Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 1
AIEA 2023

Automatic Pricing and Replenishment Strategy for Vegetable Products Based on Time-Series Analysis and BP Neural Fitting

Chongwei Ren *, Yue Yu, Jiadi Suo
Northwest University, Xi'an,China
* Corresponding Author Email: 1442791250@qq.com
Chongwei Ren, Yue Yu, Jiadi Suo are Co-first author

Abstract. To reduce the waste of vegetables and guarantee revenues for supermarkets, this paper explores the construction of a comprehensive optimization model to help the supermarket select the specific suitable replenishment scheme and pricing strategy. This sets the bedrock for it to both gain the maximum revenue and timely supply various categories of vegetables that the market demands. The expectation for further relevant research to better complete the model is also mentioned. First, the attached data were preprocessed, and no missing values and outliers were detected. Consequently, the data of multiple forms was merged, and data related to categories of vegetables was counted. In response to Question 1, descriptive statistics were conducted on different individual items and categories of vegetables. Sales volumes of categories of vegetables were visualized through sales volume statistical graphs, profit line graphs, and quarterly sales change graphs. Following, the sales volumes were verified to have satisfied the conditions for calculating the Pearson Correlation Coefficient, and heat maps were drawn for correlation analysis. For individual items, hierarchical clustering was carried out with indicators, such as sales volume, unit price, number of purchases, and wastage rate. The basis of categorization of each category of vegetable was also explored. For Question 2, average pricing was used to replace cost-plus pricing first. Then, BP Neural Net Fitting was leveraged to analyze the relation between total sales volume and average pricing of different vegetable categories. The average wholesale price of the next seven days of each vegetable category was predicted with ARIMA Model, in order to gain the profit of different categories. Finally, a nonlinear objective planning model to achieve maximum benefit for the supermarket was constructed. Corresponding constraints were given to propose a reasonable total replenishment and pricing strategy for each vegetable category. In solving Question 3, constraints were added based on the nonlinear objective planning model in Question 2, and the prediction model was optimized. In the case of meeting market demand, the replenishment volume and pricing strategy for individual items on July 1, 2023, were proposed based on a combination of factors, as a way to maximize the benefits for the supermarket.

Keywords: Pearson Correlation Coefficient; BP Neural Net Fitting; ARIMA Model; Nonlinear Objective Planning Model.

1. Research Questions

Question 1: A certain correlation may exist between different categories or different individual vegetable items. Please analyze the distribution pattern and interrelationship of the sales volume of each category and individual items.

Question 2: Considering that the supermarket makes replenishment by category, please analyze the relationship between total sales and cost-plus pricing for each vegetable category, and propose the total daily replenishment and pricing strategy for each category for the coming week (July 1 to 7, 2023) to maximize the superstore's revenue.

Question 3: Due to the limited sales space for vegetables, the supermarket wanted to further develop a replenishment plan for individual items with a total of 27 to 33 available items and a requirement that the order quantity of each single product meets the minimum display quantity of 2.5kg. Referring to the available categories from June 24 to 30, 2023, please propose the replenishment volume and pricing strategy for individual items on July 1 to maximize the revenue while trying to meet the market demand for each category of vegetable.

2. Question Restatement

2.1. Research Background

Due to the short shelf time and deteriorating quality over time of vegetables, daily replenishment plans should be conceived based on the former sales situation and demand. The supermarket always makes relevant plans without specific items and purchase costs to adapt to purchase transaction time. In terms of pricing, cost-plus pricing is often used and discounts are set for damaged and poor-quality goods. This means that accurate demand analysis is a significant determinant for replenishment and pricing strategy. In addition, considering the possible correlations between sales and time for vegetables, this paper has analyzed the sale volumes of each item and explored the correlations and distribution patterns.

Combination sales stand out because of the rich variety of vegetables from April to October but limited sale space. Storage and wastage rates should be considered as well.

3. Model Hypotheses

Model hypotheses were put forward for the construction.

1. The market where the supermarket is was assumed to be a perfectly competitive one so that pricing and replenishment volumes can be set as wish. These would not directly influence the market price.

