Information Systems and Economics (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2024.050317
ISSN 2523-6407 Vol. 5 Num. 3

Research on Automatic Pricing and Replenishment Decision for Vegetable Products Based on PSO-BP Model

Boyu Liu#, Weiran Zhang#, Yi Qin#

School of Economics and Management, Yanbian University, Hunchun, 133300, China
#These authors contributed equally.

Keywords: Dynamic Pricing Mechanism, Automatic Replenishment System, PSO-BP Model

Abstract: In the fluctuating market demand for fresh supermarket logistics, this study introduces an innovative methodology combining Spearman correlation analysis, K-means clustering, and PSO particle swarm optimization to enhance the accuracy of BP neural network predictions for sales profits. Aimed at improving the procurement process for vegetables and fresh products, our analysis reveals effective strategies for supermarkets to increase profit margins, reduce product losses, and enhance customer service quality. These strategies offer supermarkets guidance on making wiser replenishment and pricing decisions in complex market environments, maintaining a competitive edge. The application of the PSO-optimized BP neural network not only improves profit predictions but also provides flexibility in pricing and stocking decisions, allowing for rapid market adaptation and efficient inventory management. This research provides a powerful tool and theoretical support for formulating scientific strategies, helping supermarkets stay competitive in the retail industry. Future studies could explore the model's application to other product types or market conditions, extending its retail industry applicability.

1. Introduction

In recent years, China has seen significant economic and social development, with a growing demand for a green, healthy lifestyle. The country's vegetable industry has expanded by 47% from 2000 to 2022, reaching an area of 22,434 thousand hectares. However, the perishable nature of vegetables and their short life cycle lead to frequent price fluctuations, affecting both producers and consumers. Thus, the issue of accessing fresh vegetables remains a widespread concern.

Fresh vegetable pricing and stocking are unique, with purchases happening early at 3-4 am, requiring daily stock decisions without fixed prices. It's vital for merchants to find ways to keep vegetables fresh while maximizing profits [1-2].

Research on perishable products, employing an advanced SVR-LSTM-ARIMA model and PSO optimization of BP neural networks, predicts sales profits to assist in replenishment and pricing strategies [3]. Another study utilizes SARIMA and Spearman coefficients to analyze supermarket vegetable sales, automating pricing and stocking decisions through nonlinear regression and neural networks [4]. Time series models forecast weekly total replenishment volumes for vegetables, exploring the effects of cost-plus pricing on total sales, profit margins, and costs [5]. Statistical analyses and Pearson correlation coefficients shape effective pricing and replenishment strategies for supermarket vegetables. Additionally, random forest models predict future replenishment volumes and refine pricing strategies, addressing challenges in these areas [6].

This article focuses on Problem C of the 2023 China College Mathematical Modeling Competition, analyzing data correlations across categories and products using Spearman analysis and K-means clustering. The PSO algorithm optimizes the BP neural network for predicting sales profits, aiding in the planning of future stocking and pricing strategies. (Data Source http://www.mcm.edu.cn/html_cn/node/c74d72127066f510a5723a94b5323a26.html.)

2. Data Preprocessing and Correlation Exploration

2.1 Data Preprocessing

Data preprocessing involves handling missing values, managing outliers, and removing invalid entries. Initially, missing values are identified using Python's 'isnull' and 'sum' commands, and rows with missing values are removed with 'dropna', considering the minimal impact on a dataset of 870,000 records. Outlier detection is conducted through box plots for the top 20 items by purchase quantity, aiding in identifying outliers.The plot is as follows Fig 1:

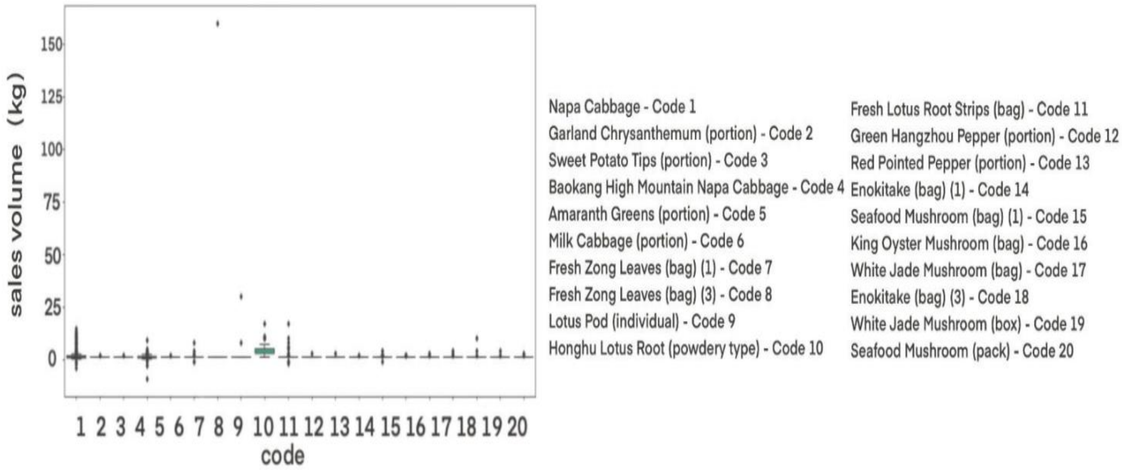


Figure 1: Top 20 Best-Selling Items

Given the dataset's assumption of a normal distribution, the 3σ rule is used to detect outliers. MATLAB's Kolmogorov-Smirnov test confirmed that selling prices are normally distributed. Using the 3σ rule, 2631 outliers were identified and subsequently removed after manual review.

2.2 Correlation Analysis

2.2.1 Spearman Correlation Analysis

When analyzing vegetable pricing and replenishment independence, correlation analysis is crucial. Assuming stable sales data and short-term market demand, initial Pearson analysis showed non-normal data distribution. Spearman analysis confirmed significant correlations among sales volumes of different product categories (p-value < 0.001).The is as follows table 1:

Observing the results, we find that edible fungi and aquatic root vegetables, as well as flower-leaf vegetables, have strong correlations, often bundled together for purchase. Conversely, tomatoes show low correlations with other categories, indicating potential for substitution. This understanding aids in replenishment strategies by leveraging these relationships to fill supply gaps.

Table 1: Spearman correlation analysis

		Cruciferous vegetables	Leafy greens	Chili peppers	Solanaceous vegetables	Edible fungi	Aquatic root and stem vegetables
Cruciferous vegetables	Correlation coefficient	1	.634**	.431**	.237**	.463**	.397**
Leafy greens	Correlation coefficient	.634**	1	.596**	.328**	.598**	.440**
Chili peppers	Correlation coefficient	.431**	.596**	1	.156**	.536**	.335**
Solanaceous vegetables	Correlation coefficient	.237**	.328**	.156**	1	-.087**	-.190**
Edible fungi	Correlation coefficient	.463**	.598**	.536**	-.087**	1	.606**
Aquatic root and stem vegetables	Correlation coefficient	.397**	.440**	.335**	-.190**	.606**	1
Note: **. Significant at the 0.01 level (two-tailed).							

2.2.2 K-means Clustering Analysis

To simplify the complex relationships among 251 individual products, the study employed K-means clustering analysis. This method categorized products based on unit price, sales volume, and wholesale price, providing more intuitive results. The picture of K-means clustering analysis is as follows Fig 2:

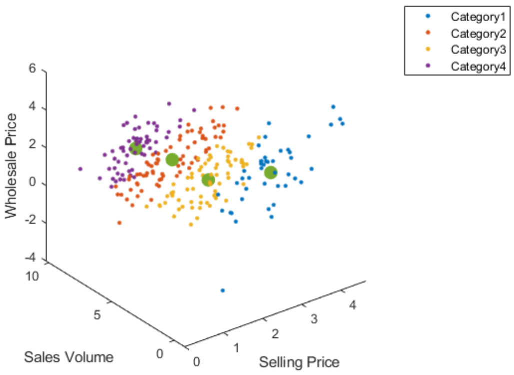


Figure 2: K-means Clustering Analysis

3. Replenishment and Pricing Strategies

Supermarkets restock fresh vegetables at dawn. Operators decide on daily restocking without precise item knowledge or purchase prices. Sales are influenced by price, season, and promotions, assuming no delay between sales and restocking. Only natural loss during procurement and storage is considered, with timely procurement to prevent shortages. No additional costs are accounted for. Cost-plus pricing adds profit margin to unit cost for setting prices.Here, the selling price of individual items can be represented by the following equation:

$$\begin{aligned} & \textit{The unit selling price of a product} \\ & = \textit{unit cost} \times (1 + \textit{markup rate}) \\ & = \textit{wholesale price of the unit} \times (1 + \textit{profit margin}) \end{aligned} \tag{1}$$

The cost-plus pricing method relies on wholesale prices and operator-set profit margins to make replenishment decisions and determine final selling prices. It's crucial to carefully set profit margins to avoid pricing issues that may impact sales or supermarket profitability.

Supposed that the supermarket plans replenishment on a category basis and devises the daily total replenishment quantity and pricing strategy for the upcoming week (July 1-7, 2023), which aims to maximize profits. The total profit for each category can be represented by the following equation:

$$Total\ profit = \sum_{t=1}^{7}(total\ sales\ revenue - total\ cost) \tag{2}$$

The total daily sales revenue for each category can be represented by the following equation:

$$\begin{aligned} &Total\ Sales\ Revenue \\ &= Category\ Sales\ Volume \times Category\ Average\ Selling\ Price \\ &= \sum^{k}(Unit\ Sales\ Volume \times (Wholesale\ Price \times (1 + Markup\ Rate))) \end{aligned} \tag{3}$$

The daily cost for each category can be represented by the following equation:

$$\begin{aligned} &Total\ Cost = Category\ Average\ Cost \times Total\ Category\ Replenishment\ Volume \\ &= \sum^{k}(Unit\ Cost \times Unit\ Replenishment\ Volume) \end{aligned} \tag{4}$$

Note: k=100 (Leafy Vegetables), 5 (Cauliflower), 19 (Aquatic Roots and Tubers), 10 (Solanaceae), 45 (Peppers), 72 (Edible Fungi).

This study utilizes key symbols as follows: the unit sale price (p_i) and average sale price per category (in CNY/kg) (P_j), the unit cost, specifically the wholesale price (c_i), and the average cost per category (in CNY/kg) (C_j), the loss rate (n_i) and profit margin, also known as the markup rate (expressed as a percentage) (l_i), the total profit per category (Π_j) and total revenue per category (in CNY) (R_j), the unit sales volume (q_i) and category sales volume (in kg) (Q_j), and the replenishment volume per item (b_i), covering 251 item names (i) and 6 category names (j). Therefore, the mathematical expression for the total profit of each category for the week can be finally written as:

$$\Pi_j = \sum_{t=1}^{7}(\sum^{k}(q_{it} \times (c_{it} \times (1 + l_{it}))) - \sum^{k}(c_{it} \times b_{it})) \tag{5}$$

3.1 Replenishment Decision Making

Separate predictions are needed for upcoming week's replenishment quantity and wholesale prices, both typical time series data.

