Construction of Data Analysis and Forecasting Model for Supply Chain Channels Based on the Internet of Things

Siyu Tian*,#
Jilin University
Changchun, 130015, Jilin, China
jiushiyanyuya@163.com
*Corresponding author
#Co-first author

Siqi Xing#
Jilin University
Changchun, 130015, Jilin, China
2432289865@qq.com
#Co-first author

Yishan Wang
Beijing Jiaotong University
Beijing, 100091, China
shanshan720919@163.com

***Abstract*—In order to solve the problem of data analysis and demand prediction in supply chain channel management, this paper presents a solution based on Internet of Things (IoT) and machine learning. Through the collection, processing, and analysis of the sales data from the e-commerce platform, the key factors that affect the sales are identified, including price, rating, and promotion efforts. On this basis, a Long Short-Term Memory (LSTM) Neural Network (LSTM) is constructed, and a system including data management, data analysis, demand prediction and visualization is designed and realized. The system uses a front end and a back end separation architecture, using Kafka, Spark Streaming to collect and process data in real time, and MySQL and MongoDB to store and manage data. Based on the development process of software engineering, the design, implementation and testing of the system are carried out. Finally, the validity of the proposed solution is verified by testing and evaluating real data, and it is very important to improve supply chain management.**

Keywords—Supply chain management, Internet of things, Data analysis, Demand forecasting, LSTM neural network

I. Introduction

As a core competence of a modern enterprise, it is essential to accurately predict the demand for the supply chain, reduce the cost, and improve the service level [1]. In recent years, the development of Internet of Things has provided a new opportunity to collect and analyze supply chain channel data. Moreover, machine learning has shown great potential in demand forecasting. [2] In this paper, we discuss how to combine the Internet of Things with Machine Learning and Machine Learning.

II. Related Theories and Technical Foundations

A. Internet of Things Technology

The Internet of Things is a network technology that connects objects to the Internet through a variety of sensing devices. Internet of things is an intelligent means of identification, positioning, tracking, monitoring and management. It has been widely used in supply chain management[3]. Internet of things is a new type of information technology, it can be in every link of the supply chain, real-time access to logistics, warehousing, transportation, distribution and other information. These studies will provide strong support for the realization of supply chain visual management and optimal decision-making. With the development of Internet of Things technology, great changes have taken place in the way of supply chain management, which enables enterprises to monitor the operation of supply chain more accurately. On this basis, the use of Internet of Things technology to analyze the problems in the supply chain, put forward solutions to improve the efficiency and adaptability of supply chain[4].

B. Supply Chain Management Theory

Supply chain management is to plan, organize, coordinate and control the whole supply chain to maximize the benefits of the whole supply chain[5]. Its core idea is to cooperate and integrate the upstream and downstream enterprises in the supply chain. Collaboration and integration between enterprises include not only information sharing, but also risk sharing and income sharing to enhance the overall flexibility and competitiveness[6]. Demand management and inventory management are the core of supply chain management. Through the accurate prediction of customer demand, supply chain managers can make optimal inventory decisions, reduce uncertainties, improve the response ability and service level of enterprises. The combination of demand management and inventory management can effectively respond to the changing market demand and improve the flexibility and competitiveness of enterprises. Supply chain management is not only an enterprise's business activities, it also involves a cross-enterprise and cross-departmental collaboration. This requires each link to cooperate closely and cooperate with each other to maximize the revenue of the whole supply chain.

C. Data Analysis and Forecasting Models

Importance of Data Analysis and Forecasting Model in Supply Chain Management[7]. Through the collection, storage, processing and analysis of a large number of data generated by various links in the supply chain, we can effectively identify the laws and trends, and discover the risks and opportunities. Traditional analysis methods include descriptive statistics, correlation analysis, clustering analysis, association rules mining and so on.

$$J = \sum_{i=1}^{k} \sum_{x \in s_i} \| x - \mu_i \|^2 \quad (1)$$

979-8-3315-3405-9/24/$31.00 ©2024 IEEE
DOI 10.1109/ICIRDC65564.2024.00068

Here, μ_i stands for the center of the group S_i, x for the points within the group, and k for the number of groups.

But data analysis alone usually reflects the past. Forecasting model is an important means to improve forecasting ability. On this basis, machine learning method and time series model are used to forecast the demand, inventory, price and other important indicators in the supply chain[8]. TThrough this research, we can help enterprises adapt to the changes of market environment and improve the predictability of supply chain. The combination of data analysis and predictive modeling is helpful for enterprises to understand the current supply chain operation and grasp the future development trend.

