Research on E-commerce Category Stock Prediction and Warehouse Management Using Model Integration and Genetic Algorithms

# Author

Sheng Gou

Chengdu Neusoft University, 23016021322@stu.nsu.edu.cn

Cheng Gao

Chengdu Neusoft University, 23016021616@stu.nsu.edu.cn

Zhongxia Li

Chengdu Neusoft University, 3380634425@qq.com

Xinru Wang

Qingdao Agricultural University, 2537588095@qq.com

Shixin Liang

Tianjin University of Finance and Economics Pearl River College, 2870138812@qq.com

Ziwei Shang\*

Chengdu Neusoft University, shangziwei@nsu.edu.cn

# abstract

This study proposes a comprehensive solution to the inventory forecasting and warehouse management problems encountered by e-commerce companies in their rapid development. The study firstly forecasts the inventory and sales of e-commerce categories by integrating multiple time series forecasting models, including ARIMA, gray forecasting model, decision tree regression, XGBoost and LSTM. Using the sliding window technique and MAPE assessment method, the model with the smallest error was selected, and the average MAPE value of the final model was 0.1051, which showed good forecasting accuracy. Next, the study constructed a multi-objective optimization model, which considered factors such as capacity utilization, total rental cost, number of warehouses, and category relevance to achieve efficient allocation of warehousing resources. In addition, the study proposes a new category warehousing scheme that allows each category to be stored in up to three warehouses, aiming to maximize category relevance and ensure that similar items are stored centrally. The optimal warehouse allocation scheme is realized by solving the model with a genetic algorithm. Finally, the study conducted a sensitivity analysis and an error analysis of the model, and presented the strengths and weaknesses of the model as well as directions for improvement. The solution of this study can significantly improve the accuracy of warehouse prediction and management efficiency, and provides powerful decision support for warehouse management in e-commerce enterprises.

# Ccs concepts

**·Information systems**→Information systems applications;

# keywords

Model fusion;time series;genetic algorithm;multi-objective optimization.

# 1 introduction

With the rapid development of e-commerce enterprises, accurate warehousing volume forecasting and warehouse management as a key link in the supply chain, is increasingly being emphasized. How to predict the cargo volume and reasonably allocate the commodity categories to different warehouses to improve management efficiency and reduce the cost of warehousing has become a concern for e-commerce users in various complex indicator situations and warehouse networks. Therefore, the establishment of a reasonable planning model and forecasting model to solve these problems can play a "cost reduction and efficiency" role. In response to the first question, the e-commerce company's warehouse network is required to establish an accurate cargo forecasting model to predict the inventory and sales volume of each category in the next three months (July-September). Considering that the sales data of most categories from 2022/10/1 to 2023/3/31 are non-existent, or even in some categories, the data is very small, with a large percentage of missing values, so the interpolation is unreasonable, so we choose to use 91 consecutive days from 2023/4/1 to 2023/6/30 to forecast respectively. Since the topic involves the prediction of multiple categories, seasonal fluctuations, promotional activities and other factors, the prediction effect of a single model in this complex situation is not good, the inclusion of factors is not comprehensive, and the generalization ability of the model is relatively low. Therefore, we used various time series forecasting models including ARIMA, gray forecasting model, decision tree regression, XGBoost, and LSTM[4,8]. by dividing the training and test sets with historical data for each category and constructing features using sliding window technique, the model calculates the MAPE on the test set in order to select the one with the smallest error. For the ARIMA model the optimal ARIMA model parameters (p, d, q) were automatically selected using the auto\_arima function in the pmdarima library and the forecasts of inventory and daily sales for the next three months were made based on the best model. After the model selection and evaluation, the final average MAPE reaches 0.1051, so the model performs better in prediction accuracy and this result is acceptable.

For Problem 2, a multi-objective optimization planning model is required to perform warehouse operations for 350 categories under the constraint of "one product, one warehouse" and taking into account other metrics. Due to the large number of constraints and the large number of warehouses, we consider using a heuristic genetic algorithm[9,10]. Due to the need to consider the economic benefits, the genetic algorithm has a very good global search ability, can avoid falling into the local optimal solution, through the setting of the relevant parameters, to achieve the optimal solution, the final results are shown in detail in the result.

For problem 3, since the problem is based on the derivation of problem two, from a product of a warehouse into a product can be more than one warehouse conditions, the maximum category relevance as the focus of the optimization objective and take into account the other indicators. metrics. We introduce the category relevance matrix. In the objective function, a penalty term is added to ensure the feasibility of the solution. The model is then solved with a genetic algorithm based on the idea of Problem 2. The optimal solution is achieved by adjusting the two important parameters, the number of hereditary generations and the mutation rate, and the final results are detailed in RESULTS.

Finally, we conducted a sensitivity analysis[11] of the model, mainly for the number of hereditary generations and the mutation rate in the genetic algorithm, and also conducted an error analysis, in addition, we also carried out a model generalization and evaluation for the present model, and listed out the strengths and weaknesses of the model, as well as the ideas for improvement.

