2025 3rd International Conference on Data Science and Network Security (ICDSNS)

Warehouse Volume Forecasting and Warehouse Segmentation Strategy based on Time Series and Machine Learning Models

Qihao Chen
College of Education
Zhejiang University of Technology
Zhejiang, China
202205720103@zjut.edu.cn

2025 3rd International Conference on Data Science and Network Security (ICDSNS) | 979-8-3315-3679-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICDSNS65743.2025.11168666

***Abstract*—With the booming development of e-commerce, the competition of e-commerce enterprises has intensified, and accurate demand forecasting and reasonable inventory strategy have become the key to improve core competitiveness. In this paper, we focus on e-commerce warehousing, construct ARIMA, decision tree and LSTM models to predict the inventory and daily sales of 350 categories in the next three months, and select the optimal prediction model for each category based on the Mean Absolute Percentage Error (MAPE) to lay a data foundation for warehousing allocation. Then, the target planning models of "one warehouse for one product" and "multiple warehouses for one product" are constructed respectively, and solved by genetic algorithms, taking into full consideration of the warehouse rental cost, warehouse capacity and capacity utilisation rate, category relevance and other factors, to arrive at the corresponding optimal warehousing scheme. These solutions help e-commerce enterprises to optimise the allocation of warehouse resources, reduce operating costs and improve warehouse management efficiency while meeting market demand.**

Keywords—e-commerce warehousing, time series model, machine learning model, binning strategy, genetic algorithm.

I. INTRODUCTION

With the development of e-commerce, online shopping has become an important way of consumption due to its convenience, wider range of choices, and less geographical influence. At the same time, e-commerce enterprises are becoming increasingly competitive, and supply chain-driven operations are accepted by more and more merchants, so accurate demand forecasting and the development of reasonable inventory strategies have become an important way to improve the core competitiveness of e-commerce enterprises [1]. Consumers will generate a large amount of online behavioural data in the process of online shopping, and through machine learning methods, we can discover the consumers' willingness to buy, and then make accurate predictions of the demand for commodities, so as to guide e-commerce enterprises to formulate a reasonable inventory strategy, reduce operating costs, and improve their core competitiveness [2].

The warehouse cluster composed of warehouses bears the pressure of e-commerce commodity storage in the region, facing problems such as large differentiation of commodity categories and large number of pieces. In this paper, we develop a reasonable binning strategy by predicting the future inventory as well as the sales volume of the storage network, and search for the optimal binning strategy under the premise of satisfying the non-violation of warehousing rules - seeking to maximise the profit. This study provides a reasonable binning strategy for warehouse clusters to cope with the impact of different needs such as multiple business objectives and different category binning needs, as shown in Figure 1.

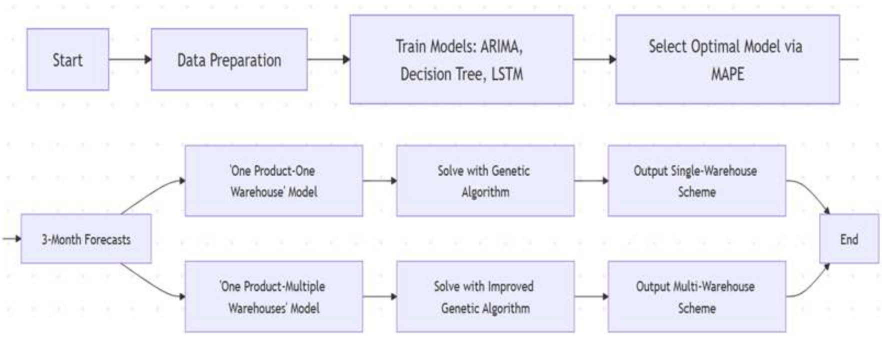

Fig. 1. Overall flow chart

II. LITERATURE REVIEW

This paper deals with demand forecasting for e-commerce platform commodities, but in real life, demand forecasting is applied to a variety of scenarios, including: spare parts and equipment demand forecasting, traffic demand forecasting, logistics demand forecasting and so on. Different scenarios of demand forecasting with reference to the impact factor will be different, the forecasting method used will also be different.

