Vegetable Pricing and Replenishment Decision Based on Correlation Analysis and Entropy Weight-TOPSIS Model

Huaze Liang*, Senlin Liu, Junkai Zhao
School of mechanical engineering, Donghua University, Shanghai, China, 201620
*Corresponding author: 18292711148l@gmail.com

Abstract. This article studies the automatic pricing and replenishment decision-making of vegetable products through statistical analysis and modeling. Firstly, a descriptive statistical analysis of the data was conducted, and it was found that the vegetable categories did not fully conform to the normal distribution relationship. The correlation between each category was also explored. Subsequently, grey correlation analysis was used to explore the correlation between individual products within each category. Then, in the overall modeling process, the entropy weight TOPSIS model is used to select sellable items, and linear regression is used to predict the pricing situation for the next day. Using ant colony algorithm to solve the maximum return portfolio pricing strategy model, a profit of 912.235 yuan and corresponding pricing strategy and replenishment volume were obtained. Finally, five factors that affect the profits of supermarkets were proposed, and relevant suggestions were given to enhance the profits of supermarkets.

Keywords: Pricing decisions, gray correlation analysis, ant colony algorithm, entropy weight-TOPSIS evaluation model, linear programming.

1. Introduction

In fresh produce supermarkets, vegetable commodities have a short freshness period and their quality declines over time, and most vegetables are no longer suitable for sale the next day after the end of the unsold day[1]. As a result, supermarkets usually restock daily based on the historical sales and demand for the commodities[2]. Due to the wide variety of vegetables and their different origins, incoming transactions usually take place between 3:00 a.m. and 4:00 a.m.. As a result, merchants must plan for the day's replenishment of each vegetable category without determining the individual items and the price at which they will be purchased. Pricing is usually based on "cost-plus pricing", and supermarkets often offer discounts for quality losses and degraded items[3]. Accurate market demand analysis is critical to replenishment and pricing decisions. From a demand perspective, the volume of vegetable commodities sold is usually time-dependent; from a supply perspective, the period from April to October is a period of abundant vegetable supply, and due to space constraints in supermarkets, a reasonable mix of sales is very important[4]. The traditional methods for vegetable price analysis and sales forecasting include time series decomposition and ensemble learning[5-6].

2. Descriptive statistical analysis of data

2.1. Data Introduction

The integrated chart shows the sales volume, sales unit price, category, and wholesale price of each individual product at different times. This data comes from the sales flow details of various products distributed by a certain supermarket. Due to space limitations, only some data is displayed here, as shown in Table 1.

Table 1 Data examples

Item Name	Sales volume (kg)	Sales unit price (yuan/kg)	category	wholesale price
Bubble Pepper (Boutique)	0.396	7.60	Chili peppers	4.32
Chinese cabbage	0.849	3.20	Florifolias	2.1
Gaogua (1)	0.251	10.00	Aquatic rhizomes	5.65
Xixia Mushroom (1)	0.217	18.00	Edible fungi	10.8

2.2. Summary and analysis of various categories of vegetables

Summarize and analyze the data of six major categories of vegetables, and calculate on a monthly basis to obtain the summary trend of changes for a total of 36 months from June 2020 to July 2023, as shown in Figure 1.