# 2 results and discussion

## 2.1 problem1

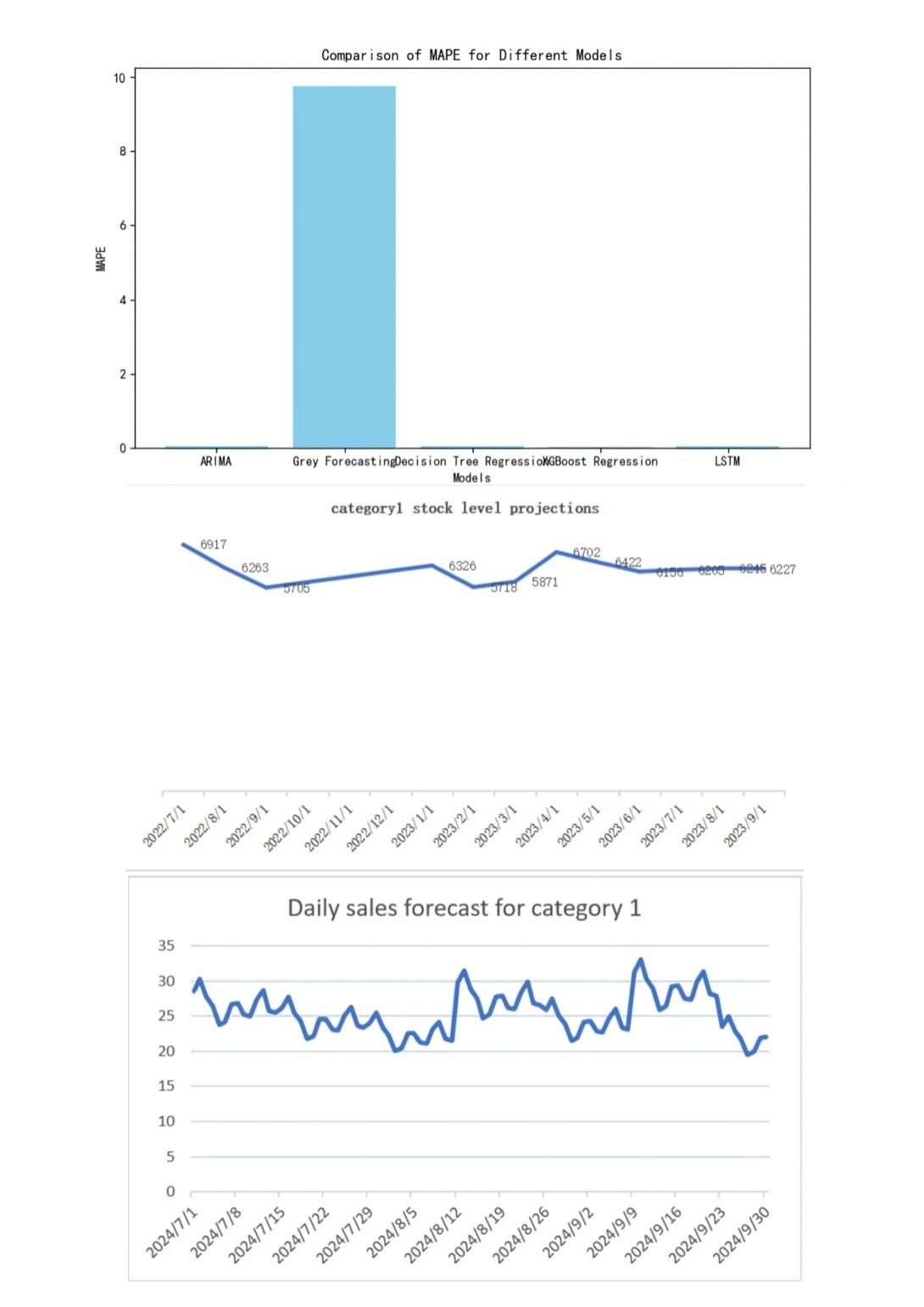
In Problem 1, our main task is to forecast the average monthly inventory and daily sales of 350 commodity categories for the next three months (July-September) to support warehouse resource planning and category warehousing program development. Accurate forecasts will help companies anticipate future needs, plan warehouse space, and outbound capacity, reducing resource redundancy and operating costs.

Given the trend, seasonal, and non-linear characteristics of inventory and sales data, we use five models: ARIMA, gray prediction model, Decession Tree Regression, XGBoost, and LSTM [1,3,4,6,8]. The ARIMA model handles linear trends and seasonal fluctuations, capturing time correlations between inventory and sales. The gray prediction model is suited for small sample time series, simplifying complex trends, especially for categories with minor changes in inventory. Decision tree regression and XGBoost identify non-linear features and irregular sales fluctuations. LSTM models manage long-term dependencies in long-series data, capturing sales fluctuations effectively.

Due to missing sales data from 2022/10/1 to 2023/3/31 and minimal data for some categories, interpolation is unreasonable. Thus, we use fusion models to balance warehouse capacity utilization, capacity, and total storage cost, meeting future space and outgoing requirements. By optimizing parameters for each category, we aim to provide a basis for centralized or decentralized storage decisions.

Given the complexity of the prediction problem, a single model is insufficient. Therefore, we propose using a fusion model to improve prediction accuracy and comprehensiveness.

