How to Set Reasonable Prices for Vegetable Products? A Comprehensive Model Study Based on Multiple Algorithms

Wenxuan Du[1,*], Yuzan Zuo[2]

[1]School of Economics and Finance, Xi'an Jiaotong University, Xi'an, China, 710049

[2]The School of Management, Xi'an Jiaotong University, Xi'an, China, 710049

* Corresponding Author Email: 3037125594@qq.com

Abstract. Due to the characteristics of a wide variety of vegetables, short shelf life, and early procurement, the formulation of vegetable category management strategies has always been a challenge for fresh supermarkets. This paper investigates the relationship between vegetable categories, individual product sales, and cost-based pricing. It formulates optimal replenishment and pricing strategies for vegetable categories and individual products in fresh supermarkets based on algorithms such as Gaussian Process Regression (GPR), simulated annealing, and regression trees. It also carefully considers the potential costs at various stages of the sales process in fresh supermarkets, making the model more applicable to real operational environments. This provides strong support for the formulation of replenishment and pricing strategies in category management for fresh supermarkets.

Keywords: Association Rules; Gaussian Process Regression; Simulated Annealing; Regression Trees.

1. Introduction

The characteristics of vegetables, including a wide variety, short shelf life, and early procurement, have posed a persistent challenge for fresh supermarkets [1-3]. Therefore, addressing the following questions is crucial for pricing and replenishment management of vegetable products: First, how to analyze the correlation between sales volume of different vegetable categories and individual products? Second, how to establish a functional relationship between the total sales volume of each vegetable category and cost-based pricing, and formulate replenishment and pricing strategies to maximize revenue in the coming week based on this relationship? Third, how to determine the optimal replenishment and pricing strategy for a given day based on the available product information, considering limited shelf space? Fourth, how to provide data collection recommendations relevant to the effective optimization of replenishment and pricing decisions for vegetable products?

In light of these challenges, this paper conducts empirical research using data from a certain supermarket's sales of six vegetable categories over the past three years, including product information, sales details, wholesale prices, and recent loss rates. First, it analyzes the correlation between vegetable categories and individual product sales. Second, it uses the Gaussian Process Regression (GPR) algorithm to establish the relationship between sales volume and cost-based pricing for each category. While constraining the relationship between category sales volume and pricing established by the GPR algorithm, it utilizes simulated annealing to determine the optimal total replenishment quantity and pricing strategy. Finally, this paper combines the Analytic Hierarchy Process and entropy weighting method to obtain composite weights for three evaluation criteria: loss rate, profit margin, and sales volume. Using the TOPSIS method, it calculates value scores for 49 sellable individual products, selecting the top 33 products for replenishment. Using a regression tree model, this paper predicts the cost and sales volume for each product on July 1, 2023, and establishes an objective function to maximize the total profit from the 33 products. Constraints include individual product display weight greater than 2.5 kg, the presence of at least 27 individual products on the shelf, and meeting market demand for each category. The result is the optimal replenishment and pricing strategy for July 1, 2023.

2. Research design

2.1. Correlation Analysis

Firstly, this paper employs the Spearman correlation coefficient to measure the correlation of sales volumes among vegetable categories. The formula for calculating the Spearman correlation coefficient is as follows:

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

Where (n) represents the total number of samples, and (d) denotes the rank difference between two data points.

2.2. Category Sales Volume-Cost Markup Pricing Relationship Model Based on Gaussian Process Regression (GPR) Algorithm

The calculations for the weighted wholesale price $P_{0_{ik}}$, weighted markup rate α_{ik}, average cost markup pricing $P_{1_{ik}}$, and total sales volume Q_{ik} of the ith vegetable category at the kth day and time from the third quarter of 2020 to the second quarter of 2023 are as follows:

$$P_{0_{ik}} = \frac{\sum p_{0_{ijk}} \times q_{ijk}}{\sum q_{ijk}} \tag{2}$$

$$p_{0_{ijk}} \times (1 + a_{ijk}) = p_{1_{ijk}} \tag{3}$$

$$\alpha_{ik} = \frac{\sum a_{ijk} \times q_{ijk}}{\sum q_{ijk}} \tag{4}$$

$$P_{0_{ik}} \times (1 + \alpha_{ik}) = P_{1_{ik}} \tag{5}$$

$$Q_{ik} = \sum q_{ijk} \tag{6}$$

Among them, $p_{0_{ijk}}$ is the wholesale price of the j vegetable item in the i vegetable category on day k, q_{ijk} is the sales volume of the j vegetable item in the i vegetable category on day k, a_{ijk} is the bonus rate in the j vegetable item in the i vegetable category on day k, $p_{1_{ijk}}$ is the sales price in the j vegetable item in the i vegetable category on day k.