III. Data Collection and Preprocessing for Supply Chain Channel Based on the Internet of Things

A. Deployment of IoT Devices and Data Collection

The deployment of Internet of Things terminal and data collection is the key of the whole supply chain system. In this project, RFID tags, sensors, GPS positioning devices are applied to each node of the supply chain to achieve automated access to logistics, warehousing, transportation and other information. For example, the warehouse is equipped with temperature and humidity sensors, which can monitor the surrounding environment in real time, and the GPS positioning device on the vehicle can accurately record the traffic path and status data. In addition, the device can be integrated with SCM systems, ERP systems, etc. to better collect various business information, such as orders, inventory, sales and so on. Through the research of this project, on the one hand, it can improve the transparency and transparency of the supply chain, on the other hand, it can provide more accurate data support for enterprises, so as to improve the operational efficiency and quality of the supply chain, so as to better meet the needs of customers and enhance the competitiveness of enterprises.

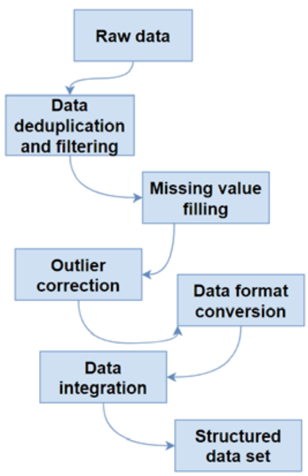

Fig. 1. Data Cleaning and Integration Process for Supply Chain Channel Data

B. Data Cleaning and Integration

There are many quality problems in the data of supply chain channel, such as noise, missing value, inconsistency, etc. It is urgent to sort out and integrate them [9]. Through repeated elimination, screening, elimination of repeated, invalid data to ensure accurate and reliable data. Aiming at the missing and outlier data, the data are filled and modified. For example, when dealing with temperature data collected by sensors, the moving average method can be used to fill in missing values to maintain data integrity. For outliers beyond the normal range, correction using the average values of neighboring time points can be applied to ensure data reliability and accuracy[9]. The cleaned heterogeneous data need to be integrated into a unified format and standard to form structured datasets, providing strong support for subsequent data analysis and mining, as shown in Figure 1.

C. Feature Engineering and Data Transformation

During feature engineering and data transformation, appropriate features are selected and transformed based on the requirements of data analysis and modeling[10]. Common feature engineering methods include feature extraction, feature selection, and feature construction. For example, for time series data, features such as trend, periodicity, and randomness can be extracted. For high-dimensional data, important features can be selected using methods such as correlation analysis and principal component analysis[11]. Regarding feature transformation, common methods include data normalization, data discretization, and data encoding. For instance, for sales data, the min-max normalization method can be used to transform it into the [0,1] interval:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

Where x is the original sales data, x_{min} and $x_{\max}$ are the minimum and maximum values of sales, and x_{norm} is the normalized data. Through feature engineering and data transformation, the quality and representation of the data can be optimized, laying the foundation for subsequent data analysis and modeling.

IV. Analysis of Supply Chain Channel Data

A. Descriptive Statistical Analysis

After conducting descriptive statistical analysis on the sales data of the e-commerce platform, the daily average sales and fluctuations of various products were identified. Taking products A and B as examples, the daily average sales of product A were \$3000, with a standard deviation of \$500, while product B had a daily average sales of \$5000, with a standard deviation of \$1000. This indicates that although the sales level of product B is higher, its sales volatility is relatively larger. By plotting a box plot, as shown in Figure 2, the sales distribution of different products can be intuitively compared. From the box plot, it can be observed that the sales distribution range of product B is relatively wide, while that of product A is relatively concentrated within a narrower range[12]. Apart from products A and B, statistical analysis was also conducted on products C and D, revealing differences in their sales ranges. Through the analysis of sales data, a better understanding of the performance of each product in terms of sales is gained, providing important references for further sales strategy formulation and inventory management.