We used five models (ARIMA, Gray Prediction, XGBoost, LSTM,Decession Tree Regression and Decision Tree Regression) to predict the monthly inventory and daily sales of category1 from 2023/4/1 to 2023/6/30. The historical data was split into training and test sets, and the Mean Absolute Percentage Error (MAPE) was calculated for each model. The model with the lowest MAPE was selected as the best prediction model for each category. Using the best model, we predicted the inventory and daily sales for the next three months. The results are stored in RESULTS, with a prediction plot for category1. The fusion model achieved an average MAPE of 0.1051, demonstrating satisfactory performance for this complex time series data, which includes multiple categories, seasonal fluctuations, and promotional activities.

Figure 1: MAPE comparisons and category1 inventory and sales forecasts for different models.

## 2.2 problem2

The objective of Problem 2 is to establish a multi-objective optimization model to determine warehouses for each category under the "one item, one warehouse" constraint. The model considers capacity utilization, total rental cost, number of warehouses per category, and category relevance to meet management needs. Binary decision variables specify category and warehouse allocation, with constraints ensuring feasibility. A genetic algorithm is used to explore the solution space and find the global optimal solution, improving warehouse resource management, logistics efficiency, and reducing operational costs [9,10].

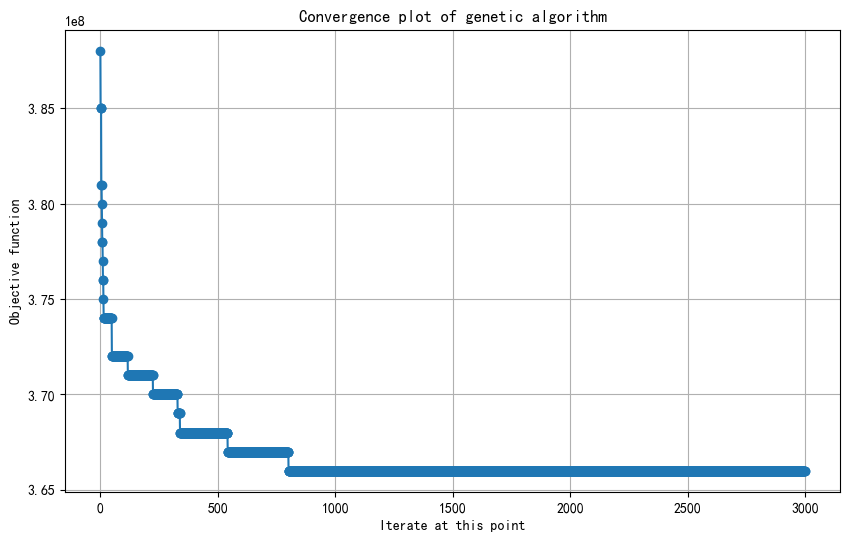
After generations of selection, crossover, and mutation, the population is updated, retaining the best-adapted individuals until a predefined number of generations is reached or fitness no longer significantly increases. Each category is then assigned to the most suitable warehouse configuration based on the coincident medium metric, ensuring rational storage and discharge.

Figure2: Convergence polt of genetic algorithm

The convergence graph shows that the genetic algorithm converged in about 1756 iterations. The objective function value stabilizes, indicating a relatively stable and suitable solution for the current warehouse planning problem. This solution meets constraints such as warehouse and production capacity, allowing us to solve for specific results.

## 2.3 problem3

In Question 3, our goal is to establish a new category warehousing scheme based on previous predictions, allowing each category to be stored in up to three warehouses. We aim to maximize category relevance and ensure that similar items are stored centrally. The model also considers warehouse capacity utilization. We define decision variables and objective functions including warehouse allocation, inventory, sales volume, and a relevance matrix. Constraints ensure that each warehouse operates within its capacity and that the allocation scheme is feasible.

Since Problem 3 builds on Problem 2, it can also be solved using genetic algorithms. The goal is to assign products to multiple warehouses, with no more than three per product. To optimize economic efficiency, similar items and high-level categories should be stored together, maximizing category relevance and considering other key indicators. This approach makes warehousing management more economical and rational for e-commerce enterprises.

After many generations of selection, crossover, and mutation, the population is updated, retaining the best-adapted individuals until a preset number of generations is reached or fitness no longer significantly improves. This ensures maximum category relevance while considering other metrics to solve the optimal binning scheme.

