Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 12, No. 2, 2024

Research on Supermarket Replenishment Pricing Decision Making Based on BP Neural Network and Greedy Algorithm

Wenjian Xiao [1,*], Yuanning Gu [2]

[1] School of Civil Engineering, Hebei University of Engineering, Handan, China
[2] School of Mathematics and Physics Science and Engineering, Hebei University of Engineering, Handan, China
[*] **Corresponding author**: Wenjian Xiao (Email: 3170108151@qq.com)

Abstract: The aim of this paper is to provide strong support for superstores through a data-driven approach. By analyzing and fitting the sales history data of each category and individual product, this paper uses BP-neural network model to predict the sales volume and transforms the replenishment problem into a knapsack problem, so as to plan the future pricing strategy and replenishment volume of the superstore. This helps to optimize the category structure of the superstore, increase the profit margin, reduce the wastage rate, and improve the service quality.

Keywords: Spearman model; Nonlinear regression model; Fitting algorithm; Neural network model; Greedy algorithm.

1. Introduction

Due to the freshness period of vegetable commodities and the limitations of the purchase transaction time, fresh food superstores usually replenish the next day's stock based on past sales experience and demand, and due to the different varieties and places of origin, planting costs and transportation costs are also different, so the cost of goods will be affected and indirectly affect the pricing of the goods. In order to reduce business risks and maximize profits, fresh produce superstores need to develop reasonable pricing strategies and replenishment totals.

In order to solve the above problems, this paper firstly visualizes the historical data to get the distribution pattern between different categories and single products in the year, quarter and important holidays, and applies the Spearman model to correlation analysis between different categories. Then, a nonlinear regression model is established to analyze the correlation between cost-plus pricing and the total sales of vegetable categories, and the relationship curve is obtained by using the fitting algorithm model. The BP-neural network model was used to predict the sales volume in the coming week and the corresponding pricing strategy was calculated. Finally, under the constraints of sales space, total number of individual items, and minimum display quantity, the pricing strategy of individual items and the replenishment quantity of individual items are analyzed, and this problem is transformed into a knapsack problem and solved by using the greedy algorithm to obtain the replenishment quantity of each individual item. The method proposed in this paper can better assist superstores in making replenishment and pricing decisions for vegetable items [1].

2. Distribution patterns and correlation analysis

2.1. Analysis of the distribution pattern of each category

In this paper, we use historical data to integrate and obtain the total sales volume of each category summarized by month and by quarter, and draw the corresponding sales line graph, which makes the data more intuitive.

A summary of monthly sales volume and quarterly sales volume for each vegetable category is shown in Figure 1 and Figure 2 below:

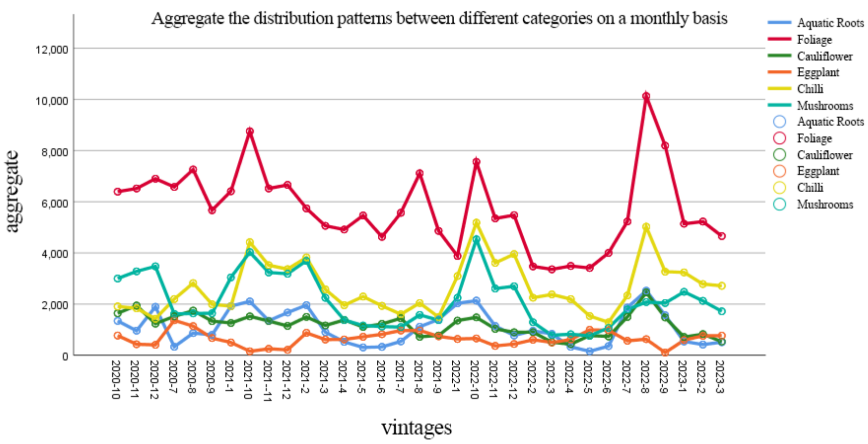

Figure 1. Monthly summary of sales volume by category

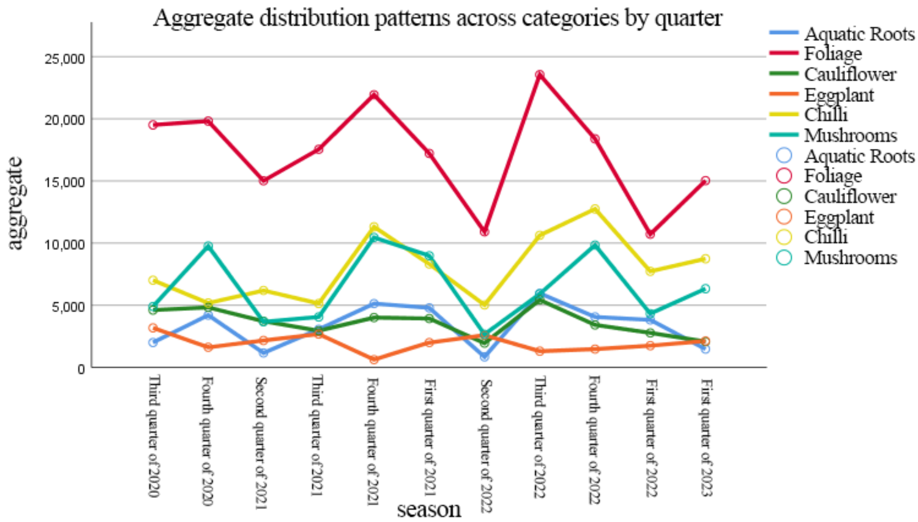

Figure 2. Quarterly Sales Volume Summary

From the line graph Figure 1and Figure 2, we can learn that the sales volume of leafy vegetables, cauliflower and chili peppers will peak in July-August and October-November every year, which means that the consumption of these three categories of vegetables is in the peak season in these two time periods, and the main sales volume of these three categories of vegetables occurs in these two time periods. The sales volume of eggplant is mainly concentrated in April and May every year, and the sales volume of eggplant rises briefly in this period due to the decline in the consumption of leafy vegetables and cauliflower, while the consumption of eggplant remains in a low state in the rest of the year. Sales of aquatic roots and mushrooms peak in August and September when they are harvested and are extremely cost-effective in terms of taste and price.

2.2. Correlation analysis by category

In statistics, the Spearman correlation coefficient shows the direction of correlation between X (independent variable) and Y (dependent variable) [2]. Spearman's correlation coefficient gives the correlation between X and Y described by the linear equation. Spearman's correlation coefficient is calculated by the formula:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{1}$$

As the absolute value of the correlation coefficient approaches 1, the correlation is stronger. The closer the absolute value of the correlation coefficient is to 0, the stronger the correlation is.

In this paper, the total sales volume of each vegetable category from July 2020 to March 2023 was extracted and calculated using Spearman analysis, and the final heat map of correlation coefficients is shown in Figure 3.