Research using machine learning methods to mine the relationship between behavioural data and demand and forecast the demand for e-commerce goods is beginning to gain attention. The study by scholars such as Alqatawna A, Abu-Salih B, Obeid N, etc. used time series forecasting techniques to predict staffing requirements and order quantities of logistic companies, and the results showed that the technique was effective in predicting the relevant data, which could be a powerful support for logistic company's human resources and business planning to provide strong

979-8-3315-3679-4/25/$31.00 ©2025 IEEE

support[3]. Ribeiro A M N C, do Carmo P R X, Endo P T and other scholars compared the application of machine learning and deep learning models in short-term and ultra-short-term enterprise-level load forecasting in warehouses, and found that different models have their own advantages and disadvantages in different scenarios, which provides a reference basis for the selection of warehouse load forecasting models[4]. Saha P, Gudheniya N, Mitra R and other scholars in the study used deep learning framework to forecast the demand of a multinational retail company, and the results show that the deep learning framework can effectively analyse the relevant data, accurately predict the market demand for multinational retail companies, assist enterprises to develop more reasonable inventory and sales strategies, and improve operational efficiency and competitiveness [5].

The purpose of inventory decision making is to develop a reasonable inventory strategy, i.e., to determine a reasonable ordering time and order quantity, so as to reduce inventory costs and improve service quality [6]. In practice, there is no certainty about the demand information, and the inventory strategy under uncertain stochastic demand requires forecasting the demand and making inventory decisions based on the results of the forecast [7,8]. For the inventory decision-making problem under stochastic demand, most of the research is based on different demand distribution laws to formulate a reasonable inventory decision-making model. Tufano A, Accorsi R, Manzini R and other scholars use machine learning methods for warehouse design prediction, and the results show that the method can optimise the warehouse design, improve the efficiency of warehouse management, and adapt to the different warehousing needs [9]. Li J focuses on improving the level of warehouse decision-making, using data-driven prediction methods and storage simulation technology to explore, and the results show that these methods can effectively use data to gain insights into the laws of warehousing, optimise the storage strategy through simulation, and provide a scientific basis for warehouse decision-making in the areas of inventory management, space utilisation, etc., and thus improve the efficiency of warehouse operations [10].

III. Proposed Method

A. *Future inventory requirements and sales volume projections*

Inventory Volume and Sales Forecasting for E-commerce Categories In e-commerce operations, accurate inventory volume and sales forecasting is critical for warehouse management and resource allocation. The goal of the first part of this paper is to predict the future inventory volume and sales volume of each category. By constructing ARIMA, decision tree and LSTM prediction models and analysing historical data, we analyse the advantages and disadvantages of different methods for different warehouses, and select the most appropriate model for each warehouse to make a prediction. This prediction not only provides a basis for subsequent warehouse allocation, but also provides data support for inventory turnover and inventory cost optimisation, thus helping e-commerce enterprises to control operating costs while meeting market demand.

Firstly, ARIMA model is built for the storage capacity of y_t at different date.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{1}$$

Where in Equation (1), ϕ_i is the coefficient of the AR term, which represents the effect of historical stock levels on current stock levels. θ_j is the coefficient of the MA term, which represents the effect of the past error on the current inventory quantity. ϵ_t is the white noise error term.

After obtaining the forecast results, the model is retrained using all available data (including the test set) to forecast stock levels for the next three months as shown in Equation (2).

$$\hat{y}_{T+k} = ARIMA(y; p, d, q) \quad (k = 1,2,3) \tag{2}$$

Next, a decision tree regression model is used to train the dataset D, which contains the feature X (e.g., historical inventory, seasonal factors, etc.) and the target variable y (inventory) as inputs; a regression tree $f(x)$ is used as output. In the input space where the training dataset is located, a binary decision tree is constructed by recursively dividing each region into two sub-regions and determining the output values on each sub-region.

