Hybrid IoT and Data Fusion Model for e-Commerce Big Data Analysis

Bing Li\textsuperscript{1,2} and Qi Lei\textsuperscript{3}

\textsuperscript{1}School of Software, Jiangxi Normal University, Nanchang, 330022 Jiangxi, China
\textsuperscript{2}Research Centre for Management Science and Engineering, Jiangxi Normal University, Nanchang, 330022 Jiangxi, China
\textsuperscript{3}Department of Nephrology, Dayuan Hospital Affiliated to Jiangxi Academy of Medical Sciences, Nanchang, 330022 Jiangxi, China

Correspondence should be addressed to Bing Li; 004970@jxnu.edu.cn

Copyright © 2022 Bing Li and Qi Lei. 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 is aimed at studying the e-commerce big data recommendation model based on data fusion and the Internet of Things. This article chooses an embedded system for the construction of an e-commerce platform and uses data fusion technology to collect, transmit, and filter useful information from various information sources. Then, the collected information and data are analyzed and integrated, and visualization algorithms are used to better present data analysis, and association rules and structural similarity methods for electronic comparison are used. This article uses the B/S architecture to design the overall framework of the data access layer, business logic layer, and user presentation layer; collects, organizes, stores, and presents the acquired consumer information; and finally analyzes the e-commerce background, platform performance, and supply and demand analysis. The experimental results show that the average clustering coefficient of the platform (0.7559) is smaller than the average clustering coefficient value of 8 items in the store (0.811) and smaller than the average network diameter and average path length of the online store (3.86, 7.7). Store products are better than store products, and the diameter and length of the product should be larger (2.71, 5.75). The recall rate of the e-commerce big data platform and model matching review method designed in this paper is 5% higher than that of the word matching model method and has a better expected effect in terms of user supply and demand.

1. Introduction

For the integration of virtual technology and real economy, e-commerce has been popularized. e-Commerce provides a network platform for corporate transactions, reduces unnecessary processes in business activities, reduces the operation time of business activities, and greatly enhances operational strength. As long as all users connected to the Internet can conduct e-commerce transactions, the scope of e-commerce users will increase and costs will be reduced. For general users, e-commerce can increase the user’s choice space, enhance the user’s initiative in consumption, and reduce the user’s capital investment and time.

e-Commerce has produced a large amount of data. How to use these data reasonably, and how to store and analyze the data has become very important, and it is also an urgent problem for major enterprises to solve. In e-commerce website data, there is a large amount of unstructured data, such as text, image, audio, and video. The indoor play is formed based on the combination of computer and traditional network technology. It includes utility computing, parallel computing, and distributed computing. It has the characteristics of network storage, virtualization, and load balancing. It represents future data management technology development direction. It can manage large amounts of data with less cost and can handle different data classifications and has good scalability.

The innovations of this article are as follows: (1) fully excavate the characteristics and shortcomings of e-commerce, propose specific improvement methods on its shortcomings,
and combine them with the Internet of Things technology according to its characteristics, so as to allow e-commerce to develop better. (2) In the process of building an e-commerce data platform, we use embedded systems in the Internet of Things, break the traditional platform design, better play the role of big data analysis, and provide better analysis results for merchants.

2. Related Work

e-Commerce has become more and more popular with the development of the Internet, so there are more and more researches on it. In Mero J e-retailing, competitors only need to click once, and prices are easy to compare. Providing high-quality customer service and mutual communication through the company’s website is an important aspect of attracting and retaining customers. An increasingly popular solution for improving customer service is the “live chat” interface, which allows consumers to have real-time online conversations with customer service agents. Since the literature on the impact of real-time communication through real-time chat is currently very limited, his research has developed and tested a model that proves that the chat service has two-way communication (that is, the core element of perceiving website interactivity) and customer satisfaction and trust degree and repurchase and word-of-mouth (WOM) intentions. The data for this study was collected through an online survey (N = 6783). The subjects of the survey were existing customers of five e-retailers that actively use real-time chat as a customer service tool. The research results show that consumers have a low level of awareness of e-commerce. However, in his research, the user’s frequency of using e-commerce is low, so it is difficult to better meet the experimental requirements [1]. In order for Maxim et al. to utilize the massive and diverse data captured and stored by the ubiquitous Internet of Things services, forensic investigators need to use evidence acquisition methods and technologies in all areas of digital forensics and may create new Internet-specific investigation processes. Although his research has developed many conceptual process models to solve the unique characteristics of the Internet of Things, many challenges remain unsolved [2]. In order to solve the reliability problem of e-commerce and record any individual business website, they recommend designing a personal business website that is used to purchase books and enable access to the domain by purchasing from the digital ocean cloud, and then, in it, install the server domain; then, they connect it to the website’s digital certificate (encrypt it) and then add some software to the client (cerbot, root) server assistant program in it, and they have connected the server and the server through Linux instructions and codes. The certifications are linked together, and the commercial website (PHP3.PHP) has been tested in the SSL laboratory. The result is that the service sites are divided into three categories: the current latest, the worst and the current worst, and the detailed information of the classification. However, they did not solve the reliability problem well [3].

3. Method of Constructing an e-Commerce Big Data Analysis Platform Based on Data Fusion and Internet of Things Technology

3.1. Internet of Things Technology

3.1.1. Concept of Internet of Things. The Internet of Things is referred to as IoT (Internet of Things), which is the inheritance and development of the Internet. The concept of the Internet of Things can be analyzed from two aspects. On the one hand, it is a connected network; on the other hand, it is a network that can connect objects to objects and objects to people [4, 5]. This article is mainly to connect the Internet of Things technology and e-commerce to build a big data analysis platform, which is to collect detailed information of goods through infinite sensor devices and then use the embedded system to store and transmit data to realize the analysis and analysis of e-commerce platform data application.

3.1.2. IoT Technology System. The Internet of Things is one of the industries with the most potential for development. It can promote the transformation and upgrading of other industries and make traditional enterprises more innovative. However, with the continuous development of the Internet of Things, my country has been lacking major technologies and related technological innovations, leading to the problem that product quality is average, but the price cost has not been reduced [6, 7]. Technical backwardness restricts development, and the lack of basic technologies such as RFID also strictly restricts the development of the Internet of Things in my country. Therefore, the focus of IoT is mainly on sensors, RFID, and cloud computing. Internet technology is applicable to many industries. In various industries, there are various requirements and technical forms. However, among various technical systems, the Internet of Things technology is mainly composed of four systems. The four main systems are perception systems, network systems, computer systems, and service and management support systems [8, 9].