2. The supermarket was assumed to have had a relatively stable sales volume of vegetables over the past three years without significant fluctuations.

3. There was enough storage for different categories of vegetables.

4. An even distribution of the wastage rate of vegetables was assumed over a short period which would not be significantly disturbed by external factors.

4. Model Variables

Table 1. Relevant model variables and their corresponding variable meanings

Model Variables	Variable Meanings
$\text{Shu\_category}_n, n = 1,2,\ldots,6$	6 Categories of Vegetables
$\text{Shu\_item}_m, m = 1,2,\ldots,251$	251 Individual Vegetable Items
$\text{Sales}_{m,l}, l = 1,2,\ldots,1085$	Selling Price of Vegetable Item m on Day l
$\text{Pifasales}_{m,l}, l = 1,2,\ldots,1085$	Wholesale Price of Vegetable item m on Day l
$\text{Xiaosales}_{m,l}, l = 1,2,\ldots,1085$	Sales Volume of Vegetable Item m on Day l
$\text{Tuihuo}_{m,l}$	Returns of Vegetable Item m on day
$\text{Lose}_n, n = 1,2,\ldots,6$	Wastage Rate of the n^{th} Vegetable Item
Profit_l	Profit of the Supermarket on Day l
$\text{Supply}_{m,l}$	Replenishment Volume of Item m^{th} on Day l
$\text{Supply}_{n,l}$	Replenishment Volume of Item n^{th} on Day l

The following table demonstrates all the relevant model variables used and their corresponding variable meanings to allow a more concise presentation of the relevant mathematical models constructed.

5. Modeling for and Solving Problem 1

5.1. Descriptive Statistics of Vegetables by Category

It was found that the sales volume of leafy category is much higher than all other categories, while the sales volume of eggplants is the lowest. The rest in descending order of sales volume are peppers and edible fungus. Aquatic ones and flower ones have similar sales, with a large gap between them and the top three categories in terms of sales volume.

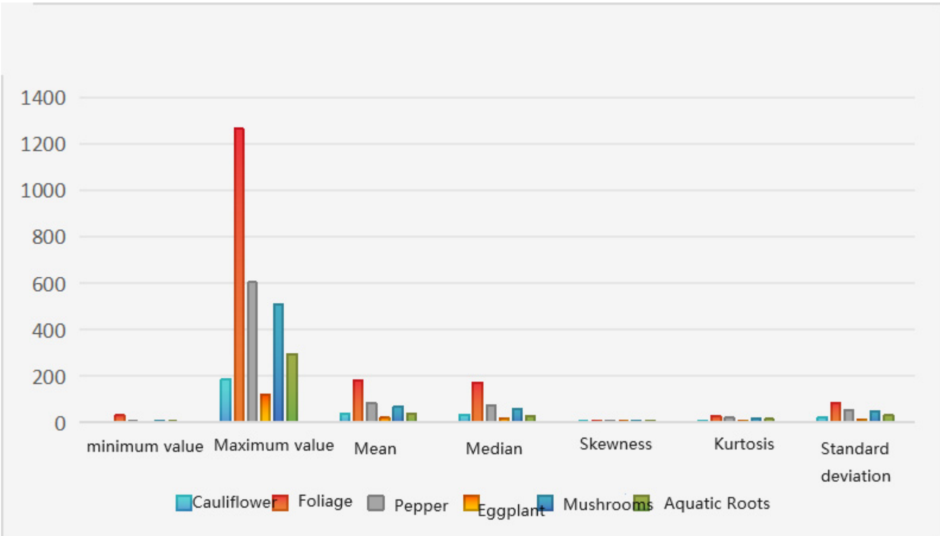

Figure 1. Descriptive Statistical Analysis by Category

The following equation of the six vegetable categories was used to calculate the average sales price. This aimed to obtain more detailed distribution relationships and regularities among the six categories and among the sales volumes when setting a quarter cycle for time change, so the overall change and trend of the total sales price and the average sales price of the six categories could be analyzed.

