

The Application of Integrating Data Mining and IoT Management Technology in Enterprise Supply Chain Information Management

Ling Gong

Accounting Institute, Chongqing College of Finance and Economics, Chongqing 402160, China

E-mail: cqyc20000219@126.com

Keywords: data mining, apriori algorithm, internet of things management technology, enterprise supply chain information management, radio frequency identification

Received: March 19, 2024

The increasing use of the Internet of Things presents both challenges and opportunities for managing enterprise supply chains. This study proposes a method of integrating data mining and Internet of Things management technology for enterprise supply chain information management. By collecting and analyzing numerous data generated by Internet of Things devices, real-time and accurate information on supply chain management can be provided, thereby helping enterprises achieve supply chain optimization and collaborative management. This study introduces the architecture and implementation method of a supply chain information management system that integrates data mining and Internet of Things management, and verifies the effectiveness of this method through case analysis. The experiment showed that the average accuracy of the enterprise supply chain information management system was 93.14%, the average sensitivity was 91.05%, and the average specificity was 91.89%. The average error rate was 7.06%, the average delay time was 0.84s, and the average accuracy was 93.21. This indicates that the performance of the enterprise's supply chain information management data system is excellent, and this method has important application value in improving supply chain efficiency and accuracy.

Povzetek: Raziskava združuje rudarjenje podatkov in tehnologijo upravljanja interneta stvari (IoT) za upravljanje informacij o oskrbovalni verigi podjetja. Z analizo podatkov, ki jih ustvarjajo naprave interneta stvari, sistem zagotavlja realnočasovne in natančne informacije o upravljanju oskrbovalne verige.

1 Introduction

With the intensification of global market competition, Enterprise Supply Chain Information Management (ESCIM) has become an important factor in enterprise competitiveness. However, ordinary Supply Chain Information Management (SC-IM) methods suffer from issues such as inaccurate information and delays [1]. The rise of Internet of Things (IoT) technology has brought new opportunities to SC-IM. The ordinary SC-IM method mainly relies on manual data collection and processing, which has problems such as inaccurate information and delay. IoT technology can monitor and collect real-time data from various links in the supply chain through sensors, devices, and other means, achieving real-time monitoring of the supply chain. Data Mining Techniques (DMT) can analyze these data and discover hidden patterns and correlations [2-3]. Applying DMT to IoT data can provide enterprises with more accurate and real-time supply chain information, help them make more reasonable decisions, reduce inventory costs, and improve supply chain efficiency. IoT devices can monitor and collect information on various aspects of the supply chain in real-time, and analyze this data through DMT to provide real-time and accurate supply chain information

[4-5]. Therefore, the integration and application of Data Mining and IoT Management (DM-IoTm) technology in ESCIM is of great significance. Integrating DM-IoTm technology can help enterprises better manage supply chain information, improve supply chain efficiency and collaboration. The application of this technology in ESCIM has important contributions and significance, helping to improve the supply chain management level of enterprises, reduce costs, enhance competitiveness, and achieve sustainable development. The context is segmented into four parts. Part 1 is a literature review that discusses and analyzes the current research status of data mining, IoT management technology, and ESCIM both domestically and internationally. Part 2 proposes an ESCIM system that integrates data mining and IoT. Part 3 verifies the effectiveness and performance of the system through experiments. Part 4 summarizes the research findings.

2 Related works

In recent years, data mining and IoT technology have developed rapidly and have been widely applied in various fields. One method to address the high turnover rate in the education technology market is to predict the turnover rate. Kiguchi et al. proposed a customer churn

prediction model based on decision trees and random forest data mining models. The proposed method was effective in determining and predicting customer churn in digital game learning [6]. In response to the widespread application of DMT in the financial services field, Plotnikova et al. proposed a standardized process for enterprise management data mining projects based on logistic regression. This method could meet the data analysis needs of financial enterprises [7]. Ganesh and Kalpana proposed an artificial intelligence-based model for analyzing social media data, which identifies and evaluates risk factors as the most important stage in supply chain risk management. The important role of data analysis in achieving accurate decision-making provided valuable insights into contemporary sustainable development issues [8]. To assist university students, improve their physical health, He et al. proposed a college student physical exercise and health management system based on the IoT era. The IoT-based physical exercise framework could strengthen the students' physical status [9]. Jiang proposed an engineering cost accounting system that combines chaotic data processing methods with IoT technology. This system had good data processing performance [10]. To address the issues of energy drift and deformation caused by environmental temperature and internal structure of prefabricated buildings, as well as high construction energy consumption costs, Wang and Jiang designed an energy management and control system for prefabricated buildings built on IoT. The root means square error convergence of energy management in this system had been proven to be good [11].

The rapid growth of IoT, big data, and cloud computing has presented ESCIM with numerous challenges and opportunities. Understanding and studying the barriers to adopting an environmental impact assessment system in the supply chain is crucial for better management of the supply chain. Deepu and Ravi proposed a grey-based

decision-making experiment and evaluation laboratory method. This study provided decision-makers with a key framework for adopting barriers in environmental impact information systems to achieve effective environmental impact information systems and better management [12]. The current cost management model was still stuck in ordinary management methods, lacking more intelligent big data analysis methods. Therefore, Mao and Chen proposed a series of practical operational methods for exploring supply chain culture dissemination enterprises using big data technology. This method could help e-commerce enterprises reduce supply chain management costs and obtain higher profit margins [13]. Due to the long-term operation of the enterprise resource planning system, it had become a platform and guarantee for completing management processes. Li and Wu proposed an enterprise logistics information management system. This system had essential application value in this system of supply chain enterprises, which can effectively improve customer satisfaction value by 86.7% [14]. Due to the negative correlation between the scale of risk transmission and operational robustness and flexibility, market information completeness, and immunity rate, Wang et al. proposed using an epidemic model to study supply chain risk transmission. This study was of great theoretical and practical significance for this field [15]. In a dual channel supply chain, manufacturers sold products through their online channels and offline retailers at optimal retail prices and service levels. In response to this issue, Yang et al. proposed incorporating the reference price effect into the hotel's utility function to determine the competitive relationship between retail prices and service levels. Numerical examples have verified the feasibility of the theoretical results [16].

Table 1 is a summary of the main findings and research limitations of the review literature.

Table 1: Summary of the main findings and limitations of the review literature

Author	Key findings	Limitations
Kiguchi et al. [6]	Decision trees and random forest models are effective for customer churn prediction in digital game learning	There is no broader application involved
Plotnikova et al. [7]	Logistic regression model can meet the data analysis needs of financial enterprises only in the field of financial services	No other industries are considered
Ganesh et al. [8]	The value of AI-based social media data analytics for supply chain risk management is limited to social media data	Not all risk factors have been fully assessed
He et al. [9]	IoT technology can improve the physical health level of college students	The broader application of health management is not considered
Jiang et al. [10]	The engineering cost accounting system combining chaotic data processing and IoT technology has a good effect	Other aspects of supply chain management are not covered
Wang et al. [11]	The IoT technology has a good effect on the energy management of prefabricated buildings	Broader supply chain management issues are not

		covered
Deepu et al. [12]	Provide a structural framework for barriers to the adoption of environmental impact information systems	Focus only on the environmental impact assessment system
Mao et al. [13]	Big data technology can reduce supply chain management costs	Slow data processing
Li et al. [14]	ERP based logistics information management system can improve customer satisfaction	A data security problem exists
Wang et al. [15]	Epidemic model has theoretical and practical significance for supply chain risk management	Lack of practical application cases
Yang et al. [16]	The competitive relationship between retail price and service level can be determined by considering the reference price effect	Weak privacy protection ability

Table 1 summarizes the main findings of the literature and highlights the limitations of each study, demonstrating the need for further research within the existing body of knowledge. The table shows that DM-IoTm technology has significant potential for application in ESCIM. However, this type of ESCIM also faces some challenges, such as data security, data processing, and privacy protection. Therefore, this study further integrates DM-IoTm technology and explores its application in ESCIM in depth.