Finally, in this paper, we use an LSTM neural network for training, where we input a sequence X (consisting of the inventory of the last months) to train the model in order to minimise the prediction error. The objective is to minimise the following loss function as shown in Equation (3).

$$\mathcal{L} = \frac{1}{m} \sum_{t=1}^{m} (y_t - \hat{y}_t)^2 \tag{3}$$

$\hat{y}_t$ is the predicted value of the model at time t and m is the sample size. After training is complete, the final hidden state h_T , is used to predict the inventory for the next three months.

B. *"One product, one warehouse "distribution programme planning*

Based on the above predictions, this paper focuses on how to optimise the allocation of categories to warehouses within the constraints of "one item, one warehouse". The goal is to allocate each category to a single warehouse to reduce the complexity of warehouse management, while minimising the warehouse rental cost and improving the utilisation of warehouse space and capacity. To this end, an optimisation model is established with the objective of minimising the total cost, taking into account factors such as warehouse rent, warehouse capacity and capacity constraints, and solved by genetic algorithms and other optimisation tools to derive an optimal one-item-one-warehouse allocation scheme, so as to maximise the efficiency of a single-warehouse layout.

1) Decision variables: The allocation of the category i to the warehouse j is quantified using an integer plan of the type $0-1$. The variable x_{ij} is used to indicate the allocation status, when the category$i$ is allocated to the warehouse , the value of x_{ij} is 1 ; when the category i is not allocated to the warehouse j , the value of x_{ij} is 0. According to the statistics, there are 350 different categories and 140 warehouses for the category i in the warehouse . Thus, the values of i and j are obtained as shown in Equation (4):

$$i = 1,2,3, \dots, 350; j = 1,2,3, \dots, 140 \quad (4)$$

2) Objective function: Minimise total warehouse rental cost: Cost = Total warehouse rental cost + the reciprocal of capacity and capacity utilisation as shown in Equation (5).

$$Z = \sum_{j \in J} R_j \cdot \left(\sum_{i \in I} I_i \cdot x_{ij}\right) \cdot 90 + \frac{1}{\sum_{j \in J} \frac{\sum_{i \in I} I_i \cdot x_{ij}}{C_j}} + \frac{1}{\sum_{j \in J} \frac{\sum_{i \in I} S_i \cdot x_{ij}}{P_j}} \quad (5)$$

3) Constraints

1. Each category can be assigned to only one warehouse as shown in Equation (6).

$$\sum_{j \in J} x_{ij} = 1, \quad \forall i \in I \quad (6)$$

2. The total inventory of each warehouse cannot exceed its maximum capacity as shown in Equation (7).

$$\sum_{i \in I} I_i \cdot x_{ij} \le C_j, \quad \forall j \in J \quad (7)$$

3. Each warehouse's outbound capacity cannot exceed its capacity ceiling as shown in Equation (8).

$$\sum_{i \in I} S_i \cdot x_{ij} \le P_j, \quad \forall j \in J \quad (8)$$

4. Whether a product can be allocated to a warehouse depends on whether the warehouse is leased or not(y_j) is the binary variable for whether warehouse j is leased or not. Equation (9) shown below.

$$x_{ij} \le y_j, \ \forall i \in I, \forall j \in J, \qquad y_j \in \{0,1\}, \ \forall j \in J \quad (9)$$

In this paper, we will moderately consider as the following factors:

(1) Minimisation of total lease costs.

(2) Maximising the utilisation of silos and capacity.

(3) Whether constraints such as single warehouse for a single category, warehouse capacity cap, capacity cap, and warehouse lease status are met.

The genetic algorithm is used to solve the problem, including the generation of the initial population, the evaluation of the fitness function, selection, crossover, mutation, and the generation of a new generation of populations. The final output of the best individual is the optimal binning scheme.

C. Optimisation strategies under "one product, multiple warehouses" allocation

Eventually, we relax the constraint of one item per warehouse to allow categories to be stored in multiple warehouses (up to three) and introduce category relevance as a new optimisation objective. This problem aims to achieve intensive storage with high correlation between categories, which further improves the accuracy of warehouse management and logistics efficiency. On the basis of retaining the objectives of warehouse rental cost and utilisation of warehouse capacity and production capacity, a relevance penalty term is added to the model, so that high relevance categories can be allocated to the same warehouse as far as possible. By comparing different warehousing solutions and analysing the performance of each business indicator, it provides e-commerce enterprises with a multi-warehouse allocation strategy that takes into account cost and management efficiency.