(1) Perception recognition technology is the most basic part of the Internet of Things technology, and it is also the foundation of the Internet of Things technology. The detection technology is mainly composed of carbon dioxide sensors, temperature sensors, humidity sensors, optical sensors, and other sensors. The identification technology mainly includes two-dimensional code tags, RFID, and GPS. Perception technology is similar to animals that use ears, eyes, nose, nose, and nerve endings to collect information and determine the source of various objects. The main function of perception technology is to collect relevant data, while the main function of recognition technology is to mark personal traces.

(2) The network layer of the network system is composed of various networks. The network layer is equivalent to the computer CPU used to calculate, analyze, and process the sent information. The
3.2. Data Fusion Technology. Data fusion refers to the process of using computer technology to process information under specific standards. This process can automatically analyze, optimize, and integrate the sensor observation results that require sufficient time to complete the required decision-making and evaluation tasks, and the collection, transmission, synthesis, filtering, correlation, and synthesis of useful information from various information sources. Data fusion is also called information fusion or multisensor fusion. The process of using computer technology to process information under certain conditions is used. This process can automatically analyze, optimize, and synthesize time series multisensor observations to complete the required set of decision-making and evaluation tasks. With the continuous development of science and technology, the data environment in life will become more complex in the future, and the amount of data will increase exponentially. At the same time, there is still a lot of uncertain and false information. In this case, the continuous development of data fusion technology will significantly improve the efficiency of information processing and provide accurate, timely, and effective information support for data decision-making.

3.3. e-Commerce

3.3.1. Definition of e-Commerce. The term e-commerce comes from the English electronic commerce, which is usually translated into e-commerce or e-business (electronic business) in China. Since e-commerce was proposed in 1995, e-commerce as a new business model has been recognized by the society, but there is no scientific and complete definition up to now. Governments, scholars, research institutions, companies, etc. have explained e-commerce from different perspectives. It can be summarized as that e-commerce is based on the network and realizes the whole process from product sales, market planning to information management [14, 15].

3.3.2. e-Commerce Model. The e-commerce model refers to a method for companies to provide customers with better and more affordable value and benefit from it in the context of e-commerce. With the rapid development of e-commerce, there are more and more business models. At present, the most conventional classification is based on e-commerce transaction objects, mainly B2B (business to business), B2C (business to consumer), B2G (government to business), C2B (consumer to business), C2C (consumer to consumer), C2G (consumer to government), G2B (government to business), G2C (government to consumer), G2G (government to government), and other nine categories [16, 17], as shown in Table 1.

(1) Business-to-business e-commerce, referred to as B2B, refers to the integration of commodity resources between enterprises and manufacturers on the Internet platform. B2B is the most popular e-commerce development model for enterprises. The integration of the supply chain is realized through the concentration of supply, the automatic realization of procurement, and the high efficiency of the supply and distribution system. Typical representatives of the B2B model include Alibaba, Made-in-China, and Huicong

(2) Business-to-consumer e-commerce, referred to as B2C, is the most intense e-commerce model in the market competition. Enterprises provide consumers with an online trading platform through the Internet, which not only saves consumers’ time but also brings huge profits to enterprises, which can be described as a win-win situation. Typical representatives of the B2C model are Amazon, Tmall, Jumei Youpin, etc. [18, 19]

(3) Consumer-to-consumer e-commerce is abbreviated as C2C, which is a business activity in which individuals conduct transactions with the help of a network platform. The C2C platform builds a bridge for both parties to the transaction and provides a series of supporting services to ensure the smooth progress of the transaction. Typical representatives of the C2C model include Taobao, Paipai, and eBay

(4) Consumer-to-business e-commerce, referred to as C2B, is a new e-commerce model that is based on consumer needs and conducts reverse transmission in the business chain. Consumers can apply actively participate in the design, production and pricing of customized products. With the help of the flourishing Internet, Internet of Things, cloud computing and big data, and other information technology, the individual needs of consumers are better and better met. Typical representatives of the C2B model include group buying websites such as Juhuasuan, Baidu Nuomi, and Meituan
In the e-commerce model, the four most common types are B2B, B2C, C2C, and C2B, and the other five types will also be applied. In addition, there is the O2O model, which realizes the perfect interaction between online and offline, and connects physical stores with online stores to meet the dual experience of consumers; the B2R model is a new model of transactions between enterprises and retailers, a typical representative is Yiwu Shopping.

3.4. Big Data

3.4.1. The Era of Big Data. As the material foundation of the cloud era, big data has attracted more and more attention. From the point of view of enterprises, some people think that big data is the total amount of data produced or exchanged in the operation of enterprises. These data are huge in quantity and have different structures. It is not yet possible to perform simple classification, calculation, and other operations on the data. The concepts of big data and cloud computing are often confused [20, 21]. In fact, big data refers to the collection of data, and cloud computing refers to the calculation method, but this calculation method is only for big data. The calculation method of big data cannot be calculated by the commonly used standalone computer. No matter how good the performance is, the host cannot receive the data processing and analysis of the supermass set. A distributed processing system must be used to measure and measure to meet the era of big data challenge.

In the era of big data, diversification of data collection methods is a major feature of e-commerce websites. Generally speaking, the data sources of e-commerce platforms can be divided into four types: (1) Platform internal data: these data include users’ browsing, search, purchase, evaluation, and transaction data within the platform, as well as merchants’ information on the sale of goods. Attribute definition and description, of course, as well as user-business interaction and evaluation data. These data can objectively and comprehensively reflect the psychological factors of both parties to the transaction, and these factors represent the key to the success or failure of the transaction and have high analytical value [22, 23]. (2) Platform external guiding data: mainly through the introduction of web advertisements, the display ranking data of the search platform, service links, and related application recommendations, etc. (3) Direct access to the data: this part of the data mainly comes from browser visits and direct visits to various types of e-commerce websites. These data can reflect the user’s preferences and habits, and more often reflect the user’s needs. (4) Wireless data: With the rapid increase in the number of mobile electronic devices, the amount of data transmitted wirelessly is also soaring. All e-commerce platforms have opened application interfaces for mobile electronic devices, allowing users to conduct e-commerce activities anytime, anywhere. Wireless devices, like wired devices, generate a large amount of data. Due to the particularity of wireless devices, these data tend to reflect the characteristics of users, and it will be more meaningful and valuable for data analysis and utilization.

