Research on Automatic Vegetable Pricing Strategies Based on Multiple Machine Learning Algorithms

Yanchen Ji [1, *], Yuze Jiang [1, #], Yunfei Zhou [2, #]

[1] School of Communication &Electronic Engineering East China Normal University, Shanghai, China

[2] School of Ecological and Environmental Sciences East China Normal University, Shanghai, China

* Corresponding Author Email: jiyanchen0707@163.com

[#]These authors contributed equally.

Abstract. This paper explores methods for optimizing product pricing and replenishment strategies in fresh supermarkets through advanced data analysis techniques. The initial data processing involves merging and cleaning, with a focus on statistical and visual analysis of sales distribution across products and categories. Descriptive statistics are derived using line charts, bar charts, and pie charts based on yearly, quarterly, and monthly data. This article analyzes relationships within the data employing the Kruskal-Wallis test, Spearman correlation coefficient, and Pearson correlation coefficient. Spectral clustering is utilized to group products, revealing patterns in their interrelationships. For pricing strategies, this paper implements a cost-plus pricing model to calculate markup rates, considering sales and fixed costs. This article apply regression algorithms, specifically XGBoost and Random Forest, to predict unit price, purchase price, and sales volume, and assess their performance for optimization. Time series forecasting techniques predict data for the upcoming week, complemented by an optimization function that visualizes this iterative process. Further data processing for optimizing purchase quantities employs regression algorithms, time series models, and genetic optimization algorithms to refine predictions, ultimately identifying strategies to maximize profit.

Keywords: Automatic pricing, Spearman correlation coefficient, Spectral clustering, PSO optimization.

1. Introduction

The operational model of fresh produce supermarkets involves daily deliveries and sales due to the perishable nature of products. Therefore, robust data analysis is essential for effective pricing and replenishment strategies. This paper investigates the interrelationships among individual products and categories through comprehensive data analysis, focusing on algorithmic approaches that enhance pricing and replenishment strategies.

In existing research, Lu et al. proposed an optimized replenishment strategy based on a dynamic matrix model. By establishing this model, they achieved intelligent replenishment on demand, which improved the overall profit of retailers [2]. Xuan et al. (2022) proposed an LSTM time series classification method by combining multi-scale convolution and attention mechanisms, significantly improving the accuracy and efficiency of time series classification [3]. Wang et al. combined Support Vector Regression (SVR) and Simulated Annealing algorithms to propose an automatic pricing and replenishment strategy for vegetable products, effectively increasing sales profits and inventory management efficiency, suitable for managing vegetable products in dynamic market environments [4]. In contrast, Hasan et al. established a non-instantaneous inventory model that considers external factors (such as improper handling, extreme weather, pests, etc.) affecting the quality of agricultural products. They proposed a strategy to separate near-spoiled products from intact ones and sell them at a discount to optimize pricing and replenishment decisions, maximizing total profits [5]. Liang et al. (2023) proposed a strategy for optimizing supermarket vegetable pricing and replenishment using an LSTM model, significantly improving profits [6]. Chen et al. (2023) used time series and ARIMA models in their analysis of replenishment and pricing strategies for vegetable products based on time

series models, optimizing the market supply chain management of different vegetable categories [7]. Paul, R.K., et al. (2022) used machine learning techniques to forecast the price of eggplant in Odisha, India. Their study demonstrated how machine learning models can effectively predict price fluctuations of agricultural products, thereby providing decision support to local farmers[8]. Kumari, P., et al. (2023) studied a Recurrent Neural Network (RNN) architecture for predicting the price of bananas in Gujarat, India. This research improved the accuracy of banana price forecasting using an RNN model, providing a basis for market decision-making [9].

