Research on Pricing Strategy of Vegetable Commodities Based on BP Neural Networks

Shitan Niu *

School of Management Engineering, Qingdao University of Technology, Qingdao, China

* Corresponding Author Email: 13053720677@163.com

Abstract. This study aims to optimize the replenishment strategy of vegetable category in fresh supermarkets and explore the relationship between vegetable sales volume and pricing by analyzing the sales data using EXCEL, SPSS, and MATLAB. In data preprocessing, we dealt with missing values and abnormal data, and found that the flower and leafy vegetables had the largest and stable sales share. After the unsatisfactory linear fitting model through SPSS, we turned to BP neural network for pricing prediction, considering the correlation between time and sales volume. Our findings provide a reference for superstores to develop more accurate pricing strategies, which can help improve their competitiveness and sustainability to meet the challenges of market changes.

Keywords: BP neural network; least square; pricing forecasting model.

1. Introduction

In today's highly competitive fresh produce supermarket industry, effective and reasonable pricing is critical to superstores' operations [1]. For the characteristics of vegetables, which are commodities with short freshness period and time-dependent sales volume, superstores must make accurate decisions before daily stocking [2]. This study aims to explore the relationship between vegetable sales volume and pricing by analyzing historical sales data to optimize the pricing strategy of superstores and improve operational efficiency and competitiveness. Combining tools such as EXCEL, SPSS, and MATLAB, we will analyze and forecast the sales data, and use methods such as BP neural network to develop more accurate pricing prediction models to adapt to changes in market demand. Through this study, we aim to provide effective operational decision support for fresh food supermarkets to achieve more robust and sustainable development.

2. Data Preprocessing

2.1. Outlier Handling

In order to analyze the correlation relationship and classification law between different categories and different individual vegetables, firstly, we preprocess the data, make up 0 for the missing values, consider the influence of outliers, calculate the total daily sales volume of each category and each individual product, and then visualize and analyze the data to get the trend and law of the different categories of vegetables over time.

According to the single product code corresponds to the name of the single product and the category name; then organize the data, select one day as the unit, get the superstore each category of vegetables a day sales volume sum and each single product a day sales total, as shown in Table 1 below.

Table 1. Total daily sales volume of each vegetable category.

Date	Cabbage	Foliage	Pepper	Eggplant	Mushroom	Aquatic root	Red cabbage
2020-07-01	46.640	205.402	76.715	35.374	35.365	4.850	0.000
2020-07-02	43.943	198.362	66.064	32.199	48.510	4.600	0.000
2020-07-03	42.076	190.779	64.253	35.896	42.442	9.572	0.000
2020-07-04	55.662	236.587	81.282	57.067	47.262	5.439	0.000
2020-07-05	55.474	223.899	98.496	61.816	73.213	4.019	0.000
2020-07-06	42.721	195.740	58.682	30.727	40.578	3.089	0.000
2020-07-07	45.749	172.052	42.213	32.112	52.560	3.117	0.000
2020-07-08	45.433	152.570	65.880	39.670	43.968	10.605	0.000
.....	.....	.....	.....	.....	.....	.....	.....

According to the statistics, the sales time of each vegetable category and single product is not continuous, which means that there are not sales every day, and the sales volume of vegetables on that day is 0. Therefore, we uniformly make up for the missing values by replacing them with zero.

After searching the data, we found that some of the raw data have anomalies, as shown in Table 2, the unit price of a certain point in time and the rest of the day there is a large gap, such as July 25, 2020 19:01:34, the unit price of sweet potato tips is 0.1 yuan / kg, significantly lower than the rest of the time the unit price of the sale of the sale of the abnormal value is ignored due to the sale of too small, so the anomaly.

Table 2. Analysis of sweet potato tip outliers.