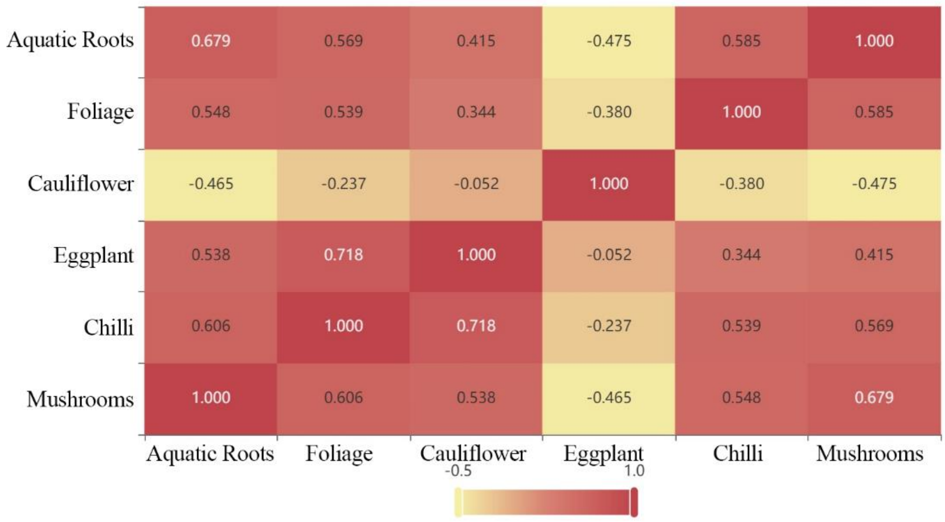

Figure 3. Heat map of correlation coefficient

From the results in Figure 3, we can see that:

(1) The sales volume of eggplant is negatively correlated with the sales volume of the other five categories, i.e., an increase in the sales volume of eggplant will reduce the sales volume of the other categories.

(2) All five categories except eggplant are positively correlated with each other to some extent.

(3) The average positive correlation between foliage and

other categories is the highest, so consumers prefer leafy and flowering vegetables in the consumption process.

3. Pricing strategy

3.1. Relationship between sales volume and cost-plus pricing

Cost-plus pricing is a method of setting the price of a product based on the unit cost of the product plus a percentage of the profit. Most companies determine the size of the profit added by the cost margin. That is:

$$P = C + C * R = C(1 + R) \quad (2)$$

Since cost-plus pricing is unit pricing, unit sales prices will be used in place of cost-plus pricing below.

In this paper, the sales volume as the dependent variable, the sales unit price as the independent variable, the establishment of multiple linear regression model to explore the relationship between the two [3]. The results are shown in Table.1.

Table 1. Linear regression analysis results

	Unstandardized coefficient		Standardized coefficient	t	p	VIF	R_2
	B	Standard error	Beta				
Constant	0.71	0.017		40.947	0.000***		0.003
Unit sales price (yuan/kg)	-0.007	0.002	-0.055	-3.784	0.000***	1	

R_2: coefficient of multiple determinations, the closer the value is to 1, the stronger the explanatory ability of the variables of the equation to y. The analysis of the results of F-test shows that the p-value is 0.000***, which is significant at the level of 0, and the original hypothesis of the regression coefficient is 0 is rejected, so the model basically meets the requirements. As for the covariate covariance, VIF is less than 10, so the model has no multicollinearity problem and the model is well constructed. The formula of the model is as follows:

$$y = 0.71 - 0.007 * P \quad (3)$$

We use the fitting algorithm to test the regression equation setting the total number of sales as Q.

$$Q = \Delta_0 + \Delta_1 * P \quad (4)$$

The Δ_0 and Δ_1 are involved in the model as constant parameters, and the estimated values of Δ_0 and Δ_1 can be obtained by fitting the model, so as to get the specific parameters of the model, in addition, this paper will also use the fitting algorithm to test the data obtained [4], so as to verify its reasonableness, in order to check the accuracy of the model more intuitively, this paper plots the effect of the scatter fitting, as shown in Figure 4.

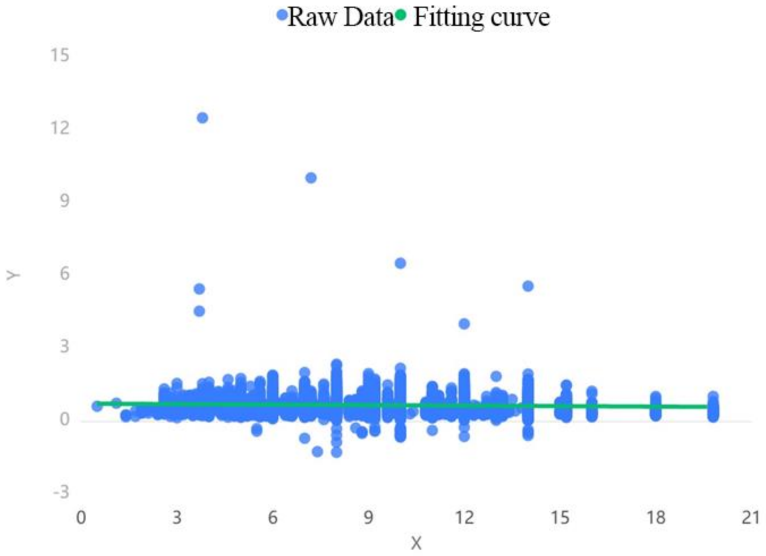

Figure. 4 Fitting effect diagram

The fitting effect Figure 4 shows that the scatter points are better combined with the predicted curves, indicating the scientific and reasonable nature of the model.