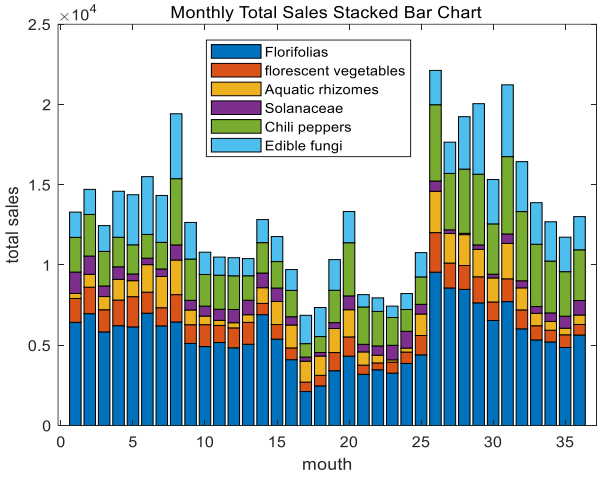

Figure. 1 Monthly Total Sales Stacked Bar Chart

According to the chart, among the six major vegetable categories, leafy vegetables showed the highest sales volume, while eggplant vegetables were at the lowest level of sales. Overall, sales data shows significant time fluctuations, especially during the period from April 2021 to June 2022, where sales generally decreased, which may be affected by external factors such as the pandemic. However, after June 2022, the data began to gradually rebound and tend towards levels close to previous years.

The period from September to January of the following year is usually a period of high sales, while March to July is usually a low season of low sales. This seasonal fluctuation is related to the growth cycle and seasonal supply of vegetables, so these factors need to be fully considered in sales forecasting and market strategy formulation.

2.3. Analysis of sales distribution of various categories of vegetables

Draw a three-year frequency distribution histogram for comparative analysis, divide sales data into multiple columnar intervals, and display the quantity of each interval to understand the concentration trend and dispersion degree of the data. Due to limited space, only one year's frequency variation chart is shown here, as shown in Figure 2.

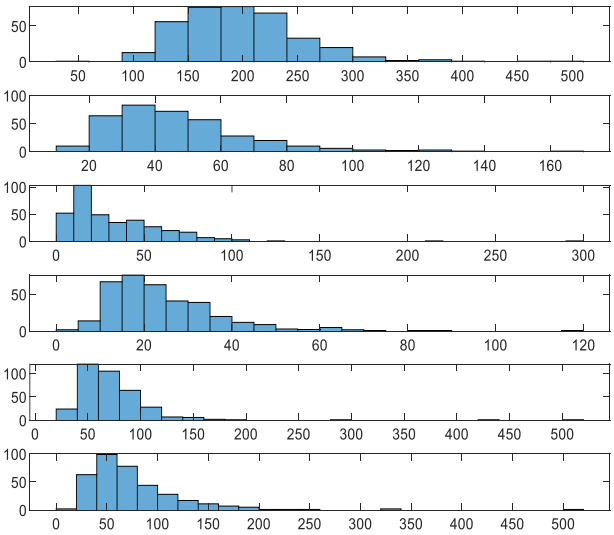

Figure. 2 Histogram of frequency distribution for each category

According to the chart data observation, the sales volume of most leafy vegetables is distributed between 120 and 300 in the first year, between 60 and 300 in the second year, and between 50 and 300 in the third year. This indicates that there has been some fluctuation in the sales volume of leafy vegetables over the past three years, but overall they remain at a relatively high level.

Unlike this, the frequency distribution of sales volume of aquatic vegetables has almost the same trend over the past three consecutive years, which means that the sales volume of aquatic vegetables has remained relatively stable during this period without significant fluctuations.

In addition, in the sales volume of chili vegetables in the next two years, the number of samples with the highest sales frequency has increased year by year. This may indicate that the market demand for chili vegetables has gradually increased in the past two years, leading to an increase in sales volume.

2.4. Correlation analysis of sales volume of various categories of vegetables

From the frequency distribution histogram in the above figure, it can be seen that the distribution trend in the dataset is to the left, showing a non normal distribution trend. Therefore, using Spearman correlation analysis[7], the steps are as follows:

Step 1:Data preparation: Assuming there are two variables X and Y, each with n observations. Firstly, sort the X and Y observations and assign rank to obtain the two ranking variables $R_x$andR_y.

Step 2:Calculate the rank difference: for each pair of rankings $d_i = R_{xi} - R_{yi}$,where i represents the i-th observation value.