$$Average\_sale_n = \frac{\sum_l \sum_m Sales_{m,l} \bullet Xiaosales_{m,l}}{\sum_l \sum_m Xiaosales_{m,l}} \tag{1}$$

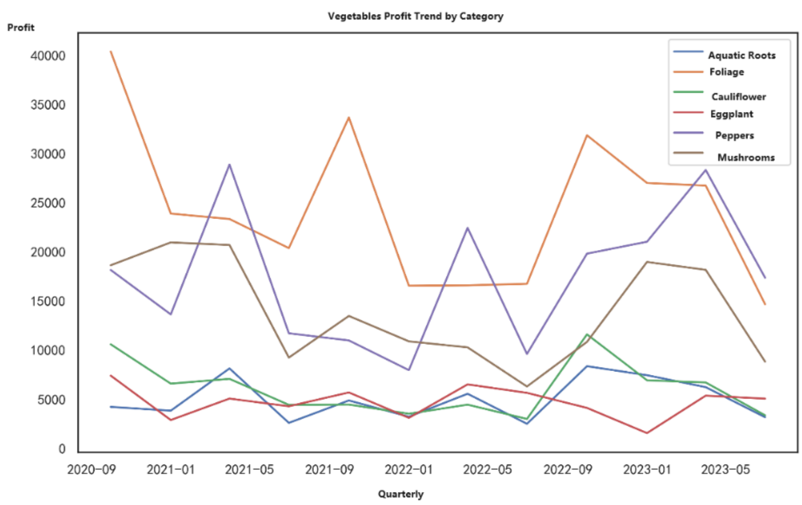

Figure 2. Overall Change Trends of Profits of Each Category over Quarters

The results of Eq.1 are demonstrated in Fig.2, showing the profit trends that the six categories almost synchronously changed with the total sales price. The largest profit of leafy vegetables was considered to be due to its large sales volume. The eggplants had the lowest profit. Flower ones and aquatic ones had similar profits, ranking low. Peppers and edible fungus ranked after leafy vegetables.

5.2. Pearson Correlation Coefficient and Corresponding Hypothesis Testing

Pearson Correlation Coefficient is a commonly used statistical indicator ranging from -1 to 1. It means that it is completely positively correlated when the correlation coefficient is 1. When the correlation coefficient is -1, it means that the two variables are completely negatively correlated. Furthermore, it means that there is no linear relationship between the two variables when the correlation coefficient is 0. The closer the absolute value of the correlation coefficient is to 1, the stronger the linear relationship between the two variables is [1].

In addition, it is important to note that the Pearson Correlation Coefficient can only measure linear relationships and is insensitive and misleading to non-linear relationships. The correlation coefficient is also affected by outliers and sample size.

The equation for Pearson Correlation Coefficient is as follows.

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_I - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_I - \overline{Y})^2}} \tag{2}$$

X_i and Y_i are the respective observed values of X and Y. $\overline{X}$ and $\overline{Y}$ are the respective averages of X and Y. n refers to the number of observed values.

Sample fulfilling a number of assumptions is required for Pearson Correlation Coefficient hypothesis test, including linear relationship, normal distribution, and equal variance. Otherwise, it may lead to inaccurate hypothesis test results.

5.3. Descriptive Statistics and Hierarchical Cluster Analysis of Individual Items

For this problem, the distribution of sales volume and correlations of individual vegetable items were analyzed. This paper studied the distribution of sales volume of individual items of each vegetable category to get a more comprehensive picture of the dynamics of the vegetable market.

The vegetables were divided into three categories by cluster analysis. The first category was cabbage, the second was broccoli, lotus root, and Wuhu green pepper. The third was other vegetables.

Cabbage was the first category due to its low unit price and a variety of cooking methods, making it popular in the market with a wide range of consumers. Cabbage is sold in large quantities and reaches a large audience.

Broccoli, lotus root, and Wuhu green peppers of the second group are common in the market and are often used in cooking methods, such as stir-frying. Their sales volume is relatively stable with little variation since most households purchase.