3 Construction of ESCIM system integrating data mining and IoT

This study is based on the Apriori algorithm to mine ESCIM data, manage supply chain information through IoT technology, integrate DM-IoTm technology, and construct an ESCIM system to achieve more efficient SC-IM.

3.1 ESCIM data mining based on Apriori algorithm

In enterprise supply chain management, DMT can help companies discover hidden patterns, patterns, and correlations in the supply chain. The Apriori algorithm is a commonly used Association Rule Mining (ARM) algorithm that can discover Frequent Itemsets (FI) and Association Rules (AR) from abundant transaction data. Therefore, applying the Apriori algorithm to ESCIM can help enterprises discover correlations in the supply chain, guide enterprise management and decision-making. ARM is a DMT used to discover meaningful associations in datasets [17-18]. The strength of AR can be measured by two metrics: support and confidence. Support represents the percentage of transactions that contain itemsets X and Y. The confidence level represents the transaction ratios that include itemset X and also include itemset Y. By setting thresholds for support and confidence, AR with sufficient strength can be filtered out. The principle of ARM is Figure 1.

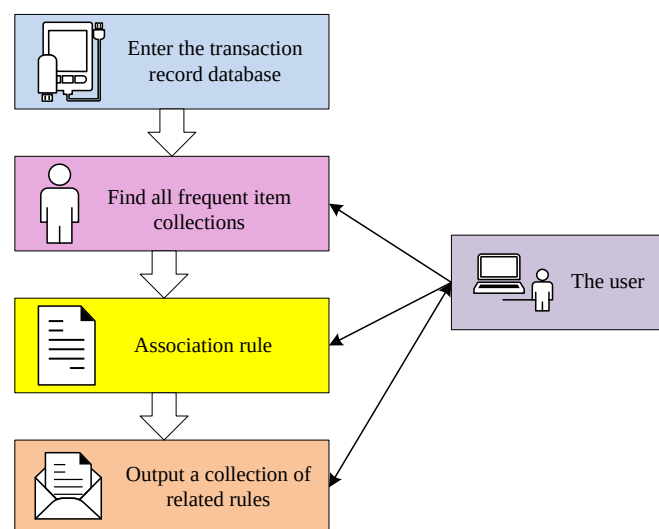


Figure 1: Principles of association rules mining

The common algorithms of ARM include Apriori and frequent pattern growth algorithms. The Apriori mainly has two stages. Stage 1 is to generate Candidate Itemsets (CI), and the second stage is to find FI. In the Stage 1, the Apriori algorithm generates CI through a layer-by-layer search method. Firstly, the set of FI identified is denoted as L1. Then, L1 is utilized to generate a candidate set, and the FI identified is denoted as L2, continuing this process until FI cannot be found. In the Stage 2, the Apriori algorithm scans the database and calculates the support of each CI to identify FI. The Apriori algorithm is a method based on FI, which finds FI and AR by generating CI and calculating support. After mining out all the FI from database D, it is relatively easy to gain the corresponding AR. That is, to generate strong AR that meet the minimum support and confidence. The support calculation of AR is equation (1).

$$\text{Support}(A \rightarrow B) = P(A, B) \quad (1)$$

The confidence calculation of AR is equation (2).

$$\begin{aligned} \text{Confidence}(A \rightarrow B) &= P(A|B) \\ &= \frac{\text{sup_port_count}(A \cup B)}{\text{sup_port_count}(A)} \end{aligned} \quad (2)$$

In equation (2), $\text{sup_port_count}(A \cup B)$ represents the quantity of transaction records containing itemset $A \cup B$. $\text{sup_port_count}(A \cup B)$ represents the amount of transaction records with itemset A . Equation (3) shows

that the degree of improvement in an AR is determined by the confidence level ratio of the antecedent to the support level of the consequent. This ratio represents the probability of both containing A and the probability of A occurring as a whole under the condition of B occurring.

$$\begin{aligned} \text{Lift}(A \rightarrow B) &= P(A|B) / P(A) \\ &= \frac{\text{Confidence}(A \rightarrow B)}{P(B)} \end{aligned} \quad (3)$$

The conditional probability here is calculated using the support frequency of the itemset. Each FI l will generate all non-empty subsets of l . All non-empty subsets s of l have output rules, as shown in equation (4).

$$\frac{\text{sup_port_count}(l)}{\text{sup_port_count}(s)} \geq \text{min_conf} \quad (4)$$

In equation (4), min_conf is the min confidence threshold. If the non empty subset s satisfies equation (4), then output rule $s(l-s)$. The mining of AR is based on the support and confidence of the itemset, as displayed in Figure 2.

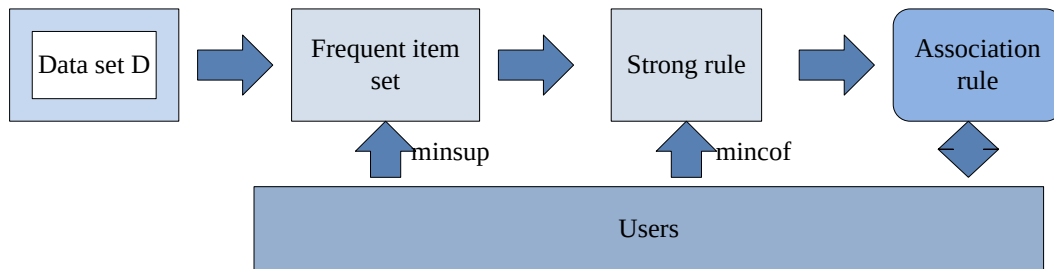


Figure 2: Association rule and output rule

In Figure 2, dataset D is the input data. The minimum support is set to obtain the frequent item set, and then to proceed to the next step. Based on the results produced in the last step and confidence setting, strong AR that meets the requirements is inferred and aggregated for validation, completing the mining process. In this process, various parameters can be set according to actual needs to guide the mining process, and the values of both can be adjusted to achieve user satisfaction. Firstly, by scanning the entire database, the first candidate obtained is the set of search results. The next step is to search for frequent

$(k+1)$ itemset $L_{(k+1)}$: first connect the frequency k itemset to itself, generate a candidate $(k+1)$ itemset $C_{(k+1)}$, and sort each item. If the previous $(k-1)$ projects are the same, then the project set self connects as shown in equation (5).

$$\begin{aligned} (I_1[1] = I_2[1] \wedge I_1[2] = I_2[2] \wedge \dots \wedge (I_1[k-1] \\ = (I_2[k-1] \wedge (I_1[k]) < I_2[k]) \end{aligned} \quad (5)$$

In equation (5), I_1, I_2 are the set of L_k , so I_1, I_2 can be connected. The Apriori process is Figure 3.

