

Research on Intelligent Supply Chain Forecasting Based on Multi-Model Hyperparameter Optimization

Yize Guo^{1,a,*,†}, Zixin Xu^{2,b,†}

¹*University of Sydney, Sydney*

²*Huaiyin Normal University, Huai'an, Jiangsu, China*

a. yizeguo99999@gmail.com, b. 1774798555@qq.com

**corresponding author*

†Yize Guo and Zixin Xu are co-first authors.

Abstract: In the e-commerce supply chain, transportation cost is one of the key factors affecting the overall supply chain cost. Reasonable planning of transportation routes and warehousing layout can effectively reduce transportation costs, thereby improving supply chain efficiency and benefits. Accurate demand forecasting plays a core role in determining the appropriate location and number of warehousing sites. Accurate forecasting can greatly enhance decision-making processes such as inventory optimization and distribution planning. This paper uses optimization algorithms and multi-model ensemble learning to accurately predict future demand, aiming to improve the adaptability and competitiveness of the supply chain. For this issue, the dataset is first standardized, and outliers are identified and removed. Subsequently, visual analysis of different data dimensions is conducted to preliminarily identify the influencing factors and features of demand forecasting. Based on the combination of merchants, warehouses, and products, demand quantity characteristics and structural features are quantified from three dimensions: central tendency, data distribution, and dispersion. The SSA Sparrow Search Algorithm-optimized KMeans clustering is used to process the quantity and structural features of historical demand, and sequence decomposition is applied to identify the demand patterns for each category, as shown in Figure 15. Finally, the paper evaluates the advantages and disadvantages of the constructed model and the handling of issues, proposes improvements to the model, and extends the model.

Keywords: Intelligent Supply Chain, SSA-KMeans, Model Fusion, Hyperparameter Optimization, Polynomial Fitting, Demand Forecasting

1. Introduction

E-commerce platforms host numerous merchants. To reduce delivery time, products are stored in the merchants' associated warehouses, enabling unified management of goods. This management method, leveraging scientific methods and intelligent decision-making, utilizes big data to intelligently drive the supply chain, significantly reducing inventory costs and ensuring timely delivery of goods.

Supply chain optimization typically involves two core aspects: demand forecasting and inventory optimization. As the foundational decision-making tool for intelligent supply chains, demand forecasting allows managers to anticipate demand conditions in various regions, enabling them to place inventory in warehouses closer to the demand points. To effectively manage inventory levels,

companies often design multiple inventory strategies to control the frequency and quantity of product replenishment, aiming for optimal inventory optimization.

2. Model Assumptions

(1) It is assumed that the supply characteristics of the central warehouse and regional warehouses are consistent.

(2) The impact of competition or cooperation with other e-commerce platforms is not considered.

(3) It is assumed that merchants undergo business transformation, where their sales items may not align with the business type.

(4) The provided data is assumed to be accurate and real, with anomalies occurring during promotional activities considered as reasonable data[1].

3. Symbol Explanation

Symbol	Explanation
$\widehat{y}_i^{(t)}$	Predicted value of the LightGBM model in the t-th round
$f_t(x_i)$	Base model in the t-th round
$X = [x_1, x_2, \dots, x_n]$	Unlabeled dataset
θ	Custom threshold
E	Model validation set error
W	Model weight
k_i	Sales trend change
x_{date}	Corresponding date of the data
Δx	Date change
f	Polynomial prediction fitting value
b_0	Demand offset
y_{final}	Final predicted result
y_{mult}	Polynomial model predicted result
$y_{SSA-BiLSTM-LGBM}$	Fusion model predicted result
w	Trust value

4. Establishment and Solution of Model for Problem 1

4.1. Data Preparation

4.1.1. Data Preprocessing

After analyzing the data related to the topic, the following treatments were applied:

Handling Duplicates:

First, the data (Appendices 1-4) was checked, and missing values were processed[2]. Duplicate data was found in Appendix 4, where the entries for "wh_21" and "wh_36" contained identical values. **These duplicates were removed**, and only one data entry was retained. After this operation, the four processed data tables were **merged** based on the columns seller_no, product_no, and warehouse_no, as shown in the table below:

Table 1: Data After Deduplication and Merging

warehouse_no	warehouse_category	warehouse_region
wh_29	Regional warehouse	East China
wh_29	Regional warehouse	East China
...	...	...
wh_38	Regional warehouse	Southwest China
wh_38	Regional warehouse	Southwest China

Central Warehouse Data Processing:

According to the literature, JD.com divides its warehouses into RDC and FDC types. Since the central warehouse shares the attribute of direct delivery to users and transfer stations with regional warehouses, the central warehouse is treated similarly to regional warehouses in this paper.

Handling Outliers:

Special holidays and merchant-designated discount sales days were not considered as outliers. For example, "Double 12" is a discount promotion day set by merchants, and the sales volume during this period was treated separately. Additionally, data from seller_3 product_1638 in warehouses wh_1/wh_31, where the shipment quantity (demand) was zero for the entire given period, was removed as outliers[3].