Date	Sweeping sales time	Product Code	Product name	Volume	Selling unit price	Sales type	Discounted or not
2020-7-25	16:11:56.447	102900005119975	Sweetpotato tip	0.402	4.7	Sales	Yes
2020-7-25	17:01:07.010	102900005119975	Sweetpotato tip	0.522	4.3	Sales	Yes
2020-7-25	17:42:09.514	102900005119975	Sweetpotato tip	0.261	4.6	Sales	Yes
2020-7-25	19:01:34.170	102900005119975	Sweetpotato tip	0.201	0.1	Sales	Yes
2020-7-25	19:11:20.733	102900005119975	Sweetpotato tip	0.221	4.5	Sales	Yes
2020-7-25	19:15:20.686	102900005119975	Sweetpotato tip	0.221	4.5	Sales	Yes

2.2. Visualization and analysis

After data preprocessing, the data set contains 878504 sales records from July 1, 2020 09:15:07 to June 30, 2023 21:40:48, a total of 1049 days. In order to analyze and discuss the correlation, distribution pattern and correlation between vegetable categories, we divided the products into six categories according to the data, and at the same time, we selected one day as the unit, summed up the sales volume of each category every day to get a summary chart and visualized the data.

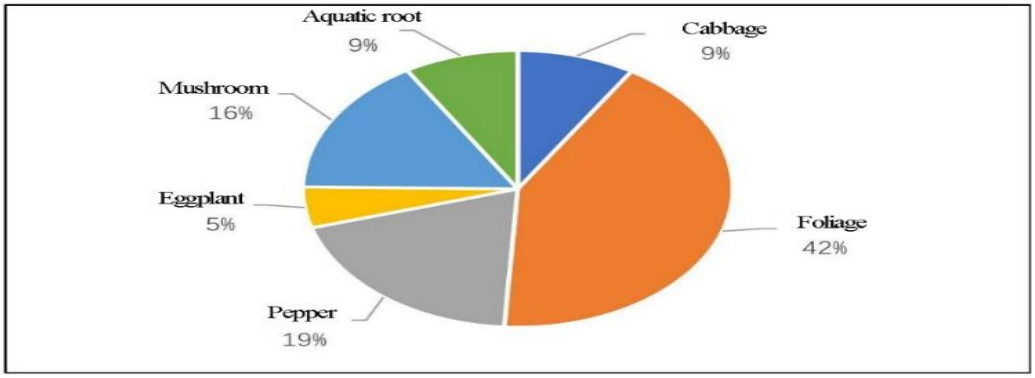

Fig 1. Sum of daily sales by category.

Comparison of Figure 1 shows that the foliage category accounted for the largest share of sales, nearly half of the total sales, while the aquatic root and cabbage categories accounted for the same share of sales, mushroom and pepper were close to each other, and eggplant had the smallest share of sales.

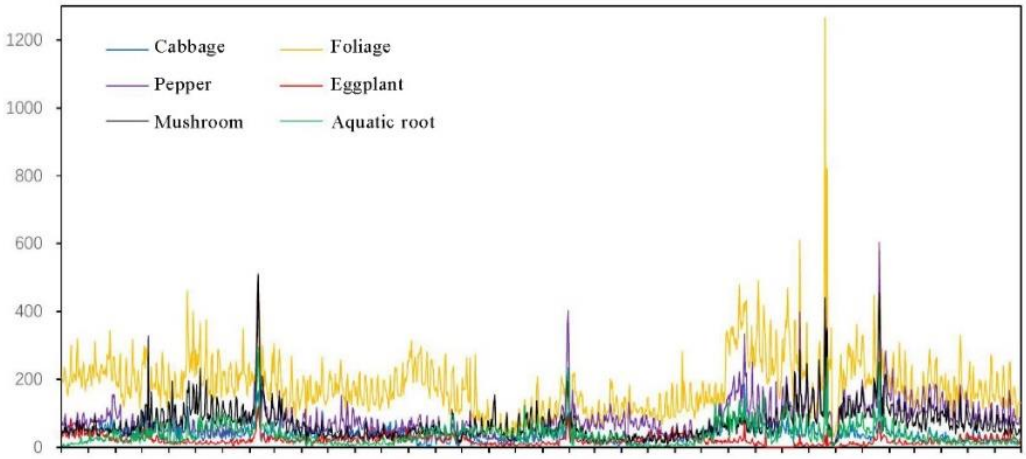

Fig 2. Daily sales volume distribution by category.