In this paper, we consider adding a penalty term to the previous paper as shown in Equation (10).

$$Penalty = \lambda \sum_{i \in I} \sum_{k \in I} A_{ik} \cdot \left(1 - x_{ij} \cdot x_{kj}\right) \quad (10)$$

If $x_{ij} \cdot x_{kj} = 1$, i.e., the categories i and k are assigned to the same warehouse, then the penalty is 0.

If $x_{ij} \cdot x_{kj} = 0$, i.e. the categories i and k are not assigned to the same warehouse, then the penalty item is added by relevance A_{ik} .

The final objective function is modified to achieve minimisation of the total warehouse rental cost: cost = total warehouse rental cost + inverse of warehouse capacity and capacity utilisation + penalty term as shown in Equation (11).

$$Z = \sum_{j \in J} R_j \cdot y_j \cdot 90 + \alpha \cdot \sum_{j \in J} \frac{\sum_{i \in I} I_i \cdot x_{ij}}{C_j} + \beta \cdot \sum_{j \in J} \frac{\sum_{i \in I} S_i \cdot x_{ij}}{P_j} + \lambda \sum_{i \in I} \sum_{k \in I} a_{ik} \cdot \left(1 - x_{ij} x_{kj}\right) \quad (11)$$

At the same time, add a correlation constraint to the above constraints to ensure that highly correlated categories are assigned to the same warehouse as much as possible. The Equation (12) shown below.

$$x_{ij} + x_{kj} \le 1 + y_{ij} \quad \forall i, k \in I, j \in J \quad (12)$$

Genetic Algorithms for Introducing Penalties

Penalty terms in genetic algorithms affect the built-in function of the algorithm.

1) Evaluation of fitness function: The penalty term is used to reduce the fitness of those individuals that do not satisfy the constraint. Individuals with low fitness have a lower probability of selection because they violate the constraints, thus guiding the population in the direction of satisfying the constraints.

2) Crossover: After crossover, the generated offspring individuals may still violate the constraints. The penalty term is applied again after crossover to evaluate the fitness of the offspring and repair strategies (e.g., constraint repair) are used to adjust these individuals so that they satisfy the constraints.

3) Mutation: Adaptive mutation mechanisms are designed to impose stronger penalties in case of constraint violations in order to motivate the algorithm to gradually reduce the number of individuals violating the constraints in subsequent generations. For example, the mutation rate can be increased when an individual violates a constraint, thus increasing the probability of finding a feasible solution.

4) Generate a new generation of population: As the iteration proceeds, the penalty coefficients are gradually adjusted to encourage the algorithm to explore multiple solutions in the early stages, and the penalty is gradually increased in the later stages to promote the convergence of the population to the optimal feasible solution.

At the same time, the penalty term adds a relevant restriction to the solution process.

1. Warehouse capacity: the penalty item ensures that the selected warehouse allocation plan does not exceed the warehouse capacity limits
2. Logistics efficiency: by adding a penalty term to the fitness function, it is possible to constrain the logistics cost of the category and optimise the distribution path.
3. Stock level constraints: If the allocation scheme results in a stock level that does not meet demand, the penalty term affects the degree of adaptation and causes it to be phased out.

IV. RESULT AND CONCLUSION

This part will present the prediction results of warehouse reserves based on time series and machine learning models, as well as the optimization schemes of two warehousing division strategies : ' one product, one warehouse ' and ' one product, multiple warehouses '. Firstly, it shows the forecast of inventory and sales of 350 categories in the next three months by ARIMA, decision tree and LSTM models, including the selection basis of different category adaptation models (based on MAPE value); then, the optimal warehousing scheme obtained by solving the two strategic goal programming models through genetic algorithm is presented, including the specific configuration of the scheme and the performance of the core indicators.

A. Results of stock level projections

In this paper, three models, ARIMA, Decision Tree and LSTM, are trained on the true values to predict the inventory levels for the next two months and compare them with the true values. Without being too general, three categories of inventory are chosen for prediction.