The novelty of this paper lies in the integration of multiple algorithms to improve predic tive accuracy. By leveraging regression models such as XGBoost and Random Forest, this p aper identifies the optimal model through comparative evaluation, significantly enhancing the effectiveness of pricing and replenishment strategies. The application of Spearman correlati on and spectral clustering methods allows for a detailed analysis of product relationships, pr oviding a scientific foundation for informed decision-making. Finally, the optimization of pr edicted data via genetic algorithms further refines pricing and replenishment strategies under varying conditions, ensuring maximum operational efficiency. (Data source: https: //univs-ne ws-1256833609. file. myqcloud. com/52Kw7CxOBXSKw/2%40Q%40dFHtlmYSo.rar)

2. Study on Sales Distribution and Relationships Among Vegetable Categories

2.1. Sales Distribution Patterns Among Categories

To analyze the sales distribution patterns among categories, the total sales of each category within a specific time period must first be calculated, followed by visualization. A bar chart clearly shows the comparison of sales volumes between categories. First, this paper uses the entire three-year period as a unit to plot bar charts and pie charts of the sales distribution of each category over the three years, allowing for an intuitive observation of sales distribution patterns.

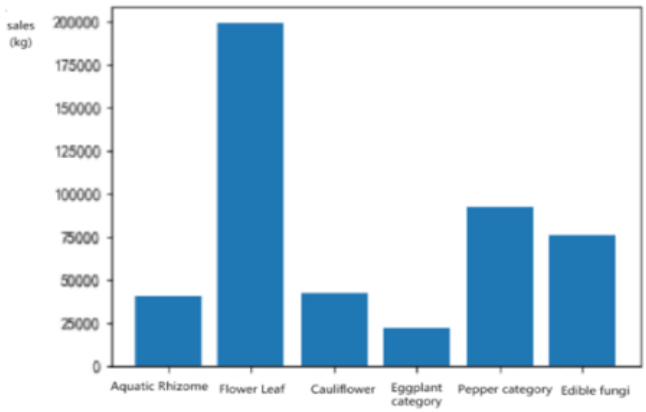


Figure 1. Histogram of Sales Distribution across Various Vegetable Categories.

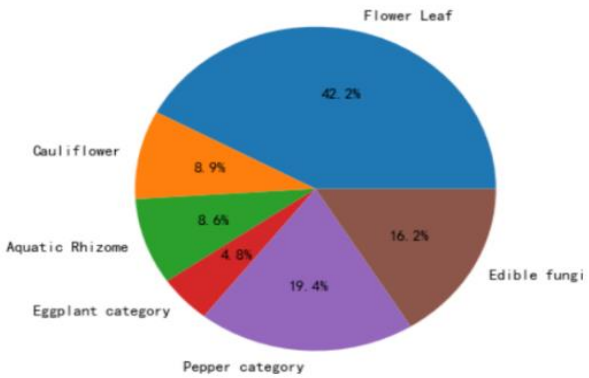


Figure 2. Pie Chart of Sales Distribution across Various Vegetable Categories.

As illustrated in Figures 1 and 2, the overall sales distribution indicates that leafy vegetables exhibit the highest sales volume, followed by chili peppers, with eggplants showing the lowest sales volume. In summary, the order of sales proportion is as follows: leafy vegetables > chili peppers > edible fungi > cauliflower > aquatic root and stem vegetables > eggplants.

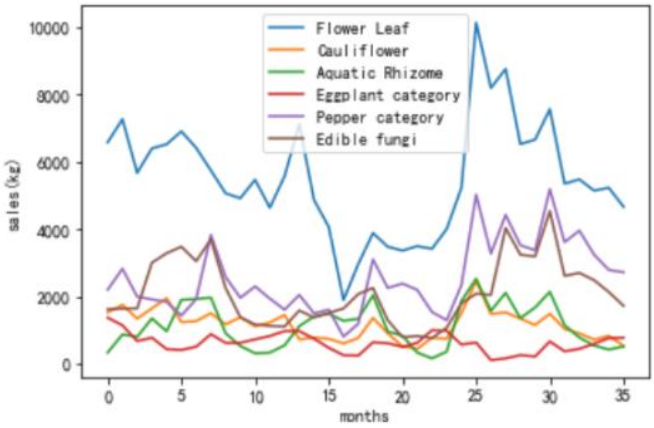


Figure 3. Line Chart of Monthly Sales across Various Vegetable Categories.