3.4.2. Big Data Analysis. The analysis of big data needs to analyze and process the data through standardized processes and tools, which can ensure that the analysis results can achieve better results. Because the diversity of unstructured data brings new challenges to data analysis, a series of tools are needed to parse, extract, and analyze data, and data visualization is the most basic requirement of data analysis tools. Predictive analysis with big data allows data analysts to make predictions on the development trend of data based on visual analysis reports and indirect data information and analyze the natural and social impacts brought about by this.

3.4.3. Big Data Processing. The collection of big data uses databases as data carriers, and these databases receive the data sent by the client. Users can perform data query and other operations in the database. In the era of big data, e-commerce will still consider the timeliness of transactions, using traditional relational databases SQL Server, MySQL, and Oracle to store and process each transaction data. In addition, some nonrelational databases. It is also commonly used for data collection, such as Redis, MongoDB, and HBase. The collection of big data has its main characteristics, which are large throughput and high concurrency. For example, ticketing websites and e-commerce websites are both business-oriented websites, so they have to face huge concurrent visits. If you want to receive and process such a huge amount of data at the same time, you need to rely on big data processing technology to ensure the integrity of the data. And considering the overall performance of the database and the utilization of resources, it needs to be load balanced.

In order to meet the needs of data analysis, statistics and analysis mainly use distributed processing systems or distributed databases to classify and process the massive data stored in them. In this regard, some real-time requirements will use the MPP system adopted by EMC’s newly acquired GreenPlum, MySQL-based columnar storage Infobright, and Oracle’s Exadata. These systems have certain advantages in decision support and data mining. Facing the statistical analysis of large amounts of data, it has brought severe challenges to system resource utilization and network resource occupancy.
3.5. Research Methods

3.5.1. Reconstruction of the Original Data Distribution. The purpose of reconstructing the distribution of the original data is to make the modified and reconstructed data still is able to perform data mining and obtain more accurate results. Reconstructing the original distribution is to reconstruct the distribution of the original data, not to rewrite the original data in the record.

Suppose the original data \( x_1, x_2, \cdots, x_n \) are the sample values of independent and identically distributed random variables \( X_1, X_2, \cdots, X_n \). In order to hide the original value, we introduce independent and identically distributed random variables \( Y_1, Y_2, \cdots, Y_n \). According to the distribution function \( f_Y \) of \( x_1 + Y_1, x_2 + Y_2, \cdots, x_n + Y_n \) and \( Y \), the cumulative distribution function \( F_X \) is estimated.

We estimate the posterior distribution function according to Bayes’ rule, such as

\[
F'_{X_i}(e) = \frac{\int_{-\infty}^{e} g_Y(q_i-z)g_X(z)dz}{\int_{-\infty}^{\infty} g_Y(q_i-z)g_X(z)dz}. \tag{1}
\]

We calculate the average value for it:

\[
F'_{X_i}(e) = \frac{1}{n} \sum_{i=1}^{n} F'_{X_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{\int_{-\infty}^{e} g_Y(q_i-z)g_X(z)dz}{\int_{-\infty}^{\infty} g_Y(q_i-z)g_X(z)dz}. \tag{2}
\]

In the same way, calculating the posterior density function can be obtained by differentiating the above value:

\[
f'_{X_i}(e) = \frac{1}{n} \sum_{i=1}^{n} g_Y(q_i-z)g_X(z)dz. \tag{3}
\]

If the calculable sample size is large enough, the \( f'_{X_i} \) obtained by the above formula (3) is closer to the true value \( f_{X_i} \). At this time, \( f_Y \) can be made public, but \( f_{X_i} \) is hidden.

Therefore, the joint distribution can be used as the initial distribution variables \( X \) have a connecting edge. Using visualization algorithms, the relationship between the objects is represented as the directed connecting edge between the nodes. The similarity between two objects depends on their structural relationship in the graph. The theoretical assumption based on the SimRank algorithm is as follows: two objects are similar if they are related to other similar objects. This assumption is similar to the assumption based on the SimFusion algorithm.

3.5.2. Visualization Algorithm. After we can automatically analyze e-commerce big data, we need to choose a visualization tool to present it to the public. According to the visualization algorithm, suppose the value of the \( i \)th point in the sequence is \( t_i \) and the value of the \( u \)th point is \( t_u \), any other data between these two data is marked as \( t_j \). The relationship that should be satisfied between them is as follows:

\[
t_u < t_x + (t_j - t_u) \times \frac{x-u}{(x-i)}, i < t < u, \tag{4}
\]

\[
\frac{(t_x - t_u)}{(x - u)} < \frac{(t_x - t_i)}{(x - i)}. \tag{5}
\]

If all the intermediate points between \( t_i \) and \( t_u \) do not cut off the straight line connecting them, then \( t_i \) and \( t_u \) have a connecting edge. Using visualization algorithms, the time series data can be mapped to form a complex network diagram.

The power-law distribution allows us to better understand the economic meaning of e-commerce, and it can also be called a discrete Pareto distribution. The expression is as follows:

\[
p(l) = \frac{l^{-\alpha}}{\beta(l-\beta)}, \tag{6}
\]

Among them, \( l \) is a positive integer that usually represents a variable of interest, \( p(l) \) represents the probability of occurrence of \( l \), and \( \beta \) represents a constant power exponent, generally between 1 and 3.

The formula of Pareto cumulative distribution is

\[
P(x \geq x_i) = \frac{\sum_{i=1}^{N} f(x_i)}{N}. \tag{7}
\]

The advantage of the cumulative occurrence probability is that it not only avoids the mediocrity in the distribution description but also does not reduce the amount of data, and each data after accumulation generally contains a lot of original data, so it will minimize statistical fluctuations.

3.5.3. Calculation Method of Structural Similarity. Based on the simple and intuitive graph theory model \( H = (B, R) \), Jeh and Widom et al. proposed the SimRank algorithm, where \( B \) represents the node in the graph and \( R \) represents the connecting edge between the nodes. The similarity between two objects depends on their structural relationship in the graph. The theoretical assumption based on the SimRank algorithm is as follows: two objects are similar if they are related to other similar objects. This assumption is similar to the assumption based on the SimFusion algorithm.

