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

DOI: 10.23977/infse.2024.050301
ISSN 2523-6407 Vol. 5 Num. 3

Optimization Strategy for Vegetable Replenishment and Pricing in Supermarkets Based on XGBoost Algorithm and GA

Wenbo Ye#, Shibo Zhu#, Boyuan Zheng#

Chang'an Dublin International College of Transportation at Chang'an University, Chang'an University, Xi 'an, 710064, China
#These authors contributed equally.

Keywords: Vegetable Pricing and Replenishment Decisions, XGBoost Algorithm, Genetic Algorithm, Goal Programming

Abstract: Vegetables as an important part of the daily diet of residents in the fresh food superstore occupies an important position, its short shelf life and the importance of the appearance and other characteristics of the vegetable superstore need to first solve the daily vegetable replenishment and pricing problems. Based on the historical sales data of a superstore, this paper analyzes the distribution pattern model and correlation of the sales volume of vegetable categories at different times, and finds that the average monthly sales volume of each category is, in descending order, as follows: anthophyllum, capsicum, edible mushrooms, cauliflower, aquatic rhizomes, solanacea, and the sales volume of each category of vegetables has a cyclical and seasonal distribution pattern; the distribution pattern of individual items is affected by the seasonal changes. The XGBoost algorithm is used to establish a model with the maximum revenue of the supermarket in the coming week as the objective function, and the pricing strategy of different categories in each day as the decision variable to establish a planning model, and the replenishment quantity and pricing strategy of six categories are obtained by genetic algorithm.

1. Introduction

In today's society, with the public's increasing attention to healthy diet and fresh food ingredients, fresh suppliers, as the main retail channel of vegetable products, occupy an important position in the food supply chain. These supermarkets are characterized by providing fresh, diverse vegetables; however, these goods have the challenges of a short shelf life, high seasonal demand, and supply diversity. Therefore, it is of great significance to deeply explore the status and management of vegetable products in fresh supermarket to improve operational efficiency, meet the changing consumer needs and promote the development of fresh retail industry [1]. Chen Jun et al. [2] used two solution ideas and designed a heuristic algorithm to solve the optimal preservation period, sales cycle and pricing. Their research provides effective solutions for the management of fresh goods. Zhao Ling et al. [3] studied joint replenishment and pricing strategies considering customer returns and fixed costs. Their research provides a strategic basis for replenishment and pricing when facing customer returns and fixed costs. Zhang Jinlong et al. [4] studied the joint decision model of new product pricing and replenishment. Their research provides decision support for fresh suppliers in new product pricing and replenishment. This paper presents a pricing and replenishment model using SVD-GA. The model can provide pricing and replenishment prediction methods for fresh supermarkets, which helps to improve the operating efficiency of supermarkets and meet consumer demand.

2. Data Analysis of Vegetable Commodity Sales

2.1 Distribution pattern of sales volume of vegetables by category

The distribution patterns of vegetable sales volume by category are complex and influenced by a variety of factors. Understanding these patterns is very important for fresh food superstores and agricultural suppliers to develop supply chain strategies and sales strategies. In this paper, we analyze the distribution law of sales volume of vegetable categories from the descriptive statistics distribution law and and time series distribution law.

2.2 Distribution pattern of descriptive statistics of sales volume of vegetables by category

The method of calculating the monthly average of the sales volume of each category of vegetables over a three-year period is to add up the sales volume of each month and divide it by the month in which it was sold, and then add up the sales volume of each category of vegetables for each day of the month in order to obtain the monthly sales volume of that category of vegetables for that month. The formula for calculating the monthly average sales volume of each category of vegetables over three years is as follows:

$$\mu = \frac{X_1 + X_2 + \cdots + X_{36}}{36} \tag{1}$$

μ denotes the monthly average of the sales volume of each category of vegetables, $X_i \ (i=1,2,\cdots,36)$ denotes the sales volume of each category of vegetables in month i ,First month of the second year $i=13$. In this order, the total number of months is 36.