Given that merchants selling products outside their typical category (e.g., sellers diversifying or expanding their business) may represent business transformations, such data was not removed. For example, seller_12 primarily sells food and beverages but also sells personal care products, as shown in the table below:

Table 2: Data After Removing Outliers

category1	category2	category3	seller_category
Personal Care	Oral Care	Oral Care_2	Food & Beverage
Beauty & Skincare	Facial Care	Facial Care_6	Food & Beverage
Household Cleaning & Paper Products	Laundry Care	Laundry Care_1	Food & Beverage

4.1.2. Data Visualization

The data was processed for visualization to uncover hidden relationships within the dataset[4].

Demand Differences Across Regions

The demand differences in regions determines the allocation of transportation resources. Accurately predicts the demand in each region, so that e-commerce companies can reasonably set up transportation sites, reduce unnecessary transportation detours, and reduce transportation costs. For East China, nearby warehousing sites can be set up to reduce long-distance transportation. A bar chart of the total sales volume by region was created, as shown below:

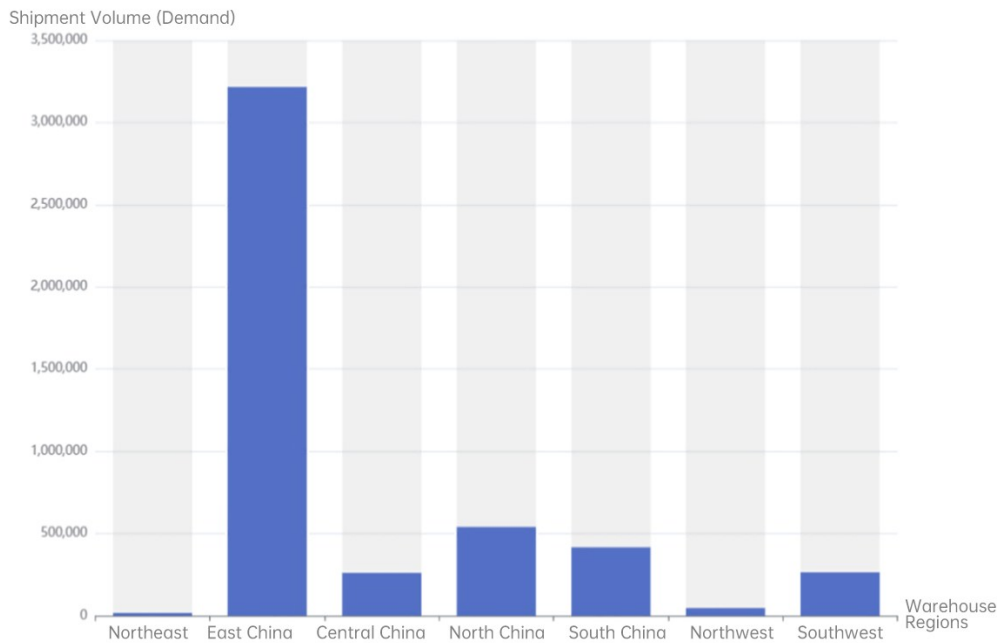


Figure 1: Regional Differences in Total Demand

Demand Differences Across Time Periods

Considering seasonal demand, periodicity, and short-term fluctuations such as merchant sales events (e.g., discount days), the data was analyzed for various time periods without removing these data points. The following charts show demand in different regions over different time periods:

