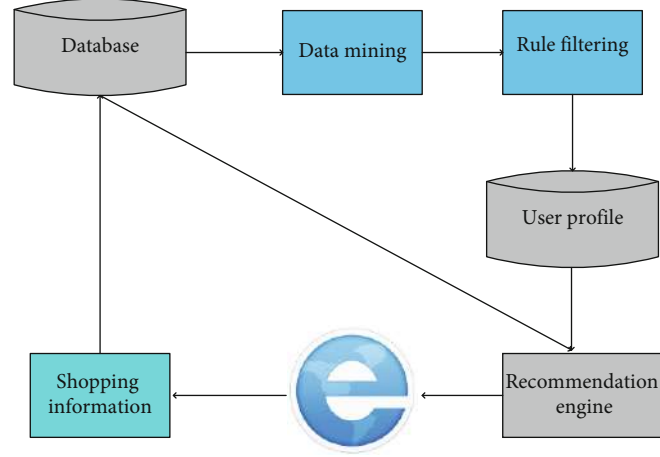


FIGURE 4: Flow chart of e-commerce recommendation system.

content-based recommendation algorithm analyzes that the restaurant users have visited or commented on and then recommends those restaurants with these same characteristics to users [23].

The most commonly used method to calculate the weight of keywords in the field of information retrieval is TF-IDF scale, that is, the value of keyword frequency/inverse document frequency. The specific definitions are as follows:  $n$  is the total number of items that can be recommended to users, and the keyword  $K_j$  has appeared in item  $n_j$ . In addition,  $f_{i,j}$  indicates the number of times the secondary keyword  $K_j$  appears in file  $d_j$ . Then,  $F_{i,j}$  represents the occurrence frequency of the keyword  $K_j$  in  $d_j$ , which is defined as [24]:

$$F_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}}. \quad (10)$$

$f_{z,j}$  is the most frequent keyword among all keywords appearing in document  $d_j$ . However, many keywords in the text information of articles are not representative. Therefore, when calculating the representativeness of keywords, it is often necessary to combine the reciprocal of document frequency (IDF) and simple keyword frequency (TF) to jointly calculate the weight of keywords. For keyword  $K_j$ , the reciprocal of document frequency is calculated as follows:

$$\text{IDF}_i = \log \frac{N}{n_i}. \quad (11)$$

Then, the TF-IDF weight of keyword  $K_j$  in file  $d_j$  is calculated as follows [25]:

$$w_{ij} = \text{TF}_{ij} \times \text{IDF}_i. \quad (12)$$

The content of  $d_j$  in the document is similar to the definition.

As mentioned earlier, what the recommendation system recommends to the user is often very similar to what he has purchased or evaluated before. Therefore, the recommendation system needs to compare the items with the items that the previous user has beaten too much and then select the items that are most similar to the items that the previous user likes for recommendation. Further analysis, we can define content-based profile ( $c$ ) to represent the preferences and interests of user  $C$ . The content-based profile ( $c$ ) can be obtained by analyzing the keyword information of items with high user scores. A more general method, the content-based profile ( $c$ ), can be expressed as a weight vector based on the content of the item, where each weight  $w_{c,i}$  represents the weight of the keyword  $k_i$  when representing the characteristics of user  $C$ . In content-based recommendation system,  $u(C, s)$  is usually defined as [26]:

$$u(c, s) = \text{score}(\text{ContentBased Profile}(c), \text{Content}(s)). \quad (13)$$

When the above method is applied to the web page recommendation system, we can take the web page URL address and web page news message as the text information of the article. The user's content-based profile ( $c$ ) and the text information of the item  $s$  can be represented by the weight vectors  $w_c$  and  $w_s$  of the TF-IDF of the keyword. The function usually calculates the cosine similarity of the weight vectors  $w_c$  and  $w_s$ :

$$u(c, s) = \cos(w_c, w_s) = \frac{\sum_{i=1}^k w_{i,c} w_{i,s}}{\sqrt{\sum_{i=1}^k w_{i,s}^2} \sqrt{\sum_{i=1}^k w_{i,c}^2}}. \quad (14)$$

$K$  in the formula represents the total number of keywords in the system.