The algorithm first represents the data object in the relational data as a node in the graph theory model, and the relationship between the objects is represented as the directed connecting edge between the nodes. The similarity between the nodes can be obtained by the SimRank calculation formula:

\[
S(b,c) = \frac{D}{|U(b)||U(c)|} \sum_{i=1}^{\frac{|U(b)|}{|U(c)|}} \sum_{j=1}^{\frac{|U(c)|}{|U(b)|}} D_{link} \left[ U_i(b), U_j(c) \right], (b \neq c). \tag{8}
\]

It can be seen from the formula that, generally speaking, the similarity between \( b \) and \( c \) is the average of the similarity between the in-degree neighbor nodes of \( b \) and the in-degree neighbor nodes of \( c \). The similarity score between the node pairs calculated by SimRank is symmetric, that is, \( S(b, c) = S(c, b) \). The formula can calculate the structural similarity score between any node pair \( b \) and \( c \) until it converges.

Cluster analysis is a method that automatically divides samples into several groups based on the measurement criteria of the correlation between samples and makes the
samples in the same group similar and different samples in different groups. In this paper, the numerical attributes and category attributes are referred to in the cluster analysis. The calculation formula for calculating the attribute similarity between objects is

\[ \text{Sim}_B(x_{ij}, x_{il}) = e^{-\|x_{ij}B - x_{il}B\|.} \]  

(9)

The sum of the differences of each attribute of the object is

\[ \|x_{ij}B - x_{il}B\| = \sum_{r=1}^{q} |x_{ijr} - x_{ilr}| + \sum_{r=q+1}^{N} \theta(x_{ijr}, x_{ilr}). \]  

(10)

The Attribute-SimRank method can calculate the similarity score and use SASimRank to express it; comprehensively considering the attribute similarity and structural similarity, the formula is expressed as follows:

\[ T_{\text{ASimRank}}(b, c) = (1 - \beta) \cdot T_{\text{attr}}(b, c) + \beta \cdot T_{\text{link}}(b, c). \]  

(11)

Among them,

\[ T_{\text{attr}}(b, c) = e^{-\sum_{r=1}^{n} \theta(x_{br}, x_{cr})}, \]  

(12)

\[ T_{\text{link}}(b, c) = \frac{D}{|U(b)U(c)|} \sum_{r=1}^{|U(b)|} \sum_{j=1}^{|U(c)|} T_{\text{link}}[U_r(b), U_j(c)]. \]  

(13)

For any pair of nodes in \( H \), there will be a SASimRank similarity score. For example, if the number of nodes of a certain type in \( H \) is \( N \), there will be a similarity score of \( N \times N \) pairs of nodes. Similarly, the SASimRank score is symmetric. That is \( S_{\text{ASimRank}}(a, b) = S_{\text{ASimRank}}(a, b) \).

We use Attribute-SimRank to calculate the authenticity and reliability of the similarity, and then, we use clustering objective evaluation indicators to verify the effectiveness of the clustering results, the formula is as follows:

\[ C_g = \frac{\sum_{i=1}^{n} \sum_{x \in C} s(x, a_i)}{\sum_{i \leq j \leq n} s(a_i, a_j)}, \]  

(14)

\[ s(a_i, a_j) = 1 - T_{\text{ASimRank}}(a_i, a_j). \]  

(15)

The above formula is defined as the ratio of compactness within clusters to isolation between clusters. A better clustering result is generally that the tightness within the cluster is manifested by the small distance between objects in the cluster, and the greater isolation between clusters is manifested by the greater distance between the centers of each cluster. The value of \( C_g \) is the smaller, the better the clustering quality.

3.5.4. Measurement Indicators of Association Rules. The degree of support in the database refers to the ratio of the number of transactions that can be included in a transaction to the number of all transactions at the same time, denoted as

\[ \text{Sup}(X \Rightarrow Y) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)}. \]  

(16)

Confidence degree can be used to measure the credibility of association rules, denoted as

\[ \text{Con}(X \Rightarrow Y) = \frac{\text{Sup}(X \cup Y)}{\text{sup}(X)}. \]  

(17)

4. System Construction of e-Commerce Big Data Analysis Platform Based on Embedded System

4.1. Demand Analysis. The target user group of this system is the relevant staff involved in e-commerce marketing activities in the enterprise, including the customer department personnel needing to understand the situation of the customer’s company store/product and show the outstanding results of the company through cases and data when communicating with customers; operations staff needing to understand the operating platform through data reports such as daily, weekly, and monthly reports and adjust and optimize operations, maintenance, and updates through data reports; and planners needing to understand excellent cases to promote inspiration, and at the same time, when insights are obtained, there must be data support to complete the planning plan more accurately; designers daily needing to check various past cases and materials; understand the layout of excellent case designs, etc., through the feedback of data reports, and can better form design output; and data analysts frequently obtaining data, performing statistics and analysis, completing various types of reports, and better assisting or guiding the output of the program through data reports. In daily work, the target user is not entirely one person responsible for a project alone and often require multiple types of personnel to be responsible for multiple customers, multiple platform/store operations, and multiple activities at the same time. The planning of product inner pages, the design of multiple design drafts, and the output of multiple data reports, so it is necessary to use or obtain data/cases/materials with high frequency and frequency.

In the design of e-commerce big data analysis platform, based on past experience, the following factors will be considered: (1) data security, customers generally only disclose the user name and password of the store to a very small number of people within the enterprise, avoiding data to a certain extent of leaks. This has led to the fact that most of the personnel responsible for the actual work in the company do not have the authority to view the data. Even if this part of the data is completely open to the outside world, generally, it can only be obtained and shared by relevant personnel with the authority, and it is shared in many cases. The information is not necessarily what other people want to
obtain. This leads to the lag and inconsistency of information acquisition. (2) Each store corresponds to a user name and password, and the data of the corresponding store can be viewed and obtained by logging in. This leads to the need to log in to different accounts when viewing the data of different stores on different platforms or the same store on different platforms, which causes a certain degree of cumbersome under the condition of ensuring security, and they cannot be done between platforms and stores. (3) The third-party data applications on most platforms, such as Taobao’s "Business Staff," have only relevant indicator data, which is relatively shallow data, and it is impossible to directly obtain reports that are more closely related to the business of the enterprise in the application. As a result, no more supplementary or instructive content can be drawn. (4) Third-party data applications on most platforms, such as Taobao’s "Business Staff," whose data is provided to all digital marketing companies, are universal and have not been expanded in terms of the business characteristics of different companies and customization.

In response to the above problems, this paper designs and implements an e-commerce big data analysis platform based on the Internet of Things technology. The establishment of the system platform involves the customer department personnel, operation personnel, planners, designers, and data analysis personnel in the enterprise, to provide these users with a data management and analysis platform that is different from the third-party applications currently used, to be closer to the business of the enterprise, and to enable users to obtain more valuable information more conveniently.