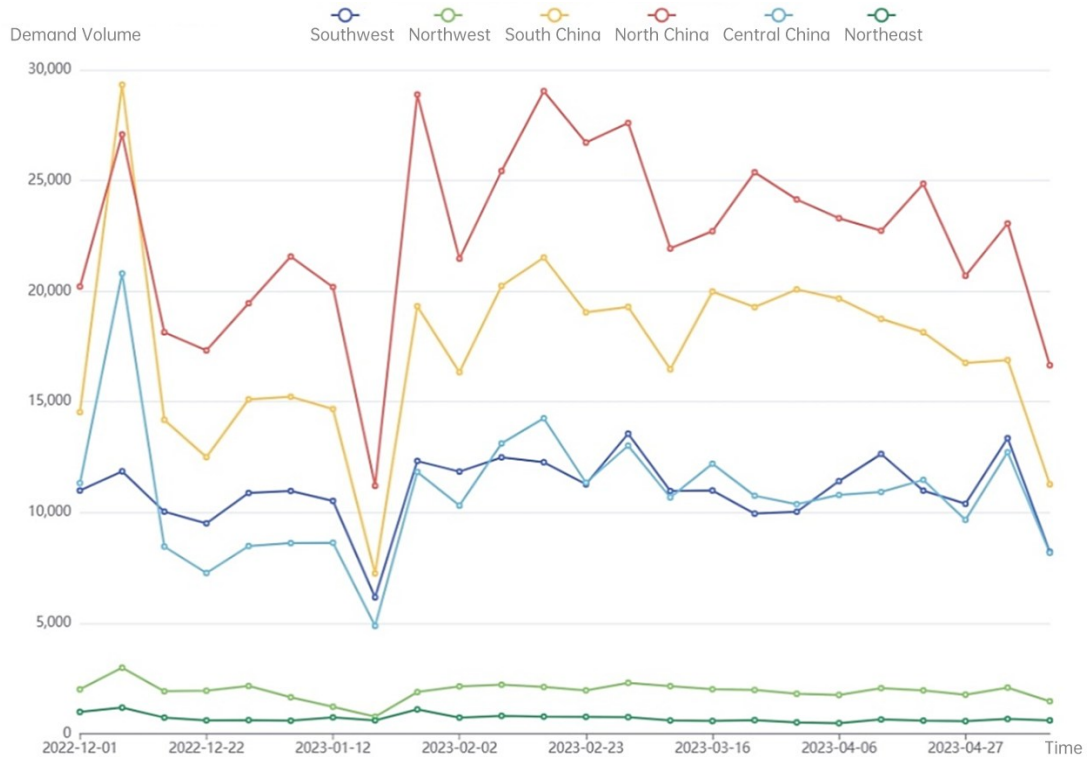


Figure 2: Demand for Products by Warehouse and Time Period

Since time series data often exhibits periodic variations, periodicity was observed during the data processing phase. To analyze this, time series decomposition was applied using an additive decomposition model[5], assuming the time series consists of three components: seasonal, trend-cycle, and residual, as follows:

$$y_t = S_t + T_t + R_t \quad (1)$$

(where y_t is the data, S_t is the seasonal component, T_t is the trend-cycle component, and R_t is the residual component).

The odd/even nature was first identified, and detrended data was calculated:

$$y_t - \hat{T}_t \quad (2)$$

The detrended values were then averaged to estimate the seasonal component for each season. These seasonal indices were adjusted to ensure their sum is zero (this constraint better illustrates the relationship between seasonality and the mean value). The monthly values were concatenated to form a yearly replicated sequence to obtain the seasonal component, denoted $\hat{S}_t$. If there is a residual component, it was computed using Equation (1)[6].

Demand Differences Across Merchant Sizes

Considering factors such as merchant popularity, target customer demographics, and marketing, demand can vary based on merchant size. The total and average demand for different merchant sizes were analyzed, as shown in the figures below:

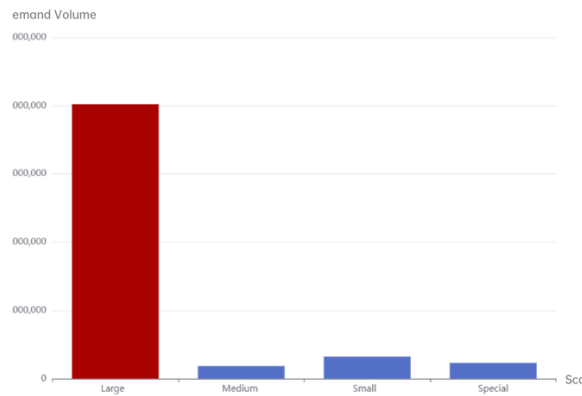


Figure 3: Total Demand for Different Merchant Sizes

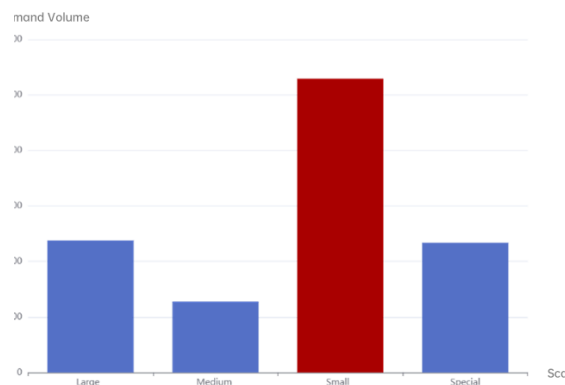


Figure 4: Average Demand for Different Merchant Sizes

Demand Differences Between Different Warehouses in the Same Region

Factors such as warehouse capacity, understanding the demand of different warehouses in the same region can help optimize the layout of storage sites in the region, avoid excessive transfer of goods

between warehouses, and thus save transportation costs. By analyzing these data, provide decision-making basis for setting up multiple stations, making the transportation route shorter and more efficient. A scatter plot of merchant sales in a particular region (South China in this case) was generated, as shown below:

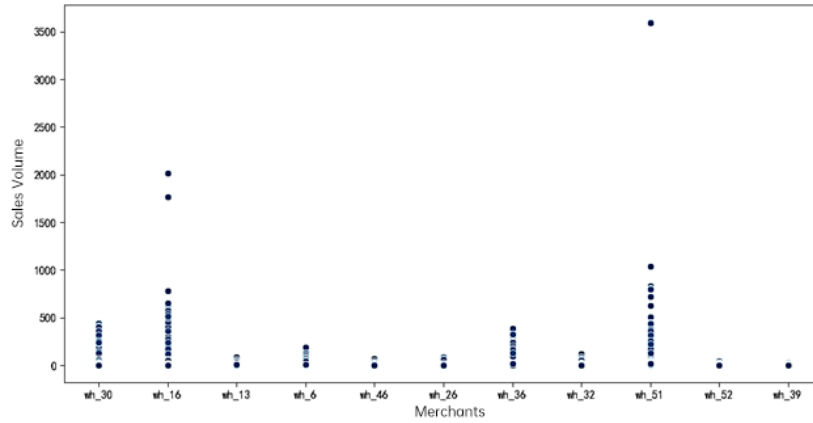


Figure 5: Scatter Plot of Merchant Sales in South China

4.1.3.Feature Construction

Based on the visualization analysis of the data above, this paper further extracts the feature values of the data, organized as follows:

Demand Quantity Series

The data is grouped by seller_no (merchant code), product_no (product code), and warehouse_no (warehouse code), and the quantity of shipments (demand) for each group is analyzed. The mean value of the qty data is calculated to identify the central tendency, and the standard deviation, coefficient of variation, and range are used to reflect the degree of data dispersion. Additionally, skewness and kurtosis are applied to explore the distribution shape of the data. The mathematical characteristics for each group are summarized in the following table:

Table 3: Mathematical Features for Each Group

label	qty_sum	qty_mean	qty_std	qty_cv	qty_range	qty_skew	qty_kurt
seller_35_product_1442_wh_35	498	3	6.157528	2.052509	60	5.525801	43.718
seller_27_product_661_wh_25	216	1.301205	1.098042	0.843865	5	0.843487	0.686909
seller_11_product_145_wh_1	738	4.445783	5.20781	1.171404	42	3.350188	17.37891
...	...	...	...	...	...	...	...
seller_9_product_1764_wh_14	219	1.319277	1.185974	0.898958	6	1.046998	1.260648

Exogenous Variables

Considering that merchants may create promotional events to increase purchasing power, this paper collects promotional event dates from the provided table and examines the variation in product demand during these periods. The trend of sales during JD.com's "Double 12" event is shown below:

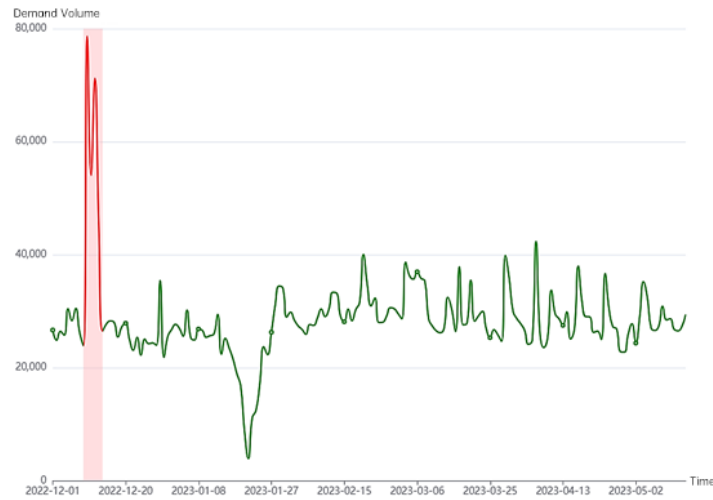


Figure 6: JD.com "Double 12" Sales Trend

Founded that the presence of promotional events before and after the event phase significantly impacts certain types of products, or multiple product types. Therefore, **this paper introduces exogenous variables (whether or not the event has an effect) as feature values when constructing data features. These variables are represented using binary coding (0/1). A value of 1 indicates that a promotional event is occurring and has a significant impact on the product, while a value of 0 indicates the opposite.**

By analyzing the start and end dates of several promotional holidays, exogenous variables were constructed to reflect the products impacted by these events. The table below shows a partial list of JD.com's promotional events:

Table 4: JD.com E-commerce Promotional Events (Partial)

Festival	Start Date	End Date
JD Member Care	2022-12-12	2022-12-12
Computer Festival	2023-01-23	2023-01-31
Spring Thunder Action	2023-03-01	2023-03-24
JD Kitchen & Bathroom Festival	2023-03-08	2023-03-12
Reading Festival	2023-04-01	2023-04-07
JD Mobile Phone Festival	2023-04-09	2023-04-19
JD Beauty Festival	2023-05-06	2023-05-19
JD Hair Care Festival	2023-06-05	2023-06-08
Refrigerator Festival	2023-06-17	2023-06-21
JD 618	2023-05-25	2023-06-20
...	...	...

4.1.4.K-Means Clustering Analysis Optimized by Sparrow Search Algorithm

This paper further performs clustering analysis on the data, and based on the data distribution shape, the demand distribution structure is derived. Based on the previous analysis, the magnitude of demand is used to categorize different demand patterns. Multiple predictive models are applied to find the best-fitting model to improve prediction accuracy.