Step 3:Calculate Spearman rank correlation coefficient: Use the following formula to calculate.

$$\rho = 1 - \frac{6\sum_{i=1}^{6} d_i^2}{n(n^2-1)} \quad (1)$$

In the equation, $d_i$represents the difference in the rank values of the i-th data pair, and n represents the total number of observed samples.

By calculating the correlation coefficient between the sales of various categories of vegetables, Figure 3 is plotted as follows.

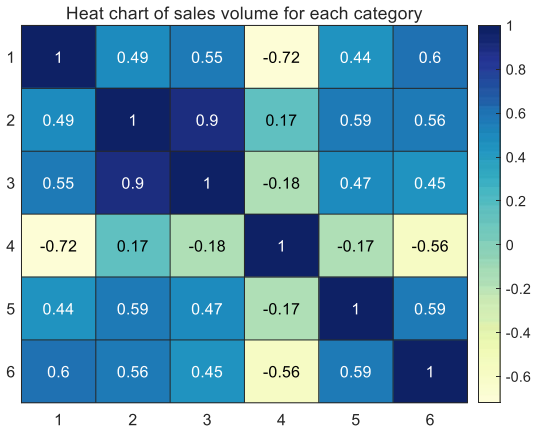

Figure. 3 Heat chart of sales volume for each category

Based on the above chart data analysis, we can draw the following conclusions:

1. There is a high positive correlation between leafy vegetables and cauliflower vegetables, with a correlation coefficient ρ is 0.9. This means that their sales show a strong positive correlation over time, with an increase in sales of one vegetable usually accompanied by an increase in sales of another vegetable.

2. There is a high positive correlation between aquatic rhizome vegetables and edible mushroom vegetables, with a correlation coefficient ρ is 0.6. This indicates that their sales volume is positively correlated to some extent and may be influenced by similar market demand or seasonal factors.

The correlation coefficient between eggplant vegetables and aquatic rhizome vegetables is -0.72, showing a negative correlation. This means that when the sales of eggplant vegetables increase, the

sales of aquatic rhizome vegetables usually decrease. This negative correlation may reflect the competitive relationship or market substitutability between the two.

2.5. Analysis of the Relationship between Vegetable Categories and Individual Products Based on Grey Correlation Analysis

Grey correlation analysis[8] is a method used to solve multi indicator decision-making problems, especially in the presence of uncertainty. It evaluates the pros and cons of different schemes or objects by calculating the degree of correlation between various indicators.

Firstly, we need to clearly define different vegetable categories as reference sequences, namely floral and leafy, cauliflower, eggplant, aquatic rhizome, chili, and edible fungi, with corresponding vegetable items becoming subsequences. Next, we use MATLAB to perform dimensionality reduction on the raw data, which helps eliminate dimensional differences between different indicators and makes them comparable for correlation analysis. The data covers 36 samples on a monthly basis, spanning from July 1, 2020 to June 30, 2023.

Reference sequence:

$$y_0 = (y_0(1), y(2), \cdots, y_0(36))^T \tag{2}$$

Subsequence:

$$x_1 = (x_1(1), x_1(2), \cdots, x_1(36))^T$$

$$x_{1091} = (x_{1091}(1), x_{1091}(2), \cdots, x_{1091}(36))^T \tag{3}$$

Hypothesis α is the minimum difference between two levels β is the maximum difference between the two poles, then there is:

$$\alpha = min(l)min(k)|y_0(k) - x_i(k)|$$

$$\beta = max(i)max(k)|y_0(k) - x_i(k)| \tag{4}$$

The formula for the correlation coefficient Y is:

$$Y_{(y_0(k),x_i(k))} = \frac{\alpha+\beta^*\rho}{|y_0(k),x_i(k)|+\beta^*\rho} \tag{5}$$

Assuming the grey correlation degree is: $Y(y_0(k), x_i(k))$, then there is:

$$Y(y_0(k), x_i(k)) = \frac{1}{n}\sum_{k=1}^{36} Y(y_0(k), x_i(k)) \tag{6}$$

Calculate the arithmetic mean root of each column, which is the degree of correlation between each sub sequence and the parent sequence. Use Matlab programming to solve, and finally visualize the grey correlation coefficients between the five sub categories and the parent sequence in the cauliflower class, as shown in Figure 4.