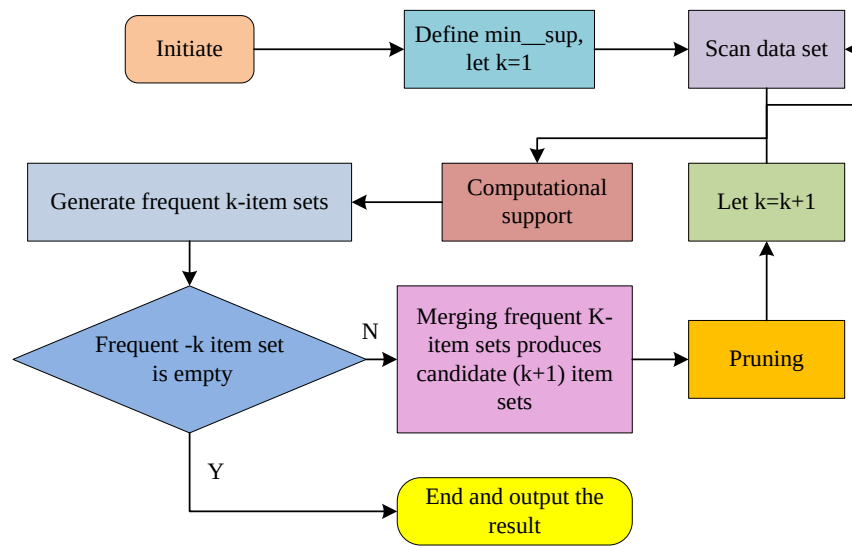


Figure 3: Process of the Apriori algorithm

The Apriori is an ARM algorithm used to discover FI in data, generate AR based on FI, and calculate their support. According to the support threshold, ARs are filtered, and FI and AR are output. The Apriori algorithm requires multiple scans of the dataset and requires lots of computation, but it can effectively discover FI and AR in the data.

3.2 ESCIM system based on IoT management technology

The integration of the IoT technology in the ESCIM system has brought unprecedented opportunities for enterprises, enabling every link of the supply chain to obtain real-time monitoring and data transmission [19]. The introduction of this technology not only enhances the transparency of the supply chain, but also greatly improves its efficiency and synergy [20]. The ESCIM based on the IoT management technology is shown in Figure 4.

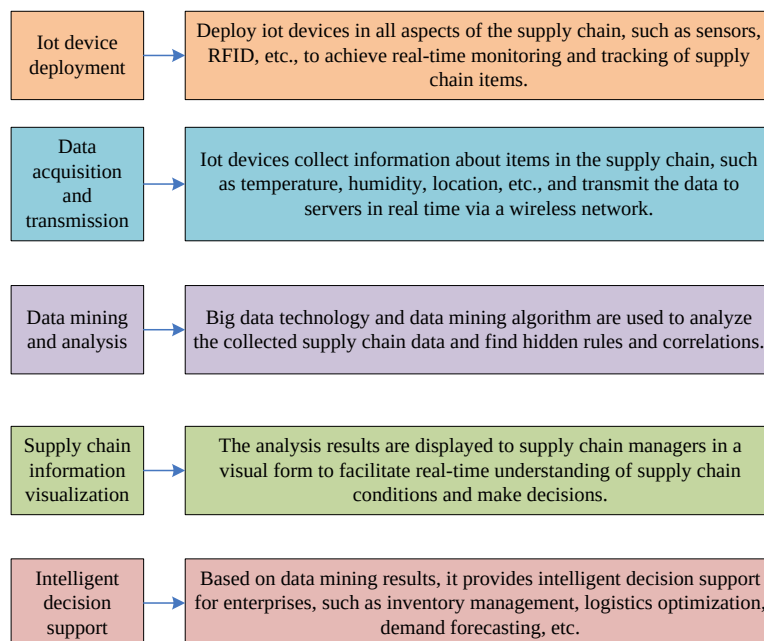


Figure 4: Enterprise supply chain information management

In an ESCIM system, IoT devices, such as Radio Frequency Identification (RFID) readers and sensors, constantly collect data, including crucial information such as the product's location, temperature, humidity, and vibration. These data streams are first aggregated into an intermediate data acquisition layer, and then subjected to specific data cleaning and preprocessing processes to remove noise and outliers in preparation for subsequent data mining and analysis. The architecture of the ESCIM system has become more intelligent and modular with the help of the IoT technology. The whole system can be roughly divided into three modules: perception layer, network layer and application layer. The sensing layer is mainly composed of various IoT devices, such as RFID tags, sensors, etc., which are responsible for collecting various data in the supply chain in real time. The network layer is mainly responsible for data transmission. The data collected by the perception layer is transmitted to the data center or cloud platform through various communication technologies. The application layer is the core of the ESCIM system, which receives data from the network layer and provides valuable insights and decision support to the enterprise through data mining, analysis

and visualization tools. RFID technology in the sensing layer is a technology that uses radio signals to automatically identify target objects and read relevant data. RFID has the characteristics of long reading distance, strong penetration ability, anti-interference, high efficiency, and numerous information. It can recognize a single specific object and process multiple labels at the same time. The anti-collision algorithm for RFID tags is mainly used to solve the conflict problem that occurs when multiple RFID tags send signals simultaneously. This study proposes an anti-collision algorithm based on label parallel recognition technology, which combines Pseudo ID Logistic Code (PILD) and Deterministic Finite State Automata (DFSA). This algorithm combines pseudo-ID code grouping and tag parallel identification technology to enhance the throughput of RFID tag identification transmission. As a result, it improves the transmission performance of RFID multi-tag identification systems while occupying the same transmission band. It effectively reduces the probability of a single label starving to death due to multiple collisions. The steps of the PILD algorithm are shown in Figure 5.

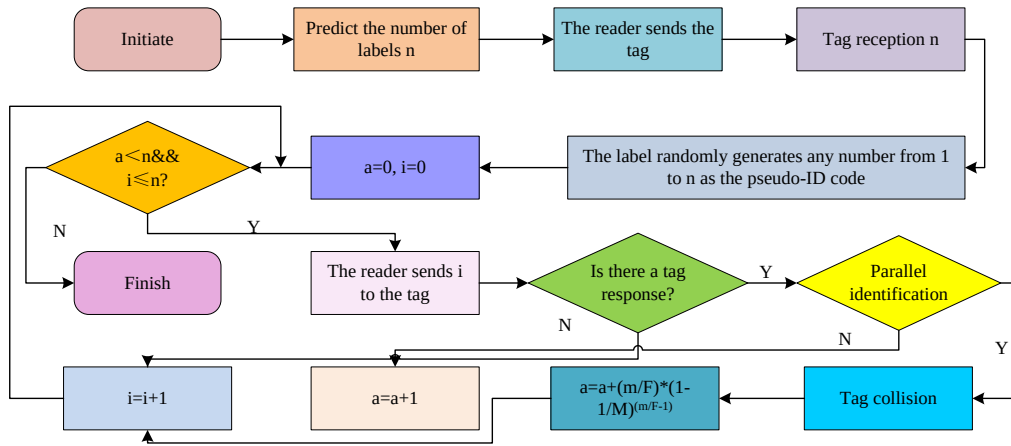


Figure 5: Steps of PILD algorithm

Assuming the reader generates a series of pseudo labels based on the estimated labels, one of which is randomly selected as its own recognition flag within the range of values. The possibility of selecting a pseudo-ID code simultaneously by m markers is equation (6).