# 3 methods

## 3.1 Modeling and Solving Problem 1

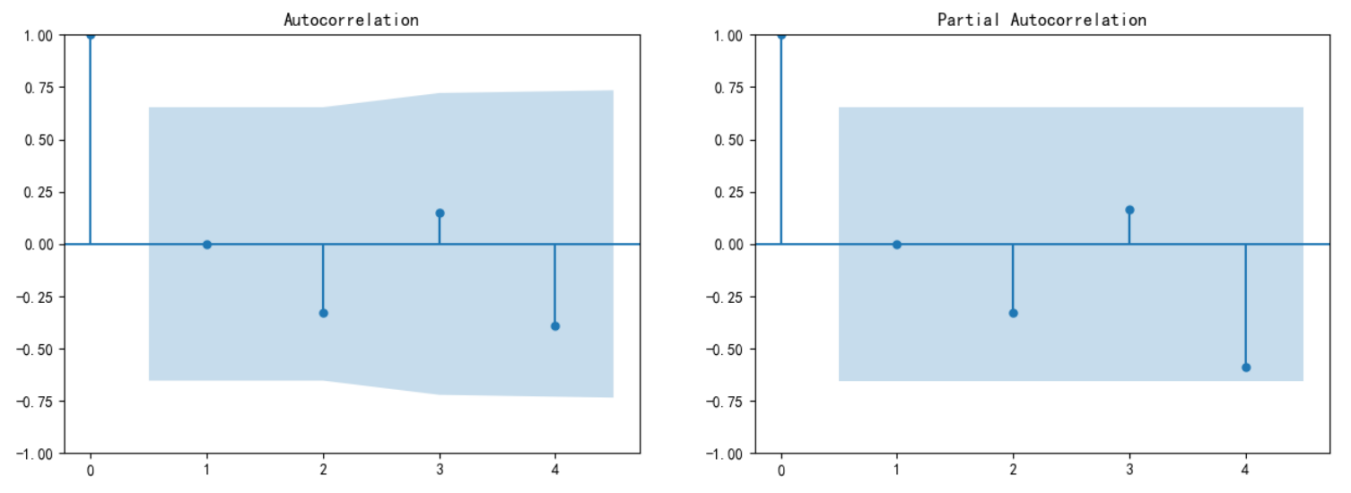
3.1.1 Data preprocessing.

To efficiently process time-series data for machine learning models, this study uses a sliding window approach to preprocess historical inventory and sales data. This method divides the time series into fixed-length windows, using the data within each window to predict future values[3]. First, we determine the window length based on the category's needs and data characteristics. Each sliding window includes data from the past day to predict the next moment's inventory levels or sales. The current window's data serves as model input, while the subsequent moment's data is the prediction target. This generates training samples for the model to learn the relationship between inventory levels and sales.

Finally, we extract features and targets from all time series data, dividing the dataset into a training set for model fitting and a test set for evaluating the model's generalization ability.

3.1.2 ARIMA model.

The time series model in this paper requires weakly smoothed data. For volatile data, smoothing should be done using differencing. The operation of differencing "d" times is called "d-order differencing," where each order builds on the previous one. The differencing should be balanced—not too few to ensure smoothness, and not too many to avoid overfitting. The data's suitability is judged by the average values of the autocorrelation and partial autocorrelation functions, stopping the differencing when these averages approach zero, indicating stability.

 Figure 3: Autocorrelation and partial autocorrelation plots.

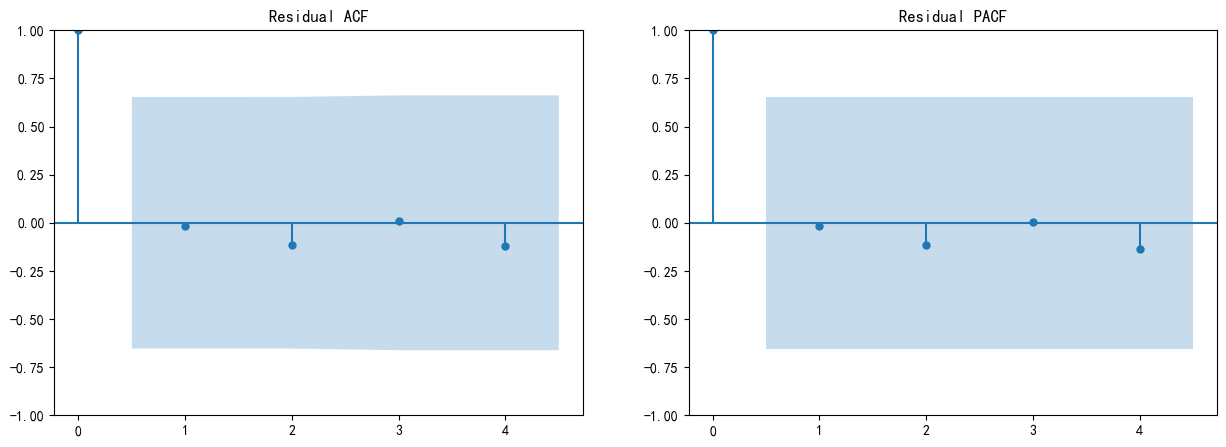


Figure 4: Residual ACF and Residual PACF plots.

Based on the requirements of Question 1 and data from Annexes 1 and 2, we combined time series images of inventory and sales for category1, category2, and category3. The chart shows that inventory and sales volumes differ across categories, with significant fluctuations in their trends. Due to the large sales volume data, we focus on category1 for relevant predictions.

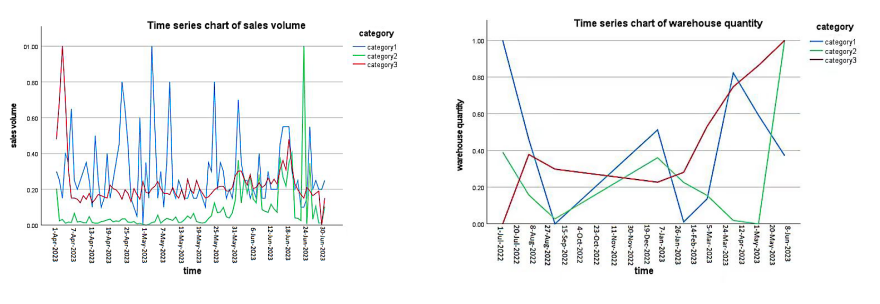


Figure 5: Residual ACF and Residual PACF plots.