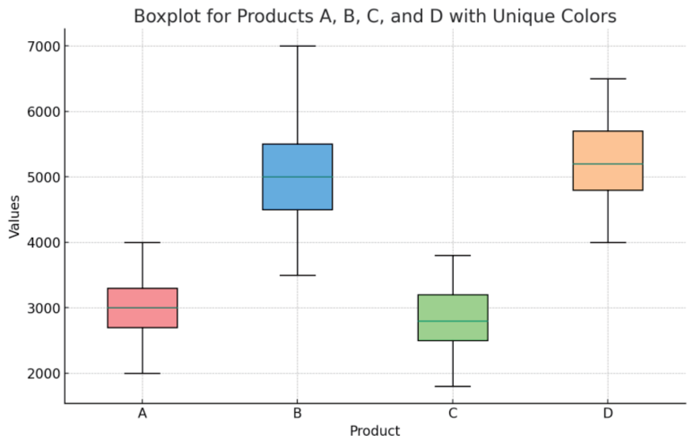

Fig. 2. Box Plot of Product Sales

B. Correlation Analysis

Correlation analysis of sales data on the e-commerce platform selected key variables such as price, sales volume, rating, and promotional intensity. Through the analysis of Pearson correlation coefficient, it is concluded that the price of the product is negatively correlated with the sales volume (correlation coefficient is-0.8), which indicates that the price of the product has a great impact on the purchase behavior of consumers, so the E-commerce platform must make a reasonable price[13]. In addition, there is a significant positive correlation between product grade and sales volume (correlation coefficient reaches 0.6), which indicates that high-quality brand products are more likely to be accepted by consumers. In addition, there is a significant positive correlation between promotion intensity and sales volume (correlation coefficient reaches 0.7), indicating that effective promotion can stimulate consumption and promote sales growth. E-commerce platform should be aimed at their own commodity characteristics, target customer groups, diversified promotion activities.

Through comprehensive analysis and consideration of these factors, the e-commerce platform can formulate more targeted and effective sales strategies, continuously optimize operational models, and improve overall sales performance[14]. Figure 3 shows the correlation coefficient matrix of some variables in the sales data of the e-commerce platform.

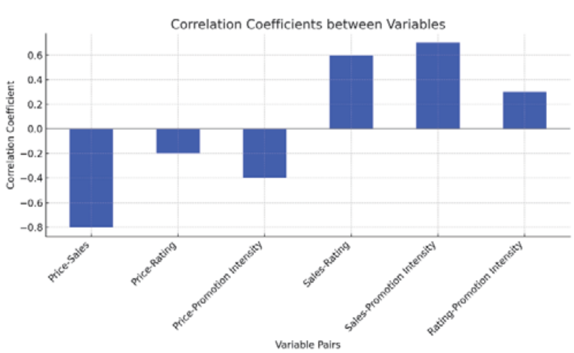

Fig. 3. Correlation Coefficient Matrix of Sales Data on the E-commerce Platform

C. Data Visualization

In the analysis of sales data on the e-commerce platform, a monthly trend chart of product sales (as shown in Figure 4) was plotted, and it is apparent from the chart that sales exhibit seasonal characteristics. Significant peaks in sales occur at promotional events such as "618" and "Double 11," which is a very notable phenomenon. Leveraging this characteristic, the e-commerce platform can more strategically plan marketing strategies, prepare inventory in advance to meet the peak demand of consumers during promotions. Through trend analysis of sales data, the e-commerce platform can better grasp market changes, allocate resources reasonably, improve sales efficiency, and adjust promotional intensity and marketing activities accurately according to different seasonal characteristics, further enhancing sales performance. Such data-driven marketing strategies not only increase sales revenue but also enhance user experience and brand competitiveness. Therefore, the e-commerce platform can fully utilize the seasonal patterns in sales data for precision marketing, injecting more vitality into the long-term development of the enterprise.

Fig. 4. Monthly Sales Trend Chart of Products

D. Anomaly Detection and Analysis

In the anomaly detection and analysis of sales data on the e-commerce platform, abnormal points in sales revenue and sales volume are identified by setting reasonable thresholds. Table I shows some examples of abnormal data, including cases of abnormal sales revenue and sales volume. Through observation of the abnormal data, it is noted that significant abnormalities occur on certain dates in terms of sales revenue or sales volume. For example, sales revenue reached 1.205 million yuan on January 3, far exceeding the usual sales level, thus being identified as an anomaly; similarly, sales revenue was only 106,000 yuan on January 5, also an abnormal point. Identifying these abnormal data points enables timely detection of anomalies in sales data, facilitating targeted analysis and processing. Analyzing abnormal data helps discover potential problems and risks, allowing for timely adjustment of sales strategies or implementation of corresponding measures to ensure the accuracy and stability of sales data. For e-commerce platforms, anomaly detection is of great significance to improve operational efficiency, reduce losses, optimize resource allocation and improve user experience. Therefore, to identify and analyze the abnormal behavior of e-commerce platform is a key link in the operation of e-commerce platform, and also an important support for the sustainable development of e-commerce enterprises.