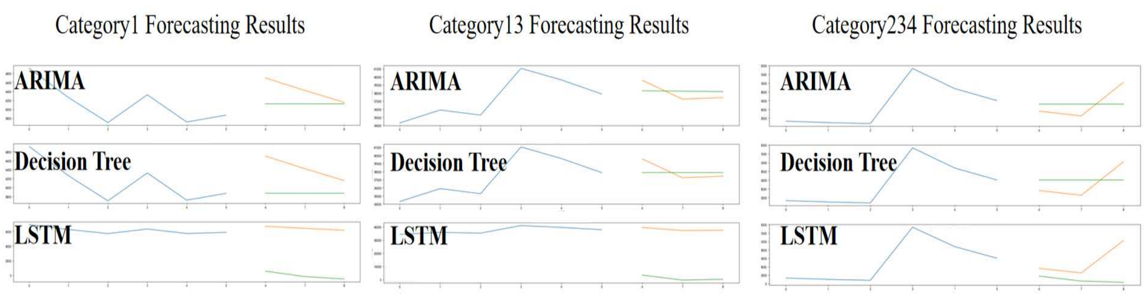

Fig. 2. Forecast results for different categories

The closer the predicted quantity is to the true value (the closer the orange line is to the green line in the figure), the more suitable the category is for the corresponding model, as shown in Figure 2.

In this paper, MAPE is introduced to calculate the percentage of absolute error between the predicted and actual values to evaluate the accuracy of the model, as shown in Figure 3.

As a classical time series model, ARIMA focuses on capturing the time series law of data, which is suitable for categories with stable trends and periodicity, but its ability to deal with nonlinear relationships is limited. The decision tree model generates rules by recursively dividing the feature space (such as historical inventory, seasonal factors, etc.).

It is highly explanatory and easy to understand, and can directly reflect the impact of various factors on inventory, but the ability to capture long-term sequences is weak. The LSTM neural network is good at dealing with long sequence dependencies. It can memorize key information through complex gating mechanisms, and has better adaptability to nonlinear and highly fluctuating data. However, the model structure is complex, the interpretation is poor, and it relies on a large amount of data training. The three models have their own emphases : ARIMA is suitable for categories with significant rules, decision tree is suitable for scenarios with clear features, and LSTM is more advantageous in complex dynamic data. This difference in characteristics provides a qualitative basis for the subsequent selection of the optimal model by category.