The time series model describes the change of an independent variable over time, using the day as the time variable. We consider four components: long-term trend (T), seasonal trend (S), cyclical trend (C), and perturbation term (I). The model can be additive or multiplicative, based on the independence and smoothness of the time series. If the series is independent and smooth, we use an additive model; otherwise, we use a multiplicative model. Observing the sales and inventory changes for category1, category2, and category3 in Fig4, we see significant seasonal and inventory fluctuations, likely due to stagnant sales in certain seasons. Thus, we use a multiplicative model to account for cyclical volatility and overall stability.

Fig. 4 shows significant seasonal fluctuations in inventory and warehouse volume for category1, category2, and category3, likely due to stagnant sales in certain seasons. We consider cyclical volatility and general stability as key factors, opting for a multiplicative model. We use an ARIMA(p,d,q) model, requiring d-order differencing. White noise testing of residuals assesses model validity; if residuals are white noise, the model fits well, otherwise, it needs improvement. Choosing appropriate p, d, and q values is crucial, and we use the auto\_arima function to optimize these parameters, fitting the model for multi-step inventory and sales forecasts.

3.1.3 Gray prediction model.

Gray Modeling (GM) uses available information to predict future system trends from incomplete or uncertain data by reducing randomness through gray generation. The main steps are:1. Given a time series.2. Generate a cumulative series.3. Construct a matrix of background values and parameters.4. Calculate parameters using the least squares method.5. Build a prediction modell

3.1.4 Decision tree regression model.

Decision trees capture nonlinear relationships by dividing rules to predict inventory levels or sales. Parameter tuning adapts the model to complex changes in different categories. They effectively handle complex inventory patterns and perform well with fluctuating warehouse capacity. The core task is recursively splitting data, choosing split points to minimize mean square error (MSE). The predicted value at each leaf node is the average of all samples within that node. The optimal split point minimizes the MSE of the divided subset by calculating the post-partition MSE sum for each feature and its value.

3.1.5 XGBoost regression.

XGBoost [3] identifies complex non-linear patterns in inventory and sales data through ensemble learning, uncovering hidden relationships between them. It accurately predicts highly volatile or strongly seasonal category data in complex warehouse scheduling environments.

3.1.6 LSTM.

LSTM [6,7] is suitable for handling long time-dependent sequence data and can utilize nonlinear features to predict daily sales, particularly for categories with high inventory demand fluctuations. It manages long histories of sales and inventory changes for accurate future demand forecasts. Key components of LSTM include:1. Forget gate: controls retention of information in the memory cell.2. Input gate: decides to add new information to the memory cell.3. Memory cell update.4. Output gate: decides the final output.

## 3.2 Modeling and Solving Problem 2

3.2.1Denetic algorithm.

In the process described above, a warehouse is first selected for each category, ensuring that the warehouse's capacity and capacity constraints are met.

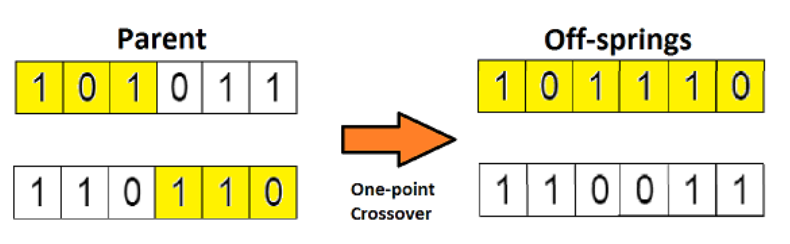
denotes the binary decision variable of whether category i is assigned to warehouse j or not. = 1 if category i is assigned to warehouse j, and = 0 if category i is not assigned to warehouse j.

In each iteration of the genetic algorithm, individuals are evaluated using a fitness function. Higher fitness scores indicate better solutions, which are more likely to be selected for reproduction. As the algorithm progresses, fitness improves, and the algorithm terminates once a satisfactory fitness value is achieved.

Initialize Population:Generate an initial set of individuals, each of which is an array of length of the number of categories, with each element being the index of the assigned repository.

Common selection methods in genetic algorithms[9,10] include the roulette method, tournament method, and truncated selection method. The tournament method is ideal for our one bin one class problem as it reduces the impact of randomness by comparing randomly selected individuals rather than relying on absolute fitness values.

This paper uses single-point(in Fig5) crossover due to its simplicity and efficiency. By exchanging parent information at a single crossover point, it effectively combines features, maintains individual structure, reduces noise, and is computationally efficient, making it suitable for handling large data sets.

The time series model describes changes in the independent variable over time, using the day as the time variable. We consider four components: long-term trend (T), seasonal trend (S), cyclical trend (C), and perturbation term (I), as shown in Fig. 3. We can use either an additive or multiplicative model based on the independence and smoothness of the time series. Independent and smooth series use an additive model; otherwise, we consider a multiplicative model. Observing sales and inventory changes for category1, category2, and category3 in Fig. 1, we note significant seasonal fluctuations and inventory variations, likely due to stagnant sales in certain seasons. Therefore, we assume cyclical volatility and overall stability are key factors and choose a multiplicative model.