As evidenced by Figure 3, at any given time point, the total sales of leafy vegetables consistently exceed those of every other category. The temporal distribution of sales among the remaining categories appears more intricate.

Categories with closer sales volumes tend to exhibit similar trends. For example, the sales fluctuation trends of chili pepper products closely resemble those of edible fungi. Conversely, there is no conspicuous correlation between eggplants and leafy vegetables as depicted in the legend, necessitating further elucidation through data analysis.

Leafy vegetables and chili pepper products exhibit 1-2 cyclical peaks each year, which may be related to their growth patterns. Chili peppers are often used as seasonings or side dishes and are typically sold in conjunction with other products. At the beginning of 2023, there was a particularly high peak in the sales of leafy vegetables, which could be attributed to holidays or social factors.

2.2. Interrelationships Among Sales of Different Categories

To describe the interrelationships among variables, statistics introduce the concept of the "correlation coefficient" for descriptive analysis. The correlation coefficient $\rho(i,j)$ ranges between -1 and 1. When $\rho=0$, the two variables are uncorrelated; when $\rho<0$, the variables are negatively correlated; and when $\rho>0$, the variables are positively correlated. The closer the value is to 1, the stronger the correlation between the two variables.

Correlation coefficients are generally categorized into three types:

Pearson Correlation Coefficient: This is commonly used when data exhibit a linear relationship and follow a normal distribution. The formula is shown in (1)

$$\rho(x,y)=\frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y} \tag{1}$$

Spearman Correlation Coefficient: This coefficient is used when there are no strict requirements on the data. The formula is shown in (2).

$$\rho(x,y)=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}} \tag{2}$$

Kendall Correlation Coefficient: This coefficient is commonly used when there are no identical elements in sets X and Y. The formula is shown in (3).

$$\mathrm{Tau}-\alpha=\frac{C-D}{\frac{1}{2}N(N-1)} \tag{3}$$

Each type of correlation coefficient has its corresponding application range, as shown by formulas (1), (2), and (3).

Therefore, to evaluate the relationship between sales of different categories, it is essential to analyze the data first and choose the appropriate correlation coefficient. The Kruskal-Wallis test is a non-parametric statistical method used to compare two or more independent samples, determining whether these samples originate from the same population distribution. Specifically, the Kruskal-Wallis test examines whether there are significant differences in the medians between multiple groups, and it is often referred to as a non-parametric one-way analysis of variance.

The basic idea of the Kruskal-Wallis test is to combine the data from each group, sort the combined data, and then reassign the sorted data points back to their original groups based on their ranks. By comparing the rank sums of each group, the test assesses whether the medians of these groups differ significantly. This paper employs the Kruskal-Wallis test to determine if there are statistically significant differences in the data.

Based on Python programming, the results showed a test statistic of 52,274 with a p-value of 0.0, where $p < 0.05$. Thus, this paper preliminarily concludes that there are statistically significant differences in sales between different categories. However, due to the large number of groups, there may be considerable errors.

The Spearman and Pearson correlation coefficients are widely applicable, making them suitable standards for assessing relationships between categories. This paper programmed to obtain the correlation coefficient matrix and plotted a heatmap for visualization, with the results as follows:

Figure 4. Heatmap of Spearman Correlations among Categories.