Model Structure Design

SSA [1]-KMeans is an optimization algorithm based on swarm intelligence, simulating the foraging behavior of sparrows. It iterates multiple times in the search space to find the optimal solution. The algorithm divides the dataset into K clusters, each with a central point, which is the mean of all points in the cluster. The goal of using the SSA algorithm [1] to optimize K-Means clustering is to minimize the sum of the distances from all points to their respective center points.

K-Means Clustering Model Solution

(1) Model Construction

K-means is an unsupervised classification algorithm, assuming the existence of an unlabeled dataset:

$$X = [x_1, x_2, \dots, x_n] \quad (3)$$

The task of the algorithm is to cluster the dataset into K clusters: $C = C_1, C_2, \dots, C_k$. The loss function to be minimized is:

$$E = \sum_{i=1}^k \sum_{x \in C_i} ||x - u_i||^2 \quad (4)$$

where

u_i is the center of cluster

$$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (5)$$

To find the optimal solution to the above problem, all possible cluster divisions must be explored. The K-Means algorithm optimized by the SSA sparrow intelligence optimization algorithm uses a greedy strategy to obtain an approximate solution. The specific steps are as follows:

(a) Randomly select k sample points from the dataset to act as the initial cluster centers

$\{u_1, u_2, \dots, u_k\}$;

(b) Calculate the distance $dist(x_i, u_j)$ between each sample point and the cluster centers, and assign each sample point to the nearest cluster

$x_i \in u_{nearest}$;

(c) Recalculate the cluster centers based on the existing sample points in each cluster;

(d) Repeat steps (b) and (c).

When improving K-Means using the SSA sparrow search algorithm, the cluster centers are treated as optimization variables in the sparrow algorithm. The fitness function is as follows:

fitness

$$fitness = \sum_{i=1}^k \sum_{x \in C_i} ||x - u_i||^2 \quad (6)$$

This fitness function is consistent with K-Means' minimum loss function. The flowchart is shown below:

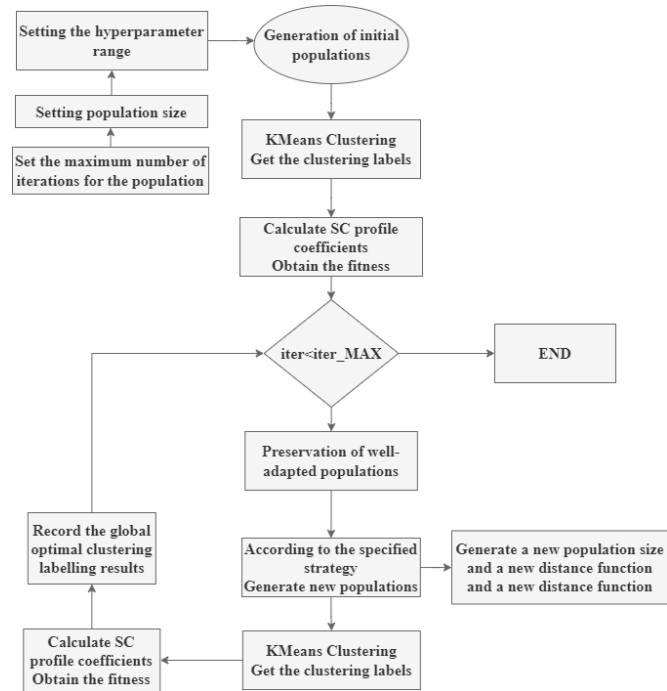


Figure 7: Flowchart of K-Means Algorithm Optimized by Sparrow Search

(2) Model Training

The hyperparameters selected for this study are as follows:

Table 5: Hyperparameters Selected for This Study

Hyperparameter	Value
Range of Optimized Clusters	[3, 20]
Sparrow Population Size	5
Number of Sparrow Population Evolutions	5

In this paper, the algorithm iterates four times to find the optimal solution, yielding a clustering silhouette coefficient of 0.85. The suitability curve of the optimization algorithm is shown below:

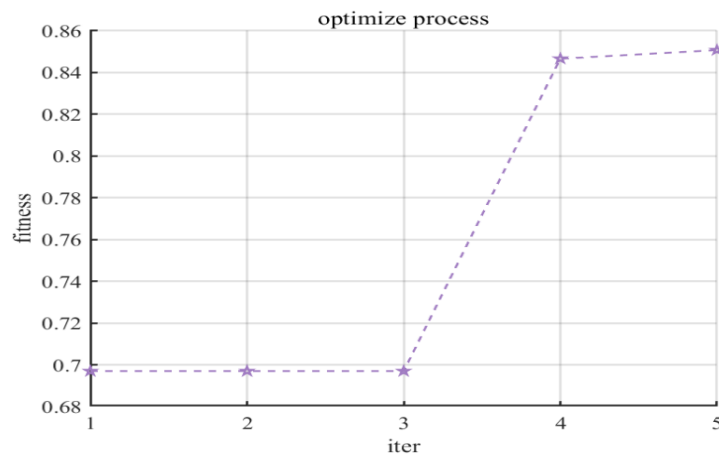
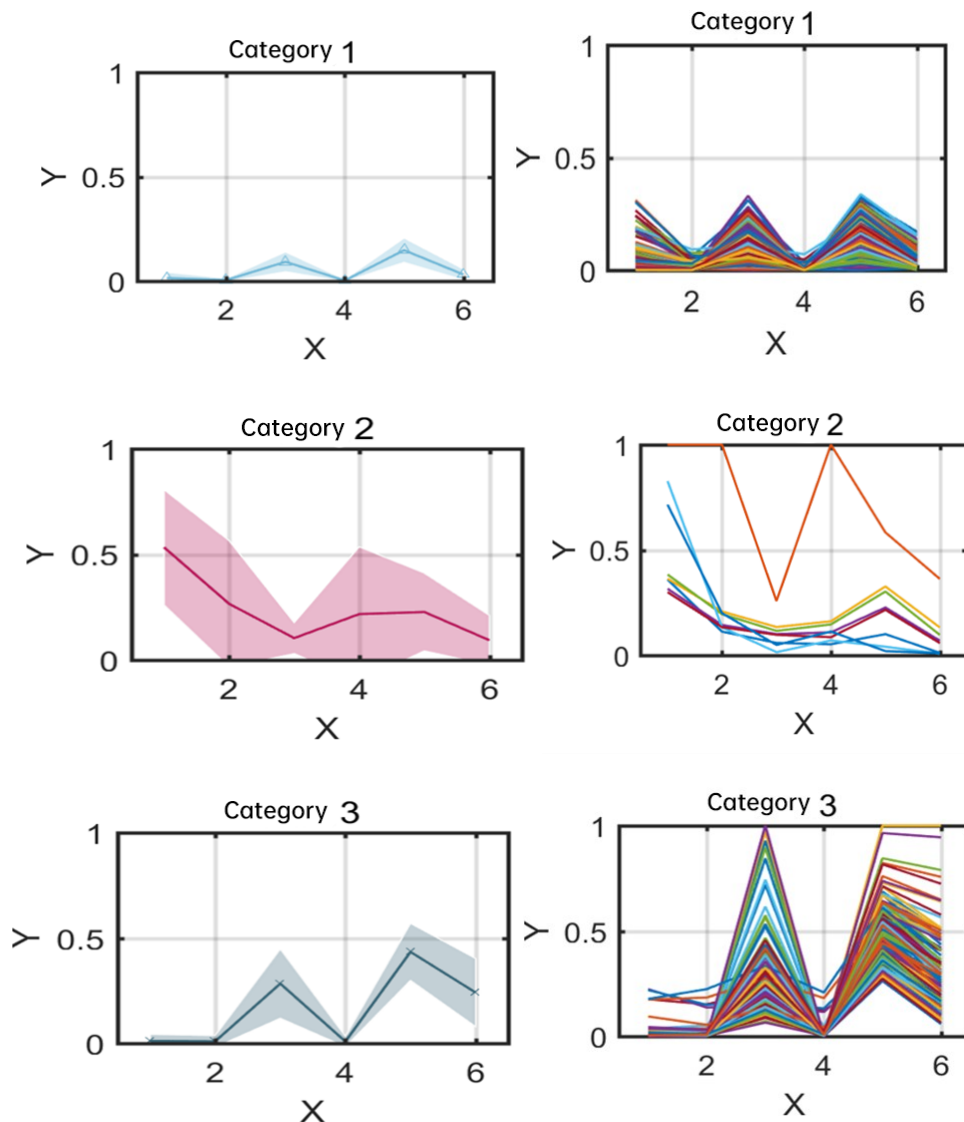


Figure 8: Model Training

(3) Experimental Results

This paper derives three categories of results through K-Means clustering [7] and analyzes them from two perspectives. The feature map is shown below:



Descriptive Feature Analysis

(a) Mean and Standard Deviation are Large: This indicates that the average demand is relatively high, and the data distribution is dispersed, showing significant volatility.

(b) Large Range: The range shows a large difference between the maximum and minimum demand values, which may imply considerable fluctuations in demand during different time periods.

(c) Coefficient of Variation is in the Middle: The coefficient of variation is the ratio of the standard deviation to the mean. Being in the middle suggests relatively stable volatility, but it can be used to measure the degree of fluctuation relative to the mean.

(d) Skewness and Kurtosis are Small: Skewness measures the asymmetry of the data distribution, while kurtosis measures the sharpness of the data distribution. Small values of skewness and kurtosis may suggest that the demand distribution is relatively uniform and close to a normal distribution.