In addition to the traditional information retrieval methods above, content-based recommendation system will also use some statistical learning methods, such as Bayesian classifier. These machine learning methods are essentially different from the traditional heuristic information retrieval methods. These machine learning methods do not use heuristic formulas to predict the similarity between the

characteristics of items and users, such as cosine similarity measurement, but use statistical learning and machine learning methods to establish a prediction model based on historical data. It uses a simple Bayesian classifier to judge whether the web pages containing  $k_1 k_2 \dots k_n$  keywords belong to a specific category  $C_n$  (specifically, it can refer to how much users like this web page) [27].

$$P(C_n | k_1 \& \dots \& k_n). \quad (15)$$

In addition, when using naive Bayes for classification, we assume that the keywords of documents are independent of each other, so the above formula can be transformed into:

$$P(C_i) \prod P(k_n | C_i). \quad (16)$$

The classification accuracy of naive Bayesian classifier is still high. According to formula (16), we calculate the probability that each page belongs to a category and then classify the web page into the category with the greatest probability.

### 2.3. Association Rules Based on Data Mining

**2.3.1. Basic Concepts of Association Rules.** Relevant rules are to investigate the internal relationship between items and find out the relationship between various transactions. For example, whether people who buy ice cream buy a cone at the same time; another example is whether people who buy toothbrushes buy toothpaste at the same time. This method is to study the relationship between goods purchased by consumers, which is beneficial to businesses to optimize the layout of goods. The processing object of association rules is composed of transaction  $T$ , roommate identifier (TID), and item set  $X$ . Suppose  $J$  is the full name of  $n$  items,  $J = \{i_1, i_2, \dots, i_n\}$ . Transaction  $T \in J$ , item level  $X \in J$ . There are two basic criteria for measuring association rules, as follows [28]:

(1) *Rule Confidence.* This is a simple measurement rule that represents the probability that both transactions will occur. It means that the probability of event  $B$  when event  $a$  occurs.

$$C_{A \rightarrow B} = \frac{|T(A \cap B)|}{|T(A)|}. \quad (17)$$

If the result is relatively large, it means that when event  $A$  occurs, the probability of event  $B$  is large, and the correlation between events  $A$  and  $B$  is large.

(2) *Rule Support.* This is a simple measure of the universality of association rules, indicating the probability of simultaneous occurrence of items  $A$  and  $B$ . The formula is as follows:

$$S_{A \rightarrow B} = \frac{|T(A \cap B)|}{|T|}. \quad (18)$$

**2.3.2. Apriori Algorithm.** Apriori algorithm is the most com-

mon and classic algorithm in correlation rule mining algorithm, which is widely used. This is a phased investigation and iterative process. Each investigation will scan the transaction database. Generating frequent item sets and generating simple correlation rules based on frequent item sets are two main parts of Apriori algorithm. Figure 5 shows the basic flow of Apriori algorithm [29].

The reason why frequent item sets are generated means that the support of item set  $C$  of project  $x$  is greater than or equal to the minimum support set by the user.

## 3. Test and Analysis

### 3.1. Design of Data Mining System

**3.1.1. System Requirement Analysis.** With the rapid development of e-commerce, the platform system has a lot of supply and demand information, but consumers often receive too much information to make correct purchase decisions. Therefore, the system needs to establish a mechanism to help consumers make choices. On the other hand, suppliers can make correct marketing decisions by analyzing product sales data and customer purchase patterns. Moreover, the system also has practical requirements for personalized recommendation and information disclosure. Therefore, the main goal of system mining system design is to enable customers to choose the required products and provide suppliers with correct marketing decision consultation.

### 3.1.2. System Structure Design

(1) *Statistical Analysis for Decision-Makers.* The structure of this part of the system is mainly divided into three levels: first, extract the data in the database, external-related data, and enterprise documents in the e-commerce system; convert them into organized data and put them into the data warehouse, and then establish a subject oriented data set according to the needs; secondly, using time series data mining technology and seasonal exponential smoothing method, the sales volume of goods is predicted; finally, the analysis results are displayed on the client through charts or other methods for reference and selection by business decision-makers. The specific structure is shown in Figure 6.