TABLE I. ABNORMAL SALES DATA ON THE E-COMMERCE PLATFORM

Date	Sales (10,000 RMB)	Quantity Sold	Abnormal
1/1/2023	50.2	1200	No
1/2/2023	48.7	1150	No
1/3/2023	120.5	3000	Yes
1/4/2023	52.3	1300	No
1/5/2023	10.6	280	Yes

V. CONSTRUCTION OF SUPPLY CHAIN CHANNEL DEMAND FORECASTING MODEL

A. Problem Description and Model Selection

In order to optimize the inventory management and improve the operation efficiency of the supply chain, it is of great importance to forecast the demand of each channel in the supply chain. On this basis, a new method is proposed to realize the prediction of product sales in the coming months. On this basis, the project intends to use LSTM (LSTM) neural network for modeling.

$$\begin{aligned} f_t &= \sigma_g\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \\ i_t &= \sigma_g\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \\ \tilde{C}_t &= tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma_g\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \\ h_t &= o_t * \tanh\left(\mathrm{C_t}\right) \end{aligned} \tag{3}$$

Where f_t, i_t, and o_t represent the forget gate, input gate, and output gate, respectively; $\tilde{C}_t$ is the current candidate memory cell; C_t is the current memory cell; h_t is the current hidden state.

The LSTM model can analyze the long-term correlation of time series data and predict the future sales volume considering other factors.

$$Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_1 \epsilon_{t-1} + \cdots + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{4}$$

Additionally, a traditional time series model such as ARIMA is chosen as the benchmark model for comparison purposes.

B. Data Preparation and Model Training

Using e-commerce platform, through the recent two years of sales data for a detailed sorting, and its pretreatment. This includes filling in missing values, normalizing data ranges, standardizing data, and ensuring that models learn from data features efficiently. Based on this, a complete set of time series data is set up, which is mainly characterized by sales volume, price and promotion, to describe its influence on sales forecast. See Table II. Training samples are divided into training samples, confirmatory samples and testing samples to achieve efficient training and verification of sample samples. By adjusting the number of hidden layers, the number of hidden layers and the network learning rate, the optimal configuration of the model is realized. Aiming at the problem of over-fitting in deep networks, this project proposes to use "early termination" and "exit" regularization methods to improve the generalization performance of the model. After a series of rigorous learning processes, we have obtained a kind of short-term memory model with excellent performance.

TABLE II. LSTM MODEL HYPERPARAMETER SETTINGS

Hyperparameter	Value
Hidden Layers	2
Hidden Units	128
Learning Rate	0.001
Batch Size	64
Dropout Rate	0.2

C. Results Analysis and Model Application

The data analysis based on LSTM (LSTM) shows that the long-short memory (LSTM) model has strong prediction ability. The average percentage error (MAPE) and root-mean-square error (RMSE) were used to evaluate the performance, and the LSTM model obtained 8.5% and 20.6% MAPE. Compared with ARIMA model, it increased 12.3% and 35.8% respectively. Finally, the LSTM model is validated to be effective in predicting the change of the market, especially during holidays and promotions. On this basis, this study proposes a new method, which can effectively promote enterprises to develop more accurate and more reasonable inventory and marketing strategies by forecasting product demand. The research results of this project will help reveal the important role of deep learning in business intelligence and provide theoretical support for rapid, accurate and accurate prediction of e-commerce platform. The model provides a strong support for data-based decision-making and creates significant economic benefits and competitiveness for enterprises. As shown in Figure 5, the figure shows the comparison between the LSTM model's prediction of market changes and actual data during holidays and promotions.