4.2. Overall Framework of the System. The data analysis system based on the e-commerce platform adopts the B/S architecture and is designed according to the three-tier architecture, including the data access layer, business logic layer and user presentation layer. (1) Data access layer: the access operation layer to the underlying data, which realizes the reading, adding, modifying, and deleting of data. The access interface of the underlying data for the last business logic layer was provided. In this system, data reading and adding are the main operations, while data modification and deletion operations are relatively few. (2) Business logic layer: for the realization of specific functions, the data is processed according to business logic and the results are returned to the presentation layer. In this system, business logic mainly includes four parts: data acquisition, data processing, storage, and data presentation. (3) User presentation layer: this is responsible for interacting with users. In this system, the user presentation layer is responsible for receiving user output, sending requests to the server, and displaying back to the front-end page.

4.3. Overall Function Module Design of the System

4.3.1. Data acquisition. The data source of the system is composed of three parts. Data is obtained from Taobao’s third-party application "Business Staff"; the crawler program crawls the basic information, material information, and buyer comments of the product; the data increase function of the system; increase the product name, links, and terminals; increase the location of monitoring materials; and increase the focus of buyers to search for keywords.

1. Use data fusion technology to obtain data from Taobao third-party applications. It is mainly based on numerical index data, which is mainly used for the call data of four front-end modules of business details, product sales, traffic sources, and marketing effects. Detailed data includes but are not limited to date, shop, terminal, page views, number of visitors, number of old visitors, per capita page views, number of old buyers, number of products paid, and number of buyers paid, payment amount, payment conversion rate, customer unit price, average length of stay per capita, average daily bounce rate, etc.

2. The crawler program obtains the data. The data that web crawlers need to crawl from the e-commerce platform is divided into three categories, the crawling of basic product information, the crawling of material information, and the crawling of buyer comments. The basic commodity information includes date and terminal, which is used to call data from the front-end module of commodity analysis. The material information includes two types, full page and point position, which are used to call data by the buyer’s attention point data processing module

3. Add data to the systems’ own functions. The system itself has designed the function of adding data, which is mainly used to provide users with personalized and customized monitored products, points, and keywords for buyers to search for. With the expansion of business, users can add corresponding information through the system. The background crawler program crawls data based on the added content, adds new data content to the database, and uses it to add functions to the three modules of product analysis, material analysis, and buyer focus. The commodity analysis module adds commodity names, links, and terminal information. The material analysis module adds the link where the monitored material is located. Buyer focus increases buyer focus search keywords

4.3.2. Data processing. Integrate and process the collected data through data fusion technology. Data processing is divided into 9 parts: (1) Raw data processing; the original data obtained from Taobao’s third-party data application has accumulated data from 2015 to 2019 and is stored in multiple tables separately according to the year. Since it is unprocessed source data, there will be attribute columns, abnormal record rows, and blank row records that are not related to this system. Therefore, according to the database design of this system, the original data is reprocessed and stored in the database data. (2) Personalized labeling of a single product: the basic product information crawled by
the crawler program is limited to the data displayed on the product detail page. In order to provide users with a better monitoring product search experience, it is necessary to add personalized labels to the products, including the category, brand, and commodity features. (3) Personalized labeling of material pictures: the material crawled by the crawler program is limited to the image and its title. In order to provide users with a better monitoring product search experience, it is necessary to add personalized tags to the material, including the category, brand, and design features of the material. (4) The construction of special dictionaries in the review field: due to the particularity of the field, the reviews will contain a large number of terms such as category, brand, product name, and special words in the review. Therefore, by obtaining all the categories and brand words on the platform, and Tmall’s word segmentation information on the title, a review field is constructed. Dedicated dictionary, auxiliary comment word segmentation improves the accuracy of word segmentation. (5) The construction of buyer focus model: the buyer attention model is used to vote and classify the attention points of buyers’ comments. Since the comments are published by buyers, there are many expressions of words. It is necessary to use words for the attention points that appear in the comments. Only by classification can you vote on the buyer’s concerns in the comments. This system uses the accumulated review corpus, after word segmentation, manually classifies the word segmentation, and builds a general buyer focus model. (6) Word segmentation and labeling: the purpose of word segmentation is to segment comments in sentence units into comments in word units, which can improve the efficiency of the algorithm. This system combines the well-structured special dictionary for the review field, the Chinese Academy of Sciences NLPIR word segmentation tool, and the common word method based on statistics for word segmentation and labeling. (7) Extraction of candidate feature words: we use the labeling information to extract the content words in each comment to construct the vector representation of the comment. This method reduces the number of words that need to be processed and further improves the efficiency. (8) Identification and statistics of buyers’ concerns: the purpose of voting for buyer concerns is to identify the distribution of different concerns in buyer comments. Combining the feature words extracted after word segmentation, the buyer focus model, and the voting classification algorithm, all buyer comments in this category are processed one by one, and the frequency of each focus category is counted, the number of valid comments in the entire category, and finally calculate the proportion of each category of concern. (9) Check the legality, repeatability and completeness of the data added by the system: due to the added functions of the system, users can add product names, links, and terminal information in the product analysis module; add links to monitor material in the material analysis module; and increase buyer focus search keywords in the buyer focus. After adding information, the web crawler will crawling the corresponding information, so the data needs to be checked before entering the database, including whether the input content is legal content, whether it is duplicated with the existing data in the database, and whether it is used as the primary key the information is not empty.

4.3.3. Data storage. The storage method of system data is divided into two parts. The first is stored in the Sql Server database, including all numeric and text data. The image material is stored in the address of the image server, and the second is stored in the image server. The pictures (.jpg, .png) crawled by the crawler program need to store the picture files in the hard disk of the picture server. Calling a picture is to mobilize the picture corresponding to the address in the picture server by obtaining the picture address in the Sql Server database.

4.3.4. Data presentation. The data presentation is divided into two parts: the presentation of the data graph (line graph and bar graph) and the presentation of the detailed table of data content. The development and design mode adopted by this system is B/S (Browser/Server, browser/server) mode. Under this structure, the user interface is realized directly through the client browser. The client sends a service request to the server, including the data request for the database. The server responds to the client request and transmits the response result to the client browser. The system front-end module is divided into 7 functional modules, which are business details, product sales, traffic sources, marketing effects, product analysis, material analysis, and buyer concerns. There are 4 modules that contain discount graph display: business details, product sales, traffic sources, and marketing effects; modules that contain bar graph display are buyers’ concerns; those that contain detailed data table display. There are 6 modules-business details, product sales, traffic sources, marketing effects, product analysis, and material analysis. Figure 1 shows the platform structure of e-commerce big data system.