(2) *Consumer Facing Recommendation System.* This section contains six modules. The user information management module is mainly used to manage user information. The user retrieval module is used to provide users with some product retrieval functions. The user feedback module is used to provide users with feedback on the products recommended by the e-commerce recommendation system. The product information management module mainly provides a page for managers to input product information and display product information to customers. The recommendation module generates a recommendation list according to user preferences. The recommendation result fusion module combines the recommendation lists generated by the three recommendation algorithms to generate the best recommendation list. This part uses the recommendation system algorithm to design.

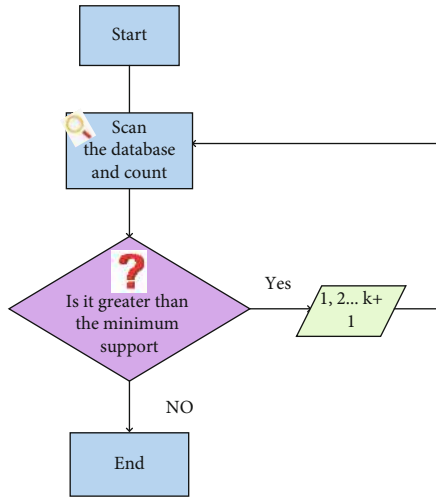


FIGURE 5: Flow chart of Apriori algorithm.

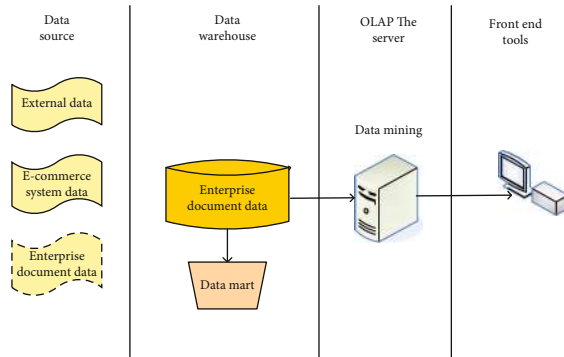


FIGURE 6: Overall structure design of the system.

3.1.3. *System Function Design.* The main functions of the system are divided into four subsystems, which are introduced as follows: the regular business subsystem mainly includes order entry, modification, query, return, exchange, and other operations of commodities in the e-commerce system. The comprehensive query subsystem mainly realizes the comprehensive query of customer management, inventory management, print management, and supplier management. The result display subsystem is divided into report making and statistical drawing, and the final results are visually displayed to users in the form of reports and graphics. The business analysis subsystem is a comprehensive system, including data warehouse, prediction subsystem, and purchase analysis subsystem. The prediction subsystem uses the improved seasonal index smoothing method to predict; that is, the best prediction factor combined with the time series index smoothing method is used to predict the sales volume of goods according to the sales data over the years, analyze the trend of historical sales data, and visually display it to users in the form of charts. The purchase analysis subsystem calculates the purchase quantity of future commodities according to the prediction results, which is convenient for coordinating the distribution of enterprise funds.

3.1.4. *Network System Logic Design.* The e-commerce platform is interconnected with the open network, so it is vulnerable to attack from the outside. In order to improve the security of the system, the website adopts the method of deploying application services in the application server and website data in the database server. At the same time, IDS is deployed in the network for intrusion detection and protection; firewall is deployed to ensure the security of the website. Figure 7 is the schematic diagram of system topology:

3.1.5. *Database Design.* The e-commerce platform system needs to dynamically add a large number of customer transaction information at any time and save a large number of customer registration information. Through the analysis of e-commerce system, we have established three subfunctional modules of system member analysis, prediction, and recommendation and system commodity analysis for the data mining system based on e-commerce. The background database includes five important tables: customer information table, order information table, and commodity information table. The specific design is shown in Tables 1–3.

3.2. *Implementation and Analysis of Data Mining System.* After the system is put into use, data is extracted from the database as needed, including sales data, sales data of goods; including time, place, sales volume, sales volume, and other information; commodity data, detailed information of commodities; including commodity name, category, price, sales quantity, and other detailed data describing commodities; time data, the time when the item was sold. Then, according to the system designed above, the sales volume of goods is estimated. This experiment is based on the data of a clothing store on Taobao.

3.2.1. *Sales Forecast Results and Analysis.* This experiment selects three years of historical data to analyze, so as to predict the sales data of the fourth year, and then compares it with the real data of the fourth year to judge the accuracy of the system’s calculated data. The following are the predicted sales results and the real data of each year based on the sales data over the years. First, the sales data of shirts are shown in Figure 8.

