

Research on Automatic Pricing and Restocking Decisions for Vegetable Products: A Case Study of a Supermarket

Hongbin Wu^{1,a,*}, Yixian Jiang^{2,b}, Weisen Shen^{3,c}

¹*College of Intelligent Build, Yango University, Fuzhou, 350015, China*

²*School of Economics and Management, Yango University, Fuzhou, 350015, China*

³*Industrial Internet College, Yango University, Fuzhou, 350015, China*

a. wuhongbin90586@163.com, b. 2174725026@qq.com, c. 2547529815@qq.com

**corresponding author*

Abstract: For fresh supermarkets, it is crucial to formulate restocking decisions and pricing strategies that maximize profits while meeting market demand. This paper mainly studies the automatic pricing and restocking decisions for vegetable products. Through model assumptions, correlation analysis, and time series methods, the study investigates these decisions. Based on the obtained sales data, Pearson correlation coefficients are calculated to analyze the relationships between different categories and individual product sales. A Long Short-Term Memory (LSTM) time series model is built to forecast future costs, with the model parameters estimated using the least squares method. The optimal restocking and pricing strategies are derived using Random Forest and Autoregressive Integrated Moving Average (ARIMA) models. Additionally, the restocking and pricing problems are treated as a combinatorial optimization problem and solved using a mixed-integer linear programming model to maximize revenue. This paper emphasizes the importance of comprehensive data collection in automatic pricing and restocking decisions to address challenges in the industry.

Keywords: Automatic pricing, Correlation analysis, Random forest model, ARIMA model

1. Introduction

In fresh supermarkets, the shelf life of vegetable products is typically short, and their quality gradually deteriorates as sales time progresses, leading to the fact that most products may no longer be saleable the following day. Therefore, supermarkets must make accurate restocking decisions daily, based on historical sales data and demand trends for vegetable products. Supermarkets offer a wide range of vegetable products, sourced from various origins. While meeting the diverse needs of consumers, they must also strive to maximize profits. For retailers, accurately predicting future demand, and formulating reasonable restocking and pricing decisions is crucial, especially when they do not have precise information on the cost price of individual products. This paper establishes appropriate model assumptions based on data provided in the 2023 “Higher Education Press Cup” National College Students Mathematical Modeling Competition (C problem) [1] and conducts in-depth research on automatic pricing and restocking decisions for vegetable products through algorithm optimization and other methods.

2. Investigating the Correlation of Sales Volumes for Different Categories and Individual Products

2.1. Distribution Patterns of Sales Volumes for Different Categories and Individual Vegetable Products

The relevant data for vegetable products in a supermarket were subjected to missing value processing, data normalization, and standardization preprocessing steps. Similar items were merged, and outliers were removed to ensure the data met the requirements of Principal Component Analysis (PCA) [2]. PCA decomposes the sample matrix or sample variance matrix of process variables, and the selected principal components are uncorrelated. These principal components can be derived from the process variables through linear combinations. After integrating the data, this study obtained the sales volumes for various vegetable categories and individual products, using descriptive statistical analysis for further examination. Initially, the focus is on the overall sales volume of vegetable categories, followed by an analysis at the individual product level. The histograms of sales volumes for vegetable categories and individual products are shown in Figures 1 and 2 below.