Analysis of the Three Clusters' Features from K-Means

(a) Category 1 Feature Description - Highly Variable Demand Model and Its Management

This type of demand experiences large fluctuations in different time periods, and the data distribution is close to normal distribution, uniform and without obvious skew or spikes. This requires

the supply chain to have strong responsiveness. Therefore, **through accurate prediction, temporary or supplementary stations can be set up in areas or time periods with large demand fluctuations to ensure timely supply of goods and reduce the high costs caused by emergency transportation**, ensuring the controllability of transportation costs.

(b) Category 2 Feature Description - Low Volatility Demand Model Prediction and Management

From the above figure, it is observed that this demand model exhibits very low volatility, with the mean and standard deviation of demand being very small, and the differences between data points are minimal. The demand data distribution is highly uneven, showing clear skewness and kurtosis, which may indicate a long-tail distribution or peak characteristics.

Therefore, for areas where demand is relatively concentrated but in small quantities, small transit stations can be set up. These stations are responsible for handling specific types of goods, avoiding the waste of resources caused by large stations when handling small batches and special-demand goods, thereby effectively reducing transportation costs and ensuring stable and efficient operation of the supply chain.

(c) Category 3 Feature Description - Low Volatility Demand Model Prediction and Management

From the above figure, it can be observed that this type of demand model has extremely small range, mean and standard deviation, indicating that the volatility of demand is low and the average demand is relatively stable, which facilitates the long-term planning of fixed storage sites. This type of demand model shows obvious skewness and kurtosis, indicating that the data distribution is very asymmetric and may show long-tail distribution or peak characteristics.

Therefore, it is recommended to adopt specific strategies when dealing with demand fluctuations to ensure the stability and efficient operation of the supply chain. This type of demand is relatively stable. **By reasonably setting up an appropriate number of sites, stable transportation routes and frequencies can be maintained, avoiding idleness or waste of transportation resources, and achieving continuous reduction in transportation costs.** From the above figure, it can be observed that this type of demand model has extremely small range, mean and standard deviation, indicating that the volatility of demand is low and the average demand is relatively stable, which facilitates the long-term planning of fixed storage sites. This type of demand model shows obvious skewness and kurtosis, indicating that the data distribution is very asymmetric and may show long-tail distribution or peak characteristics.

Therefore, it is recommended to adopt specific strategies when dealing with demand fluctuations to ensure the stability and efficient operation of the supply chain. This type of demand is relatively stable. **By reasonably setting up an appropriate number of sites, stable transportation routes and frequencies can be maintained, avoiding idleness or waste of transportation resources, and achieving continuous reduction in transportation costs.**

4.1.5. Time Series Decomposition

Based on the clustering results mentioned earlier, this paper uses the results grouped by seller_no, product_no, and warehouse_no as index markers, and finally divides them into three categories, such as seller_11_product_180_wh_16 (Index Marker) for Category 1, seller_14_product_194_wh_1 for Category 2, and seller_32_product_1110_wh_29 for Category 3. This paper utilizes EEMD (Ensemble Empirical Mode Decomposition) for decomposition.

Through multiple iterations, the IMFs (Intrinsic Mode Functions) under different noise levels are obtained, and the final decomposition results of the signals for different categories are summarized as follows. The decomposition result for Category 1 is shown below (only the decomposition results for Category 1 are presented in this paper):

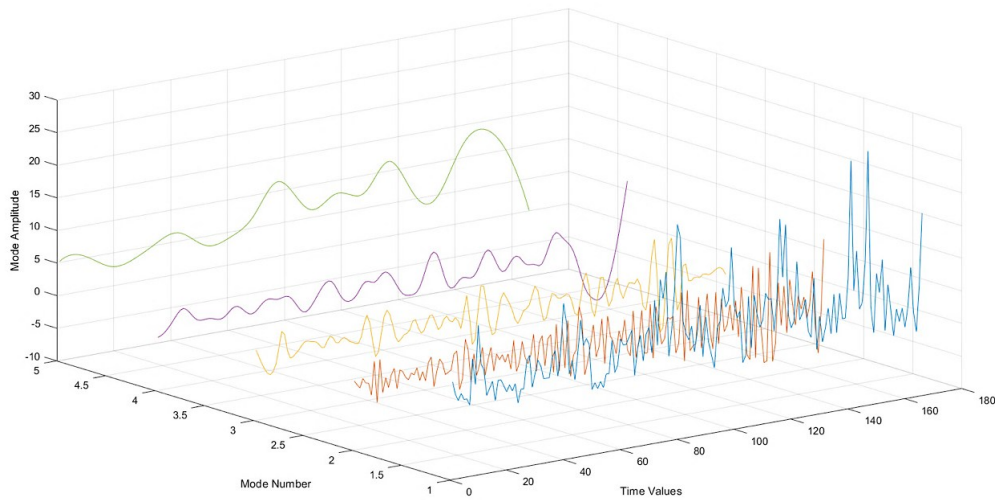


Figure 15: Time Series Decomposition for Category 1

(Note: Blue line - Original waveform, Red line - High-frequency decomposition, Yellow line - Higher-frequency decomposition, Purple line - Medium-frequency decomposition, Green line - Low-frequency decomposition.)