$$P(L, n, m) = C_n^m \cdot \left(\frac{1}{L}\right)^m \cdot \left(1 - \frac{1}{L}\right)^{n-m} \quad (6)$$

In equation (6), L represents the number of pseudo-ID codes generated by the reader. n represents the total number of tags to be identified within the recognition range. m is the number of tags selected for the current pseudo-ID code. When the ratio of the expected and total number of pseudo-ID codes selected for a single label is taken to the limit, the result is calculated by equation (7).

$$P(L, n, m) = C_n^m \cdot \left(\frac{1}{L}\right)^m \cdot \left(1 - \frac{1}{L}\right)^{n-m} \quad (7)$$

In equation (7), the number of successfully identified pseudo codes has reached its maximum value. In fact, the number of n in the equation is very large, so 1 can be ignored, and $L = n$ is taken. Before using the Logistic DFSA algorithm to recognize labels, the reader first estimates the total number of labels within the recognition range n , and then sends n to the labels. The random number generator of the tag generates a number between $1 \sim n$ as a pseudo-ID code. At this point, there are several scenarios for pseudo-ID codes. The pseudo-ID code has no label selection, i.e., $m = 0$. There is one tag that selects the pseudo-ID code, which is

$m=1$. When the number of labels for selecting the pseudo-ID code is greater than or equal to 2, i.e., $m \geq 2$, a collision phenomenon in ordinary algorithms occurs. When $m=0$ occurs, the probability of an empty ID code appearing during the recognition process is shown in equation (8).

$$P_{(m=0)} = (1 - \frac{1}{L})^n \quad (8)$$

When $m=1$, the probability of successfully identifying the pseudo-ID code can be obtained as shown in equation (9).

$$P_{(m=1)} = C_n^1 \cdot (\frac{1}{L}) \cdot (1 - \frac{1}{L})^{n-1} = \frac{n}{L} \cdot (1 - \frac{1}{L})^{n-1} \quad (9)$$

When $m \geq 2$ is present, the pseudo-ID code is the usual collision ID code, and the collision probability P_m of the label when $m \geq 2$ is shown in equation (10).

$$P_m = 1 - P_{(m=0)} - P_{(m=1)} = C_n^m \cdot (\frac{1}{L})^m \cdot (1 - \frac{1}{L})^{n-m} \quad (10)$$

Assuming a pseudo-ID code is selected by m tags simultaneously and conflicts occur, the tag parallel recognition algorithm is used to provide an algorithm for tag m conflicts, and the number of tag queries and throughput are counted. Logistic DFSA is a parallel recognition algorithm based on Logistic mapping and DFSA algorithm. This algorithm combines the chaotic characteristics of Logistic mapping with the determinacy and finiteness of DFSA to achieve efficient parallel pattern recognition. Logistic mapping is a nonlinear dynamic system with chaotic characteristics, where small initial condition changes may lead to completely different results. This chaotic characteristic gives Logistic mapping an advantage in dealing with complex pattern recognition problems. The logistic regression function is Figure 6.

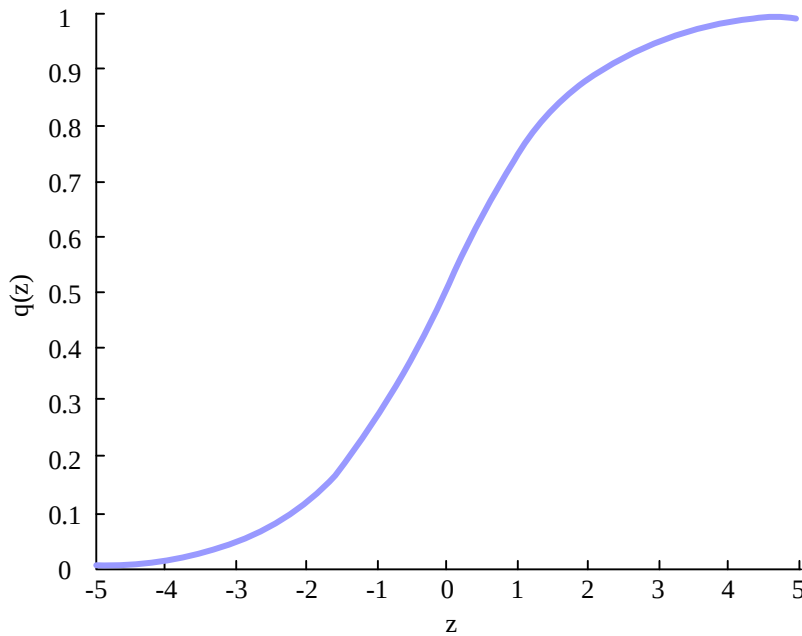


Figure 6: Logistic regression function

DFSA is a finite state automaton that performs operations based on a defined set of rules and state transitions. The advantages of DFSA are determinism and finiteness, where each input leads to a clear output and the state transition is finite. This feature makes DFSA efficient and predictable in dealing with pattern recognition problems. The Logistic-DFSA algorithm uses Logistic mapping to preprocess input data to generate a set of chaotic patterns. Then, these chaotic patterns are used as inputs for DFSA for pattern recognition. Due to the chaotic nature of Logistic mapping and the determinacy and finite nature of DFSA, Logistic-DFSA algorithm can achieve efficient

pattern recognition in parallel environments. Overall, the Logistic-DFSA algorithm is an effective and parallelizable pattern recognition algorithm. It combines the advantages of Logistic mapping and DFSA algorithms, can handle complex pattern recognition problems, and has predictability. The number of labels in the time slot of the Logistic-DFSA algorithm is equation (11).

$$n_{slot} = n / F \quad (11)$$

In equation (11), F is the initial frame length, and in general, F is taken as 8. The probability of identifying the number of m labels in a certain time slot is

equation (12).

$$P(m, M) = \frac{C_M^1 (M-1)^{m-1}}{M^m} = \left(1 - \frac{1}{M}\right)^{m-1} \quad (12)$$

When the spreading code is M , the number of recognizable labels is equation (13).

$$n_{cg} = n_{sx} \left(1 - \frac{1}{M}\right)^{n_{sx}-1} = \frac{n}{F} \left(1 - \frac{1}{M}\right)^{\frac{n}{F}-1} \quad (13)$$

In equation (13), n_{cg} represents the number of recognizable labels, and n_{sx} represents the number of labels. The interaction between IoT components and the data mining process is key to ESCIM systems. First, the data provided by IoT devices is the basis for data mining. This data is preprocessed and fed into various data mining algorithms to uncover patterns, trends, and associations. At the same time, the results of data mining can also be fed back to IoT components to optimize their data acquisition strategies. In summary, the integration of IoT technology in the ESCIM system enables a close interaction between data flow, system architecture, and data mining processes. This provides enterprises with more intelligent, efficient, and accurate solutions for supply chain management.

4 Analysis of ESCIM system integrating DM-IoTm technology

This study conducted experiments using actual enterprise supply chain data to verify the effectiveness of the ESCIM system based on Apriori algorithm and Logistic-DFSFA algorithm, and conducted a comprehensive analysis of the ESCIM system.