Through equation (1), the calculation was repeated six times to obtain the average monthly sales volume of each category of vegetables over a three-year period. The results are shown in Figure 1.

As can be seen from the figure, the average monthly sales volume of anthophyllum vegetables was the highest and that of solanacea vegetables was the lowest. The ranking of the six categories of vegetables in terms of sales volume may be influenced by the following factors: firstly, the anthophyllum vegetables may be ranked first due to their diverse varieties and popular healthful attributes, which appeal to a wide group of consumers. Second, capsicum vegetables are commonly used to add flavor and therefore have a high demand in sales. Thirdly, edible mushroom vegetables ranked third due to their unique taste and nutritional value. Cauliflower vegetables may have high sales volume in certain seasons or uses. Aquatic rhizomes vegetables may rank fifth due to seasonal availability or specific culinary demand. Finally, solanacea vegetables may rank lower in sales volume, influenced by local market preferences or seasonal availability.

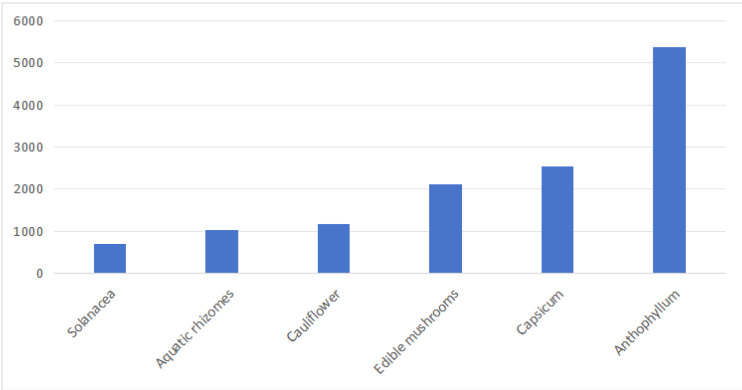

Figure 1: Average monthly sales of vegetables by category over a three-year period

Calculate the variance of the monthly mean of the sales volume of each category of vegetables; to calculate the data you need the sales volume of each category of vegetables for each month. This means $X_i,\ (i=1,2,\cdots,36)$, the formula for calculating the variance is as follows:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n-1} \tag{2}$$

S^2 is the variance of the monthly mean of the sales volume of each category of vegetables, n is the number of months (36 in total).

By using equation (2), the variance of the monthly mean of the sales volume of each category of vegetables is obtained, and the results are shown in Figure 2:

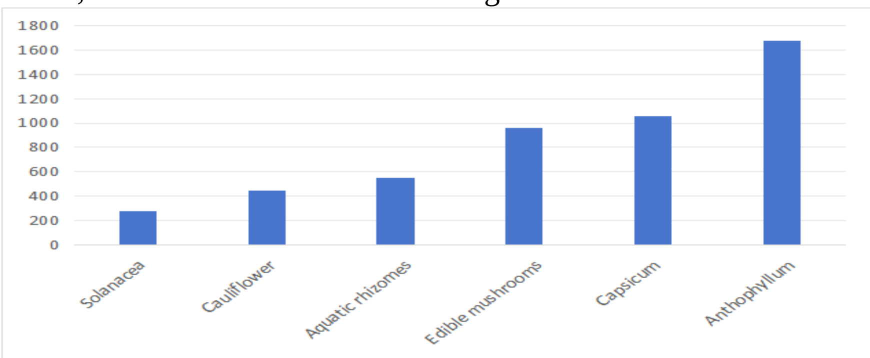

Figure 2: The variance of the monthly mean of the sales volume of each category of vegetables

From the figure, it can be seen that the variance of the monthly average sales volume of anthophyllum vegetables is the largest and that of solanacea vegetables is the smallest, which implies that the sales volume of anthophyllum vegetables fluctuates a lot in different time periods, whereas the sales volume of eggplant vegetables is relatively stable. The larger variance in sales volume for anthophyllum vegetables may be due to the fact that these vegetables are more affected by fluctuations in demand during different seasons or time periods. For example, the demand for anthophyllum vegetables may change considerably during different seasons or festivals, resulting in higher fluctuations in sales volume. In contrast, solanacea vegetables have less variance in sales volume, probably because the demand for these vegetables is relatively stable and not strongly influenced by seasonal or festive factors. Consumers may be more inclined to purchase solanacea vegetables at different times throughout the year, making their sales volume less volatile.