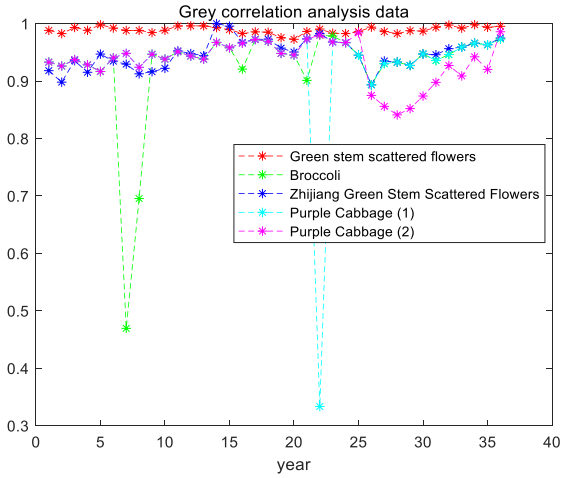

Figure. 4 Grey correlation analysis data

In the 36 month data analysis, we averaged the different distributions of the grey correlation coefficient and obtained the ranking of the grey correlation coefficient in each vegetable category. The following is the ranking result of the grey correlation coefficient for cauliflower and other five major categories of vegetables, as shown in Table 2.

Table 2 Grey relational coefficient

Peppers	florescent vegetables	Florifolias	Aquatic rhizomes	Solanaceae	edible fungi
Wuhu Green Pepper (1) (0.9728)	Green stem scattered flowers (0.9889)	Shanghai Qing (0.9687)	Clean Lotus Root (1) (0.9864)	Purple Eggplant (2) (0.9872)	White Jade Mushroom (Bag) (0.9722)
Screw pepper (0.9637)	Zhijiang Green Stem Scattered Flowers (0.948)	Yellow Cabbage (2) (0.9661)	Honghu Lotus Root (Pink Lotus Root) (0.9535)	Green Eggplant (1) (0.9747)	Xixia Mushroom (1) (0.9682)
Red pepper (1) (0.9575)	Purple Cabbage (2) (0.9347)	Milk cabbage (0.9543)	Water chestnut (0.9377)	Round eggplant (2) (0.9566)	Golden Needle Mushroom (Box) (0.9471)
Xiaomi Pepper (portion) (0.9568)	Purple Cabbage (1) (0.9294)	Baby Cabbage (0.9623)	Gaogua (2) (0.9370)	Long Line Eggplant (0.9487)	Tremella fuciformis (flower) (0.9467)
Green Pepper (0.9548)	Broccoli (0.9248)	Spinach (0.9623)	Gaogua (1) (0.9344)	Da Long Eggplant (0.9346)	Apricot Abalone Mushroom (Bag) (0.9447)

3. Establishment and Solution of Entropy Weight-TOPSIS Model

3.1. A sellable item screening model based on entropy weight Topsis

After conducting preliminary data screening to obtain 49 individual products, an entropy weighted Topsis[9] model was established, with the following algorithm.

Step 1: Forward processing for selecting importance indicators in single category vegetables

Very large indicators (benefit based indicators)

$$r'_{ij} = \frac{r_j - r_{min}}{r_{max} - r_{min}} \tag{7}$$

Very small indicators (cost indicators)

$$r'_{ij} = \frac{r_{max} - r_j}{r_{max} - r_{min}} \tag{8}$$

After processing, it can form a data matrix $R = (r_{ij})_{m \times n}$

According to the above formula, the importance of evaluating different single category vegetables needs to consider the following factors:

Total sales volume of single category vegetables r_1 (Benefit based indicator): Calculate the total sales volume of these 49 single category vegetables. The larger the total sales volume, the more important it is.