4.1 Performance analysis of Apriori algorithm and Logistic-DFSFA algorithm

To analyze the enterprise supply chain data, the first step is to prepare a transaction dataset. This dataset should be derived from the actual transaction records of a large retail enterprise and should cover the entire chain of transactions from suppliers to final consumers. The dataset comprises hundreds of thousands of transaction records, each detailing a complete transaction, including the item name, quantity, transaction time, and location. To ensure data quality and consistency, a series of operations such as data cleaning, deduplication, missing value filling and outlier detection are carried out in the data pre-processing stage. A unified coding process is

implemented to ensure that each item has a unique identifier, so that Apriori algorithm can accurately identify and calculate the support and confidence of each set in the subsequent data mining process.

When implementing Apriori algorithm, the research focused on the following two parameters: minimum support and minimum confidence. The minimum support was a threshold used to measure how frequently the item set appears in the data set. It was considered a "FI" only if the item set's support was above this threshold. For this experiment, a minimum support of 0.05 was set, indicating that the item set had to appear in at least 5% of transactions to be considered 'frequent.' Minimum confidence was a threshold used to evaluate the strength of AR. Only when the confidence level of the rule was above this threshold was it considered a "strong rule". In this experiment, the minimum confidence level was set to 0.7, meaning that the probability of the rule head appearing in the case of the rule body needed to be at least 70%. By setting appropriate minimum support and minimum confidence thresholds, the algorithm could only generate FI and AR with practical significance.

The implementation of Apriori is as follows: First, the algorithm scanned the entire database, calculated the support degree of each item, and selected the item that met the minimum support degree to form a frequent 1-item set. Then, based on frequent 1-item sets, the algorithm generated candidate 2-item sets. Then, the algorithm scanned the database again, calculated the support degree of candidate 2-item sets, selected the item sets that met the minimum support degree to form frequent 2-item sets. At the same time, according to the nature of Apriori, the candidate set was pruned to reduce unnecessary calculations. The algorithm then continued this process, generating candidate 3-item sets, candidate 4-item sets, etc., until it could not be regenerated into larger frequent item sets. After finding all frequent item sets, the algorithm generated AR based on these item sets and calculated their confidence levels. Only rules with higher than minimum confidence were retained. Under the above experimental background, the performance of Apriori algorithm and Logistic DFSFA algorithm was compared in this study, and accuracy, sensitivity, specificity, bit error rate and delay time were selected as evaluation indicators for analysis. The comparison results are shown in Table 2.

Table 2: Index evaluation of enterprise supply chain information management data system

Serial number	Accuracy rate (%)	Sensitivity (%)	Specificity (%)	Bit error rate (%)	Delay time (s)	Accuracy (%)
1	89.1	92.3	84.7	12.9	0.6	91.2
2	91.3	88.1	87.3	8.7	0.3	93.5
3	93.9	79.6	93.2	6.1	0.8	89.7
4	97.6	93.4	96.8	2.4	0.2	93.1
5	92.8	91.9	97.5	7.2	1.3	94.2
6	96.2	92.6	91.1	3.8	1.5	91.5
7	89.5	94.7	89.9	10.5	0.7	94.9
8	93.7	96.1	92.3	6.3	1.2	88.7
9	92.0	95.5	94.7	8.0	0.4	93.6
10	95.3	86.3	90.4	4.7	0.5	94.3
Mean value	93.14	91.05	91.89	7.06	0.84	93.21

In Table 2, the average accuracy, sensitivity, and specificity of the ESCIM system based on Apriori and Logistic-DFSA were 93.14%, 91.05%, and 91.89%, respectively. The average error rate was 7.06%, the average delay time was 0.84s, and the average accuracy was 93.21. This indicates that the performance of the

Apriori-based ESCIM system is excellent. This study further analyzed and compared the gross of time slots and throughput of the Logistic-DFSA algorithm, and conducted simulation verification using Matlab. The comparative algorithms include logistic regression algorithm and decision tree algorithm, as shown in Figure 7.

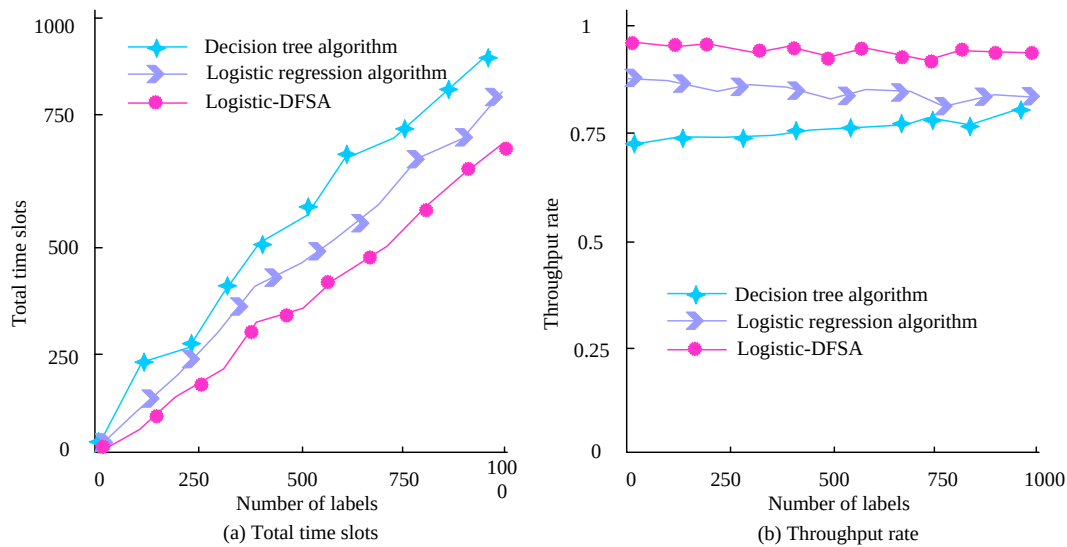


Figure 7: Comparison of total time slots and throughput

Figure 7 (a) is a comparison of the overall time slots for the three algorithms. Logistic-DFSA had a lower time slot than the other two algorithms, indicating better performance. The comparison of throughput rates among various algorithms in Figure 7 (b) shows that Logistic-DFSA had better throughput rates than the other two algorithms, with a stable throughput rate of around 96%. Based on the Logistic-DFSA algorithm, the computational efficiency was high and the algorithm was simple, only requiring prefix judgment on the reader without significantly increasing label overhead. It was a practical and feasible method. To more intuitively

evaluate the performance of ESCIM data systems based on Apriori and Logistic-DFSA, this study used ESCIM systems based on these two algorithms (Method 1), SC-IM data systems based on support vector data description algorithms (Method 2), SC-IM data systems based on error backpropagation algorithms (Method 3), SC-IM data systems based on genetic algorithms (Method 4), and SC-IM data system based on particle swarm optimization (PSO) algorithm (Method 5) to make a comparison. The comparison of the average accuracy of these five algorithms is Figure 8.