According to the time and sales volume of category sales volume distribution chart can be found, the sales volume of foliage in the whole year is significantly higher than the other five categories, almost in every point in time is higher than the other categories. Eggplant sales volume throughout the year is lower than the other five categories, 97% of the sales value is less than 50 kg, and eggplant has the most 0 value, and the same conclusions as the pie charts. Meanwhile, it is found that the sales volume of the six categories on February 11 and 12, 2021, January 30, 2022, November 2 and 4, 2022, November 30-December 3, 2022, and January 21, 2023 are all zero, with peaks appearing at intervals in the graph, especially in the foliage, mushroom, and aquatic root, whereas cabbage sales volume is relatively stable. Considering the seasonality of vegetables, the taste and growth degree of foliage, mushroom, and aquatic root may change a lot due to seasonal changes, and people will selectively purchase them.

3. Relationship between Total Category Sales Volume and Cost Plus Pricing

3.1. Calculate Cost Plus Pricing

When restocking, the merchant does not know the day's individual categories and purchase prices, so the cost-plus pricing method is used to price the vegetables to make a restocking plan for the day. The cost-plus pricing model is,

$$P = C \times (1 + \alpha) \tag{1}$$

P Is the price, C is the average cost, and α is the cost-plus ratio. The model is suitable for pricing products with relatively complete historical cost information.

At each point in time, different categories of vegetables sales unit price and wholesale unit price data, due to the large amount of data, in order to save time and cost, while smoothing price and cost fluctuations, the calculation of the average wholesale unit price and the average sales unit price of the six categories for each date, which helps to maintain a relatively stable price over a period of time, but also reduces the negative impact of price fluctuations on the business. The formula for the average wholesale unit price of each category of vegetables is as follows.

$$C = W_d = \frac{\sum_n W}{N} \tag{2}$$

W Is the wholesale price, W_d is the average wholesale price, and N is the number of items in a category on the same day.

The formula for the average unit price of vegetables sold in each category is as follows.

$$S_d = \frac{\sum_n S}{N} \tag{3}$$

S Is the selling price, S_d is the average selling price.

$$\alpha = \frac{(S_d - W_d)}{S_d} \times 100\% S_d = \frac{\sum_n S}{N} \tag{4}$$

The cost-plus ratio expresses the percentage difference in price from the average purchase price to the average selling price and reflects the pricing of a product or service that covers expenses and profits other than the original cost. The formula solves for the pricing of 6 categories of vegetables for one day as follows.

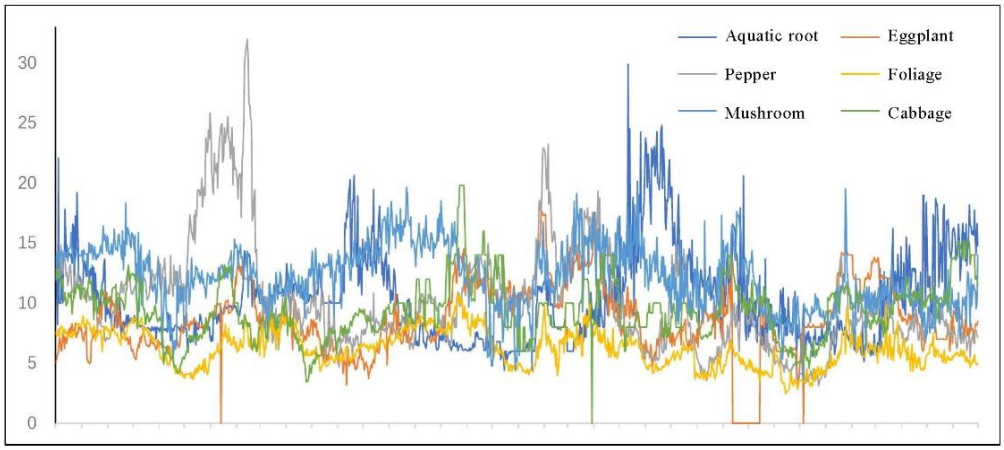

Fig 3. Cost-per-day plus pricing by category.