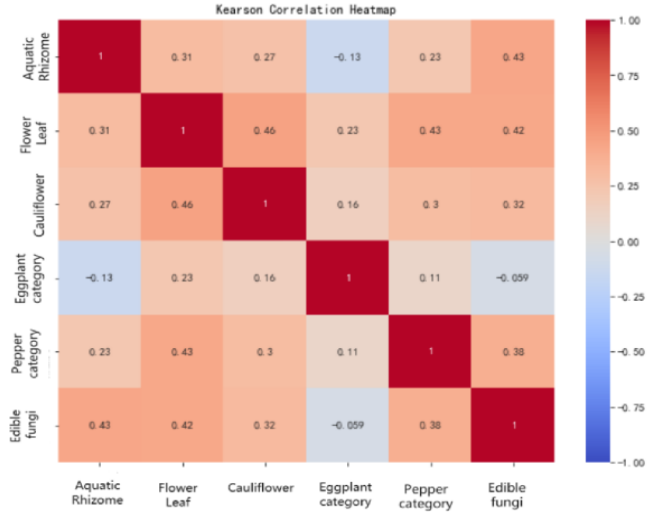


Figure 5. Heatmap of Pearson Correlations among Categories.

The corresponding correlation coefficients and correlation distributions are shown in Table 1:

Table 1. Distribution of Correlation Strength.

Very weak or no correlation	Weak correlation	Moderate correlation	Strong correlation	Very strong correlation
0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0

Based on Table 1 as well as Figures 4 and 5, it is evident that despite differences in the correlation coefficient values, they all indicate the same trend. The sales of eggplants are negatively correlated with those of edible fungi and aquatic root and stem vegetables, but the negative correlation with edible fungi is less pronounced, which aligns with the preliminary patterns observed in Figure 4. Leafy vegetables show a certain degree of positive correlation with other categories, while the correlation between eggplants and other categories remains relatively low.

2.3. Distribution of Individual Product Sales

Given the large number of individual products, many of which appear infrequently and have low sales volumes, this paper will initially select the top 20 and top 50 products by sales volume for visualization analysis.

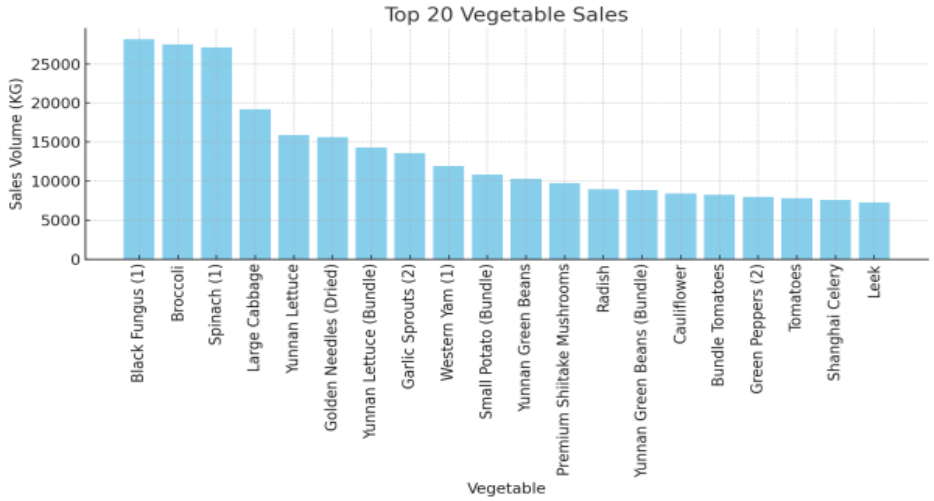


Figure 6. Top 20 Products by Sales Volume.