5. Application Research Analysis of e-Commerce Big Data Platform Based on Data Fusion Technology and Internet of Things Technology

5.1. Background Analysis of e-Commerce. As shown in Table 2, the e-commerce market continues to increase in 2019, and the penetration rate of global online shopping users is high (the penetration rate of online shopping users refers to the ratio of the number of online shopping users to the number of Internet users in a certain period of time). At present, the penetration rate of online shopping users in China is 56%, ranking 4th in the world; and due to China’s large population base, online shopping users have reached 374 million, ranking first in the world. Combining the penetration rate of online shopping and the number of netizens, China’s e-commerce market still has a lot of room for development.

Judging from the ranking of e-commerce platforms in Figure 2, the top three e-commerce platforms are Taobao, JD, and Tmall. From the ranking status in recent years, it can be concluded that China’s e-commerce industry has entered a “duopoly.” In the era, the two major forces of Ali
(Taobao + Tmall) and http://JD.com divided the market. In the platform model, e-commerce platforms can be divided into B2C platforms, independent shopping malls, C2C platforms, O2O platforms, etc., among which B2C platforms account for the largest proportion. On the one hand, most companies choose the B2C platform as their first choice for online marketing platforms, and because there are so many B2C platforms on the market, in order to expand their market share, companies generally deploy their online marketing business to multiple platforms. On the other hand, the number of users who use mobile clients for online shopping is increasing, and there is even a trend to catch up with the PC side. Therefore, companies also take into account the business of the PC side and the mobile side.

The biggest advantage of e-commerce over traditional retail is that all sales can be controlled and changed through data. Operators check the quantified marketing situation through indicator data, including page views, number of visitors, transaction amount, transaction conversion rate, bounce rate, and number of shopping carts, through the user’s shopping list, favorite list, and shopping cart list, and so on, to analyze the user profile of the target user; buyer’s rating and comment data, and get the buyer’s feedback on the product. By monitoring data changes and further analyzing the data, you can effectively understand the marketing status of stores, products, and activities and analyze the concept of data, overall sales, user portraits, and the results of activities. Others improved valuable conclusions and help operators improve store operations and increase profits. Therefore, through the use of e-commerce to provide services to consumers and traders, we obtain data and use data to further provide services to consumers and

| Country      | Online shopping penetration rate | Online shopping user scale |
|--------------|----------------------------------|---------------------------|
| China        | 3.75                             | 56.01                     |
| United Kingdom | 0.45                           | 92.54                     |
| Russia       | 0.88                             | 61.34                     |
| United States | 1.79                           | 71.68                     |
| Japan        | 0.66                             | 45.58                     |
| Germany      | 1.48                             | 55.42                     |
| France       | 0.72                             | 56.28                     |
| Korea        | 1.69                             | 42.29                     |
| Canada       | 0.88                             | 43.88                     |

Table 2: The size of Internet users in various countries and their penetration rate distribution.

Figure 1: e-Commerce big data system platform.
Figure 2: 2019 e-commerce platform rankings.

![E-commerce platform rankings](image)

Figure 3: Changes in e-commerce research hotspots.

![Research hotspots changes](image)

Table 3: Detection performance under the influence of different impurities.

| Performance Testing   | All traffic | Detect event Number of cycles $= 40$, impurity influence $= 0.05$ | Detect event Number of cycles $= 40$, impurity influence $= 0.01$ | Detect event Number of cycles $= 40$, impurity influence $= 0.001$ |
|-----------------------|-------------|------------------------------------------------------------------|------------------------------------------------------------------|------------------------------------------------------------------|
| Number of events      | 7648823926  | 1382485298                                                       | 949241947                                                        | 680325591                                                        |
| BOT events            | 2394884311  | 1139873281                                                       | 714024972                                                        | 461692487                                                        |
| Number of events for logged-in users | 4057881285 | 27319423                                                        | 17423987                                                        | 15145264                                                        |
| Number of new detection events | —      | 242612018                                                       | 235226976                                                        | 218634104                                                        |
| Login rate            | —           | 2.01%                                                            | 1.81%                                                            | 2.20%                                                            |
| New detection rate    | —           | 10.11%                                                           | 9.82%                                                            | 9.10%                                                            |
traders in this cycle. Data analysis in the field of e-commerce has become particularly important.

It can be seen from Figure 3 that the research hotspots on e-commerce have been in the process of change. In the past, there was some concern about payment security performance. The attention rate reached 20.01% in 2015. However, with the development of science and technology, security with the improvement of technical means, the attention to safe payment has dropped to 7.73%; however, the attention to the database has not decreased but has increased. In the past, everyone did not understand the role of the database, but now everyone has a better understanding of the database. The database can well grasp the consumer’s consumption habits and consumption psychology and analyze the consumption data required by the merchants. Therefore, it has also increased the attention to the database. In 2019, it reached 19.24% of the attention. Therefore, it is necessary and realistic for the research of e-commerce big data analysis platform.

5.2. Performance Analysis of e-Commerce Big Data Platform.

We divided the detection events into different impurity situations and then analyzed the e-commerce big data analysis platform system. The results are shown in Table 3. It can

| Table 4: Number of old and new rules detected. |
|-----------------------------------------------|
| **Quantity** | **Number of events** | **BOT events** | **Number of events for logged-in users** |
| All traffic | 7648823926 | 2394884311 | 4057881285 |
| Number of old rules detected | 479534246 | 301234764 | 15677342 |
| Number of existing rule detections | 1382485298 | 1139874281 | 27319423 |
| Increase the number of detections | 902951055 | 838638527 | 11643082 |
| Add detection ratio | — | 279% | — |

| Table 5: System performance test results. |
|------------------------------------------|
| **The amount of data** | **Days** | **1 day** | **5 days** | **10 days** |
| **Rows** | **2916238268** | **16514201586** | **3476679874** |
| Data extraction | | | | |
| Real time | 10'39 | 47'13 | 53'03 |
| CPU time | 7d2h | 30d23h | 62d19h |
| Data cleaning | | | | |
| Real time | 21'45 | 1h27'34 | 3h3'16 |
| Data cleaning | | | | |
| Real time | 14d2h | 68d4h | 117d11h |
| Supply and demand analysis | | | | |
| Real time | 15'23 | 18'54 | 20'10 |
| Total time | | | | |
| Real time | 37'47 | 2h33'1 | 4h14'29 |
| Total time | | | | |
| CPU time | 22d21h | 101d5h | 182d19h |

**Figure 4:** Experimental results of the two methods.

- Number of valid comments manually identified
- Effective review recall rate of word matching model method
- Effective review recall rate of model matching review method
be seen from the table that the lowest false alarm rate and the highest BOT event detection rate can be achieved after the thresholds of various combinations of parameters are combined.