Based on the relationship between total sales volume and cost-plus pricing for each vegetable category, in this section, a bp neural network model is used for predicting future data using sales date as x, sales volume of foliage, sales volume of pepper, sales volume of tomato, sales volume of edible mushrooms, sales volume of aquatic roots and tubers, and sales volume of cauliflower as y. In this problem, we can use BP neural network for predicting future sales volume and pricing strategy based on this model. In this problem we can use BP neural network to model the relationship between total sales and cost-plus pricing of vegetable category and based on this model we can predict the future sales and pricing strategy [5].

The sales data up to the last 30 days is fitted and predicted by BP neural network, and the data for the next 7 days is predicted directly.

Table 2. Table of correlation between categories

	RMSE	MAE	MAPE	R_2
Training set	3.088	2.69	15.871	0.774
Test set	8.011	6.405	53.652	-0.531

We evaluated the model first, from the results shown in

Table.2, the RMSE, MAE, MAPE, R_2 of the training set take smaller values than the values of the test set, then the accuracy of the numerical model with the training set is higher [3].

From the comparison, it can be seen that the model fit of the training set is better than that of the test set, so the training set model is used, and the final sales volume prediction results are obtained as shown in Table.3, and the pricing strategy is shown in Table.4.

Table.3. Forecasted sales volume

	Foliage	Cabbage	Aquatic roots	Tomatoes	Peppers	Edible mushrooms
July 1	152.374	19.9373	20.85	23.763	95.952	55.13958
July 2	115.9889	19.369	15.82115	18.89291	65.97154	38.28986
July 3	98.108	15.48477	12.83722	16.665	56.84513	33.03559
July 4	115.7406	13.87391	14.18126	18.28795	69.6	40.21797
July 5	123.5264	16.4125	17.2546	19.5421	77.6548	45.2156
July 6	137.2546	18.2456	19.5526	22.1456	89.6589	50.5896
July 7	136.3263	18.2563	19.00021	21.6359	85.3369	47.9522

Table 4. Pricing Strategy

	Foliage	Cabbage	Aquatic roots	Tomatoes	Peppers	Edible mushrooms
July 1	0.6814	0.405	0.49109	0.53781	0.64409	0.54361
July 2	0.64675	0.512502	0.57377	0.524326	0.60185	0.598898
July 3	0.628	0.5252	0.5349	0.5475	0.56621	0.58996
July 4	0.665	0.58984	0.552481	0.625143	0.254896	0.584611
July 5	0.65895	0.58421	0.57698	0.59632	0.59632	0.56241
July 6	0.64897	0.58412	0.56324	0.56412	0.60125	0.61520
July 7	0.58976	0.59632	0.57412	0.58662	0.59426	0.57623

4. Replenishment volume decisions

Because of the limited sales space of vegetable goods, the superstore wants to further develop the replenishment plan of single product to ensure the maximum words of superstore revenue, so first we first extracted out the available varieties in June 24-30, 2023, and for the single product type of which, we calculate its average sales, average profit, average wholesale unit price and loss rate.

The average sales volume is calculated by the formula:

$$S_i = \frac{\sum_{i=1}^{n} S}{n_i} \tag{5}$$

S is the sales of each individual item, S_i is the average sales per individual item?

The average profit is calculated by the formula:

$$L_i = \frac{\sum_{i=1}^{n} S-P_i}{n_i} \tag{6}$$

L is the profit per individual product, L_i is the average profit per individual item?