3.2. Relationship Between Category Sales Volume and Pricing

Using the least squares method to solve the relationship between total sales volume and pricing for each vegetable category [3], the linear relationship between total sales volume (Y) and cost-plus pricing (X) was obtained, and the linear fitting equations for the six vegetable categories were obtained by SPSS as shown in Table 3.

Table 3. Regression equations for each vegetable category.

Category	Fitted equation	R^2
Aquatic root	$y = 10.313 - 0.027x$	0.081
Mushroom	$y = 135.4 - 7.297x$	0.062
Eggplant	$y = 29.94 - 1.113x$	0.038
Pepper	$y = 204.908 - 2.374x$	0.013
Foliage	$y = 6.218 - 0.004x$	0.053
Cabbage	$y = 8.875 - 0.025x$	0.069

As shown in Table 3, the goodness-of-fit R^2 of the fitted equations by the least-squares method is lower than 0.1, which indicates that the fitting effect is poor.

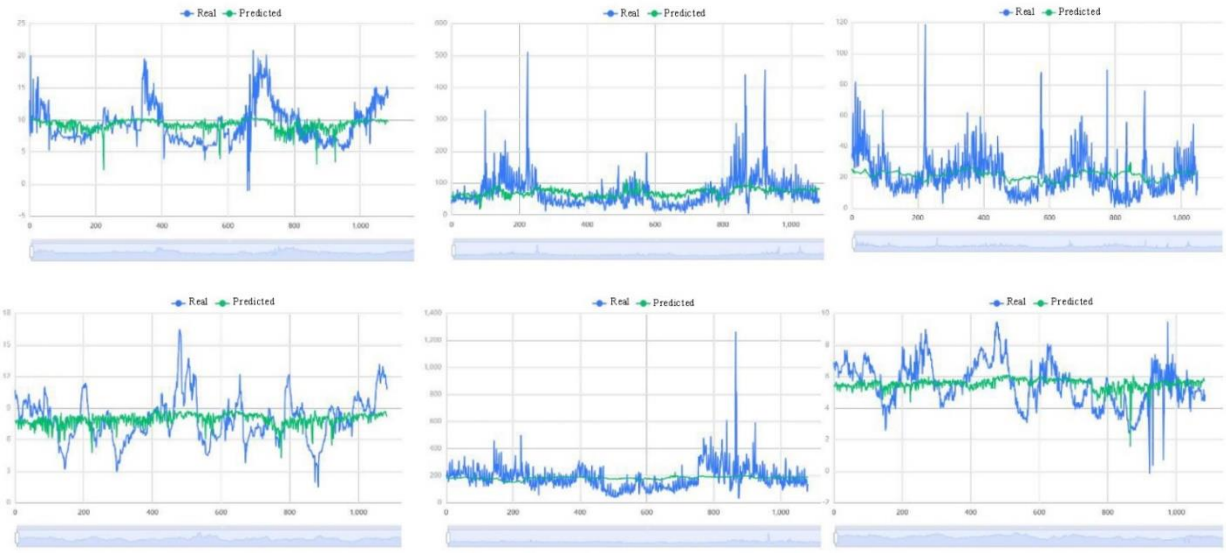

Fig 4. Effectiveness of fitted curves for 6 categories of vegetables (From left to right, top to bottom are aquatic root, mushroom, eggplant, foliage, pepper and cabbage, respectively).

Combined with the image analysis, the blue curve represents the true value, the green curve represents the predicted value, observing the image, the true value curve is fluctuating greatly, indirectly appearing peaks, while the predicted value curve is relatively smooth, and there is a big gap with the true value, which indicates that the linear regression under the method of the fitting effect is poor.

To analyze the reasons for the poor effect, the least squares method may not be suitable for handling large-scale data, when dealing with large-scale data sets [4], BP neural networks can usually accelerate the training process through parallel computing and other methods; in addition, the least squares method is usually used for linear model fitting, if the relationship between the data is nonlinear, the effect of using a linear model for the fitting may be poorer. BP neural networks are able to capture more complex nonlinear relationships, so they may be more suitable in this case.