2.3. Optimization Model for Future Replenishment and Pricing Strategies of Various Vegetable Categories

Taking into account the varying degrees of losses that occur during transportation, storage, and other processes for different vegetable products, some products with higher levels of loss are discarded on the same day, while others with lower loss levels are sold at discounted prices on the same day. Therefore, this paper establishes the total profit for the i vegetable category on the k day as R_{ik}.

$$R_{ik} = (P_{1_{ik}} - P_{0_{ik}}) \times (1 - L_{ik}) \times Q_{ik} + (r_i \times P_{1_{ik}} - P_{0_{ik}}) \times D_i \times Q_{ik} - (L_{ik} - D_i) \times P_{0_{ik}} \times Q_{ik} \tag{7}$$

$$L_{ik} = \frac{\sum l_{ijk} \times q_{ijk}}{\sum q_{ijk}} \tag{8}$$

$$d_{ij} = \frac{n_{ij}}{N_{ij}} \tag{9}$$

$$D_i = \frac{\sum d_{ij}}{t_i} \tag{10}$$

Among them, l_{ijk} is the loss rate of the j vegetable item in the i vegetable category on the k day, L_{ik} is the sales weighted loss rate of the i vegetable category on the k day, n_{ij} is the number

of times that the j vegetable item in the i vegetable category has been discounted in the sales details in the past three years, N_{ij} is the total number of times the j vegetable item in the i vegetable category appears in the sales details in the past three years, d_{ij} is the sales discount ratio of the j vegetable item in the i vegetable category in the past three years, t_i is the number of individual vegetable types in the i vegetable category, D_i is the sales discount ratio of the i vegetable category in the past three years.

In addition, considering the significant variation in discount rates among different vegetable categories in different months, to ensure model accuracy, this paper uses the average ratio of post-discount sale prices to pre-discount sale prices for the i vegetable category's products in July over the past three years as the discount rate r_i for predicting the discount rates for the upcoming week (July). When formulating replenishment and pricing strategies for each vegetable category for the upcoming week (July 1-7, 2023) to maximize the total profit for the supermarket, this paper establishes the following optimization model:

$$max\, Z = \sum_k \sum_i R_{ik} \tag{11}$$

Where the total sales volume Q_{ik} for the i vegetable category on the k day, along with the average cost markup pricing $P_{1_{ik}}(i = 1,2,\ldots,6)$ across 6 vegetable categories, conforms to the category sales volume-cost markup pricing relationship model constructed based on the Gaussian Process Regression (GPR) algorithm, as outlined in Section 5.2.2.

2.4. Evaluation Model for Saleable Vegetable Products based on Entropy Weight-AHP

To establish a scientifically sound ranking of saleable vegetable product categories, this paper analyzes the factors influencing these categories. Drawing upon the research of Zhu et al. (2020)[1], Ma et al. (2018)[4], we construct an objective selection model using the AHP-Entropy Weight method.

2.5. Vegetable Product Wholesale Price Prediction Model Based on Regression Tree Algorithm

The Regression Tree algorithm, also known as the Fine-Grained Tree, is employed in this paper to address regression problems. Specifically, it is used for time series estimation of wholesale prices for individual vegetable products. In this model, each vegetable product is characterized by three features: year, month, and day. The algorithm partitions the feature space at various nodes and calculates the average prediction within each region R_J using the samples within that region.

$$\hat{y}R_j = \frac{1}{n}\sum j \in R_j y_j \tag{12}$$

This allows to obtain the sum of squared errors for each feature space as:

$$RSS = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}R_j)^2 \tag{13}$$

2.6. Data Preprocessing

To ensure the accuracy of the vegetable product data indicators, this study used sales details from a certain convenience store over the past three years as samples. It matched the product ID with the product name and category name to complete the sales detail data for vegetable products. Considering the possibility of anomalies such as sensor failures and data entry errors during the data collection process, this study employed box plots to handle outliers in sales volume. Sales volumes falling within the range of [Q1 - 1.5×IQR, Q3 + 1.5×IQR], i.e., ["-0.388", 1, 42], were labeled as outliers to avoid affecting the accuracy of the model analysis, resulting in the removal of 10588 records.