.

Figure 5: Schematic diagram of a single-point crossover

Selection of mutant individuals:1. randomly select an individual I from the current population for mutation.2. randomly select a gene position k, ranging from 1 to 350, indicating the selected category.3. perform the mutation operation, and for the selected gene position k, flip its value: if =0, change to 1. if =1, change to 0.4. return the mutated individual I'.

## 3.3 Modeling and Solving Problem 3

3.3.1 Dynamic planning model with one product and multiple warehouses.

Decision variable we define the warehouse allocation to indicate whether the category is allocated to the warehouse , if it is allocated, then =1, otherwise =0. The warehouse allocation indicates the amount of storage of the category in the warehouse .

Objective Function - Maximize Correlation

Where: category C=: contains the set of all categories,W= denotes the set of all warehouses, inventory denotes the maximum inventory of category , sales Si denotes the average daily sales of category ,warehouse capacity limit denotes the storage limit of warehouse , capacity limit denotes the daily outflow limit of warehouse , warehouse rent daily cost denotes the daily rent of warehouse , category correlation matrix denotes the correlation between category and.

Category Relevance Matrix

If >0 means that category and category ck are related, <0 means that category and category are not related, we usually consider that the relevance satisfies the following table:

|  |  |
| --- | --- |
| correlation coefficient | correlation interval |
|  | high |
|  | middle |
|  | low |
|  | null |

Constraints: 1. The total inventory of each warehouse must not exceed its maximum capacity. 2. 2. the daily outgoing quantity of each warehouse must not exceed its capacity limit. 3. each category is assigned to a maximum of 3 warehouses. 3. Each category is assigned to a maximum of 3 warehouses. 4. 4. When a category is assigned to a warehouse, the corresponding stock level must be greater than 0. 5. Only categories with similar piece types or high-level categories may be stored in the same warehouse. 6. 6. The stock level of each category in the warehouse must be a non-negative integer. 7. the warehouse allocation variable is a binary variable indicating whether the category is allocated to a warehouse or not. We sort out the constraints to get the entire planning model:

|  |
| --- |
|  |
|  |

3.3.2 Genetic Algorithm Solving Models.

Based on the introduction to the genetic algorithm, we can optimize by focusing on maximum category relevance. Using the model from the previous section, we solve the problem with the following steps:1. Use the genetic algorithm for solving.2. Apply the same parameter settings as in the second question.

|  |  |
| --- | --- |
| parameters | Numerical Selection |
| Population size | 100 |
| generations | 3000 |
| Mutation rate | 0.01 |
| tournament size | 5 |
| Penalty factor | 1e6 |

Individual Codes Each individual code is a list of length 350 (the number of categories), with up to 3 warehouses assigned to each category. The list element indicates the warehouse to which the category is assigned, unassigned is indicated by -1. Fitness Function Fitness = Total Correlation - Penalty Term Total Correlation Calculation: maximizes the correlation between categories. The penalty term is calculated[9,10] as follows: P1: If the inventory of a warehouse exceeds its capacity limit P2: If the outgoing quantity of a warehouse exceeds its capacity limit P3: If the number of assignments of a category in a warehouse exceeds 3 The penalty term is used to ensure the feasibility of the solution.

1. Individual representation: Each individual is a binary sequence or array of length 350, indicating whether the category is assigned to a warehouse. 2. Select Parents: Randomly select two individuals as parents from the current population. 3. 3. Select crossover point: randomly select a crossover point k, ranging from 1 to N-1, and generate two children from the crossover point. 4. Generate offspring: Generate two offspring by crossover and add them to the new population. 5. Mutation operation: Select mutation individual: randomly select an individual from the current population for mutation. Setting mutation rate: decide whether to mutate each gene of an individual. Check mutation gene by gene: check each gene and perform mutation operation if the mutation conditions are met. Mutation operation: randomly selects a new repository number and assigns it to the gene.[9,10]

## 3.4 Methodology and testing of the model

Due to the large number of parameters of the genetic algorithm, we refer to the literature and find that: the most important parameters affecting the algorithm are: the mutation rate and the number of genetic generations. Considering the waste of time and space and the large overhead of computational resources caused by the tuning parameter, we choose these two important parameters for sensitivity analysis.[11]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Population size | Mutation rate | generations | Penalty facto | Convergence number |
| 100 | 0.01 | 1000 | 1e6 | 913 |
| 100 | 0.01 | 2000 | 1e6 | 1367 |
| 100 | 0.01 | 3000 | 1e6 | 1678 |
| 100 | 0.01 | 4000 | 1e6 | 3024 |