3.1.1 K-means Clustering Analysis

The replenishment quantity relies on future sales volume, which correlates with daily replenishment quantity and losses. Establishing a time series model to predict next week's sales volume allows calculation of replenishment quantity. It can be represented by the following equation:

$$Replenishment\ Volume = \frac{Predicted\ Sales\ Volume}{1 - Loss\ Rate} \tag{6}$$

Below is an example of constructing a time series model using sales volume data of Aquatic Roots and Tubers. The table displays ADF test results, including variables, differencing orders, T-test results, AIC values, etc., used to assess time series stationarity. The results indicate that when the differencing order is 0, the p-value is 0.026**, significant at the 5% level, confirming stationarity. When the differencing order is 1 or 2, the p-values are both 0.000***, significant at the 1% level, reaffirming stationarity. Q statistic test results for Q6 are insignificant, supporting the model's residuals as white noise sequences. Moreover, the model's goodness-of-fit R^2 is 0.689, indicating satisfactory performance. Sales volume forecasts for six vegetable categories for the upcoming week have been obtained based on the time series model, which are as follows Table 2:

Table 2: Forecasted Sales Volume (kg) for the Next Week

order(time)	Solanaceous Vegetables	Cruciferous Vegetables	Chili Peppers	Ornamental Leaf Vegetables	Edible Fungi	Aquatic Rhizomes
1	22.6240154	22.8639324	83.885727	125.365952	52.9878004	20.1364258
2	20.8350680	20.4237387	85.2909564	116.100363	42.0046738	21.0826471
3	17.6596465	19.2753294	85.7888124	118.831927	32.2763486	21.4670938
4	17.3255105	18.7265544	85.8894286	121.384362	43.5609894	21.8415974
5	16.6677499	18.4561241	85.8959691	125.167777	51.5695576	22.2064153
6	18.4070779	18.3148989	85.8979353	126.506361	44.47866716	22.561798
7	19.3969191	18.2336495	85.9069378	125.4934794	45.40518078	22.90798936

Given that the loss rate does not vary with time, study have categorized and summarized the loss rates of 251 individual items, obtaining the average loss rate for each product category as follows Table 3:

Table 3: Loss Rate for Each Product Category

Category Name	Ornamental Leaf Vegetables	Cruciferous Vegetables	Aquatic Rhizomes	Solanaceous Vegetables	Chili Peppers	Edible Fungi
Category Name	11.158	9.168	9.441	6.751	8.004	8.297

Calculating the replenishment quantity for the next seven days based on the relationship formula replenishment quantity = forecasted sales volume / (1 - loss rate), the results are as follows Table 4:

Table 4: Predicted Replenishment Quantity (kg) for the Next Week

order(time)	Solanaceous Vegetables	Cruciferous Vegetables	Chili Peppers	Ornamental Leaf Vegetables	Edible Fungi	Aquatic Rhizomes
1	24.2619389	25.1716713	91.1841025	141.111132	57.7819705	22.2356980
2	22.3434761	22.4851801	92.7115922	130.681844	45.8051251	23.2805653
3	18.9381618	21.2208576	93.2527636	133.756475	35.1966114	23.7050914
4	18.5798352	20.6166927	93.3621338	136.629480	47.5022512	24.1186381
5	17.8744543	20.3189670	93.3692433	140.888068	56.2354095	24.5214891
6	19.7397054	20.1634874	93.3713806	142.394770	48.5029575	24.9139213
7	20.8012087	20.0740372	93.3811664	141.254676	49.5132992	25.2962039

3.1.2 Wholesale Price Prediction

Wholesale price prediction also employs a time series model, illustrated with cauliflower. ADF testing indicates stationary time series with p-values below 0.01 for various differencing orders. Residual analysis confirms white noise sequence, achieving an R^2 value of 0.929, signifying excellent model performance.

Similarly, study can obtain the wholesale price forecasts for six categories of vegetables for the upcoming week as follows Table 5:

Table 5: Predicted Wholesale Quantity (kg) for the Next Week

order(time)	Solanaceous Vegetables	Cruciferous Vegetables	Chili Peppers	Ornamental Leaf Vegetables	Edible Fungi	Aquatic Rhizomes
1	4.44339527	7.80210275	5.94740327	3.68685334	4.68606852	11.6559138
2	4.44403937	7.87704190	5.94635030	3.67576613	4.69224771	11.5613609
3	4.44468348	7.81987691	5.94529732	3.67839829	4.69076831	11.440233
4	4.44532758	7.86426422	5.94424435	3.67539636	4.68825788	11.3233680
5	4.44597168	7.83058532	5.94319137	3.67470816	4.68560865	11.2106157
6	4.44661579	7.85691818	5.94213840	3.67306978	4.68294073	11.1018316
7	4.44725989	7.83711826	5.94108543	3.67182162	4.68027030	10.9968758