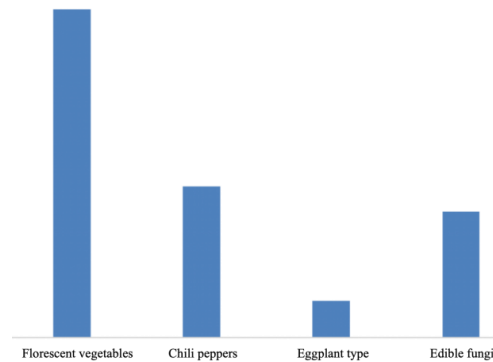


Figure 1: Sales Volume by Vegetable Category

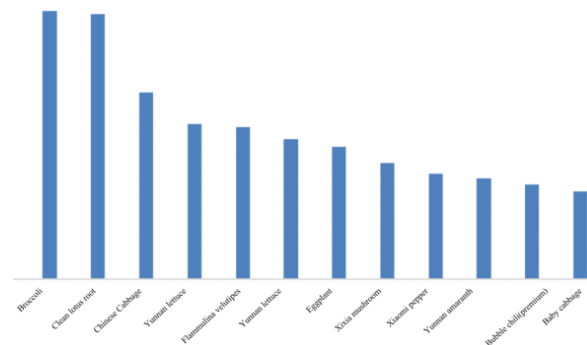


Figure 2: Sales Volume by Individual Vegetable Product

According to the statistical analysis results, leafy vegetables ranked highest in sales volume, followed by chili peppers and edible fungi. In terms of individual vegetable product sales, WuHu green peppers (1), broccoli, and clean lotus roots (1) ranked in the top three. Only 11 individual products had sales volumes exceeding 10,000 kilograms, accounting for nearly 5% of the total vegetable product sales.

Based on the in-depth analysis of the sales data, this paper aims to gain a more comprehensive understanding of market trends and forecast sales movements, which will inform reasonable restocking decisions. To achieve this, a time series analysis method is employed to explore the trends and seasonal fluctuations in sales volumes. The specific analysis results are shown in Figure 3 below.

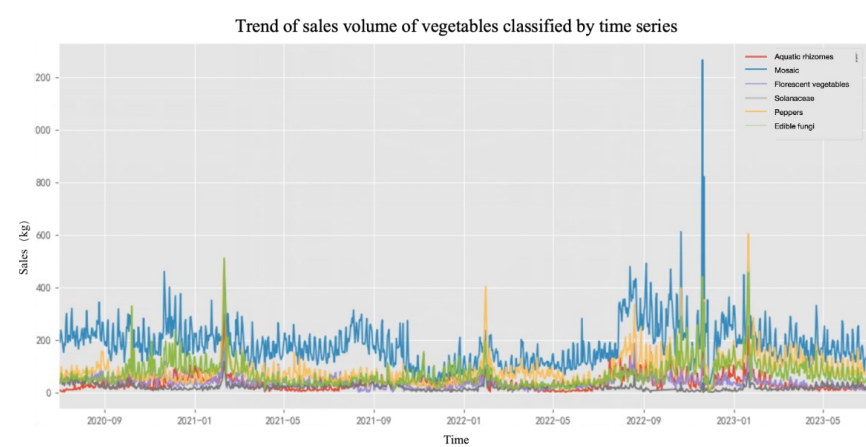


Figure 3: Sales Volume Trend for Vegetable Categories in Time Series

According to the model analysis results, sales volumes for vegetable categories show significant seasonality and cyclicity. During the periods of September 2021 and January 2022, sales volumes for all vegetable categories reached their peaks. Particularly, aquatic root vegetables and cauliflower experienced two distinct sales peaks between September 2021 and January 2022.

The analysis suggests that, in general, sales volumes for all vegetable types appear to show an upward trend, likely driven by long-term demand increases due to factors such as population growth and economic improvement. Demand in the individual vegetable product market is relatively dispersed, with consumers exhibiting diverse and personalized preferences. Leafy vegetables are likely to be more popular in the market, potentially due to their broader applications and consumer demand. Leafy vegetables reach their peak during specific times of the year, which may be related to high consumer demand, holidays, or seasonal events.

2.2. The Interrelationship of Sales Volumes for Different Vegetable Categories and Individual Products

Based on the distribution patterns described above, correlation analysis is used to determine the relationships between different vegetable categories. Specifically, the Pearson correlation coefficient [3] is calculated to study the associations between sales volumes of different vegetable categories. The correlation coefficient matrices for vegetable categories and individual products, as well as the heatmap of the correlation coefficients between individual product sales volumes, are shown in Figures 4 and 5 below:

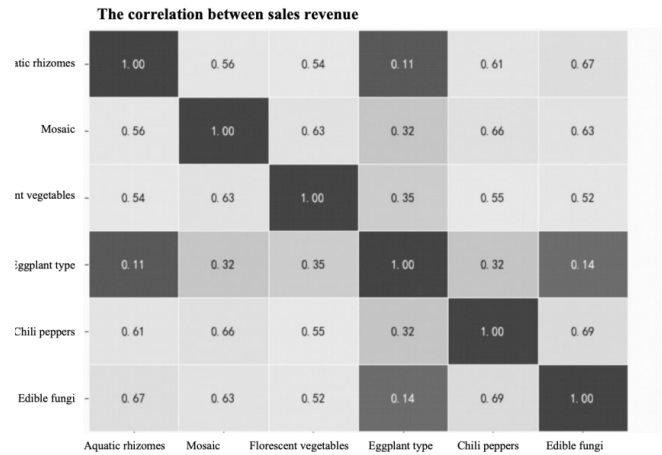


Figure 4: Correlation Matrix Heatmap for Vegetable Categories

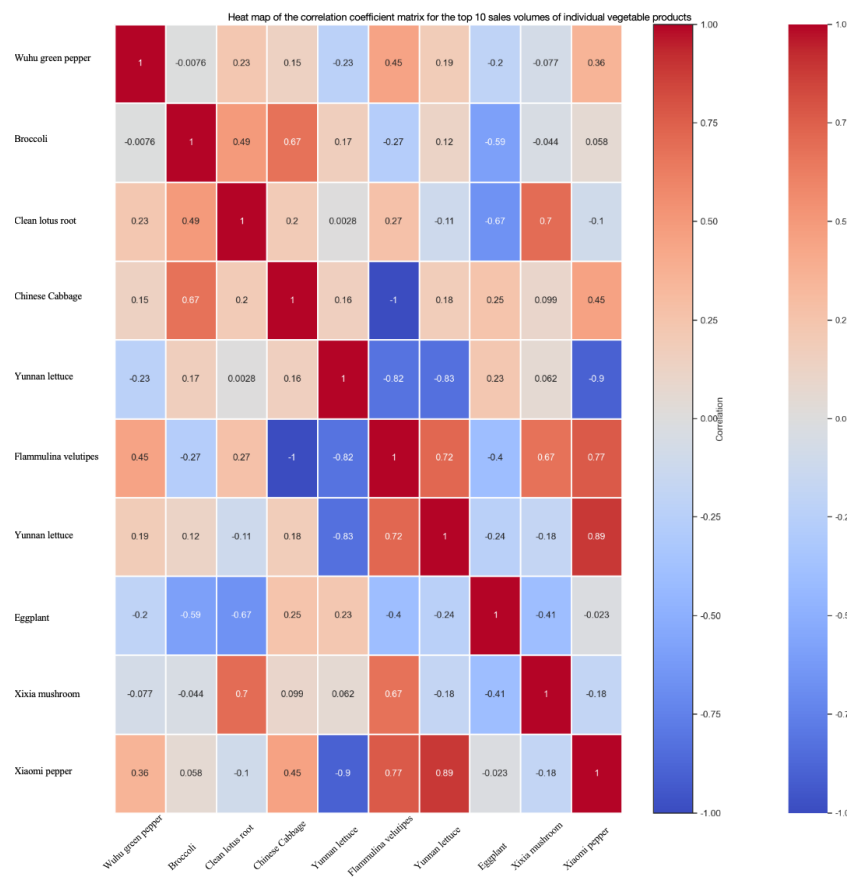


Figure 5: Correlation Matrix Heatmap for Top 10 Individual Vegetable Product Sales

The Pearson correlation coefficient is used to measure the linear correlation between two datasets, X and Y, with values ranging from -1 to 1 [4]. The analysis results indicate that there is a high positive correlation between chili peppers and edible fungi, suggesting that as consumers increase their purchases of chili peppers, the sales of edible fungi may also rise. The correlation between eggplant and most other vegetable categories is relatively low, indicating that the sales pattern of eggplant may be independent. A strong positive correlation exists between enoki mushrooms, Yunnan lettuce, and small chili peppers, while a significant negative correlation is observed with Chinese cabbage. This could be due to certain combinations or meal sets, or perhaps consumer purchasing habits.

3. Mathematical Model Development

3.1. Random Forest Regression

Given that supermarkets must both satisfy consumer demand and maximize profits, but the number of individual products available for sale each day is limited, it is difficult to meet all consumer demands. Therefore, a comprehensive evaluation of customer satisfaction and profitability is required. Since the characteristics of each product vary, relying solely on sales volume to measure consumer demand may overlook products with lower individual sales but higher purchase frequency. Thus, by considering both the sales volume and the frequency of sales, we can more comprehensively assess consumer demand for individual products and make more reasonable sales decisions.