The setting of the genetic generation has a significant impact on the convergence speed and quality of the genetic algorithm's solution, when the genetic generation is set to 1000, the genetic algorithm converges around about 913 generations; while when the maximum generation is increased to 3000, the convergence occurs at about 1678 generations, which may be due to the fact that the algorithm retains better-adapted individuals within more generations, thus improving the convergence efficiency at an early stage. An appropriate maximum number of generations can balance the global exploration and local exploitation of the algorithm, thus finding a better balance between cost and warehouse utilization, and effectively improving the overall efficiency of warehouse management.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Population size | Mutation rate | generations | Penalty facto | Convergence number |
| 100 | 0.01 | 3000 | 1e6 | 815 |
| 100 | 0.05 | 3000 | 1e6 | 1256 |
| 100 | 0.075 | 3000 | 1e6 | 2376 |
| 100 | 0.1 | 3000 | 1e6 | 2814 |

Based on the sensitivity analysis of the above parameters, the number of convergence iterations of the algorithm gradually increases from 815 to 2814 as the variability increases from 0.01 to 0.1. This indicates that a lower variability of 0.01 can reach convergence more quickly but may result in a lack of diversity in the solutions, whereas a higher variability of 0.1 increases the range of exploration of the solution, allowing the algorithm to find possible optimal solutions, but requires more iterations. For this problem, a balance needs to be found between higher solution diversity and convergence efficiency to ensure that the model can both converge quickly and find the optimal binning scheme while satisfying constraints such as bin capacity and production capacity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Population size | Mutation rate | generations | Penalty facto | Convergence number |
| 100 | 0.01 | 1000 | 1e6 | 856 |
| 100 | 0.01 | 2000 | 1e6 | 1645 |
| 100 | 0.01 | 3000 | 1e6 | 796 |
| 100 | 0.01 | 4000 | 1e6 | 2130 |

As the number of genetic generations increases from 1000 to 4000, the number of convergence iterations of the algorithm exhibits unstable fluctuations: it is 856 at generation 1000, increases to 1645 at generation 2000, but significantly decreases to 796 at generation 3000, and then rises to 2130 at generation 4000. For Problem 3, the appropriate number of genetic generations should be able to balance the quality of the solution with the computational efficiency to ensure the optimality of the binning scheme and the reasonableness of the computational time while satisfying the constraints of storage capacity and production capacity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Population size | Mutation rate | generations | Penalty facto | Convergence number |
| 100 | 0.01 | 3000 | 1e6 | 742 |
| 100 | 0.05 | 3000 | 1e6 | 1576 |
| 100 | 0.075 | 3000 | 1e6 | 846 |
| 100 | 0.1 | 3000 | 1e6 | 2301 |

As the variation rate increases from 0.01 to 0.1, the number of convergence iterations shows fluctuating changes: the convergence is faster when the variation rate is 0.01, only 742 iterations are needed; it increases to 1576 at 0.05; when the variation rate is 0.075, the number of convergence iterations is reduced to 846; but when the variation rate reaches 0.1, the number of convergence iterations is greatly increased to 2301. This indicates that a moderate variability rate can increase the diversity of solutions and help avoid local optimization.

# 4. Evaluation and Improved Extension of Models

## 4.1 Advantages of the model

The ARIMA model handles linear trend time series data well and requires few exogenous variables.The Gray prediction model is suitable for small samples and incomplete data, with simple calculations.Decision tree regression is easy to understand, interpret, and handles nonlinear relationships.XGBoost regression is powerful, manages nonlinear data well, and effectively prevents overfitting.The LSTM model captures long-term dependencies in time series, suitable for complex forecasting.The genetic algorithm has strong global search ability, doesn't rely on gradient information, is suitable for complex, discontinuous optimization problems, and supports parallel computing. Its stochasticity and diversity help avoid local optima and it can be combined with other algorithms for enhanced effects, making it suitable for various optimization scenarios.

## 4.2 Disadvantages of the model

ARIMA performs poorly on nonlinear data and has complex parameter selection.Gray prediction models are ineffective for long time series and rapidly changing data.ecision tree regression is prone to overfitting and may produce unstable predictions.XGBoost regression has a long training time and complex parameter tuning.LSTM requires high data volume, computational resources, and has long training times.Genetic algorithms have slow convergence and high computational costs due to many objective function evaluations.

## 4.3 Improvements to the model

Improvements in genetic algorithms focus on enhancing search efficiency, avoiding premature convergence, and increasing diversity. Key methods include:Adaptive parameter tuning: Dynamically adjusting population size, crossover rate, and mutation rate to meet different search needs.Elite retention strategy: Ensuring individuals with higher adaptability are preserved to maintain optimal solutions.Combining selection strategies: Using methods like roulette and tournament selection to enhance diversity and stability.Multi-population co-evolution: Parallel searching in different regions with regular individual exchanges to improve global search capability and reduce premature convergence risk.Enhanced crossover and mutation operations: Designing smarter operations to explore solution space more effectively.Refined local search and heuristic strategies: Performing local optimization based on genetic algorithm solutions and using prior knowledge to initialize the population and reduce search space.These improvements help genetic algorithms find optimal solutions faster and more reliably, especially in complex optimization problems.

## 4.4 Extension of the model

The model combines ARIMA, gray forecasting, decision tree regression, XGBoost, and LSTM to select the optimal model for data prediction, enhancing accuracy but increasing computational complexity and time.Genetic algorithms are crucial in:Machine learning: Feature selection, hyperparameter optimization, neural network structure optimization, and AutoML, reducing manual design workload.Computer vision: Feature selection and extraction, optimizing segmentation parameters, evolutionary adjustment of image processing algorithms, and improving template matching and classifier performance.