The third category of vegetables is others that were not categorized. The frequency of use of them is highly related to the average unit price. They were set in the third category due to low demand and relatively low sales [2].

For Question 1, this paper statistically analyzed the sales volume, sales unit price, average unit price, and number of occurrences of each vegetable category and individual item to explore the correlations. Hierarchical clustering was also used to reclassify them and probe the internal relationship of different vegetables. Based on this, Question 2 fitted the relationship between total sales and cost-plus pricing of different vegetable categories with the BP neural network. A time-series model was then implemented to predict the average wholesale price for the next 7 days. In addition, a nonlinear objective planning model was constructed to propose a reasonable replenishment and pricing strategy with the goal of maximum profits.

6. Modeling for and Solving Problem 2

6.1. Prediction of Neural Network Model for the Relationship between Sales Pricing and Total Sales Volume

In order to analyze the vegetable category, the average daily sales pricing was calculated first based on actual sales data, rather than simply using a cost-plus pricing method. The exact equation is as follows.

$$Average\_sale_n = \frac{\sum_l \sum_m Sales_{m,l} * Xiaosales_{m,l}}{\sum_l \sum_m Xiaosales_{m,l}} \tag{3}$$

In this way, it was able to determine the average sales pricing of each category based on actual sales, which is more in line with market demand and competitive environment. This is because it directly took into account the actual sales price and sales volume, reflecting the purchasing behavior of consumers.

The equation for calculating the total daily sales for each vegetable category is as follows.

$$Xiaosales_{m,l} = \sum_{m \in n} Xiaosales_{m,l}$$

6.2. Prediction of Average Wholesale Prices Using ARIMA Model

The average wholesale price of the vegetable categories was also predicted to obtain the profit figures, with the equation below.

$$Average\_Pifasales_{m,l} = \frac{\sum_{m \in n} (1 + Lose_n) * Xiaosales_{m,l} * Pifasales_{m,l}}{\sum_{m \in n} (1 + Lose_n) * Xiaosales_{m,l}} \tag{4}$$

Based on Eq.4, changes in monthly average wholesale price were plotted.

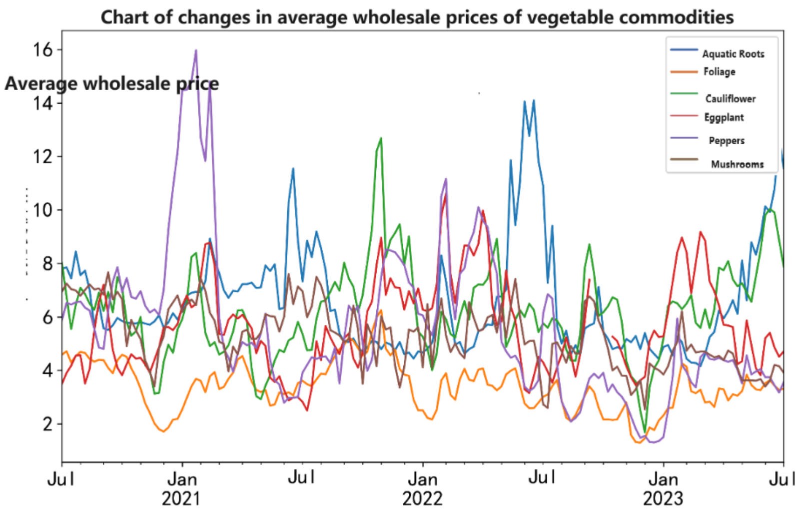

Figure 3. Change in Average Wholesale Prices of Vegetables

Next, an ARIMA model was built to come up with a prediction of the average wholesale price of vegetables for the next 7 days.

It can be used to describe and predict trends and seasonal variations in time series data, applicable for data that have a certain regularity and change over time. The core idea of the ARIMA model is to capture and model trends in time series data by performing autoregressive, difference, and moving averages [4].

The ARIMA model in this research was built with autoregressive order (p) of 1, difference order (d) of 2, and moving average order (q) of 2. This selected ARIMA model was utilized to predict the average wholesale price of vegetables for the next 7 days. It can be used to capture trends and cyclical variations in the data by fitting it to the historical one to be applied to future prediction.