Average loss rate of single category vegetables r_2 (Cost based indicator): Calculate the average loss rate of these 49 single category vegetables. The larger the average loss rate, the higher the cost and the less important it is.

Number of discounted sales of single category vegetables r_3 (Cost based indicator): Calculate the number of times 49 single category vegetables have been discounted for sale within three years. Considering that the more discount sales, it indicates that the single category vegetable needs to be discounted to promote sales. Customers have a low purchase rate when not discounted, so it is believed that the greater the discount sales, the lower its importance.

Average selling price of single category vegetables r_4 (Benefit based indicator): Calculate the average pricing of 49 single category vegetables sold within three years. The higher the average pricing, the greater the profit. Therefore, the higher the pricing, the greater its importance.

Step 2: Standardize data processing

$$z_i = \frac{r_i}{\sqrt{\sum_{i=1}^{n} r_i}} \tag{9}$$

Step 3: Determine positive ideal solution and negative ideal solution
Define the maximum value of each indicator, that is, each column, as r_j^+

$$r_j^+ = max(r_{1j}, r_{2j} \dots, r_{nj}) \tag{10}$$

Define the minimum value of each indicator, i.e. each column, as r_j^-

$$r_j^- = min(r_{1j}, r_{2j} \dots, r_{nj}) \tag{11}$$

Step 4: Calculate the weight of indicators using the entropy weight method
Calculate the weight of each indicator using the entropy weight method

$$S_i = \sum_{j=1}^{n} w_j r_{ij}$$

The results are shown in Table 3.

Table 3 Weight of evaluation factors

Evaluation factors	Total sales volume	Average loss rate	Discount sales times	Average pricing
weight	0.5669	0.1052	0.0375	0.2904

Step5: Calculate Euclidean distance
Define the distance between the i-th evaluation object and the maximum value as d_i^+

$$d_i^+ = \sqrt{\sum_{j=1}^{n} \left(r_j^+ - r_{ij}\right)^2} \tag{12}$$

Define the distance between the i-th evaluation object and the minimum value as d_i^-

$$d_i^- = \sqrt{\sum_{j=1}^{n} \left(r_j^- - r_{ij}\right)^2} \tag{13}$$

Step6: Calculate ratings

$$\text{Score}_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{14}$$

The final evaluation results of the importance of 49 single category vegetables are shown in the chart. Due to limited sales space in supermarkets, only 27 to 33 single category vegetables can be purchased. Therefore, the scores and rankings of the top 30 single category vegetables are shown in Figure 5.