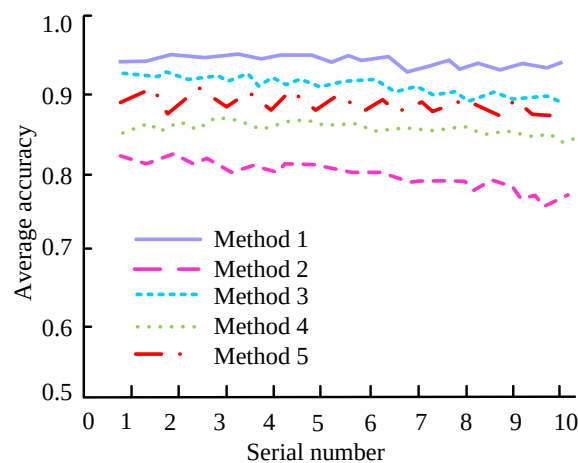


Figure 8: Comparison of average accuracy of five algorithms

In Figure 8, the average accuracy of methods 2, 3, 4, and 5 was 84%, 91%, 89%, and 90%. The average accuracy of Method 1 was 96%, which is higher than the average accuracy of all four models. Therefore, the ESCIM system based on Apriori algorithm and Logistic DFSA algorithm performed better.

4.2 Performance testing analysis of ESCIM system

To analyze the practical application effect of the proposed ESCIM system, this study compared the overall

performance of the ESCIM system integrating DM-IoTm technology with the general international transportation and logistics information management system. Performance and cost-effectiveness were used as comparative indicators. This study set the same testing conditions and dataset. The experimental environment was mainly divided into hardware environment and software environment, and Table 3 shows the specific parameters.

Table 3: Specific experimental environment of enterprise supply chain information management system performance test

Environmental classification		Description	Disposition
Hardware environment	Server	It is used to deploy a SC-IM system	CPU: Intel Xeon Silver 4216, Memory: 128GB DDR4, Storage: 1TB NVMe SSD
	Network equipment	It ensures the stability and speed of the network connection	Gigabit Ethernet switches, routers, firewalls, etc
	Terminal equipment	Used to test the response and interaction performance of the system	Multiple computers and mobile devices with different configurations
	Operating system	Used for server operation and management	CentOS 8.2
Software environment	Database management system	Used to store and manage supply chain information data	MySQL 8.0.23
	Middleware	Used to support the operation and interaction of the system	Apache Tomcat 9.0, Redis 6.0
	Development tools and environment	Used for system development and testing	Java 11, Python 3.8, IntelliJ IDEA, Git, etc

Table 3 provides detailed experimental environment configuration information, including hardware devices and software environments. This configuration information is crucial for conducting performance testing and evaluating the actual application

effectiveness of the system. The performance of the two systems recorded and analyzed in the above experimental environment on key performance indicators such as response time and CPU usage is Figure 9.

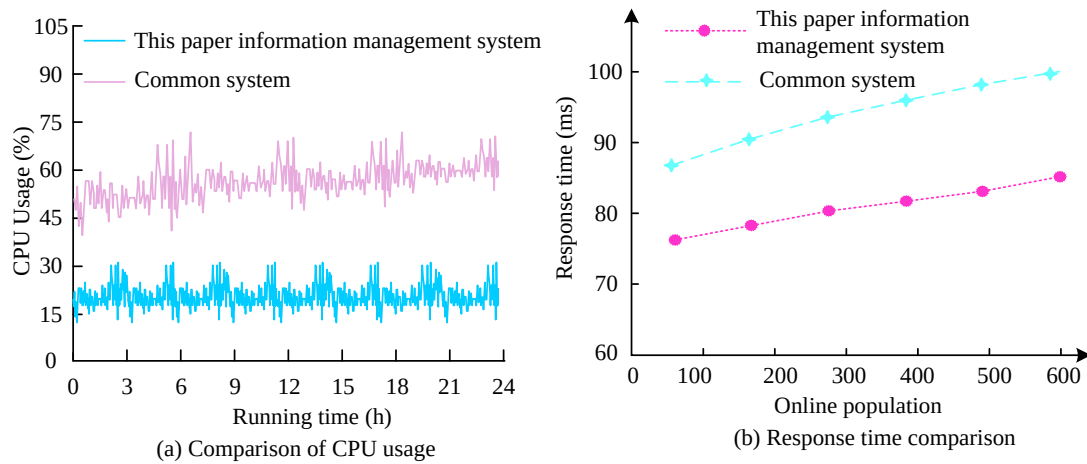


Figure 9: Performance test comparison between the two platforms

In Figure 9 (a), the CPU usage of the proposed ESCIM system fluctuated between 15% and 30% within 24 hours. The fluctuation range of ordinary SC-IM systems was 45% to 75%. In Figure 9 (b), the response time of the proposed ESCIM system was 85 ms when the number of people was 600, and 99.5 ms for a regular SC-IM system. The response time of ESCIM was much lower than that of ordinary SC-IM. The constructed ESCIM system had better CPU usage and response time than ordinary

information management systems, fully reflecting the platform's high computing speed, high network bandwidth, and other characteristics. After verifying the superior performance of the ESCIM system, this study also analyzed its economic benefits, mainly comparing the total investment cost and benefits. Figure 10 shows the comparison results of various indicators of economic benefits of the ESCIM system.

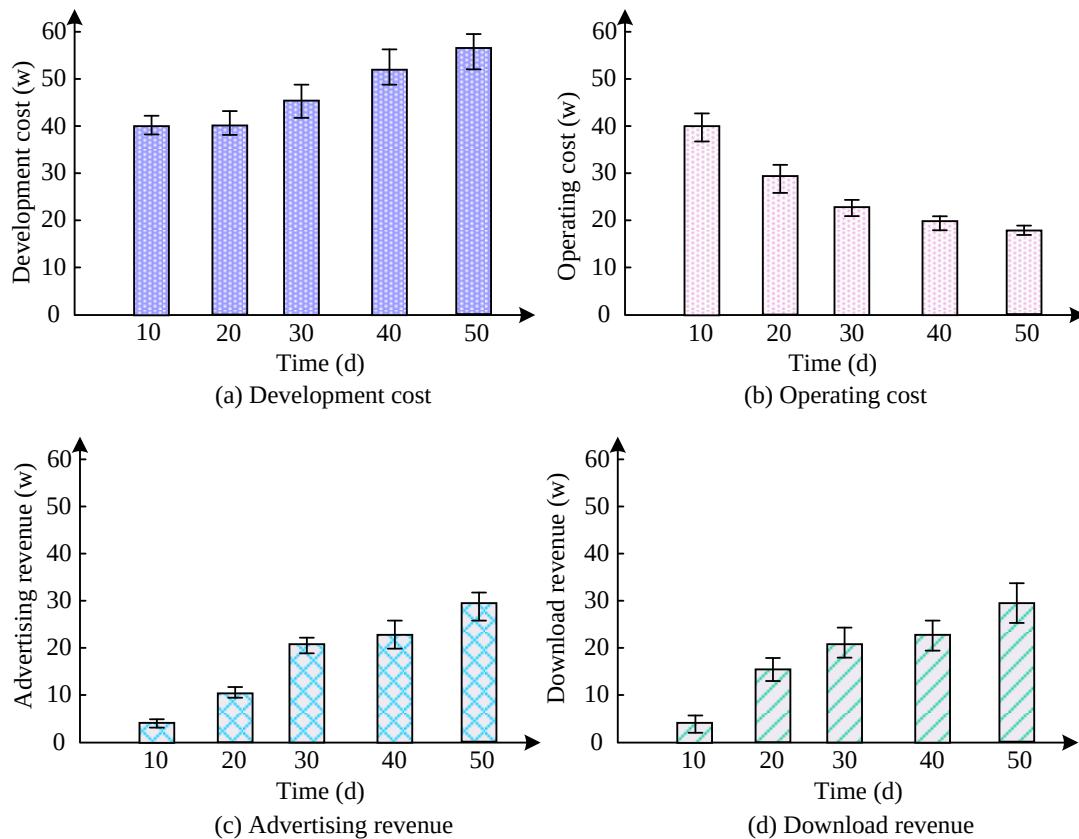


Figure 10: Comparison of economic benefit index of enterprise SC-IM system

In Figure 10 (a), the development cost of the ESCIM system increased over time. In Figure 10 (b), the operating cost of the ESCIM system decreased over time. In Figure 10 (c), the advertising revenue of the proposed ESCIM system showed an upward trend over time. In Figure 10 (d), the download revenue of the ESCIM system generally increased over time. The above results

indicate that the advertising and download revenue of the ESCIM system are increasing over time, and can achieve good cost-effectiveness in the future. Next, the scalability analysis of the system in different enterprise environments was carried out, and the indicators were normalized, as shown in Figure 11.