Then there is the sales data of T-shirt products, as shown in Figure 9.

From the above shirt and T-shirt sales data, it can be seen that the predicted sales volume in the fourth year is highly coincident with the sales volume curve predicted by the system, which proves that the prediction of the data is relatively accurate, but the calculation error using the algorithm is more rigorous. Therefore, the average absolute percentage error is used for error analysis, and the average absolute percentage error of the prediction data is calculated according to the following formula:

$$MAPE = \frac{|\text{actual value} - \text{Estimate}|}{\text{Estimate}} \times 100\%. \quad (19)$$

According to the results calculated by the formula, the average absolute percentage error of shirt prediction data is

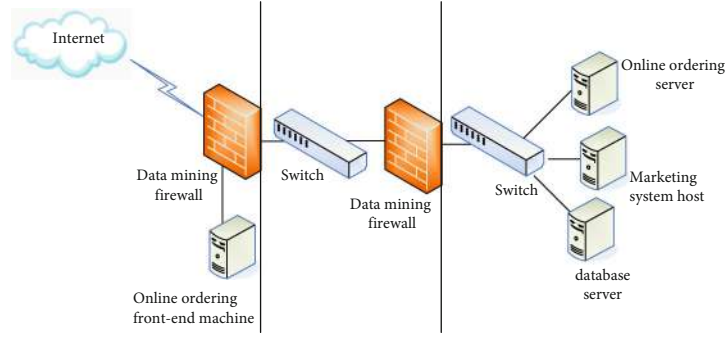


FIGURE 7: Logic design diagram of e-commerce platform network system.

TABLE 1: Customer information.

Data name	Data type	Data length	Explain
userid	Varchar(10)	10	Customer number
Username	Varchar(20)	10	Name
Sex	char	2	Gender
useraddr	rchar(30)	20	Address
Email	Varchar(30)	20	E-mail
usertel	Varchar(20)	20	Telephone

TABLE 2: Commodity information.

Data name	Data type	Data length	Explain
Item_id	Varchar(20)	24	Item number
Item_name	Varchar(50)	20	Trade name
Stock_qty	Float	8	Inventory quantity
Date	Datetime	10	Date
Price	Float	10	Unit price

TABLE 3: Order information table.

Data name	Data type	Data length	Explain
orderid	Varchar(10)	16	Order No
Order_user	User	32	Customer information
Type	Varchar(10)	10	Order type
Order_date	Datetime	10	time
orderzt	int	6	Order status
Is_modify	char	4	Can it be changed

8.6%, and the average absolute percentage error of T-shirt prediction data is 5.3%. It can be seen that the error is relatively small, so the prediction accuracy of the system is relatively high, which verifies the rationality of the system design. Therefore, businesses can predict the sales volume from next year according to this system, take this as the basis to purchase goods, make better decisions, and provide better services for customers.

**3.2.2. Recommendation Module Satisfaction Analysis.** The system will recommend the customer according to the items purchased by the customer. If it is added to the shopping

cart, collection, or purchase, it represents user satisfaction. If the user chooses not to recommend again or has not viewed it, it means that the user is not satisfied.

Click in to view the recommended products, but it is not added to the shopping cart, collection, or purchased. Experiments verify the accuracy of the system, that is, the user's satisfaction with the products recommended by the system. Figure 10 shows the statistical results of consumer satisfaction. The first is the statistical results before the application of the recommendation system, and the second is the survey results of satisfaction after the use of the recommendation system.

It can be seen from the above two figures that with the increase of the number of respondents, the probability of consumer satisfaction is gradually increasing and the dissatisfaction is gradually decreasing, but the decline and increase are small before using the recommendation system. After using the recommendation system, not only the probability of satisfaction is higher than before but also the probability of consumer satisfaction is greatly increased with the increase of the number of respondents; the maximum value can be increased to 37.2%, which is 25% higher than the maximum value before use. The corresponding dissatisfaction decreased from 24.6% to 11.5% and 13.1% after using the recommendation system. It can be seen that the recommendation system can greatly increase consumers' satisfaction and increase consumers' experience effect.