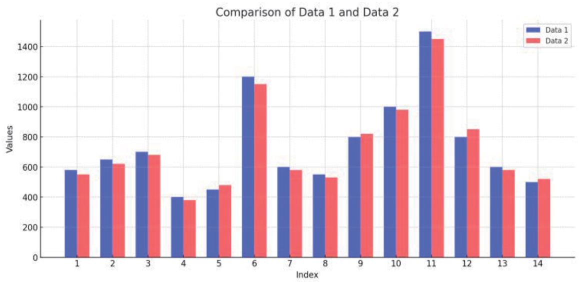

Fig. 5. Comparison between LSTM Model Predictions and Actual Sales Volume

VI. System Design and Implementation

A. System Architecture Design

In order to realize the supply chain channel data analysis and prediction system based on Internet of things, a supply chain channel data analysis and prediction system framework is proposed. The front end uses Vi. js architecture to visualize and interact data, and the back end uses Spring Boot framework to process data and business logic through RESTful API. In the data collection layer, data from various stages of the supply chain is collected in real-time through RFID, sensors, and other IoT devices. The data is decoupled and buffered through message queues such as Kafka, and then processed and transformed in real-time through Spark Streaming. The processed data is stored in relational databases such as MySQL and NoSQL databases such as MongoDB to meet different query and analysis requirements. In the data analysis layer, Python's scientific computing libraries (such as NumPy, Pandas) and machine learning libraries (such as Scikit-learn, TensorFlow) are used to model and predict data.

B. Database Design

In designing the database for supply chain channel data analysis and forecasting, emphasis has been placed on constructing a comprehensive relational database architecture. This architecture encompasses core entity tables such as products, warehouses, orders, and logistics, with carefully designed relationship tables like order-product tables to reveal the complex many-to-many relationships among these entities. Such a design not only ensures logical coherence and consistency among the data but also provides a solid foundation for subsequent data analysis and forecasting. To optimize data query performance, indexes have been created for some key fields, significantly enhancing the speed and efficiency of data retrieval. Considering the diversity of supply chain data, particularly the presence of a large amount of semi-structured and unstructured data in logistics information, it was decided to use MongoDB to handle this part of the data. MongoDB's flexibility and powerful document query capabilities make storage, management, and querying of unstructured data more efficient and convenient. This hybrid data management strategy not only fully utilizes the advantages of relational and non-relational databases but also provides a comprehensive and efficient data support platform for the entire supply chain management system, ensuring the system can handle various data types and meet complex data analysis and forecasting requirements.

C. Functional Module Design

When constructing a Supply Chain Channel Data Analysis and Prediction System, a modular design approach was adopted, dividing the system into four core functional modules: Data Management, Data Analysis, Demand Forecasting, and Visualization Display, to meet various business needs. The Data Management module serves as the foundation of the system, responsible for initial processing of the entire data flow, including data importation, cleaning, transformation, and storage, ensuring high standards of data quality and laying a solid foundation for subsequent analysis and prediction tasks. Based on the high accuracy and high consistency data, the in-depth multi-dimensional statistics and mining of consumers are carried out to obtain more important information, such as sales trends and commodity popularity, so as to help enterprises to grasp market trends. One of the high-level functions is demand forecasting. It accurately predicts future product demand by leveraging historical sales data and advanced machine learning algorithms, generating easily understandable visualized forecasting results to help enterprises make more precise plans and decisions. Finally, the Visualization Display module presents all complex data and forecasting information in intuitive and understandable charts and interactive formats, enhancing user experience and reinforcing the effectiveness of information communication. This enables decision-makers to quickly grasp the overall picture and trends of supply chain operations, effectively supporting the decision-making process. The close collaboration and mutual supplementation of these four modules collectively constitute a comprehensive, operationally convenient, and rapidly responsive system, greatly enhancing the efficiency and accuracy of supply chain channel data analysis and prediction, and bringing significant business value to enterprises. Figure 6 illustrates the functional module design of the system.

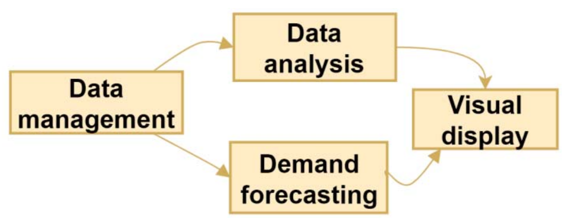

Fig. 6. System Functional Module Design

D. System Implementation and Testing

Based on the architecture design, database design, and functional module design outlined earlier, the system was developed and implemented. The backend was built using the Spring Boot framework, providing RESTful API interfaces to implement core functionalities such as data management, data analysis, and demand forecasting. The frontend was developed using the Vue.js framework, integrated with visualization libraries like Echarts to realize data display and user interaction functionalities. During the development process, the agile development principles were followed, continuously refining system functionalities through iterative development. On this basis, a complete test scheme is proposed, including unit test, integration test and system test. Test frameworks such as JUnit are used to automate testing of critical functional modules and interfaces, while tools such as Postman are used for interface invocation and response validation. After repeated debugging, the performance indicators basically meet the design indicators. Table III lists several sample test cases and their results.