The average wholesale unit price is calculated as:

$$P_i = \frac{\sum_{i=1}^{n} P}{n_i} \tag{7}$$

P is the wholesale unit price per individual item, P_i is the average wholesale unit price per individual item?

We need to clean the data further, as we need to meet the minimum display quantity of 2.5 kg for each item. Due to the short shelf life of vegetables, most vegetables are sold on the same day and cannot be sold the next day. Therefore, we introduce the loss rate data to correct the shelf life and establish the correlation equation between the shelf life and the loss rate, which is used to further filter the unqualified items, i.e., the longer the shelf life of the items, the lower the loss rate, so in the case that we use $F = 2.5$ as the threshold, the unqualified items will be further eliminated.

The loss rate is analyzed and the box line plot of loss rate is obtained as shown in Figure 5.

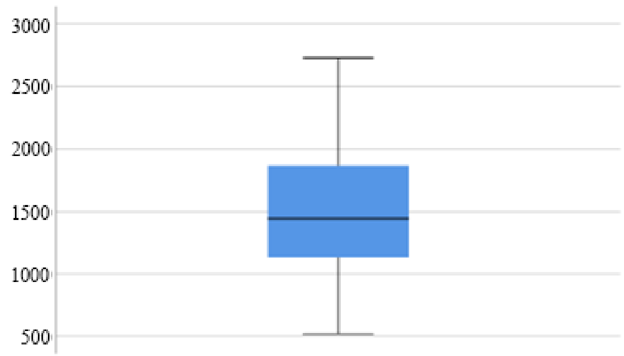

Figure 5. Loss rate boxplot

Modeled accordingly:

$$C = \begin{cases} 1, k > 9.56\% \\ -\frac{k}{k_m} + 3, k \le 9.56\% \end{cases} \tag{8}$$

Where the wastage rate is K, the average wastage rate of the sample is K_m, and C is the corrected shelf life.

In the process of merchandise display, we sometimes need to divide, categorize and proportion merchandise and other techniques to better meet the demand for display space minimization. For example, when displaying a large number of commodities, we can use standard-sized shelves or display racks as a benchmark to divide and categorize commodities according to their sizes and shapes, so as to maximize the number of commodities displayed on each shelf or display rack. At the same time, we can also use the matching technique, the different sizes and shapes of goods with each other, so that the display space can be more reasonable use.

This transforms the problem into a typical knapsack problem. Sales, profit, unit price, are the typical elements of a knapsack problem.

The knapsack problem is a classic combinatorial optimization problem, which can be described in the following terms: given n items, each with its own weight and value, the requirement is to select a number of items within a finite total weight so as to maximize the total value of the items. Greedy algorithms are used in fractional knapsack problems, where the global optimal solution is achieved by continuously selecting the currently optimal item. The backtracking search solves all variants of the knapsack problem by enumerating all possible combinations to find the optimal solution [6].

Finally, based on the processed data, we finally get 28 items to choose from in a comprehensive comparison. Considering the case of using the total description of the category as a backpack by the question, the individual products within each category are calculated independently, i.e., a total of six backpacks, after which the optimal selection method is obtained through the greedy algorithm in the case that it is able to obtain a better result [7]. The results are shown in Table.5.