3.2 Pricing Strategy

This article employs PSO particle swarm optimization to predict profits using a BP neural network in Matlab R2022b. BP neural networks suffer from slow convergence and local optima due to manually setting or randomly generating initial weights and thresholds, significantly impacting performance. Optimizing initial weights and thresholds improves BP neural network performance. [7]The study utilizes particle swarm optimization to enhance network performance, offering more accurate profit predictions.

The total profit from vegetable sales is influenced by various factors, often with nonlinear relationships. The BP neural network, with its nonlinear mapping capabilities, is better suited to capture complex data patterns. Optimizing with the PSO particle swarm algorithm enhances its ability to find global optimal solutions faster [8]. In conclusion, the PSO-optimized BP neural network model is suitable for predicting total vegetable sales profit in supermarkets, leveraging its nonlinear fitting, global search, and parameter optimization abilities to enhance prediction accuracy and stability [9-10].

After optimization with the PSO particle swarm algorithm, the formulas for the BP neural network are as follows:

The formula for the forward propagation of the BP neural network is as follows:

$$a^l = \sigma\left(w^l \times a^{l-1} + b^l\right) \tag{7}$$

The formula for the error calculation of the BP neural network is as follows:

$$E = \frac{1}{2}\sum_{i=1}^{N}(yi - \hat{y}i)^2 \tag{8}$$

The formula for the weight updating of the BP neural network (BP) is as follows:

$$w_{jk}^l = w_{jk}^l - \eta\frac{\partial E}{\partial w_{jk}^l} \tag{9}$$

The formula for the bias updating of the BP neural network (BP) is as follows:

$$b_j^l = b_j^l - \eta \frac{\partial E}{\partial b_j^l} \tag{10}$$

The formula for the velocity updating of the particle swarm algorithm (PSO) is as follows:

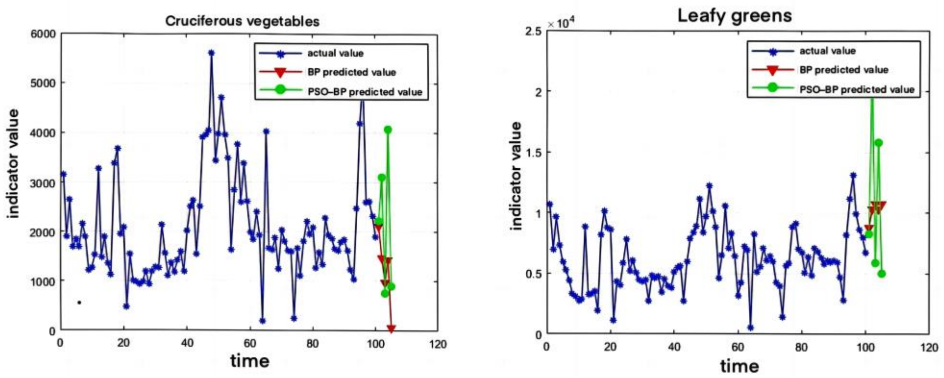
$$v_{id}^{t+1} = wv_{id}^t + c_1 \times rand(\ ) \times \left(p_{id} - x_{id}^t\right) + c_2 \times rand(\ ) \times (p_{gd} - x_{id}^t) \tag{11}$$

The formula for updating the weights and biases of the BP neural network optimized by the PSO algorithm (PSO + BP) is as follows:

$$\textit{The updated weights and biases} = \textit{the weights and biases before the update} + v_{id}^{t+1} \tag{12}$$