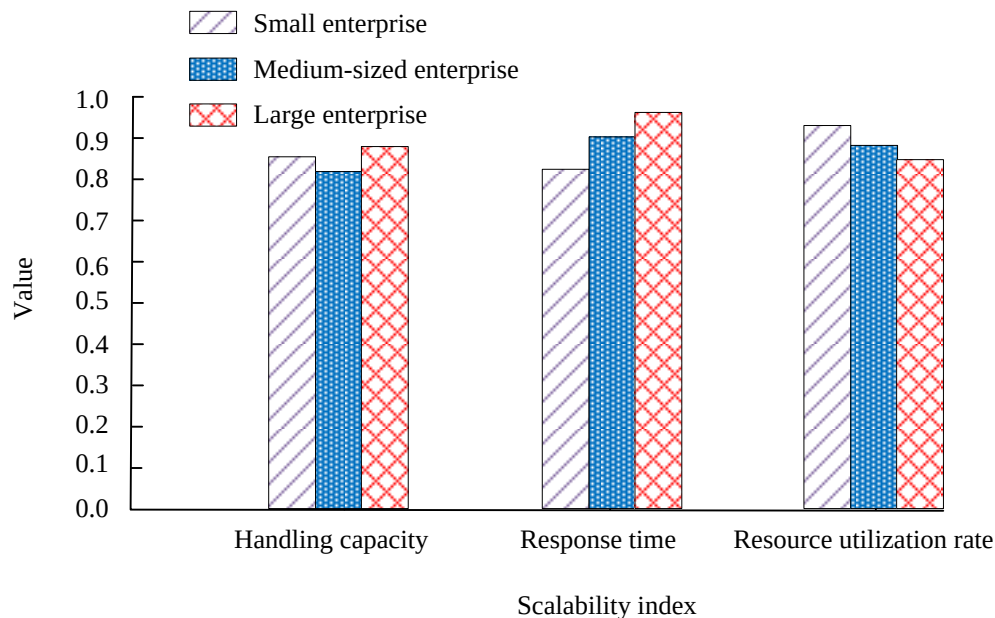


Figure 11: Scalability analysis of systems in different enterprise environments

Figure 11 shows that the scalability indicators, including throughput, response time, and resource utilization, were all above 0.8 for small, medium, and large enterprises. This suggested that the system was feasible in various

enterprise environments. Finally, the statistical significance test of the decision space and preference space of the system was carried out, as shown in Figure 12.

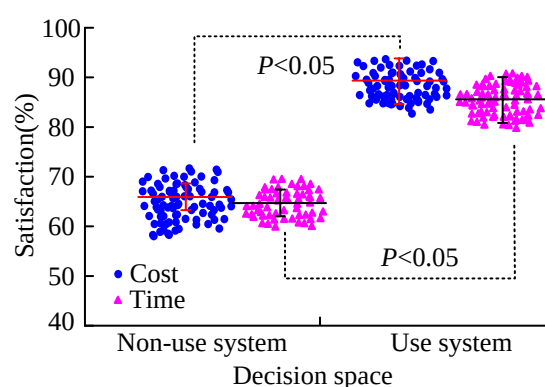


Figure 12: Statistical significance test

In Figure 12, the difference in time and cost before and after using the SC-IM system is statistically significant, indicating that the system has brought substantial improvement to the enterprise.

5 Discussion

First of all, from the perspective of accuracy and other evaluation indicators, the ESCIM system based on Apriori algorithm and Logistic-DFSA algorithm shows excellent performance. In particular, Logistic-DFSA

algorithm is superior to other algorithms in throughput because of its high computational efficiency, simple algorithm and no obvious increase in label cost. This result is similar to that obtained by Gao et al. in the study of ESCIM system [21]. In addition, by comparing the average accuracy of different algorithms, this study further verifies the superiority of the SC-IM system based on Apriori algorithm and Logistic-DFS algorithm. This study compares the performance of the ESCIM system integrated with data mining and IoT technology with that of the general international transportation logistics information management system. The results show that the ESCIM system outperforms the latter in key performance indicators such as CPU utilization rate and response time. This shows that the ESCIM system proposed in this paper has higher computing speed and better network performance in practical application, and can better meet the needs of enterprises. This result coincides with the conclusion reached by Amini's team in 2023 [22]. The results of economic benefit analysis show that although the development cost of ESCIM system increases with the passing of time, its operating cost gradually decreases, while the advertising revenue and download revenue gradually increase. This shows that the system can achieve good cost efficiency in long-term operation and bring real economic benefits to enterprises. Zhang et al. also reached a similar conclusion when conducting research on ESCIM system [23]. In terms of scalability analysis, this study normalizes scalability indicators under different enterprise environments, and finds that throughput, response time, resource utilization and other indicators of small, medium and large enterprises are satisfactory. This shows that the ESCIM system proposed in this study can show good performance and strong scalability in different scale enterprise environments. This is similar to the research results obtained by Sawe's team in 2021 [24].

To sum up, this study has made a novel contribution. By considering multiple evaluation indicators, comparing algorithm performance, analyzing practical application effects and economic benefits, and conducting scalability analysis, this research validates the superiority of the ESCIM system based on the Apriori and Logistic-DFS algorithms. It also provides new ideas and methods for related fields. These results are significant both theoretically and practically as they promote the development and optimization of SCIM systems. They also help improve the supply chain management level of enterprises, reduce costs, improve competitiveness, and achieve sustainable development.

6 Conclusion

To provide more accurate and real-time supply chain information for enterprises and improve their competitiveness, this study explored the application of integrated DM-IoTm technology in ESCIM. It included aspects such as supply chain data analysis, inventory

management, order management, and supplier management. The research results indicated that the gross time slots of the Logistic-DFS algorithm was lower than other algorithms, and the throughput was also better than other algorithms, with a stable throughput of around 96%. The CPU usage of the proposed ESCIM system fluctuated between 15% and 30% within 24 hours, and the response time was 85 ms when the number of people was 600. This indicated that the constructed ESCIM system had better CPU usage and response time than ordinary information management systems, fully reflecting the platform's high computing speed, high network bandwidth, and other characteristics. Due to the complexity of the source of supply chain data, the quality of the data can be uneven. Additionally, the sensitivity of supply chain data is high, and if leaked or abused, it can cause significant losses to enterprises. At the same time, with the widespread application of IoT technology, its potential security vulnerabilities also bring new risks to enterprise data. To address the aforementioned issues, the following improvements can be made: Firstly, enhancing data encryption and anonymization processes. Secondly, implementing a rigorous data access control mechanism. In addition, in view of the security vulnerability problem of IoT devices, it is recommended that enterprises should regularly conduct security risk assessment and vulnerability scanning to discover and repair potential security problems in a timely manner. At the same time, close cooperation should be maintained with IoT equipment suppliers to jointly address security threats and ensure enterprise data security. Finally, it is emphasized that enterprises need to strengthen the data security awareness training of employees.