To sum up, the average absolute percentage error of the prediction data of the system is relatively small, whether in the sales prediction of shirts or T-shirts, and consumer satisfaction has been greatly improved after the use of the system, which proves that the use effect of the system is good.

## 4. Discussion

E-commerce enterprises hope that enterprises can obtain economic benefits by improving services, take data mining method as a means of e-commerce application, track and analyze users' preferences, and realize personalization. Data mining can combine the knowledge of database, artificial intelligence, statistics, and other fields to sort out, summarize and mine the huge transaction data of e-commerce, extract the potentially useful information and knowledge that people do not know in advance, help enterprises make decision analysis, guide business action, and provide more



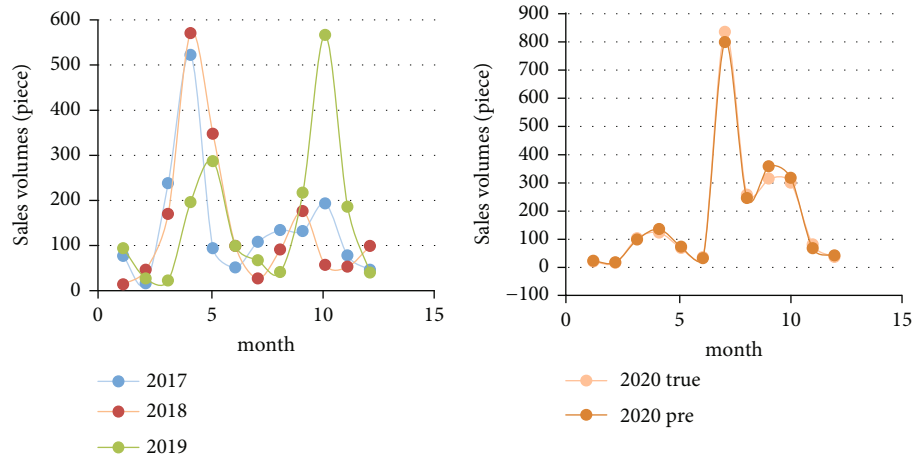


FIGURE 8: Shirt sales data.

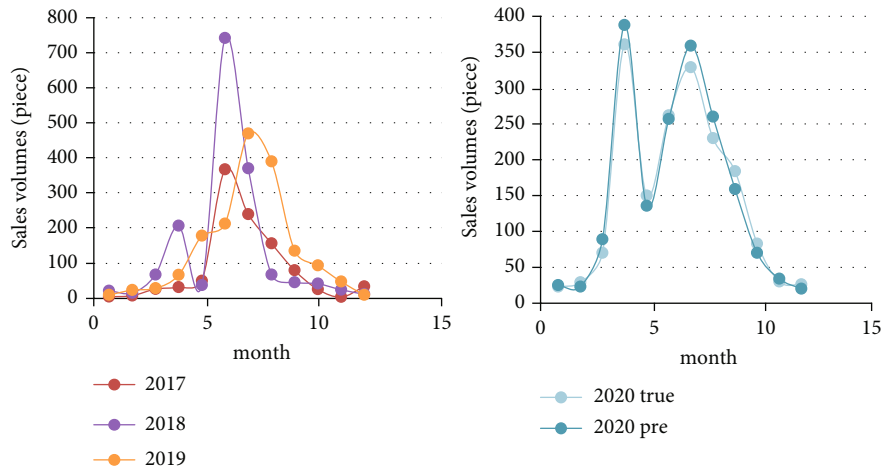


FIGURE 9: T-shirt sales data.