Table 4 shows the comparison of the number of detections under the new and old rules. When the number of cycles = 40, the impact of impurities = 0.05, and the cycle length = 5, the experimental results are obtained by comparing the old rules with the current rules. It can be seen that the total number of BOT events for all traffic has increased, and the detection ratio has increased by 279%.

The system performance test data is actual user behavior data, and the data volume of 1 day, 5 days, and 10 days are used for testing. As shown in Table 5, the amount of data in 1 day is about 2.9 billion rows, about 16.5 billion rows in 5 days, and 34.7 billion rows. The system performance tests of data extraction, data cleaning, supply and demand analysis and mining are performed on these data, and the results are the test results shown in Table 5. The results are divided into actual time and CPU time. Actual time refers to the actual time used by the test result, and CPU time refers to the CPU time consumed by the system.

It can be seen from Figure 4 that the effective review recall rate of the word matching model method is more than 67%, and the effective review recall rate of the model matching review method is more than 72%. Among the two methods, “beer” has the worst recall rate, while “Olay” has the best recall rate. In terms of overall performance, the method in this article has a higher recall rate than previous research methods, with an average of 5% higher.

The results of the recognition accuracy of buyers’ attention points from each category review are shown in Figure 5. Overall, the trend of the recognition accuracy of the two methods in each category remains the same, but there are also some categories with large fluctuations, including Budweiser’s “gift/trial/reward/benefits,” “smell,” “appearance,” “expiration date/generation date,” “return,” Pampers’s “color,” “origin/manufacturer”; gum care toothpaste, the “purchase channel,” “manufacturer/origin,” and “new product/style” of the shaver’s “specification/content/concentration/capacity/area/weight/quantity.”

As shown in Figure 5, the recognition accuracy of each category of the “Olay” two methods is above 45%. The top 5 of the categories with the best performance of the word matching model method are texture, expiration date/prodution date, taste, manufacturer/origin, and true and false; the TOP5 of the categories that perform best in the model matching comment methods are texture, expiration date, taste, manufacturer, and true and false.

Table 6: Comparison of user search and demand analysis.

| User raw search data          | Demand analysis results |
|------------------------------|------------------------|
| Cybershot Sony DSC-F707 charger | Sony battery charger   |
| 15 204 Mercedes man united kit | Mercedes kit           |
| Weather york ambient mercury  | Air sensor             |
| Yamaha 50 00 mm trt           | Yamaha piston trt      |
| Flex iPhone 5s case           | iPhone 5s case         |
| Aquarium filter pharmaceuticel| Aquarium filter        |
| Calor trocador                | Exchanger heat         |
| Dome e30 light                | bmw light              |
| Eay62810801                   | Lg power supply        |
| Kei yoki                      | African violet leaf    |

Figure 5: Comparison of the accuracy of each classification of the two identification methods of buyers’ concerns.
5.3. e-Commerce Supply and Demand Analysis. The results of the demand analysis are tested by sampling, and the mining of user search data meets the expected results. The results of random extraction are shown in Table 6. It can be seen that through the mining of user search behaviors, the supply and demand analysis module can draw conclusions that are conducive to merging and understanding result.

5.4. Sales Distribution and Similarity Distribution. The regression coefficient beta means that when the other variables remain unchanged, for each additional unit of the independent variable, the dependent variable “increases” by $b$ units on average. When $b > 0$, it means that the dependent variable and the independent variable are positively correlated. When $b < 0$, it means that the dependent variable and the independent variable are negatively correlated. The absolute value of $b$ describes the degree of correlation.

Figure 6 lists the coefficient values of each variable when the linear regression model is applied to the memory stick data set. For buyers who purchase new memory products, the cumulative consumption amount has the greatest impact on the purchase of different grades of products by new product buyers. The correlation coefficient value is 0.0619133. Based on this, we know that the more customers who...
purchase new memory products have historically spent on the website, the higher the price of their memory purchases, that is, the higher the grade.

From the numbers shown in Figures 7 and 8, the clustering coefficients of each network are relatively large and between 0.7 and 0.8. It can be seen that the visualization graph constructed by the time series of commodity sales has a higher clustering coefficient. On the whole, the average clustering coefficient (0.75593) of the 10 products in the online store is smaller than the average clustering coefficient of the 8 products in the store (0.811). From the average network diameter and average path length in the figure, the network diameter is between 5 and 7, and the average path length is also relatively small. However, the average network diameter and average path length (3.86, 7.7) of online store products are larger than those of the store product network (2.71, 5.75).

In summary, it can be inferred that the network constructed by the time series of merchandise sales in online stores has a smaller world effect than the network constructed by the time series of merchandise sales in stores. From the network degree correlation column in the figure, we can see that the Pearson correlation coefficient $r$ value of the store product network is less than 0, and the correlation coefficient value of the online store product network is positive or negative. From this, it can be inferred that the
network constructed by the time series of merchandise sales in stores is negatively correlated, and nodes with large degrees in the network tend to be connected with nodes with small degrees. The degree correlation of the network constructed by the time series of online store product sales is related to the specific product, and the degree correlation is not consistent.

It can be seen from Figure 9 that when user similarity calculation adopts the method Attribute-SimRank in this paper, the average hit rate of product recommendation for cluster center users is in most cases higher than that of SimRank and Attribute methods. When the value of $N$ changes from 10 to 100, only when $N = 30$ and $N = 90$, the Avg_HitRate (10.7% and 5.08%) corresponding to the Attribute-SimRank method is slightly lower than the Attr_HitRate (12% and 12%) corresponding to the Attribute method. In general, Attribute-SimRank increased the average hit rate of SimRank by about 2.6 percentage points, and Attribute-SimRank increased the average hit rate of Attribute by about 1.85 percentage points. Therefore, the results of this experiment further indirectly and effectively evaluate the similarity calculation method proposed in this paper. Because in the recommendation application, if the target user’s similar interest user acquisition is more accurate, the recommendation result is more meaningful, and the accuracy of finding users with similar interests depends on whether the similarity acquisition between similar users is accurate.