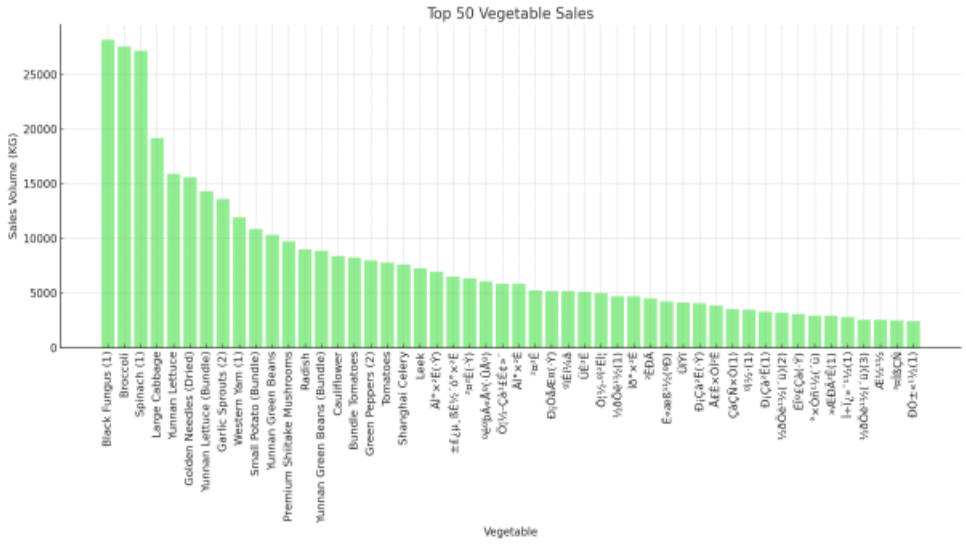


Figure 7. Top 50 Products by Sales Volume.

From Figures 6 and 7, it can be observed that the rate of change in the ranking of individual product sales slows down gradually from highest to lowest. Among these, leafy vegetables dominate, with 9

out of the top 20 products belonging to the leafy vegetable category, and 25 out of the top 50 products falling into this category. Additionally, boxed products outperform bagged and bulk products in terms of sales.

To analyze the trend in vegetable sales, a monthly cycle is selected to visualize the sales volume trends of the top 10 vegetable products, as shown in Figure 8:

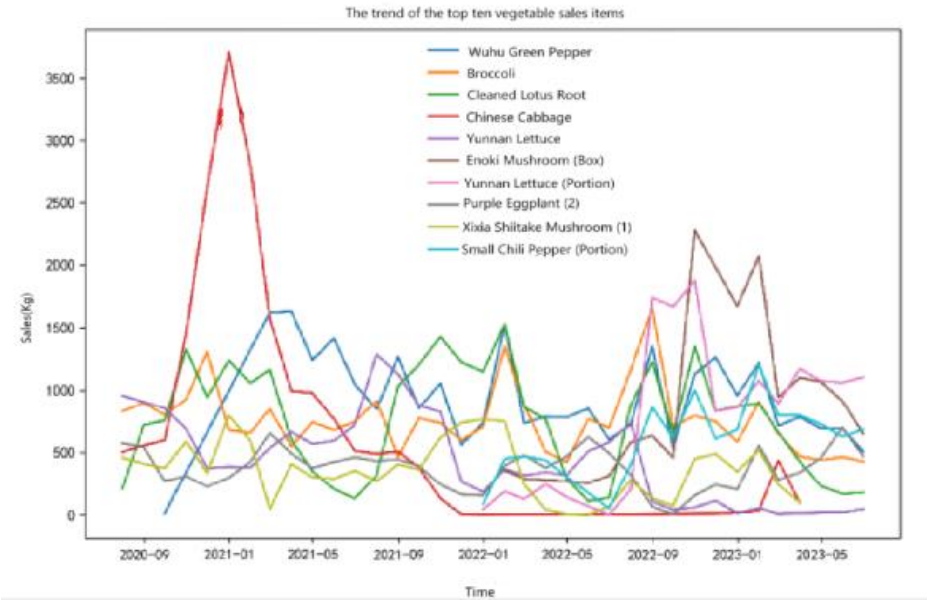


Figure 8. Sales Trend of the Top 10 Vegetable Products.