The symbols involved in the formulas mentioned above include, but are not limited to, the activation value of the first layer (a^l), the activation value of the previous layer (a^{l-1}), the activation function (σ), weights and biases (w^l, b^l) of various layers, error measurement (E), sample size (N), actual and predicted (yi , $\hat{y}i$) outputs, learning rate (η), and the partial derivatives of error with respect to weights and biases ($\frac{\partial E}{\partial w_{jk}^l}$, $\frac{\partial E}{\partial b_j^l}$), among others. These symbols are key elements in the model's learning process. Additionally, parameters within the Particle Swarm Optimization (PSO) algorithm, such as the velocity update of particles (v_{id}^{t+1}), inertia weight (w), acceleration coefficient (c_1 , c_2), personal best and global best positions of particles (p_{id} , p_{gd}), the current position of particles in specific dimensions (x_{id}^t), and random number generation ($rand(\ )$), are crucial for understanding and implementing the neural network and PSO algorithm. They play a central role in enhancing the accuracy of predictions for sales and inventory management strategies.

The final predictions for each individual product's total profit are as follows. The following six graphs compare future data and actual values of total profit across various product categories using BP neural network and PSO-BP optimized predictions. These models utilize particle swarm optimization for the BP neural network. Each graph presents three datasets: actual profit values (asterisks), profit values predicted by the BP neural network (red lines), and profit values predicted by the PSO-BP neural network (green lines), as shown in Fig 3.



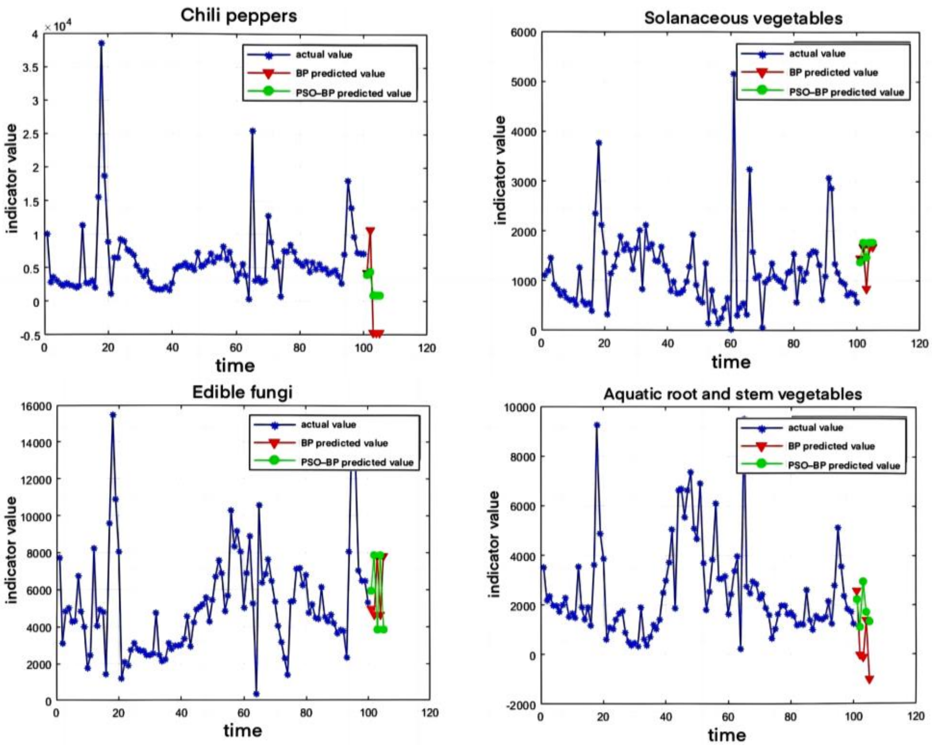


Figure 3: A Comparative Analysis of Future Data Prediction between BP Neural Network and PSO-BP Optimized Model versus Actual Values

The study examines the profit dynamics of various vegetable categories. Cauliflower profits are volatile, accurately predicted by the PSO-BP model. Leafy vegetable sales fluctuate with seasonal changes, requiring inventory adjustments. Chili profits vary consistently, aligned with PSO-BP predictions. Tomato profits cycle, matching PSO-BP forecasts, especially during declines. Edible mushroom profits fluctuate periodically, with effective PSO-BP predictions. Aquatic vegetable profits decline significantly due to seasonal shifts, as indicated by PSO-BP. Adjusting inventory during profit declines can mitigate losses.