Vegetable products are typically influenced by factors such as supply and demand, seasonal changes, origin, and quality. Prices may experience some fluctuations, but supermarkets generally

maintain prices within a reasonable range. This study conducted a survey of the average market prices of 251 individual vegetable products over the past three years. It identified anomalies where, for example, the average market price for a certain type of mushroom was 8.24 yuan per kilogram, while the selling price in the sales details was 0.5 yuan per kilogram. These anomalies contradicted considerations of market reference prices and supermarket profits, resulting in the removal of 12 records. In total, 10600 abnormal records were removed.

3. Results

3.1. Correlation in Vegetable Sales

As indicated in Table 1, this study revealed varying degrees of correlation among sales volumes of different vegetable categories. Specifically: positive correlations were observed between leafy vegetables and flower vegetables, capsicum annuum, and edible fungi, flower vegetables showed significant positive correlations with aquatic root vegetables and edible fungi, aquatic root vegetables exhibited significant correlations with solanaceae and edible fungi, a significant negative correlation was found between solanaceae and edible fungi, capsicum annuum displayed a significant positive correlation with edible fungi.

These correlations provide valuable market sales insights; however, they should be analyzed and considered in conjunction with other factors for comprehensive decision-making.

Table 1: Spearman Correlation Coefficients for Sales Volume Among Six Vegetable Categories

			Leafy Vegetables	Flower Vegetables	Aquatic Root Vegetables	Solanaceae	Capsicum Annuum	Edible Fungi
Spearman Correlation Coefficient	Leafy Vegetables	Correlation Coefficient	1.000	0.617**	0.285	0.051	0.609**	0.451**
		Significance (Two-tailed)	.	0.000	0.092	0.769	0.000	0.006
		N	36	36	36	36	36	36
	Cauliflower	Correlation Coefficient	0.617**	1.000	0.360*	0.047	0.314	0.458**
		Significance (Two-tailed)	0.000	.	0.031	0.784	0.063	0.005
		N	36	36	36	36	36	36
	Aquatic Root Vegetables	Correlation Coefficient	0.285	0.360*	1.000	-0.477**	0.268	0.708**
		Significance (Two-tailed)	0.092	0.031	.	0.003	0.114	0.000
		N	36	36	36	36	36	36
	Solanaceae	Correlation Coefficient	0.051	0.047	-0.477**	1.000	-0.156	-0.462**
		Significance (Two-tailed)	0.769	0.784	0.003	.	0.364	0.005
		N	36	36	36	36	36	36
	Capsicum Annuum	Correlation Coefficient	0.609**	0.314	0.268	-0.156	1.000	0.492**
		Significance (Two-tailed)	0.000	0.063	0.114	0.364	.	0.002
		N	36	36	36	36	36	36
	Edible Fungi	Correlation Coefficient	0.451**	0.458**	0.708**	-0.462**	0.492**	1.000
		Significance (Two-tailed)	0.006	0.005	0.000	0.005	0.002	.
		N	36	36	36	36	36	36

Furthermore, due to the large number of individual vegetable products in this fresh supermarket, this paper presents the dependency indicators among the top ten vegetable products in terms of total sales volume for Leafy Vegetables over the past three years. These indicators are shown in Figure 1.