Overall, genetic algorithms enhance accuracy and efficiency in image analysis and technological development.

# 5 conclusion

In Problem 1, our main task is to predict the average monthly inventory and daily sales of 350 items for the next three months (July through September). We used five models, ARIMA, Gray Prediction Model, Decision Tree Regression, XGBoost, and LSTM, to deal with the trending, seasonal, and nonlinear characteristics of the inventory and sales data. Due to missing data and implausible interpolation, we used a fusion model to balance warehouse capacity utilization, capacity, and total storage cost. The average MAPE of the fusion model is 0.1051, which indicates that the fusion model performs satisfactorily for complex time-series data containing multiple categories, seasonal fluctuations, and promotions.

In Problem 2, our objective is to build a multi-objective optimization model to determine the warehouse for each category under the constraint of “one item, one warehouse”. The model considers capacity utilization, total leasing cost, number of warehouses per category, and category relevance. A genetic algorithm is used to explore the solution space and find a globally optimal solution. The genetic algorithm converges after approximately 1,756 iterations and the objective function is stable, suggesting that this is a relatively stable and suitable solution for the warehouse planning problem at hand.

In Problem 3, our goal is to build a new categorization warehousing scheme based on previous predictions that allows each categorization to be stored in up to three warehouses. We aim to maximize the relevance of the categories and ensure that similar items are stored centrally. The model also considers warehouse capacity utilization. After multiple generations of selection, crossover and mutation, the population is renewed, retaining the most adapted individuals until a predefined number of generations is reached or adaptation is no longer significant. This ensures that category relevance is maximized while other metrics are considered to solve the optimal binning scheme.

By exploring these three issues, we not only improve the accuracy of inventory forecasting, but also optimize the allocation of warehouse resources and provide an effective warehouse management strategy for e-commerce enterprises. Future research will be devoted to further optimizing the model parameters, improving the accuracy of forecasting, and exploring more innovative warehouse management strategies.

# 6 references

[1]Bai, L., Lu, K., Dong, Y. *et al.* Predicting monthly hospital outpatient visits based on meteorological environmental factors using the ARIMA model. *Sci Rep* **13**, 2691 (2023). <https://doi.org/10.1038/s41598-023-29897-y>

[2]Zahra Zamanzadeh Darban, Geoffrey I. Webb, Shirui Pan, Charu Aggarwal, and Mahsa Salehi. 2024. Deep Learning for Time Series Anomaly Detection: A Survey. ACM Comput. Surv. 57, 1, Article 15 (January 2025), 42 pages. <https://doi.org/10.1145/3691338>

[3]Yuzhe Huang, Linrui Wu, Hao Wang, and Zhiyi Li. 2024. Commodity intelligent pricing and replenishment decision based on dynamic programming model based on sliding window and XGBoost regression algorithm. In Proceedings of the 2023 5th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI '23). Association for Computing Machinery, New York, NY, USA, 109–114. <https://doi.org/10.1145/3653081.3653100>

[4]Jingbin Zhang, Yanfang Zhang, Fan Tang, Yuanpeng Song, Yuanhaonan Deng, and Songlin He. 2024. E-commerce Retail Merchandise Based on Optimized K-means Algorithm and Multi-model Fusion Demand Forecasting Research. In Proceedings of the 2024 International Conference on Smart City and Information System (ICSCIS '24). Association for Computing Machinery, New York, NY, USA, 512–516. <https://doi.org/10.1145/3685088.3685178>

[5] Yucui, Li, Huang Zhiming, and Wei Jinling. "Analysis of the development trend of cross-border e-commerce between China and Vietnam based on ARIMA model [J]." Foreign Trade and Economic Cooperation 7 (2023): 6-10.

[6] Gao, Zhenyu, et al. "A novel multivariate time series prediction of crucial water quality parameters with Long Short-Term Memory (LSTM) networks." Journal of Contaminant Hydrology 259 (2023): 104262.

[7]Jiang, Keruo, et al. "Deep belief improved bidirectional LSTM for multivariate time series forecasting." Math. Biosci. Eng 20 (2023): 16596-16627.

[8]Liu Jiaying. Research on air quality prediction based on model fusion [D]. Lanzhou. Lanzhou University,2020.

[9] Jia-Rong Kang and Meng-Huai Chang. 2024. The Application of Elite Genetic Algorithm in Sustainable Agricultural Transportation. In Proceedings of the 2024 International Conference on Information Technology, Data Science, and Optimization (I-DO '24). Association for Computing Machinery, New York, NY, USA, 6–11. https://doi.org/10.1145/3658549.3658551

[10] Li M Q, et al. Basic theory and application of genetic algorithm. Ke xue chu ban she, 2002.

[11]Martin Fyvie, John Mccall, Lee Christie, and Alexander Brownlee. 2023. Explaining a Staff Rostering Genetic Algorithm using Sensitivity Analysis and Trajectory Analysis. In Proceedings of the Companion Conference on Genetic and Evolutionary Computation (GECCO '23 Companion). Association for Computing Machinery, New York, NY, USA, 1648–1656. https://doi.org/10.1145/3583133.3596353