Based on the data for available products from June 24 to 30, 2023, the scope of restocking for July 1 is determined. This paper employs a Random Forest Regression model. Random Forest is a statistical learning method that uses the bootstrap resampling technique to extract multiple samples

from the original dataset. A decision tree model is built for each bootstrap sample, and the predictions of all the decision trees are combined through voting to yield the final prediction. This method is robust to outliers and noise, with good tolerance, and is less prone to overfitting [5], offering higher stability and reliability. The results are shown in Figure 6.

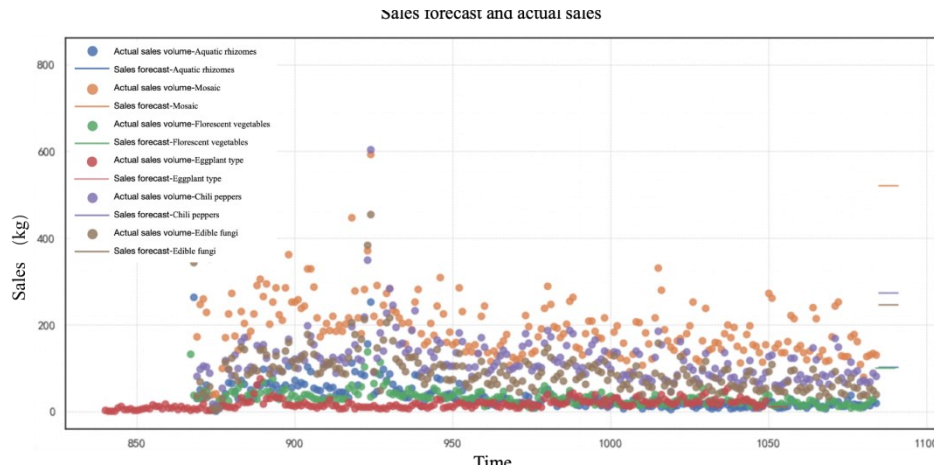


Figure 6: Scatter Plot of Random Forest Model

By calculating the Mean Squared Error (MSE) between the predicted and actual values, we evaluated the accuracy of the model. As shown in Figure 7 below:

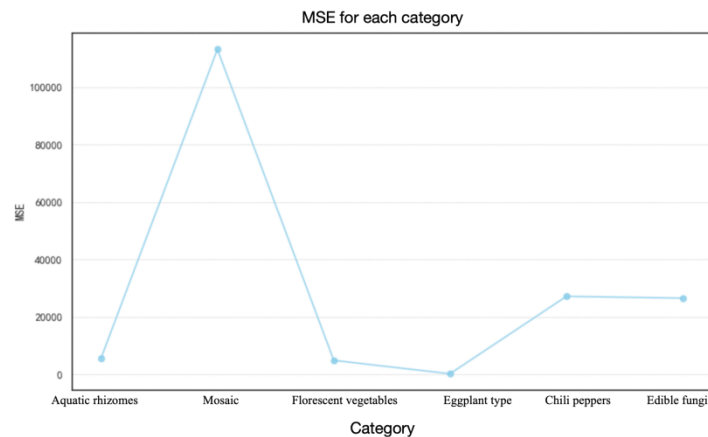


Figure 7: Comparison of MSE Values for Different Categories

As seen in Figure 7, the Random Forest model demonstrates high accuracy in predicting the sales volumes of cauliflower and eggplant, while its performance is relatively poor in predicting the sales volumes of leafy vegetables. Although some errors are present in the predictions, they remain within an acceptable range.

3.2. Autoregressive Integrated Moving Average (ARIMA) Model

To improve the overall performance of the model, feature engineering and model adjustments are performed for the categories with poor performance, and an Autoregressive Integrated Moving Average (ARIMA) model is used for prediction. The ARIMA model [6] can handle stationary time series and capture both seasonality and trends. The model can be expressed as the following formula:

$$\hat{p}^{\{t\}} = p_0 + \sum_{j=1}^p \gamma_j p^{\{t-j\}} + \sum_{j=1}^q \theta_j \varepsilon^{\{t-j\}}$$

To develop a pricing strategy, we consider the wholesale price, average loss rate, and expected sales volume for each vegetable category. A “cost-plus pricing” strategy is employed, taking into account the product cost and expected sales volume to determine a suitable selling price. Based on the formula, the adjusted model is used to predict the price of the products. We selected 33 individual products with the highest expected returns, with a daily total expected revenue of approximately 1,483.93 yuan.