From the figure above, it can be seen that the original waveform data exhibits significant fluctuations, indicating that **all the information and features are unstable**. The high-frequency component typically captures noise or rapid fluctuations in the data. The higher-frequency decomposition shows relatively rapid but lower-frequency fluctuations, suggesting a certain periodicity or rapid changes in the data. The medium-frequency decomposition reflects moderate-speed fluctuations, indicating **the presence of certain trends or periodicity in the data**. The low-frequency component corresponds to relatively slow fluctuations, indicating **the presence of long-term trends or periodicity**.

5. Model Evaluation

5.1. Advantages of the Model

(1) The model can effectively capture long-term dependencies and contextual information in time series data, enabling better prediction of future demand and more comprehensive capture of data complexity.

(2) It can process input data step by step, extracting features from the sequence and making predictions. Powerful guidance for enterprises in setting up transportation sites through accurate demand forecasting, effectively avoiding the increase in transportation costs caused by blind transportation or unreasonable warehouse layout, and reducing supply chain costs from the root.

(3) Voting Model:

(a) Improved Accuracy: Parallel predictions from multiple models can reduce the bias and errors of individual models.

(b) Enhanced Robustness: Parallel predictions from multiple models can reduce over-reliance on specific models.

(c) Handling Diverse Data: Parallel predictions from multiple models allow for focusing on specific types of data.

5.2. Disadvantages of the Model

- (1) The model requires substantial computational resources and time for training, which may limit its scalability and real-time capabilities.
- (2) The ensemble model includes multiple hyperparameters, and tuning these hyperparameters may require significant experimentation and computational resources.

5.3. Model Improvements

(1) Voting Model:

Consider dynamically adjusting the model combination strategy during runtime to adapt to different types of input data or environmental changes, which can improve the model's adaptability and flexibility.

(2) Model Fusion Strategy:

The model constructed in this paper is an ensemble model, which can easily be combined and extended with other models to achieve better results, especially for application in predictive technologies.

6. Conclusion

This paper focuses on demand forecasting in smart supply chains, using optimization algorithms and multi-model ensemble learning methods to accurately predict future demand. The dataset was first standardized, identifying and removing outliers, and through visual analysis of various data dimensions, the influencing factors and features for demand forecasting were preliminarily determined. Based on combinations of sellers, warehouses, and products, demand sequence features were quantified from central tendency, data distribution, and dispersion. The SSA Sparrow Search Algorithm optimized KMeans was applied for clustering historical demand, and sequence decomposition was used to identify different demand patterns. The constructed model shows significant advantages, effectively capturing long-term dependencies and contextual information in time series data, comprehensively capturing data complexity, and improving accuracy, robustness, and diversity handling through parallel predictions from multiple models. Helps reduce supply chain costs by optimizing transportation. Allow for strategic site - setting, cutting transportation costs and thus overall supply chain costs. However, it also has drawbacks, such as requiring substantial computational resources and time for training, as well as difficulties in hyperparameter tuning. To address these, improvements are suggested, such as dynamically adjusting the combination strategy of the voting model during runtime and utilizing fusion models that allow for easier integration with other models. This study provides effective methods and improvement directions for demand forecasting in smart supply chains, contributing to enhanced adaptability and competitiveness in supply chains, while still requiring further exploration in resource utilization and model optimization.

References

- [1] Xue, Jiankai and Bo Shen. "A novel swarm intelligence optimization approach: sparrow search algorithm." *Systems Science & Control Engineering* 8 (2020): 22 - 34. <https://doi.org/10.1080/21642583.2019.1708830>
- [2] Li, Yanbo, Yin, Pei, Chen, Junshuo, et al. "Short-term power load forecasting based on improved residual network and Bi-LSTM." *Journal of Harbin Institute of Technology*, 2023, 55(08): 79-86.
- [3] Li, Weikang, Deng, Xing, Shao, Haijian. "Deep learning ECG classification method design and research based on multi-model voting." *Computer Simulation*, 2023, 40(08): 339-344.
- [4] Wang, Jie, Li, Jinze, Wang, Zitong, et al. "Improved whale optimization LightGBM explainable heart disease risk prediction model." *Journal of Beijing University of Posts and Telecommunications* [2023-11-02], 1-6. <https://doi.org/10.13190/j.jbupt.2023-015>.
- [5] *Intelligent Supply Chain: Forecasting Algorithm Theory and Practice [M]*, Beijing: Electronic Industry Press, 2023.

- [6] Makridakis, S., Spiliotis, E., Assimakopoulos, V. "The M5 Accuracy competition: Results, findings and conclusions." *International Journal of Forecasting*, 2020, 36(1): 224-227.
- [7] Makridakis, S., Spiliotis, E., Assimakopoulos, V., et al. "The M4 Competition: 100,000 time series and 61 forecasting methods." *International Journal of Forecasting*, 2020, 36(1): 54-74.