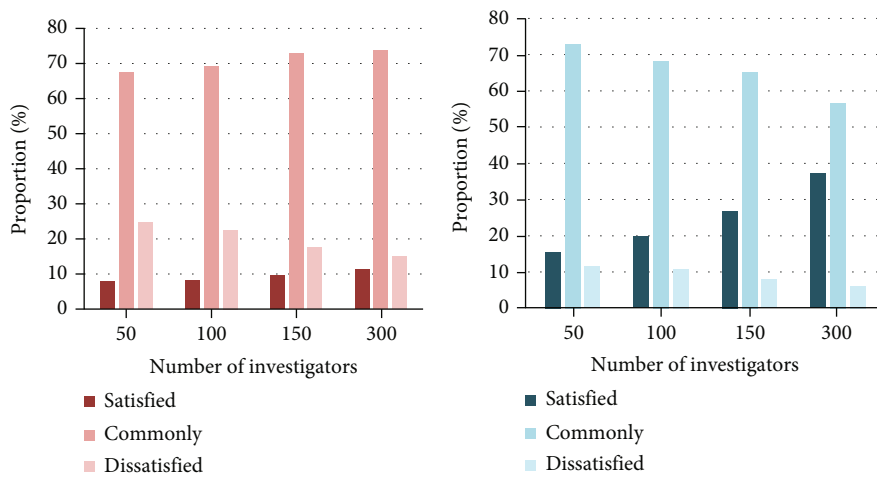


FIGURE 10: Consumer satisfaction survey results.

valuable services for customers. Data mining is very important for the development of modern enterprises in the era of e-commerce. Due to the late start of this technology in

China, it is not as advanced as foreign technology in many aspects. Its development is relatively slow and has not formed the form of collectivization. However, with the

development of this technology in China, data mining technology began to receive enough attention, and many scholars and operators began to focus on this research.

## 5. Conclusions

This paper introduces the related concepts and background of data mining. It can be seen that it develops very rapidly and is widely used and introduces the methods related to data mining and uses some of them to design an e-commerce system based on data mining, so as to help decision-makers make decisions and better recommend products for consumers. The details are as follows: (1) analyzes the needs of system users from the perspectives of decision-makers and consumers. (2) The seasonal exponential smoothing method is used to design the systematic analysis and prediction data plate. Using the sales data of three years to predict the sales data of the fourth year and then comparing with the real data of the fourth year, the average absolute percentage errors of shirts and T-shirts are 8.6% and 5.3%, respectively, which proves that the accuracy of the system is high. (3) The consumer recommendation section is designed by using the system recommendation algorithm, and the satisfaction survey is carried out. The results show that the satisfaction has been greatly improved, the maximum value of satisfaction probability has been increased by 25%, and the maximum value of dissatisfaction probability has been reduced by 13.1%.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## References