6. Conclusion

This paper mainly studies the construction and application research of e-commerce big data analysis platform based on data fusion technology and Internet of Things technology. This article uses an embedded system to build an e-commerce data analysis platform, solves the current problems of e-commerce from the management and technical level, and processes massive amounts of data reasonably to provide reliable, fast and convenient analysis conclusions. The innovation of this article is the use of a combination of quantitative analysis and qualitative analysis, and a combination of theoretical analysis and empirical research to build a better e-commerce data analysis platform. The shortcomings of this article are that the selection of experimental standards and experimental objects is random and not representative, and the experimental results need to be viewed more comprehensively. With the development of the Internet of Things in the future, e-commerce companies must have more powerful data computing and processing capabilities to better meet the needs of users and provide users with high-quality and high-level services and experience.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no potential conflict of interest in this study.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (72161020), Jiangxi Social Science Foundation Project (21GL44), and Youth Fund Project of Humanities and Social Sciences in Colleges and Universities of Jiangxi Province (GL19223).

References

[1] J. Mero, "The effects of two-way communication and chat service usage on consumer attitudes in the e-commerce retailing sector," Electronic Markets, vol. 28, no. 2, pp. 205–217, 2018.
[2] C. Maxim, Z. Sherali, and B. Zubair, "Internet of Things forensics: The need, process models, and open issues," IT Professional, vol. 20, no. 3, pp. 40–49, 2018.
[3] M. A. Sayal, M. H. Alameady, and S. A. Albermany, "The use of SSL and TLS protocols in providing a secure environment for e-commerce sites," Webology, vol. 17, no. 2, pp. 503–523, 2020.
[4] M. A. Ziadat, M. Al-Majali, and A. A. Muala, "Factors affecting university student’s attitudes toward E-commerce: case of Mu’tah University," International Journal of Marketing Studies, vol. 5, no. 5, pp. 88–93, 2018.
[5] B. Palese and A. Usai, "The relative importance of service quality dimensions in E-commerce experiences," International Journal of Information Management, vol. 40, no. JUN., pp. 132–140, 2018.
[6] L. Vanneschi, D. M. Horn, M. Castelli, and A. Popović, "An artificial intelligence system for predicting customer default in e-commerce," Expert Systems with Applications, vol. 104, no. aug., pp. 1–21, 2018.
[7] Y. Qi and P. Tapio, "Weak signals and wild cards leading to transformative disruption: a consumer Delphi study on the future of e-commerce in China," World Futures Review, vol. 10, no. 1, pp. 54–82, 2018.
[8] F. Arnold, I. Cardenas, K. Sörensen, and W. Dewulf, "Simulation of B2C e-commerce distribution in Antwerp using cargo bikes and delivery points," European Transport Research Review, vol. 10, no. 1, pp. 1–13, 2018.
[9] L. Jia, G. Xue, Y. Fu, and L. Xu, "Factors affecting consumers’ acceptance of e-commerce consumer credit service," International Journal of Information Management, vol. 40, no. JUN., pp. 103–110, 2018.
[10] J. Nan, J. Tao, and J. Du, "Personalized recommendation based on customer preference mining and sentiment assessment from a Chinese e-commerce website," Electronic Commerce Research, vol. 18, no. 1, pp. 1–21, 2018.
[11] S. Karim and E. Gide, "Barriers to adopting E-commerce with small to mid-sized enterprises-SMEs in developed countries: an exploratory study in Australia," Global Journal of Information Technology Emerging Technologies, vol. 8, no. 2, pp. 43–54, 2018.
[12] M. Chen, D. Wang, and Z. Hu, "Considerations of constructing quality, health and safety management system for agricultural products sold via e-commerce," International Journal of Agricultural & Biological Engineering, vol. 11, no. 1, pp. 31–39, 2018.
[13] A. J. Lin, E. Y. Li, and S. Y. Lee, "Dysfunctional customer behavior in cross-border E-commerce: a justice-ffect-
behavior model," Journal of Electronic Commerce Research, vol. 19, no. 1, pp. 36–54, 2018.

[14] S. Ackermann, I. Adams, and N. Gindele, "The role of e-commerce in the purchase of agricultural input materials," Landtechnik, vol. 73, no. 1, pp. 10–19, 2018.

[15] L. F. Chen, "Green certification, e-commerce, and low-carbon economy for international tourist hotels," Environmental Science and Pollution Research, vol. 26, no. 18, pp. 17965–17973, 2019.

[16] F. A. Nantembelele and S. Gopal, "Assessing the challenges to e-commerce adoption in Tanzania," Global Business & Organizational Excellence, vol. 37, no. 3, pp. 43–50, 2018.

[17] K. Kardoyo, N. Farliana, and M. Feriady, "Pelatihan Pemasaran Hasil Pertanian dan Perkebunan Lembaga Permasyarakat Terbuka Kendal Melalui E-commerce Berbasis web content management system," E-Dimas Jurnal Pengabdian kepada Masyarakat, vol. 12, no. 1, pp. 81–89, 2021.

[18] D. T. L. Jahalia and L. Naﬁati, "Niat Menggunakan E-commerce Dengan technology acceptance model (TAM) Pada UMKM," Jurnal Bisnis dan Kewirausahaan, vol. 17, no. 1, pp. 24–31, 2021.

[19] A. Lorenc and A. Burinskiene, "Improve the orders picking in e-commerce by using WMS data and big data analysis," FME Transactions, vol. 49, no. 1, pp. 233–243, 2021.

[20] D. Santi, E. Magnani, M. Michelangeli et al., "Seasonal variation of semen parameters correlates with environmental temperature and air pollution: a big data analysis over 6 years," Environmental Pollution, vol. 235, no. apr., pp. 806–813, 2018.

[21] J. H. Bae and E. A. Burm, "Understanding of rehabilitation nursing based on academic big data analysis," The Korean Journal of Rehabilitation Nursing, vol. 23, no. 2, pp. 105–111, 2020.

[22] P. Shukla and B. Radadiya, "Study of artificial intelligence techniques to permit the automated big data analysis in egovernance projects to promote digitalization," International Journal of Scientific Research, vol. 8, no. 4, pp. 1–3, 2019.

[23] R. Massobrio, S. Nesmachnow, A. Tchernykh, A. Avetisyan, and G. Radchenko, "Towards a cloud computing paradigm for big data analysis in smart cities," Programming and Computer Software, vol. 44, no. 3, pp. 181–189, 2018.