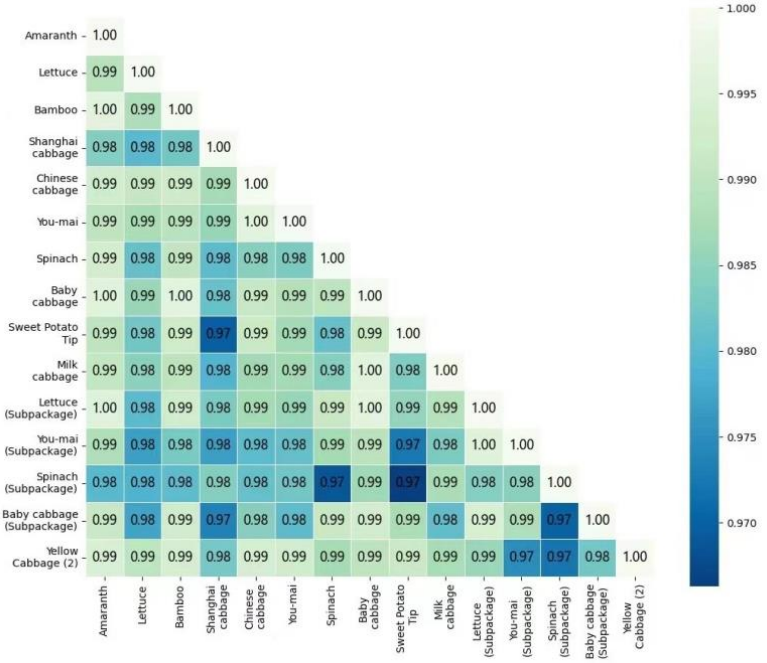

Figure 1: Dependency Analysis of the Top Ten Leafy Vegetable Sales Items

3.2. Optimization of Replenishment and Pricing Strategies

Due to the interrelationships between different categories and the influence of time on sales, this paper selected the selling prices and time as independent variables for the six categories. The sales volume of a specific category was considered as the dependent variable, and a Gaussian Process Regression (GPR) model was established to determine the correlation between the sales volume of that category and the selling prices of these six categories. Ultimately, based on historical prices for each category, the upper and lower limits of prices for each category were determined.

Within this price range, a simulated annealing method was employed to solve the multivariate optimization problem in this study. It recorded the replenishment quantities and selling prices corresponding to the day when the total profit for the six categories was maximized. The scatter plot of predicted values obtained through the exponential GPR method in the Gaussian Process Regression algorithm and the actual values is not presented here due to space constraints.

To address the multivariate optimization problem in this study using the simulated annealing method, the sales volume for each category constituted a multivariate function of the prices of all categories. The model selected a decay parameter of 0.98, an initial annealing temperature of 100, a step size of 0.05, and a final temperature of 1. The model was solved using Matlab software, and the daily replenishment totals and pricing strategies are described in Table 2.

Table 2: Daily Replenishment Total and Pricing Strategy for Each Vegetable Category

Vegetable Categories	Daily Replenishment Total	Sales Prices
Leafy Vegetables	150.6403	5.7229
Flower Vegetables	36.93176	13.28785
Aquatic Root Vegetables	23.9494	14.98756
Solanaceae	20.42049	8.818586
Capsicum Annuum	90.31632	6.920576
Edible Fungi	49.62495	7.631802

3.3. Ranking of Saleable Vegetable Categories by Value

Based on the entropy-weighted AHP (Analytic Hierarchy Process) evaluation model for saleable vegetable products, the scores of 47 saleable vegetable products are ranked, and the top 33 scores are selected as the list of products requiring replenishment, as shown in Table 3.

Table 3: List of the Top 33 Products Requiring Replenishment Based on Scores

Product Name	Vegetable Category	Comprehensive Value Score	Ranking
Bell Peppers (2)	Capsicum Annuum	100	1
Honghu Lotus Stems	Aquatic Root Vegetables	93.7846	2
Wuhu Green Peppers (1)	Capsicum Annuum	85.3279	3
Water Bamboo (2)	Aquatic Root Vegetables	84.5097	4
Yunnan Lettuce (Per Serving)	Leafy Vegetables	83.9268	5
Xixia Flower Mushrooms (1)	Edible Fungi	83.5051	6
…	…	…	…

3.4. Final Pricing of Vegetable Categories

Based on decision tree algorithms, this paper forecasts the wholesale prices of the 33 available vegetable products on July 1, 2023. The results are presented in Table 4 and graphically illustrated below:

Table 4: Wholesale Price Predictions for the 33 Saleable Vegetable Products

Product Name	Vegetable Category	Predicted Purchase Price
Bell Peppers(2)	Capsicum Annuum	2.873
Honghu Lotus Stems	Aquatic Root Vegetables	14.385
Wuhu Green Peppers(1)	Capsicum Annuum	4.2
Water Bamboo(2)	Aquatic Root Vegetables	9.733
Yunnan Lettuce (Per Serving)	Leafy Vegetables	4.908
Xixia Flower Mushrooms(1)	Edible Fungi	1.827
…	…	…

Taking into account the varying degrees of loss for different vegetable products during transportation and storage, this paper utilizes the nonlinear programming module of Lingo software to determine the optimal individual product replenishment quantities and pricing strategies for July 1, 2023, as shown in Table 5.

Table 5: Optimal Individual Product Replenishment Quantities and Pricing Strategies for July 1, 2023

Product Name	Vegetable Category	Product Replenishment Quantity	Sales Pricing
Bell Peppers(2)	Capsicum Annuum	2.873	35.386
Honghu Lotus Stems	Aquatic Root Vegetables	14.385	2.5
Wuhu Green Peppers(1)	Capsicum Annuum	4.2	22.998
Water Bamboo(2)	Aquatic Root Vegetables	9.733	2.5
Yunnan Lettuce (Per Serving)	Leafy Vegetables	4.908	2.5
Xixia Flower Mushrooms(1)	Edible Fungi	1.827	36.122
…	…	…	…

4. Conclusion

This paper explores the relationships between both vegetable category and individual product sales volumes and cost-based pricing. It formulates optimal replenishment and pricing strategies for vegetable categories and individual products in fresh supermarkets using Gaussian Process Regression (GPR), simulated annealing, and regression tree algorithms. Leveraging a variety of machine learning algorithms and evaluation models to address issues in vegetable category management strategy formulation, this paper also meticulously considers the potential costs at various

stages of the sales process in fresh supermarkets, making the model closely align with actual operational environments. This provides robust support for the development of replenishment and pricing strategies in category management for fresh supermarkets.

References

[1] Yubing, Zhou., (2022), Research on fresh product inventory control and dynamic pricing based on intensive learning [D]. Central South University, 2022. DOI: 10.27661/d.cnki.gzhnu.2022.004101.

[2] Mingming, Zeng., (2022) Research on dynamic pricing strategy of fresh community supermarket based on time situation a [D]. Southwestern University of Finance and Economics, 2022. DOI: 10.27412/d.cnki.gxncu.2021.002858.

[3] Jing, Lu,. (2021) Inventory control and dynamic pricing of fresh agricultural products [D]. Tianjin University, 2021. DOI: 10.27356/d.cnki.gtjdu.2019.000182.

[4] Zhu, Y., Tian, D., & Yan, F. (2020). Effectiveness of entropy weight method in decision-making. Mathematical Problems in Engineering, 2020, 1-5.

[5] Ma, L., Chen, H., Yan, H., Li, W., Zhang, J., & Zhang, W. (2018). Post evaluation of distributed energy generation combining the attribute hierarchical model and matter-element extension theory. Journal of Cleaner Production, 184, 503-510.

[6] Yali, L. (2023), Joint Replenishment and pricing decision-making considering investment in preservation technology. Chongqing University

[7] Xiuxiu, Zhen (2018). Study on the two-channel supply chain inventory strategy of fresh agricultural products considering the effect of fresh-keeping level. Xidian University

[8] Fang, Zhang, Jianbiao. Chen. (2020). Fresh agricultural products inventory control research. China's storage and transportation. 2020(06): 126-127. DOI: 10.16301/j.cnki.cn12-1204/f.2020.06.048.

[9] Yifei, Wang, Meng, Shang. (2021) Study on the optimization of fresh agricultural product inventory management based on ABC classification and AHP taking a supermarket as an example. China, 2021(21): 127-128. DOI: 10.13939/J. CNKI. November 127,2021.

[10] Yufei, Zhang. Study on optimization of replenishment strategy for grain, oil and food of Company D during promotion period [D] . Hebei University of Technology, 2023. DOI: 10.27105/d.cnki.ghbgu.2021.001421.