4. Model Analysis

4.1. Advantages of the Model

Based on relevant literature, the Autoregressive Integrated Moving Average (ARIMA) model is used to forecast data, introducing a time series model for effective modeling and prediction. This approach provides high interpretability and visualizability. A linear optimization model is employed, considering multiple constraints and objective functions, to correct predicted sales and offer the best solution.

4.2. Directions for Improvement

When predicting sales, it is important to recognize that price fluctuations alone may not fully capture market dynamics. Ignoring the impact of price changes in substitute and complementary products could lead to prediction biases. In future research, the cross-price effects on product sales will be incorporated as key variables, allowing for the development of a more precise and comprehensive sales forecasting model.

5. Conclusions and Recommendations

5.1. Research Conclusions

This paper constructs an automatic pricing and restocking decision model for fresh produce in supermarkets, and the following conclusions are drawn:

(1) There is varying degrees of positive correlation between different vegetable categories, with certain individual products showing significant positive or negative correlations. These correlations are closely tied to the characteristics of the vegetables themselves and reveal consumer purchasing habits and potential product bundle effects.

(2) The Random Forest Regression model and the Autoregressive Integrated Moving Average (ARIMA) model established in this study can ensure that consumer demand is met while improving daily supermarket revenues.

(3) The study also considers the interrelationship between sales and pricing. The ARIMA model is used to adjust the predicted sales, helping supermarket managers gain deeper insights into market dynamics and consumer behavior.

5.2. Policy Recommendations

(1) Optimize Purchasing Time and Quantity for Vegetable Products

In supermarket management, vegetables have a short shelf life, and their quality deteriorates as time passes. Optimizing purchasing time and quantity is crucial as it directly impacts operational

efficiency and cost control. By collecting sales data for vegetable products and using time series analysis to predict sales over a future period, supermarkets can combine this with inventory data, customer demographic data, and purchasing behavior data to determine optimal purchasing times and quantities. Additionally, considering the shelf life and transportation conditions of vegetables will improve the rationality of procurement arrangements. The goal is to reduce inventory costs, minimize product losses, and optimize purchasing plans to maximize supermarket revenue.

(2) Improve Inventory Management and Logistics Efficiency

In supermarket management, optimizing inventory and logistics is essential. Introducing artificial intelligence (AI) technology can help monitor sales and inventory data in real-time, enabling supermarkets to accurately track vegetable inventory levels and provide scientific inventory management strategies. This can help avoid sales losses caused by inventory shortages or surpluses. Moreover, AI can be used to analyze geographic location, store size, vegetable shelf life, transportation conditions, and origin data, allowing for optimized logistics routes through algorithms. This can shorten logistics time, reduce vegetable losses, improve logistics efficiency, and enhance the supermarket's market competitiveness, creating greater commercial value.

(3) Develop Rational Pricing Strategies and Promotional Activities

Due to the short shelf life and perishability of vegetables, it is essential for supermarkets to develop reasonable pricing strategies and promotional activities. Through market research to understand customer demand and price sensitivity, as well as closely monitoring competitors' strategies, supermarkets can better meet consumer needs. A pricing strategy and promotional activities that both satisfy consumer demand and improve customer satisfaction should be formulated, increasing market share and revenue.

References

- [1] National Undergraduate Mathematical Contest in Modeling Committee. (2023, December 7). 2023 "Higher Education Press Cup" National Undergraduate Mathematical Contest in Modeling Problems [Z/OL]. Retrieved from http://www.mcm.edu.cn/html_cn/block/8579f5fce999cdc896f78bca5d4f8237.html
- [2] Li, E., & Yu, J. (2001). Application of PCA in process fault detection and diagnosis. *Journal of East China University of Science and Technology (Natural Science Edition)*.
- [3] Xiao, X., Jiang, B., Ren, Q., et al. (2019). Photovoltaic data cleaning based on interpolation and Pearson correlation. *Information Technology*, 43(5), 5.
- [4] Ji, D., Jin, F., Dong, L., et al. (2022). Photovoltaic power station data restoration based on Pearson correlation coefficient. *Proceedings of the CSEE*, 42(4), 42.
- [5] Tang, Y. (2024, August 22). Research and application of time series analysis [Master's thesis]. Northeast Agricultural University.
- [6] Li, C., & Chen, Y. (2019). Improved ARIMA model and its prediction. *Fujian Quality Management*, 000(011), 268.