3.3. Pricing Strategy for the Week Ahead

Analyzing the correlation between time and category sales volume in terms of day, the following Table 4 is obtained.

Table 4. Correlation coefficient between time and category sales volume.

	Cabbage	Foliage	Pepper	Eggplant	Mushroom	Aquatic root
July 1	0.127 ***	-0.08	-0.016	0.109 ***	−0.134 ***	−0.497 ***
July 2	0.233 ***	0.01	0.092	0.05	-0.006	0.312 ***
July 3	−0.125 ***	-0.012	0.03	0.03	-0.098	0.416 ***
July 4	0.085	-0.14	-0.033	-0.056	0.042	0.236 ***
July 5	−0.17 ***	−0.139 ***	0.053	-0.082	-0.004	0.235 ***
July 6	0.03	−0.136 ***	−0.133 ***	-0.078	0.029	0.428 ***
July 7	0.066	-0.079	-0.071	0.229 ***	0.034	−0.26 ***

Note: ***, **, * represent 1%, 5%, 10% level of significance respectively

According to Table 4, most of the correlation coefficients are less than 0.1 and show negative correlation, which indicates that the correlation between time and sales volume is low if the days are taken as the basic unit, so the time range is extended to take the year as the unit.

Because one day's sales data may be affected by many transient and short-term factors, such as seasonality, weather, promotional activities, etc., these factors may lead to fluctuations in sales, and it is difficult to establish a stable correlation. Expanding the time unit to a year smooths out these fluctuations and makes it easier to see how sales relate to longer-term trends, market changes, and other factors.

Use a BP neural network to analyze pricing strategies for each vegetable category for the week ahead. Neural networks automatically learn features from the data, eliminating the need to manually extract features [5]. This is useful for high-dimensional data and large-scale datasets because it reduces the effort of feature engineering.

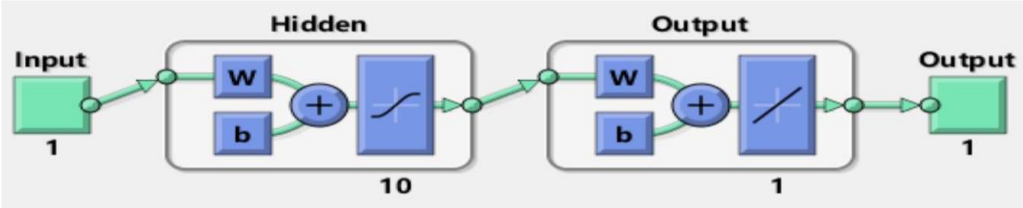

Fig 5. Flow chart of neural network algorithm.

The main parts in the Fig. 5 include.

Input Data: The input feature vector accepted by the neural network.

Input Layer: The first layer that receives the input data, each neuron corresponds to a feature.

Hidden Layer: The middle layer of the neural network, which usually contains multiple neurons and is used to learn complex relationships between features.

Output Layer: Generates the output of the neural network, which usually corresponds to the prediction or classification of the problem.

Output data: The final output of the neural network.

In a neural network, each neuron is connected to every neuron in the previous layer, passing information through weights [6]. The neural network is trained by a back-propagation algorithm, which constantly adjusts the weights to minimize the loss function, thus enabling the network to make accurate predictions or classifications.