In summary, the descriptive statistical distribution pattern of sales volume of each vegetable category is shown:

The average sales volume of each vegetable category by month, in descending order, was: anthophyllum, chili peppers, edible mushrooms, cauliflower, aquatic rhizomes vegetables, and solanacea; the stability of each vegetable category by sales volume, in descending order, was: solanacea, cauliflower, aquatic rhizomes vegetables, edible mushrooms, chili peppers, and anthophyllum.

2.3 Temporal Distribution Patterns of Vegetable Sales Volume by Category

Statistics of monthly sales of various categories of vegetables, made of line graphs to reflect the trend of sales of various categories of vegetables, in order to analyze the distribution of sales of various categories of vegetables in the chronological pattern, as shown in Figure 3

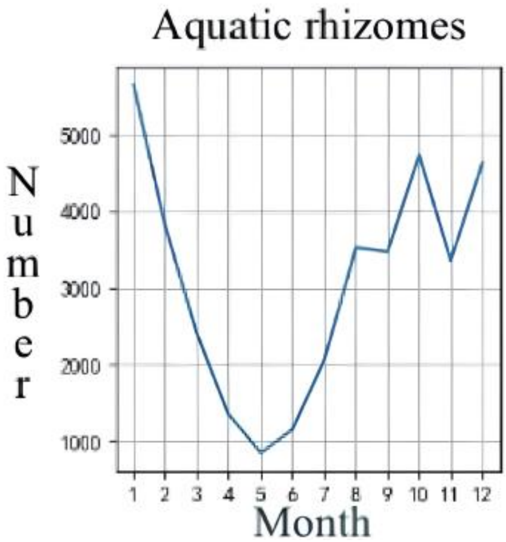

Figure 3: Trends in sales volume of vegetables by category over a one-year period (An example of aquatic rhizomes)

As can be seen from the graph, the sales volume of the two categories of vegetables, aquatic rhizomes and edible mushrooms, showed an overall upward trend from June to December, and a downward trend from January to May, and the sales volume reached its lowest in April and May, probably because these two categories of vegetables usually require a warmer climate for growth; on the contrary, the growth of vegetables slows down due to the cold weather from winter to early spring, with a lower supply, and thus the sales volume Decline.

Sales of leafy vegetables, cauliflowers and chili peppers show an upward trend from June to August and a downward trend from September to December, with a high level of sales in January followed by a downward trend from February to May. The reason could be that vegetable categories grow differently in different seasons. Usually, they are harvested in the warm summer to early fall season, when supply is higher and prices are relatively lower. As a result, June through August is the period when their sales volume rises. September through December is the end of the growing season for many vegetable categories, when supply may decrease, which can lead to higher prices. As a result, sales may drop at this time of year. January is usually accompanied by New Year's and other holidays where people buy ingredients to prepare festive dishes. This may lead to an increase in sales volume in January. The months of February through May are winter and spring months with cooler temperatures, which may lead to a decrease in production and supply chain impacts for certain vegetable categories, and therefore a decrease in sales volume.

The sales volume of solanacea is mainly concentrated from March to September, with an upward trend in March to July and a downward trend from July to December, but there is a higher sales volume in January and February. It may be due to the fact that eggplant vegetables usually grow better in the warmer seasons, and thus the supply of eggplant vegetables is usually larger and of better quality during the warmer seasons from March to September. In January and February, there may be some special factors that lead to higher sales. For example, some regions celebrate the Chinese New Year or other important festivals during this time, which may prompt people to buy more ingredients to prepare festive dishes, including eggplant vegetables.