References

- [1] Y. Li, R. K. Shyamasundar, and X. Wang, "Special issue on computational intelligence for social media data mining and knowledge discovery," *Computational Intelligence*, vol. 37, no. 2, pp. 658-659, 2021. <https://doi.org/10.1111/coin.12457>
- [2] A. W. Al-Khatib, "Internet of things, big data analytics and operational performance: the mediating effect of supply chain visibility," *Journal of Manufacturing Technology Management*, vol. 34, no. 1, pp. 1-24, 2023. <https://doi.org/10.1108/JMTM-08-2022-0310>
- [3] Z. Shao, S. Yuan, J. Xu, and Y. Wang, "A statistical feature data mining framework for constructing scholars' career trajectories in academic data," *Applied Soft Computing*, vol. 118, no. 1, pp. 108550-108561, 2022. <https://doi.org/10.1016/j.asoc.2022.108550>
- [4] T. Wang, B. Ren, C. Li, K. Guo, J. Leng, and P. Zhou, "Monolithic tapered Yb-doped fiber chirped pulse amplifier delivering 126 μ J and 207 MW femtosecond laser with near diffraction-limited beam quality," *Frontiers of Optoelectronics*, vol. 16,

- no. 3, pp. 30-30, 2023.
- [5] F. H. Awad, and M. M. Hamad, "Big data clustering techniques challenged and perspectives," Informatica, vol. 47, no. 6, pp. 203-218, 2023. <https://doi.org/10.31449/inf.v47i6.4445>
- [6] M. Kiguchi, W. Saeed, and I. Medi, "Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest," Applied Soft Computing, vol. 118, no. 1, pp. 108491-108511, 2022. <https://doi.org/10.1016/j.asoc.2022.108491>
- [7] V. Plotnikova, M. Dumas, and F. P. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," Data and Knowledge Engineering, vol. 139, no. May, pp. 102013.1-102013.17, 2022. <https://doi.org/10.1016/j.datak.2022.102013>
- [8] A. D. Ganesh, and P. Kalpana, "Supply chain risk identification: a real-time data-mining approach," Industrial Management and Data Systems, vol. 122, no. 5, pp. 1333-1354, 2022. <https://doi.org/10.1108/IMDS-11-2021-0719>
- [9] L. He, Y. Cao, and J. Mao, "Exploring college students' fitness and health management based on Internet of Things technology," Journal of High-Speed Networks, vol. 28, no. 1, pp. 65-73, 2022. <https://doi.org/10.3233/JHS-220679>
- [10] Y. Jiang, "Project cost accounting based on internet of things technology," Journal of Interconnection Networks, vol. 22, no. 3, pp. 1-20, 2022. <https://doi.org/10.1142/S0219265921450122>
- [11] L. Wang, and D. Jiang, "Energy management control system of prefabricated construction based on internet of things technology," International Journal of Internet Protocol Technology, vol. 14, no. 2, pp. 86-92, 2021. <https://doi.org/10.1504/ijipt.2021.116256>
- [12] T. S. Deepu, and V. Ravi, "Modelling of interrelationships amongst enterprise and inter-enterprise information system barriers affecting digitalization in electronics supply chain," Business Process Management Journal: Developing Re-Engineering Towards Integrated Process Management, vol. 28, no. 1, pp. 178-207, 2022.
- [13] H. Mao, and L. Chen, "E-Commerce enterprise supply chain cost control under the background of big data," Complexity, vol. 2021, no. 6, pp. 1-11, 2021. <https://doi.org/10.1155/2021/6653213>
- [14] Q. Li, and G. Wu, "ERP system in the logistics information management system of supply chain enterprises," Hindawi Limited, vol. 2021, 2021. <https://doi.org/10.1155/2021/7423717>
- [15] J. Wang, H. Zhou, and X. Jin, "Risk transmission in complex supply chain network with multi-drivers," Chaos Solitons and Fractals, vol. 143, no. 5439, pp. 110259-110269, 2021. <https://doi.org/10.1016/j.chaos.2020.110259>
- [16] H. Yang, S. Zhao, and J. Peng, "Optimal retail price and service level in a dual-channel supply chain with reference price effect," Journal of Industrial and Management Optimization, vol. 19, no. 6, pp. 3883-3912, 2023. <https://doi.org/10.3934/jimo.2022115>
- [17] G. Liu, C. Li, W. Wei, W. Li, and H. Zhen, "Data mining analysis of gene prognostic markers of metastatic skin cancer based on the elastic network method," Mathematical Problems in Engineering, vol. 25, no. 1, pp. 6636058.1-6636058.12, 2021. <https://doi.org/10.1155/2021/6636058>
- [18] H. Wang, "Analysis and prediction of CET4 scores based on data mining algorithm," Complexity, vol. 2021, no. 12, pp. 1-11, 2021. <https://doi.org/10.1155/2021/5577868>
- [19] G. Mehdi, H. Hooman, Y. Liu, S. Peyman, and R. Arif, "Data mining techniques for web mining: A survey," Artificial Intelligence and Applications, vol. 1, no. 1, pp. 3-10, 2022. <https://doi.org/10.47852/bonviewAIA2202290>
- [20] M. N. Faisal, "Role of Industry 4.0 in circular supply chain management: a? Mixed-method analysis," Journal of Enterprise Information Management, vol. 36, no. 1, pp. 303-322, 2023. <https://doi.org/10.1108/JEIM-07-2021-0335>
- [21] Q. Gao, S. Guo, X. Liu, G. Manogaran, N. Chilamkurti, and S. Kadry, "Simulation analysis of supply chain risk management system based on IoT information platform," Enterprise Information Systems, vol. 14, no. 9, pp. 1354-1378, 2020. <https://doi.org/10.1080/17517575.2019.1644671>
- [22] M. Saratchandra, and A. Shrestha, "The role of cloud computing in knowledge management for small and medium enterprises: a systematic literature review", Journal of Knowledge Management, vol. 26, no. 10, pp. 2668-2698, 2022. <https://doi.org/10.1108/JKM-06-2021-0421>
- [23] X. Zhang, P. Sun, J. Xu, X. Wang, J. Yu, Z. Zhao, and Y. Dong, "Blockchain-based safety management system for the grain supply chain," IEEE Access, vol. 8, no. 3, pp. 36398-36410, 2020. <https://doi.org/10.1109/ACCESS.2020.2975415>
- [24] F. B. Sawe, A. Kumar, J. A. Garza-Reyes, and R. Agrawal, "Assessing people-driven factors for circular economy practices in small and medium-sized enterprise supply chains: Business strategies and environmental perspectives," Business Strategy and the Environment, vol. 30, no. 7, pp. 2951-2965, 2021. <https://doi.org/10.1002/bse.2781>