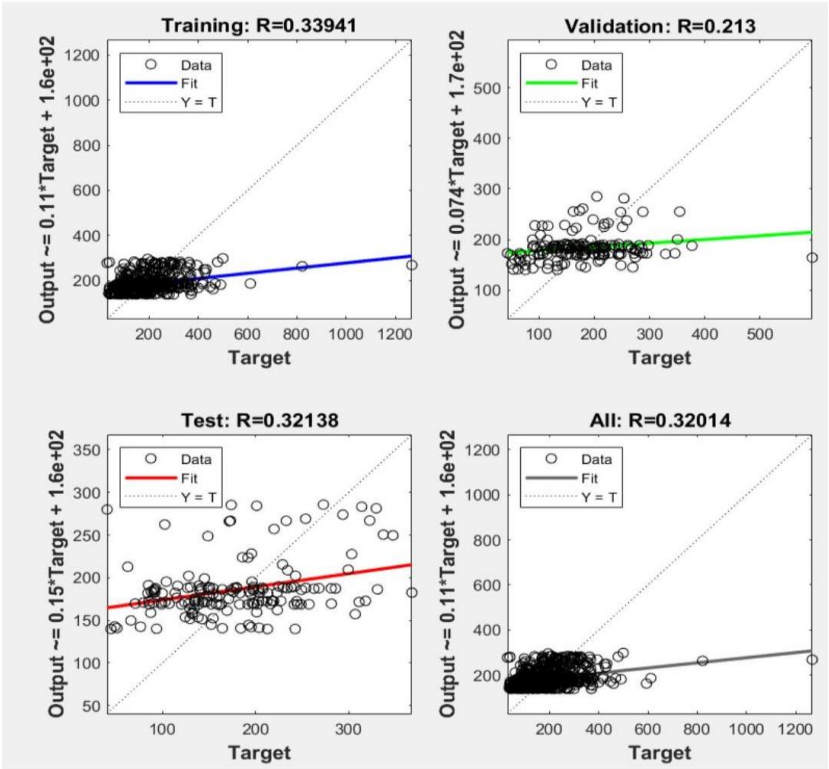

Fig 6. Scatter plot of neural network fitting.

According to the scatter plot in Fig. 9, it can be observed that there is a certain gap between the predicted value and the real value, which indicates that the neural network fitting effect is general, and the R-value is more than 0.1, so the neural network is used to predict the pricing of vegetable categories. The neural network prediction of pricing for six categories is shown in the following Table 5.

Table 5. Pricing decisions in the vegetable category.

Date	Cabbage	Foliage	Pepper	Eggplant	Mushroom	Aquatic root
July 1	10.495	4.773	6.757	9.353	11.810	13.973
July 2	11.121	4.982	7.225	7.956	11.471	15.536
July 3	11.351	3.990	6.995	8.359	11.417	14.448
July 4	9.998	4.636	7.415	8.441	11.539	14.871
July 5	10.335	4.233	7.021	7.864	11.404	15.463
July 6	8.972	4.973	7.277	9.537	11.702	15.546
July 7	11.094	4.773	6.890	8.370	11.400	15.578

4. Summary

In this study, we successfully developed a sales forecasting and pricing optimization model for vegetable categories in fresh produce superstores. We found that the sales volume of different vegetable categories showed different trends over time, and the sales proportion of foliage was the largest and more stable. In terms of pricing, through the application of BP neural network model, we realized accurate prediction of pricing for vegetable categories, and the BP neural network model has better prediction effect compared with linear fitting method. Our study also found an obvious correlation between time and daily sales volume, which provides an important reference for superstores to develop reasonable pricing strategies. Taking market demand and supply into account, our model can effectively guide the replenishment decisions of superstores, maximize their operational efficiency and competitiveness, and achieve sustainable development.

References

[1] Martinez O, Rodriguez N, Mercurio A, et al. Supermarket retailers' perspectives on healthy food retail strategies: In-depth interviews [J]. BMC Public Health, 2018, 18: 1-16.

[2] Li H, Liu J, Qiu J, et al. ARIMA-Driven Vegetable Pricing and Restocking Strategy for Dual Optimization of Freshness and Profitability in Supermarket Perishables [J]. Sustainability, 2024, 16(10): 4071.

[3] Menke W. Review of the generalized least squares method [J]. Surveys in Geophysics, 2015, 36: 1-25.

[4] De Brabanter K. Least squares support vector regression with applications to large-scale data: a statistical approach [J]. Faculty of Engineering, KU Leuven, Katholieke Universiteit Leuven, 2011.

[5] Jia F, Lei Y, Lin J, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data [J]. Mechanical systems and signal processing, 2016, 72: 303-315.

[6] Cilimkovic M. Neural networks and back propagation algorithm [J]. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, 2015, 15(1).