In summary, the temporal distribution pattern of sales volume of vegetable categories is cyclical and seasonal, reflecting the seasonal demand of consumers and seasonal changes in vegetable supply.

3. Establishment of model

3.1 Establishment of the relationship model between total sales and cost-plus pricing

Cost-plus pricing is a common pricing strategy that is commonly used in both retail and wholesale businesses, especially for rapid consumables such as vegetable products. The formula for cost-plus pricing is as follows:

$$P = e \times (1+\beta) \tag{3}$$

This is the selling price P, and the wholesale price e, indicating the bonus rate or the expected profit margin β. The relationship between total sales volume and cost-plus pricing is the relationship between total sales volume and sales price.

In order to study the relationship between them, we made a preliminary scatter plot, and made a linear fitting by using the matlab fitting toolbox, as Figure 4:

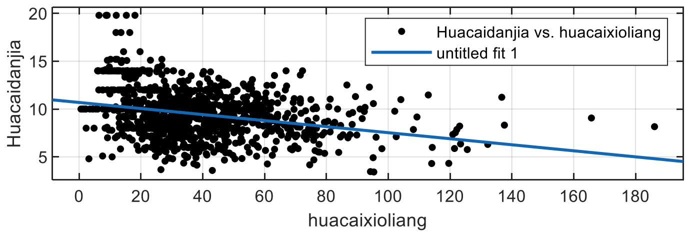

Figure 4: Linear fit diagram of total sales volume and sales price

The scatter plot and the fitting effect of the six categories are similar. This paper only shows the graphics of cauliflower here, and the graphics of the remaining five categories are included in the appendix. We put the specific fitting function and goodness of fit in the following table 1:

Table 1: Fit function and goodness of fit table for each vegetable category

Class	Fits the function	R^2
Flower vegetables	$y = -0.03158x + 10.69$	0.0831
Flowers and leaves	$y = -0.003739x + 6.906$	0.0418
Capsicum	$y = -0.009534x + 11.05$	0.0131
Solanum	$y = -0.03468x + 9.636$	0.033
edible mushrooms	$y = -0.01944x + 13.5$	0.1229
Aquatic rhizomes	$y = -0.04781x + 12$	0.1213

Observation of the chart shows that the fitted linear relationship is in line with the actual situation, that is, the sales volume is negatively correlated with the sales price, which is consistent with the price elasticity demand theory. However, the goodness of fit is too small, and the maximum is only 0.1229, and the distribution of the scatter plot has no obvious rule, so it is not suitable to predict with the fit. This paper uses XGBoost algorithm to analyze the relationship between total sales and sales price and predict the sales volume [5].

(1) Establish the square-loss function:

$$l\left(y,\hat{y}^{(t)}\right) = \left(y-\hat{y}^{(t)}\right)^2 \tag{4}$$

$\hat{y}^{(t)}$ is the prediction result of the model after training times, y is the true value of the sales volume

(2) Establish the model complexity term:

$$\Omega = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} {w_j}^2 \tag{5}$$

γT is the L1 regular term and $\frac{1}{2}\lambda\sum_{j=1}^{T} {w_j}^2$ is the L2 regular term.

(3) Minimize the function using the gradient lifting method and quickly find the parameters:

$$Obj^{(t)} = l\left(y,\hat{y}^{(t)}\right) + \Omega \tag{6}$$

$Obj^{(t)}$ is the scoring function of the model, $Obj^{(t)}$ the smaller the value, the better the effect of the model.

Every time a tree is generated, it will change in the parameters in the model, and then the continuous optimization calculation will occur. When a new tree is built, the parameter values will change. Through the continuous update of the parameter values, the optimal solution will be found.