6.3. Determining Capture Pricing Strategies with Nonlinear Objective Programming

After obtaining the average wholesale price and pricing-sales volume model, a nonlinear planning model was constructed to solve for the optimal total replenishment and pricing strategy. The constructed nonlinear planning model is as follows Objective function:

$$max(Profit_{n,l} = Sales_{n,l} \bullet Xiaosales_{n,l} - Supply_{n,l} \bullet Pifasales_{n,l}),$$

$$\text{s.t.} \quad \begin{aligned} & Supply_{n,l} > XiaosaleSales_{n,l} + Lose_n \bullet Supply_{n,l}, \\ & Sales_{n,l}, Supply_{n,l} > 0, \\ & Xiaosales_{n,l} = func(Sales_{n,l}), \\ & Sales_{n,l} > Pifasales_{n,l}, \end{aligned}$$

The objective function refers to the maximum benefits of the supermarket on the l^{th} day in the vegetable category n, meaning the total sales price - total wholesale price. $Supply_{n,l}$ is the replenishment volume of n on the l^{th} day, and $Sale_{n,l}$ is the sales volume of n on the l^{th} day. Constraint Function 1 represents that the replenishment volume is greater than the sales volume plus loss. Constraint Function 2 refers to that the sales price is greater than the wholesale price. Constraint Function 3 denotes that the replenishment volume and sales pricing are greater than 0. Constraint Function 4 is the pricing-sales volume function to determine sales volume from sales prices, through which data were input to obtain the daily replenishment volume and the pricing strategy of vegetables for the next 7 days attached in the appendix.

7. Modeling for and Solving Problem 3

7.1. Total Daily Replenishment and Pricing Strategy for an Individual Item for the Coming Day

The number of items that could be restocked on July 1 had been assumed to be roughly the same as the number of categories that could be sold in the previous week. Figure 12 presents the sales of all vegetable items from June 24-30, 2023, in which there were a total of 49 purchasable items.

Similar to Question 2, the research first used mathematical models to predict the wholesale price of each vegetable category in the coming day. The former past average method was used for the prediction of the wholesale price of each item $Pifasale_{m,l}$. At the same time, the following definition was carried out which was the nonlinear programming model.

Objective function : $\max(Profit_l = select_l \bullet Sales_l \bullet Xiaosales_l - select_l \bullet Supply_l \bullet Pifasales_l),$

$$select_l \bullet Supply_{m,l} > select_l \bullet Xiaosales_{m,l} + select_l \bullet Lose_m \bullet Supply_{m,l},$$

$$s.t. \quad Sales_{m,l} > Pifasales_{m,l},$$

$$select_l * Supply_{m,l} > select_l,$$

$$27 \le sum(select_l) \le 33,$$

$$select_l * Sales_{m,l} = Sales_{m,l},$$

$$select_l * Sales_{m,l} = Sales_{m,l},$$

$$select_l * Xiaosales_{m,l} = funv(Sales_{m,l}),$$

$$num(n) = 6,$$

$$select_l = [random.rand\,\mathrm{int}(0,1)\, for\ \_\ inrange(49)]$$

Table 2. Relationship between the pricing of each vegetable and the volume of sales