TABLE III. SAMPLE TEST CASES AND RESULTS

Test Case	Expected Result	Actual Result
Data Import	Successfully import 1000 sales records	Passed
Sales Stats	Statistical results consistent with manual calculation	Passed
Demand Forecast	Mean Absolute Percentage Error (MAPE) of forecast results < 10%	Passed
Chart Display	Charts displayed correctly, consistent with backend data	Passed

VII. CONCLUSION

This paper studies the data analysis and demand forecasting in supply chain channel management by combining Internet of Things technology with machine learning. Through collecting, sorting out and analyzing the sales data on e-commerce platform, identify the main factors affecting sales, such as pricing, pricing, promotion, etc. Based on this, a new demand forecasting method is proposed, which uses LSTM neural network to forecast the demand. Experimental results show that the proposed method can effectively improve the accuracy of supply chain channel demand forecasting. The research results of this paper can provide data support for inventory optimization and marketing strategy of enterprises, and it is of great significance to improve the overall management level of supply chain.

REFERENCES

[1] Zhou M, Liu F, Lu Y .A Supply Chain Model of Petrochemical Products Based on Internet of Things[J]. 2020.

[2] Wu Z, Wang S, Yang H ,et al. Construction of a Supply Chain Financial Logistics Supervision System Based on Internet of Things Technology[J]. Mathematical Problems in Engineering, 2021.

[3] Zhao L, Tian Z, Yang Y .Research on Building Supply Chain Collaboration Model Based on BIM and Internet of Things Technology and Its Application[J]. IOP Conference Series: Earth and Environmental Science, 2020, 455(1):012203 (6pp).

[4] Deng Y, Dun R .Study on the Construction of Intelligent Supply Chain System Based on Internet of Things Technology[J]. Springer, Singapore, 2022.

[5] Emroozi V B, Roozkhosh P, Modares A ,et al.A new supply chain design to solve supplier selection based on internet of things and delivery reliability[J]. Journal of Industrial and Management Optimization, 2023, 19(11):7993-8028.

[6] Cao T, Wu H, Ji T .Bioinformatics-based construction of prognosis-related methylation prediction model for pancreatic cancer patients and its application value[J]. Frontiers in Pharmacology, 2023, 14.

[7] Yang F, Fan J .Construction of OBE Concept Autonomous Learning Mode in University Teaching Based on the Internet[J]. J. Cases Inf. Technol. 2022, 24:1-20.

[8] Norrdine A, Motzko C .An internet of things based transportation cart for smart construction site[C]//IEEE International Conference on Green Computing and Communications; IEEE International Conference on Internet of Things; IEEE International Conference on Cyber, Physical and Social Computing; IEEE International Conference on Smart Data; IEEE Congress on Cybermatics.2020.

[9] Sodangi M .Modeling the Constraints to the Utilization of the Internet of Things in Managing Supply Chains of Off-Site Construction: An Approach toward Sustainable Construction[J]. Buildings, 2022, 12.

[10] Ali Y, Saad T B, Rehman O U .Integration of IoT technologies in Construction Supply Chain Networks; CPEC a Case in Point[J]. Sustainable Operations and Computers, 2020.

[11] Rane S B, Thakker S V .Green procurement process model based on blockchain–IoT integrated architecture for a sustainable business[J]. Management of Environmental Quality An International Journal, 2019, ahead-of-print (ahead-of-print).

[12] Katsma C P, Moonen H, Hillegersberg J .Supply Chain Systems Maturing Towards the Internet-of-Things: A Framework[J]. 2022.

[13] Peraković, Dragan; Periša, Marko; Zorić, et al. Challenges and Issues of ICT in Industry 4.0[C]//International Conference on Design, Simulation, Manufacturing.2020.

[14] Abbasi W A, Wang Z, Zhou Y, et al. Research on measurement of supply chain finance credit risk based on Internet of Things[J]. International Journal of Distributed Sensor Networks, 2019, 15(9):155014771987400.