Based on Figure 8, it is evident that the sales of Chinese cabbage saw a significant increase in the first quarter of 2021 but gradually declined thereafter, reaching very low levels throughout 2022. Within the same category, individual products may influence each other's sales due to intra-species competition, where a decline in the sales of one product might lead to an increase in another. Further detailed data analysis is necessary to gain deeper insights.

2.4. Interrelationships Among Individual Product Sales

First, a Kendall's Tau test was conducted on the individual product data, yielding a p-value of 0.0, which is below the significance level, indicating the presence of correlation. Based on the characteristics of the sales data for individual vegetable products, this paper used the Spearman correlation coefficient to depict the relationships between them.

Since the Spearman coefficient only reveals pairwise correlations between individual products, this paper attempted to perform clustering analysis based on the Spearman coefficient to identify correlations among multiple products.

Observing that individual products exhibit graph-like features under the Spearman coefficient, this paper chose to employ spectral clustering for analysis. Spectral Clustering is a clustering algorithm based on graph theory, widely used in the fields of data mining and machine learning. The main idea is to treat data samples as nodes in a graph and use spectral analysis of the graph for clustering. The specific steps include: first, constructing a similarity graph by calculating the similarity between data samples (e.g., Euclidean distance or Gaussian kernel function), and then constructing a similarity graph based on the similarity matrix, where nodes represent data samples and edge weights represent the similarity between samples; next, calculating the Laplacian matrix from the similarity graph, with both normalized and unnormalized Laplacian matrices being commonly used; then, performing eigenvector decomposition on the Laplacian matrix to obtain eigenvectors and their corresponding eigenvalues, selecting the leading eigenvectors as clustering features; subsequently, using these eigenvectors as inputs, clustering the data with algorithms like K-Means, where the number of clusters is typically specified by the user; finally, obtaining the cluster labels for each data point, thereby completing the clustering analysis.

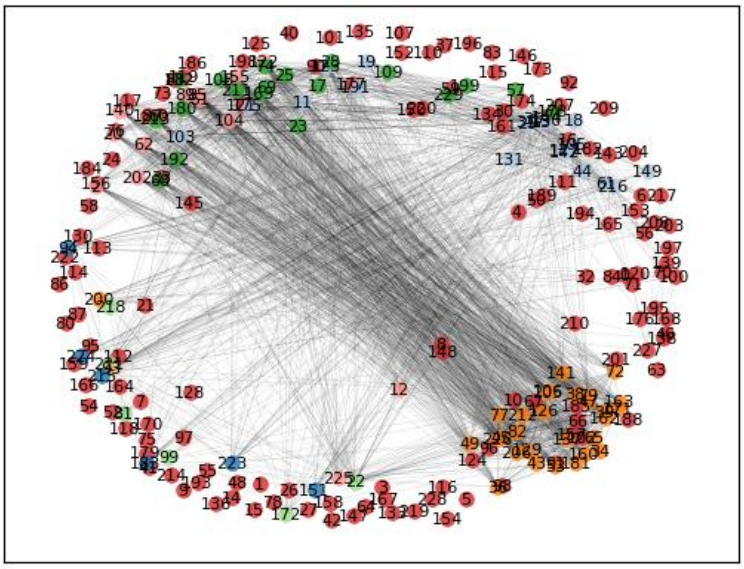


Figure 9. Visual Representation of Spectral Clustering Results.

Set all Spearman coefficient absolute values less than 0.8 to zero, and focus on analyzing the products with strong correlations. After tuning, the clustering results are most satisfactory when the number of clusters is set to eight. The spectral clustering results are shown in Figure 9.

From the clustering results, the following conclusions can be drawn: In clusters with a larger number of members, the vegetables from the six major categories are generally covered, indicating that customers tend to purchase vegetables across different categories. In clusters with fewer members, such as fresh rice dumpling leaves and black porcini mushrooms, it was found that black porcini mushrooms appear simultaneously in another cluster, indicating a strong negative correlation between fresh rice dumpling leaves and other individual vegetable products in that cluster. Additionally, the same product with different packaging appears in two different clusters, demonstrating a negative correlation between them, meaning they are in a competitive state in the market.