Item Name	Purchase Quantity	Sals Price
Mixed Pepper(2)	28.45577654	18.89444457
Bok Choy	3.45032879	8.00123
Yunnan Romaine (Portion)	20.67015428	4.1235654
Yunnan Lettuce (Portion)	14.1742567	4.541200312
Lotus Root (1)	0	0
Agaricus Bisporus (Box)	0	0
Round Eggplant (2)	28.57311023	6.14444444
Crowndaisy Chrysanthemum	0	0
Milk Cabbage	8.549208123	4.813
Ginger, Garlic, Millet and Pepper Combo (Small)	20.97023215	4.8012445
Baby Chinese Cabbage	5.614456798	6.56523489
Wrinkled Pepper (Portion)	0	0
Small Chili Pepper (Portion)	14.08725645	723164445
Brassica Chinensis (1)	9.45764235	5.2702
Malabar Spinach	15.975461235	5.115678942
Zhijiang Green Stem Cauliflower	0	0
Honghu Lotus Root	0	0
Seafood Mushroom (Packet)	24.64523987	2.54987511
Bengal Dayflower Herb	12.485632758	3.743218975
Purple Eggplant (2)	0	0
Red Pepper (2)	5.634569871	1887564987
Red Lotus Root Belt	0	0
Sweet Potato Tips	0	0
Wuhu Green Pepper (1)	7.364898423	5.202
Amaranth	0	0
Ornamental Cabbage	0	0
Spinach	4.7564822031	14
Spinach (Portion)	4.23265997	5.234598752
Water Chestnut	29.874569511	14
Cordyceps Flower (Portion)	9.695621238	3.6235
Screw Pepper	16.08453246	11.28345697
Screw Pepper (Portion)	0	0

In the equation, $m \in n$. The objective function was the effect- and benefit-maximizing sales profit of the supermarket, which was the sum of the sales profits of each item.

In the mathematical model, $select_l$ was used as a control list for the number of purchases of all categories of vegetables. It was a list of lengths up to 49 with a value of 1 or 0, where 1 meant to purchase the item and 0 meant not to purchase the item. The function of the constraint is 9. $Supply_l$ was used as a list of replenishment volumes whose length size was 49. The purchase quantity of each category of vegetable item indicated its value. $Sales_l$ was taken as a list of sales prices, whose length was also 49, and the sales unit price of each category of vegetable item represented its value [5].

The constraint function was used to find the relationship between the pricing and sales volume of each kind of vegetable item. This aimed to maximize the objective of multiplying the unit price of each item by the total sales volume through the equation. It can be obtained from the following table 2.

8. Evaluation and Improvement of the Model

8.1. Advantages:

1. Data Driving: The model is able to make decisions based on actual data, rather than on assumptions or guesswork.

2. Comprehensive Consideration: The integrated approach helps to holistically analyze the impact of various factors on decision-making, thus making more comprehensive decisions.

3. Forecasting and Optimizing: The combination of forecasting and optimizing can help supermarkets better grasp market trends and thus make more informed decisions.

4. Pricing Strategy: A cost-plus pricing methodology was used, which takes into account the impact of transportation losses and quality degradation on pricing, helping maintain a reasonable level of profitability. This pricing method is based on the actual situation, being conducive to maintaining the competitiveness of the supermarkets and having a reasonable level of profit.

5. Meeting Market Demand: This approach to meeting market demand helps supermarkets better satisfy consumer demand, thereby increasing sales and revenue.

8.2. Disadvantages:

The validity of the model is limited by assumptions and data quality, so it needs to be applied to real life with care. Adjustment and improvement are necessary as real-life situations change.

8.3. Improvement:

The model has many advantages, including a large amount of data as a core driver and multiple considerations. Furthermore, it is important to recognize the limitations of the model, which is highly dependent on data quality and is subject to constraints and limitations of assumptions, market uncertainty, and the risk of human intervention.

References

[1] Yao Haoli, Wang Peihui, Ru Liming. A Station Phase Identification Method Based on Pearson Correlation Coefficient Algorithm.

[2] Gan Wenyan, LI Deyi, WANG Jianmin. A Hierarchical Clustering Method Based on Data Field[J]. Journal of Electronics, 2006, 34(2):258-262.

[3] He Yong, Xiang Liguo. Research and application of BP neural network model based on fuzzy clustering[J]. Systematic Engineering Theory and Practice, 2004, 24(2):79-82.

[4] Jiang Ming, WU Chunming, Zhang Min, et al. A Comparative Study of Time Series Models in Network Traffic Prediction [J]. Journal of Electronics, 2009(11):6-10.

[5] Tang Jiafu, Wang Dingwei, Gao Zhen, et al. A Hybrid Genetic Algorithm for Nonlinear Planning Problems[J]. Journal of Automation, 2000, 26(3):4-8.