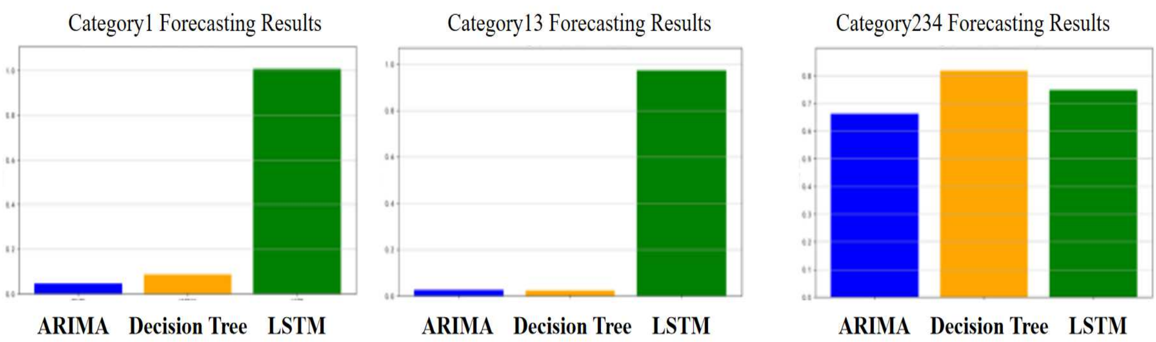

Fig. 3. Different model MAPE

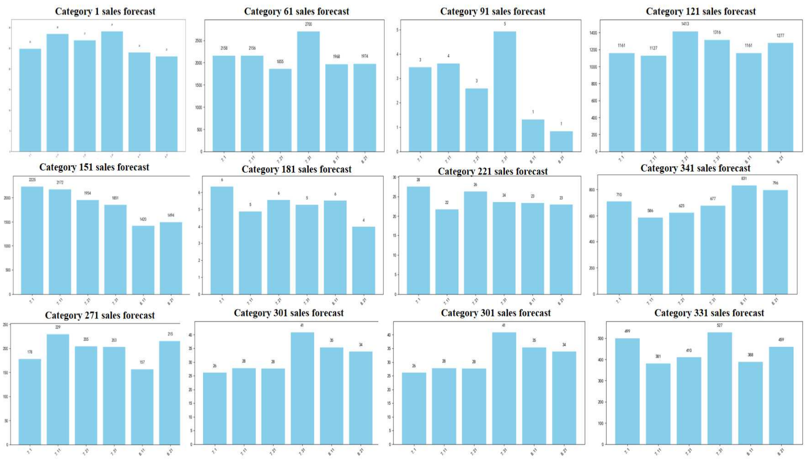

Fig. 4. Visualisation of forecast results

Eventually, all daily sales are predicted, and some of the daily sales are extracted and visualised, as shown in Figure 4.

B. Results of the "one product, one warehouse" distribution programme

Firstly, the objective function to be optimised is specified as minimising the total warehouse rental cost, and then the genetic algorithm is used to solve this objective function, initialising a population, i.e. generating a set of initial solutions, with each solution representing a warehouse allocation scheme, and then carrying out the fitness assessment, which evaluates each individual's fitness by calculating the value of the objective function. The curve of change of fitness was obtained, as shown in Figure 5.

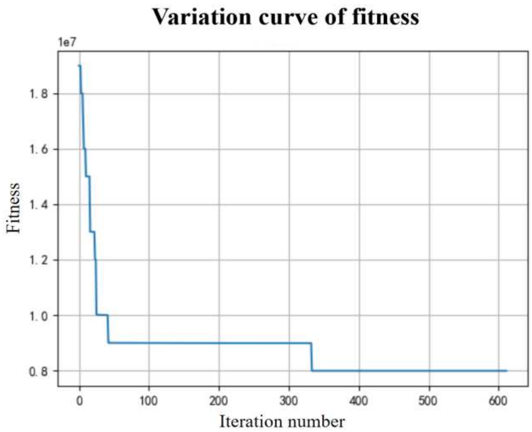

Fig. 5. Change of fitness curve

Table 1 illustrates a programme for one product, one warehousing as shown below.

TABLE I. "ONE PRODUCT, ONE WAREHOUSE" WAREHOUSING PROGRAMME

category	warehouse	category	warehouse
categoryl	warehouse41	category121	warehouse94
category31	warehouse26	category151	warehouse5
category61	warehousel	category181	warehouse65
category91	warehouse4	category211	warehouse41

C. Results of the "One Product, multiple Warehouses" distribution programme

After adding the penalty term, the solution was performed again using the improved genetic algorithm. The curve of adaptation change was obtained, as shown in Figure 6.

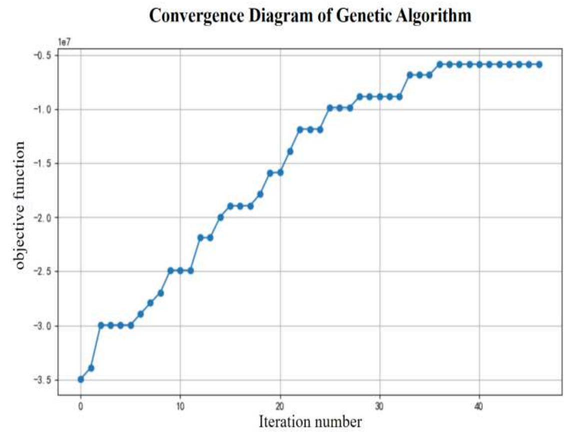

Fig. 6. Adaptation change curve

Table 2 represents a programme for one product, multiple warehousing as demonstrated below.