- [1] M. Meng and J. Yu, "In-depth analysis of financial market based on iris recognition algorithm of MATLAB GUI," *Neural Computing and Applications*, vol. 33, no. 11, pp. 5659–5674, 2021.
- [2] X. Zheng, G. Zhu, N. Metawa, and Q. Zhou, "Machine learning based customer meta-combination brand equity analysis for marketing behavior evaluation," *Information Processing and Management*, vol. 59, no. 1, article 102800, 2022.
- [3] H. Zhou, G. Sun, S. Fu, J. Liu, X. Zhou, and J. Zhou, "A big data mining approach of PSO-based BP neural network for financial risk management with IoT," *Access*, vol. 7, article 154035, 2019.
- [4] Y. Yan, M. L. Shyu, and Q. Zhu, "Supporting semantic concept retrieval with negative correlations in a multimedia big data mining system," *International Journal of Semantic Computing*, vol. 10, no. 2, pp. 247–267, 2016.
- [5] L. Xu, C. Jiang, Y. Chen, J. Wang, and Y. Ren, "A framework for categorizing and applying privacy-preservation techniques in big data mining," *Computer*, vol. 49, no. 2, pp. 54–62, 2016.
- [6] W. Hu and B. Y. Tan, "Data mining and big data," *IEEE Transactions on Knowledge & Data Engineering*, vol. 26, no. 1, pp. 97–107, 2016.
- [7] M. Batran, M. Mejia, H. Kanasugi, Y. Sekimoto, and R. Shibasaki, "Inferencing human spatiotemporal mobility in greater Maputo via mobile phone big data mining," *International Journal of Geo-Information*, vol. 7, no. 7, p. 259, 2018.
- [8] D. S. Atta-ur-Rahman, "Big data analysis for teacher recommendation using data mining techniques. International," *Journal of Control Theory and Applications*, vol. 10, no. 18, pp. 95–105, 2017.
- [9] A. Jha, M. Dave, and S. Madan, "A review on the study and analysis of big data using data mining techniques," *International Journal of Latest Trends in Engineering and Technology*, vol. 6, no. 3, pp. 94–102, 2016.
- [10] W. Cong, "Study of financial warning ensemble model for listed companies based on unbalanced classification perspective," *International Journal of Intelligent Information Technologies*, vol. 16, no. 1, pp. 32–48, 2020.
- [11] D. Das, A. S. Sadiq, N. B. Ahmad, and J. Lloret, "Stock market prediction with big data through hybridization of data mining and optimized neural network techniques," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 29, no. 1–2, pp. 157–181, 2017.
- [12] C. S. Mayo, M. L. Kessler, A. Eisbruch et al., "The big data effort in radiation oncology: data mining or data farming?," *Advances in Radiation Oncology*, vol. 1, no. 4, pp. 260–271, 2016.
- [13] X. Zhang and P. Guo, "Research on e-commerce logistics and traditional industry integration mode based on big data," *Journal of Physics Conference Series*, vol. 1744, no. 4, article 042052, 2021.
- [14] G. Chen, Y. Lu, Y. Meng et al., "FUSO: fast multi-path loss recovery for data center networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 3, pp. 1376–1389, 2018.
- [15] H. Goyal, C. Sharma, and N. Joshi, "An integrated approach of GIS and spatial data mining in big data," *International Journal of Computer Applications*, vol. 169, no. 11, pp. 1–6, 2017.
- [16] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *Journal of Big Data*, vol. 4, no. 1, p. 20, 2017.
- [17] Z. Liu, "Fast kernel feature ranking using class separability for big data mining," *Journal of Supercomputing*, vol. 72, no. 8, pp. 3057–3072, 2016.
- [18] D. Dahmani, S. A. Rahal, and G. Belalem, "Improving the performance of data mining by using big data in cloud environment," *Journal of Information & Knowledge Management*, vol. 15, no. 4, article 1650038, 2016.
- [19] C. Kuo, C. H. Lin, and M. H. Lee, "Analyze the energy consumption characteristics and affecting factors of Taiwan's convenience stores-using the big data mining approach," *Energy and Buildings*, vol. 168, pp. 120–136, 2018.
- [20] F. Kong and X. Lin, "The method and application of big data mining for mobile trajectory of taxi based on MapReduce," *Cluster Computing*, vol. 6, pp. 1–8, 2018.
- [21] S. Shadroo and A. M. Rahmani, "Systematic survey of big data and data mining in Internet of Things," *Computer Networks*, vol. 139, pp. 19–47, 2018.
- [22] E. E. Pratama, H. Sastypratiwi, and Yulianti, "Analisis kecenderungan informasi terkait COVID-10 berdasarkan big data

- sosial media dengan menggunakan metode data mining,” *Jurnal Informatika Polinema*, vol. 7, no. 2, pp. 1–6, 2021.
- [23] A. Ghenabzia, O. Kazar, A. Merizig, Z. Sayah, and M. Zoubeidi, “Multi-agent approach for data mining-based bagging ensembles to improve the decision process for big data,” *International Journal of Information and Communication Technology*, vol. 17, no. 4, p. 380, 2020.
- [24] H. Zhou, G. Sun, S. Fu et al., “A distributed approach of big data mining for financial fraud detection in a supply chain,” *Computers, Materials & Continua*, vol. 64, no. 2, pp. 1091–1105, 2020.
- [25] M. Nie, Z. Xiong, R. Zhong, W. Deng, and G. Yang, “Career choice prediction based on campus big data—mining the potential behavior of college students,” *Applied Sciences*, vol. 10, no. 8, p. 2841, 2020.
- [26] Y. Guan, “Data mining system and construction analysis of skilled talents based on big data,” *Journal of Physics Conference Series*, vol. 1744, no. 3, article 032021, 2021.
- [27] K. N. Neeraj and V. Maurya, “A review on machine learning (feature selection, classification and clustering) approaches of big data mining in different area of research,” *Journal of Critical Reviews*, vol. 7, no. 19, pp. 2610–2626, 2020.
- [28] Y. Du and T. Zhao, “Network teaching technology based on big data mining and information fusion,” *Security and Communication Networks*, vol. 2021, Article ID 6629563, 9 pages, 2021.
- [29] M. Asim, S. Ahmad, M. Shuaib, and O. K. Harsh, “Adaptation of artificial intelligence in big data mining and its impact: a study,” *Solid State Technology*, vol. 63, no. 5, p. 2322, 2020.