Table 5. Final results

	Yunnan Lettuce (servings)	Peppers (portions)	Yunnan Oil Wheat Vegetable (servings)	Enoki Mushroom (box)	Wuhu Green Pepper	Brussels sprouts	Broccoli
Replenishment	17.711	15.288	11.895	7.566	52.696	7.589	19.856
Profit ($/kg)	3.152	2.633	2.898	3.656	2.290	1.896	3.365
Profit ($)	56.162	36.589	28.989	23.899	117.589	15.113	65.896
	Screw Peppers (servings)	Small wrinkled skin (portions)	Purple Eggplant	Mini-sized variety	Agaricus bisporus (box)	Amaranth greens	King cobra
Replenishment	11.356	5.508	16.656	16.563	7.885	5.412	11.188
Profit ($/kg)	2.165	1.7855	2.96	1.823	1.952	1.863	3.158
Profit ($)	24.656	9.123	50.152	29.165	14.982	9.685	36.598
	Cabbage	Net root	Bok choy	Xixia Mushroom	Sweet potato tip	Long term eggplant	Shanghai Youth
Replenishment	6.173	20.589	3.586	2.727	6.029	2.981	8.236
Profit ($/kg)	1.896	2.789	2.458	7.586	2.561	3.962	3.026
Profit ($)	10.968	47.362	8.36	21.365	15.36	12.103	23.976
	Green Eggplant	Yunnan lettuce	Chinese flowering cabbage	Broccoli	Yunnan oilseed rape	White Mushroom (Bag)	Red Pepper
Replenishment	3.656	17.121	4.757	5.869	11.188	3.764	4.82
Profit ($/kg)	2.586	3.252	2.653	3.869	2.989	2.635	6.89
Profit ($)	8.898	53.626	15.689	21.588	30.156	9.865	31.658

5. Conclusion

In order to optimize the category structure of superstores, improve profitability and reduce risk, this paper analyzes the sales parameters of each category and each single product of superstores, so as to provide more favorable support for them. For the correlation relationship between the sales volume of different categories and individual products of vegetables and their distribution patterns, this paper uses the Spearman model for correlation analysis. In order to maximize the profit of the superstore, a nonlinear regression model is established to carry out correlation analysis between the cost-plus pricing and the total sales volume of vegetable categories, and the relationship curve is obtained by using the fitting algorithm model, and the sales volume is predicted by using the BP-neural network model, and the corresponding pricing strategy is calculated. Under the condition of limited space for vegetable sales, this paper transforms the replenishment planning problem into a knapsack problem and solves it using the greedy algorithm to obtain the replenishment quantity of various vegetable individual items. The final results show that the model proposed in this paper has high accuracy and robustness, thus providing strong support for superstores.

References

[1] Zhan Mengting. Research on strategic cost management of Yonghui Supermarket under new retail mode [D]. Jiangxi University of Finance and Economics, 2023.

DOI: 10. 27175/ d. cnki. gjxcu. 2023. 001487.

[2] Yu Qun,Huo Xiaodong,He Jian et al. Trend prediction of power outages in China based on Spearman correlation coefficient and system inertia[J]. Chinese Journal of Electrical Engineering, 2023,43(14):5372-5381.

DOI:10.13334/j.0258-8013.pcsee.220035.

[3] Yang XL,Wang MX,Li XJ et al. Machine learning-multiple linear regression technique to predict water ecological benchmarks for copper[J/OL]. China Environmental Science, 1-11[2024-03-14].

https://doi.org/10.19674/j.cnki.issn1000-6923.20240312.002.

[4] LIU Junli,ZHANG Xu,LI Quanyang et al. Selection of CMM fitting algorithm in engineering measurement[J]. Micromotors, 2023, 56(12):60-62+ 66.

DOI: 10. 15934/ j. cnki. micromotors. 2023. 12.011.

[5] A.D. Zhong,Z.Y. Wu,Z.X. Xie et al. Research on photovoltaic power prediction based on hybrid algorithm improved BP neural network[J/OL]. Journal of Yunnan University for Nationalities(Natural Science Edition), 1-7[2024-03-14].

http: // kns. cnki. net/ kcms/ detail/ 53. 1192. n. 20240308. 1602. 002.html.

[6] S. F. Zhuang,Y. Ching Wang,J. Luo et al. An optimised simulated annealing algorithm for solving the knapsack problem[J]. Computer Programming Skills and Maintenance, 2023,(11):53-56.

DOI: 10. 16184/j.cnki.comprg.2023.11.023.

[7] Xiao Jian,He Zhicheng,Hu Xin et al. Tree WSN low-power routing algorithm based on greedy algorithm[J]. Internet of Things Technology,2024,14(01):48-51+54.

DOI:10.16667/j.issn.2095-1302.2024.01.013.