TABLE II. "ONE PRODUCT, MULTIPLE WAREHOUSES" WAREHOUSING PROGRAMME

category	warehouse	warehouse	warehouse
category1	warehouse127	warehouse41	warehouse20
category31	warehousel12	warehouse57	warehouse58
category61	warehouse10		
category91	warehouse124	warehouse99	warehouse103
category121	warehouse20	warehousel10	warehouse71
category151	warehouse37	warehouse68	warehouse43

V. CONCLUSION

This paper focuses on e-commerce warehousing problems, using ARIMA, decision tree and LSTM models to predict the inventory and daily sales of 350 categories in the next three months, based on the MAPE value to determine the optimal prediction model for each category, which provides a reliable data support for the allocation of warehousing. On this basis, we constructed "one warehouse for one product" and "multiple warehouses for one product" target planning models, and solved them with the help of genetic algorithms, which resulted in the corresponding optimal warehousing schemes. These solutions take into account factors such as warehouse rental cost, warehouse capacity and capacity utilisation, category relevance, etc., which can help e-commerce enterprises to effectively plan warehouse resources, reduce operating costs and improve warehouse management efficiency while meeting market demand.

The computational complexity of the model in this paper is high, and when dealing with large-scale data and multi-warehouse scenarios, the difficulty of solving increases, which affects the efficiency of real-time decision-making; the prediction model relies on historical data, and the prediction error may increase when the market environment is changing rapidly or when facing a new on-shelf category. In addition, the model is based on specific assumptions, which limits its generality, and the parameter settings have a large impact on the results, requiring repeated debugging.

Future research can introduce particle swarm optimization, ant colony algorithm, etc. in algorithm optimization to improve the speed and quality of the solution; in model improvement, integrate a variety of prediction models to improve the prediction accuracy, and adopt deep learning architectures such as Transformer to enhance the ability to process complex data; and also establish an automated parameter adjustment mechanism to achieve adaptive adjustment of parameters by using machine learning technology to adapt to different market environments, further improving model performance and better serving e-commerce warehouse management.

REFERENCES

[1] Kalkha H, Khiat A, Bahnasse A, Ouajji H. Enhancing warehouse efficiency with time series clustering: a hybrid storage location assignment strategy[J]. IEEE Access, 2024.

[2] Zohdi M, Rafiee M, Kayvanfar V, Salamiraad A. Demand forecasting based machine learning algorithms on customer information: an applied approach[J]. International Journal of Information Technology, 2022, 14(4): 1937-1947.

[3] Alqatawna A, Abu-Salih B, Obeid N, Salamiraad A. Incorporating time-series forecasting techniques to predict logistics companies' staffing needs and order volume[J]. Computation, 2023, 11(7): 141.

[4] Ribeiro A M N C, do Carmo P R X, Endo P T, Rosati P, Lynn T. Short- and very short-term firm-level load forecasting for warehouses: a comparison of machine learning and deep learning models[J]. Energies, 2022, 15(3): 750.

[5] Saha P, Gudheniya N, Mitra R, Das D, Narayana S, Tiwari MK. Demand forecasting of a multinational retail company using deep learning frameworks[J]. IFAC-PapersOnLine, 2022, 55(10): 395-399.

[6] Berns F, Ramsdorf T, Beecks C. Machine learning for storage location prediction in industrial high bay warehouses[C]//Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part IV. Springer International Publishing, 2021: 650-661. 2021: 650-661.

[7] Mashayekhy Y, Babaei A, Yuan X M, Xue A. Impact of Internet of Things (IoT) on inventory management: a literature survey[J]. Logistics, 2022, 6(2): 33.

[8] Wang C C, Chien C H, Trappey A J C. On the application of ARIMA and LSTM to predict order demand based on short lead time and on-time delivery requirements[J]. Processes, 2021, 9(7): 1157.

[9] Tufano A, Accorsi R, Manzini R. A machine learning approach for predictive warehouse design[J]. The International Journal of Advanced Manufacturing Technology, 2022, 119(3): 2369-2392.

[10] Li J. Improving Decision Making in Warehouse: Data-Driven Forecasting and Storage Simulation [D]. University of Twente, 2023.