The PSO-BP neural network accurately predicts profit fluctuations, offering insights for pricing strategies: 1. Raise prices during peak profit periods to maximize margins. 2. Implement discounts during profit downturns to stimulate sales and prevent inventory buildup. 3. Maintain flexible pricing strategies to optimize profits in response to market changes. Regarding replenishment strategies: 1. Increase inventory during predicted profit rises to meet demand. 2. Reduce inventory during unfavorable profit forecasts to mitigate losses. 3. Dynamically manage inventory levels based on sales forecasts.

4. Conclusions

The research optimizes fresh vegetable replenishment in supermarkets to boost profits and satisfaction. Using methods like Spearman correlation, K-means clustering, and PSO-BP neural networks, it forecasts profits. Sales exhibit cyclic fluctuations, influenced by seasonality, demand, and promotions. PSO-BP predicts accurately, especially in extreme conditions. Recommendations include pricing adjustments and inventory optimization for market adaptability and efficiency. This offers valuable insights for managers. Future research could refine forecasting for better efficiency and satisfaction.

References

[1] Li, Dong, and Wang, Xuefeng. Dynamic supply chain decisions based on networked sensor data: An application in the chilled food retail chain [J]. International Journal of Production Research, 2021, 59(7), 2142-2156.
[2] Zhang, Huifeng; Yan, Ziheng. Automatic Pricing and Replenishment Decisions of Vegetable Products Based on Machine Learning [J]. Highlights in Business, Economics and Management, 2024, 25: 228-236.
[3] Gu Aihua, Yan Zhongzhen, Zhang Xixi, Xiang Yongsheng. Research on the Modeling of Automatic Pricing and Replenishment Strategies for Perishable Goods with Time-Varying Deterioration Rates [J]. Axioms, 2024, 13(62)
[4] Zheng Luoyi, Lou Zhiyan. Automated Pricing and Replenishment Decision Modeling for Supermarket Vegetables [J]. Transactions on Computer Science and Intelligent Systems Research, 2023, 2: 46-54.
[5] Deng Baoluo, Liu Taotao, Zhang Ruizhi. Replenishment Forecasting of Vegetable Commodities Based on Time Series Modeling [J]. Highlights in Business Economics and Management, 2023, 20: 784-789.
[6] Wang Kaile, Su Keming, Li Hao. RandomForest-Based Restocking and Pricing Prediction for Vegetable Items [J]. Frontiers in Business Economics and Management, 2023, 11(2): 352-356.
[7] Shen Xueli, Zhang Hongyan, Zhang Jisuo. Improved Particle Swarm Optimization Algorithm for BP Neural Network Optimization [J]. Computer Systems Applications, 2010, 19(2): 57-60.
[8] Zong Chensheng, Zheng Huanxia, Wang Linshan. Grain Yield Prediction Based on BP Neural Network Optimized by Improved Particle Swarm Optimization [J]. Computer Systems & Applications, 2018, 27(12): 204-209.
[9] Yasanur Kayikci, Sercan Demir, Sachin K. Mangla, Nachiappan Subramanian, Basar Koc. Data-driven optimal dynamic pricing strategy for reducing perishable food waste at retailers [J]. Journal of Cleaner Production, 2022, 344: 131068.
[10] Miguel Pina, Pedro Dinis Gaspar, and Tânia Miranda Lima. Decision Support System in Dynamic Pricing of Horticultural Products Based on the Quality Decline Due to Bacterial Growth [J]. Applied System Innovation, 2021, 4(4): 80.