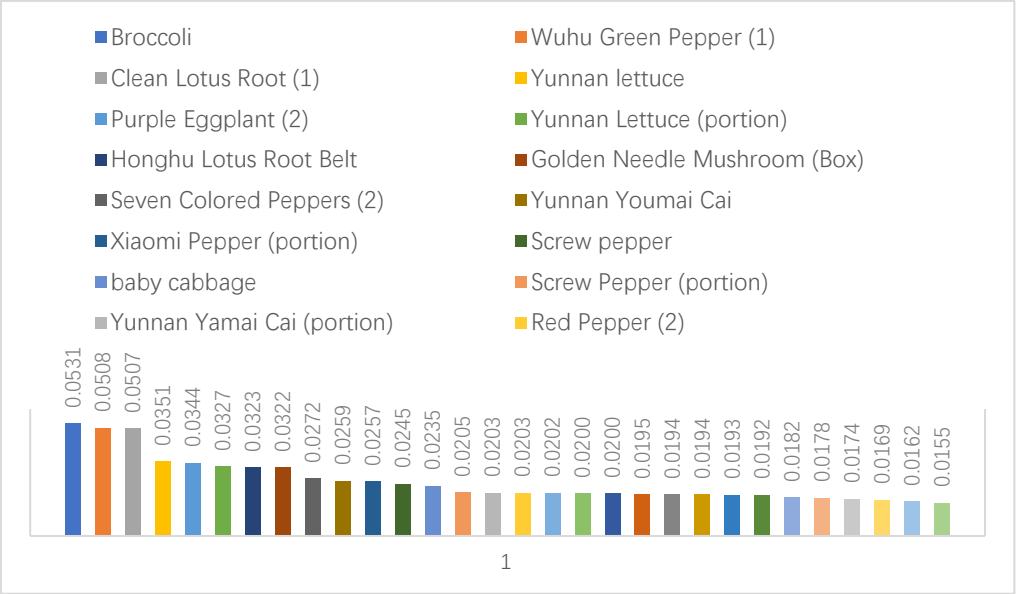

Figure. 5 Single product comprehensive score index bar chart

In the subsequent planning model, this article will display the replenishment quantity and pricing strategy of the top 30 single category vegetables as decision variables, in order to obtain the replenishment plan and sales strategy under the maximum profit margin.

3.2. Solution of Pricing Linear Programming Model Based on Ant Colony Algorithm

Decision variable: Based on the above analysis, we establish decision variable p: pricing strategy

Using ARIMA to predict the cost of these 30 single category vegetables, and replacing the loss rate on July 1st with the average loss rate, using linear regression[10] and ant colony algorithm.

Objective Function: Maximum Supermarket Return

$$(P_1 - C_1) \times S_1 \tag{15}$$

$$C_1 = (1 + \partial) \times C_2^* \tag{16}$$

C_1represents cost, C_2^* represents the cost predicted through linear regression, S_1represents sales volume, and partial represents loss rate.

Constraint 1: Do not consider return situations

Constraint 2: The order quantity of a single bottle meets the requirement of a minimum display quantity of 2.5 kilograms

Constraint 3: Control the total number of sold items between 27 and 33

Constraint 4: The order quantity and pricing strategy of each individual product need to meet the supply and demand relationship of the market for each category of vegetable products

In summary, constructing a combination pricing model that maximizes the revenue of supermarkets and supermarkets

$$max[P_2 - (1 + \beta) \times C_2^*] \times S_2 \tag{17}$$

$$\begin{cases} 0 \leqslant S_2 \leqslant N_{max}(1 - \beta) \\ 0 \leq \partial \leq 1 \\ l_1 = -1.83x + 44.31 \\ l_2 = -0.85x + 39.35 \\ \cdots\cdots \\ l_6 = 0.12x + 8.38 \end{cases} \tag{18}$$

Finally, the maximum supermarket profit for 30 single category vegetables is 912.235 yuan, and the corresponding optimal file is obtained: pricing and replenishment quantity for each category, as shown in Table 4.

Table 4 Pricing and Replenishment Volume

Item Name	Pricing/yuan	Replenishment quantity/KG	Item Name	Pricing/yuan	Replenishment quantity/KG
broccoli	20.46	2.51	Red Pepper (2)	19.37	2.51
Wuhu Green Pepper (1)	42.25	2.67	spinach	14.29	4.76
Clean Lotus Root (1)	21.23	2.53	Shanghai Qing	8.59	3.54
Yunnan lettuce	16.56	5.49	Gaogua (2)	24.36	2.51
Purple Eggplant (2)	14.76	2.78	Cauliflower	11.52	4.48
Yunnan Lettuce (portion)	11.94	15.49	Rhombic horn	12.69	2.56
Honghu Lotus Root Belt	40.32	2.68	Bamboo leaf vegetable	16.68	3.75
Golden Needle Mushroom (Box)	12.86	3.48	Wild Flour Lotus Root	23.56	5.46
Seven Colored Peppers (2)	18.56	4.29	Echinacea officinalis	5.76	7.89
Yunnan Youmai Cai	13.73	2.53	Gaogua (1)	15.62	4.47
Xiaomi Pepper (portion)	11.49	2.63	Long line eggplant	19.34	3.58
Screw pepper	17.28	2.82	Zhijiang Green Stem Scattered Flowers	29.45	4.67
baby cabbage	8.03	2.61	Spinach (portion)	5.76	4.38
Screw Pepper (portion)	11.94	2.57	Amaranth	4.49	5.72
Yunnan Yamai Cai (portion)	17.26	2.53	Red Lotus Root Band	14.38	2.67