The sales volume forecast result through the XGBoost algorithm is shown in the Table 2:

Table 2: Sales volume forecast table of each category

Class	July 1	July 2	July 3	July 4	July 5	July 6	July 7
Flower vegetables	66.707	68.231	65.963	66.736	65.976	65.796	65.425
Flowers and leaves	9.082	12.839	8.281	9.238	10.285	10.741	11.542
Capsicum	142.85	148.12	144.94	139.58	138.27	140.21	142.32
Solanum	13.91	14.32	15.95	12.59	13.24	14.31	13.53
edible mushrooms	44.48	43.24	44.31	46.25	41.94	40.77	45.32
Aquatic rhizomes	34.61	35.21	31.98	34.29	33.87	37.82	32.38

3.2 Establishment of target planning model

The revenue of the supermarket needs to consider multiple characteristics comprehensively. We considered the selling price, wholesale price, sales volume and the commodity loss rate caused by various reasons. The wholesale prices and loss rates of all vegetable categories are averaged by using historical data. The calculation formula is as follows:

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} \tag{7}$$

$\bar{e}$ represents the average wholesale price of vegetables in this category, and n represents the type of vegetables in this category.

$$\bar{s} = \frac{\sum_{i=1}^{n} s_i}{n} \tag{8}$$

$\bar{s}$ represents the average loss rate of vegetables in this category, n indicates the vegetable juice of this category.

$$\max \sum_{t=1}^{7}\left(\sum_{j=1}^{6} X_{i,j} \times \left(1-\bar{s}\right) \times \left(x_{i,j}-\bar{e}\right)\right)$$

$$s.t. \begin{cases} x_{t,j} > \bar{e} \\ x_{t,j} > 0 \\ \bar{e} > 0 \end{cases} \tag{9}$$

The optimal value was solved using a genetic algorithm and set to a population size of 50, crossover probability of 0.5, variation probability of 0.2, and number of iterations of 40. The daily sales quantity and daily sales unit price corresponding to the daily revenue maximization are regarded as the daily replenishment quantity and the daily pricing strategy of vegetable categories. The maximum profit of the supermarket in the next week will be 37,660.764 yuan. The daily replenishment volume, pricing and earnings of the supermarket for the next week are shown in the Table 3:

Table 3: Sales of vegetables in the coming week

In the coming week	**Sales unit price**	**Sales quantity**	**Business super income**	**Class**
Day 1	15.766	92.067	677.237	edible mushrooms
Day 2	16.796	107.967	895.679	edible mushrooms
Day 3	11.920	139.410	536.287	edible mushrooms
Day 4	16.574	116.168	940.146	edible mushrooms
Day 5	16.574	116.168	940.146	edible mushrooms
Day 6	12.655	115.685	522.535	edible mushrooms
Day 7	11.925	139.410	536.870	edible mushrooms
The largest profit for the supermarket			37660.764	

4. Conclusion

Overall, this article analyzes the distribution pattern of various categories of vegetables and the optimal pricing and replenishment strategies. Based on the distribution pattern and pricing replenishment strategies, consumers' consumption preferences at different time periods can be determined, which can help supermarkets develop more accurate pricing strategies and more effective promotion and marketing strategies, and help improv sales volume and profitability.

References

[1] Gu Sihong. Study on dynamic pricing of fresh products in H retailers considering freshness changes [D]. And Donghua University, 2023.

[2] Chen Jun, Kangsha. Joint decision on pricing and inventory replenishment of agricultural products for dual-channel sales [J]. Industrial Engineering, 2023, 26 (03): 39-46.

[3] Zhao Ling, Liu Zhixue. Joint replenishment and pricing strategy considering customer returns and fixed costs [J]. Operations Research and Management, 2022, 31 (06): 105-110 + 124.

[4] Zhang Jinlong, Wu Xiang, Xu Haoxuan. Joint decision model of perishable new product pricing and replenishment [J]. Journal of Systems Engineering, 2018, 33(01).

[5] Wang Shijie, Wang Xingfen, Yue Ting. Multiple influencing factors selection and prediction of cotton prices based on XGBoost and TCN Attention [J/OL]. Computer System Application: 1-12 [2023-09-10].