4. Conclusion

In the first part, statistical analysis and correlation analysis and test were conducted for each category of vegetables and each single product. For example, the correlation coefficient between eggplant vegetables and aquatic root vegetables was -0.72. Finally, gray correlation analysis was used to find single products with strong correlation between similar single products.

In the second part, a total of 49 single products available for sale were firstly screened out, and then four indicators, namely, total sales volume, average loss rate, number of discounted sales, and average pricing, were selected to construct a single product category vegetable selection model based on entropy weight-TOPSIS, and the weight situation of the four indicators was calculated by using entropy weighting method, and the most weighted one was the total sales volume, which accounted for 0.5669. meanwhile, the top 30 single product category vegetables were selected to be ranked in the top 30, and the top three of the combined rankings were selected to be ranked in the top 30. The top three comprehensive rankings are, in order: broccoli, Wuhu green pepper (1), net lotus root (1). The maximum revenue combination pricing strategy model is constructed and solved, and the final

profit is 912.235 yuan, and the pricing strategy and replenishment quantity corresponding to the 30 individual categories of vegetables are given.

In the construction of this model, we have not yet considered the impact of factors such as seasonal factors and service quality on the actual superstore profits. Therefore, in the subsequent improvement, we will consider adding the time series decomposition of sales volume to the model to determine its seasonal fluctuation as well as the influence of service quality and so on through the covariance analysis method in case of insufficient data.

References

[1] Gong Xuemei Research on the Control and Governance of the Value Chain of Vegetable Circulation in Supermarkets [D]. Shandong University, 2017.

[2] The palace is famous Research on the Optimization of Supply and Demand Connection of Vegetable Products in Shandong Province [D]. Shandong University of Technology, 2020.

[3] Liu Lihong Research on the Mechanism of Vegetable Price Formation and Regulation Policies in Beijing [D]. Chinese Academy of Agricultural Sciences, 2016.

[4] Lan Tingting Research on Optimization Strategies for Vegetable Supply Chain Management Mode [D]. Changchun University of Technology, 2015.

[5] Zhao Anping, Wang Dashan, Xiao Jinke, et al. Decomposition and Analysis of Vegetable Price Time Series: Based on Data from Beijing from 2002 to 2012 [J]. Journal of Huazhong Agricultural University (Social Science Edition), 2014 (01): 49-53.

[6] Zhang He Vegetable sales prediction based on ensemble learning [D]. Yunnan Normal University, 2021.

[7] Xu Liqin, Zhou Qing, Liu Xia. Analysis and suggestions on the expression of "Spearman related" in scientific papers [J]. Journal of Editors, 2019,31 (04): 398-400.

[8] Zhou Ziyi. A Study on the Competitiveness Evaluation of Listed Food Manufacturing Enterprises Based on Grey Relational Analysis [J]. Renting and Selling Intelligence, 2023 (07): 82-84.

[9] Zhang Yong. Evaluation of Water Resources Carrying Capacity of Xingfu River Lake in Anhui Province Based on Entropy Weight Method and TOPSIS Model [J]. Shaanxi Water Resources, 2023 (09): 45-47.

[10] Fu Yewei, Zhang Wenkan, Xu Shihao, et al. Study on Anti seepage and Precipitation Measures for Linshui Foundation Pit Based on Multiple Linear Regression Analysis [J]. Hydropower and New Energy, 2023,37 (08): 13-17.