3. Multivariable Regression Model for Predicting Pricing and Replenishment

3.1. Data Preprocessing

The markup rate plays a crucial role in the cost-plus pricing method [10]. Cost-plus pricing refers to setting a product's price to compensate for production and sales costs while ensuring a reasonable return. This involves multiplying the original cost by a coefficient to adjust the cost price, making it more reasonable amid fluctuations.

The general calculation formula is as follows:

$$P_i' = C \times (1 + \omega) \tag{4}$$

$$\omega_i = \frac{P_i - f_i}{f_i} \times 100\% \tag{5}$$

Where ω_i represents the markup rate, P_i represents the category's selling price, and f_i represents the wholesale price.

In the data, the "edible fungi" category was selected, and the "sales volume" was summed and accumulated based on the sales date. The "sales price (yuan/kg)", "wholesale price (yuan/kg)", and "loss rate (%)" were averaged, and the markup rate was calculated for each day.

3.2. Training and Prediction of the Regression Model

Regression analysis is a statistical modeling method used to study and quantify the relationship between dependent variables (targets) and independent variables (predictors). This technique is widely applied in various fields, including forecasting future events, analyzing time series data, and

exploring causal relationships between different variables. The model fits a curve to the discrete data points, and a series of evaluation metrics are used to assess the quality of the training.

In this paper, the data from existing categories were used to predict the procurement and pricing strategies for the upcoming week. Given the large number of independent variables, including pricing, purchase price, sales volume, etc., and since this paper is predicting a single dependent variable, it aligns well with the scope of regression models. Therefore, this paper selected a regression model for the analysis.

The choice of regression models includes linear regression, random forest, decision trees, etc., with different models potentially affecting the prediction results. To evaluate the model's performance, the Mean Absolute Percentage Error (MAPE) was used, which is a suitable metric for time series data. A lower MAPE value generally indicates more accurate predictions.

The calculation formula is shown in (6):

$$\text{MAPE} = \frac{\sum_{i=1}^{n} \frac{|\widehat{y_i} - y_i|}{|y_i|}}{n} \tag{6}$$

Using the model functions from the Python `sklearn` library, this paper screened various regression models. By comparing the R^2 and MAPE metrics, this paper comprehensively concluded that:

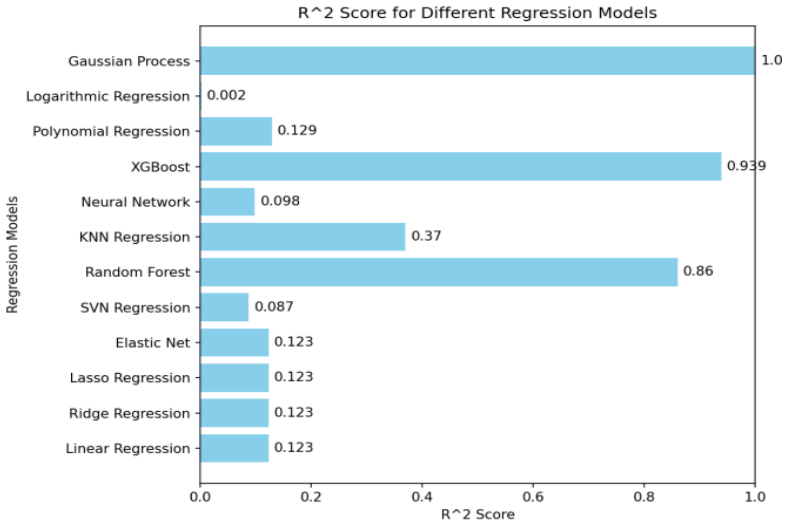


Figure 10. R^2Metrics.

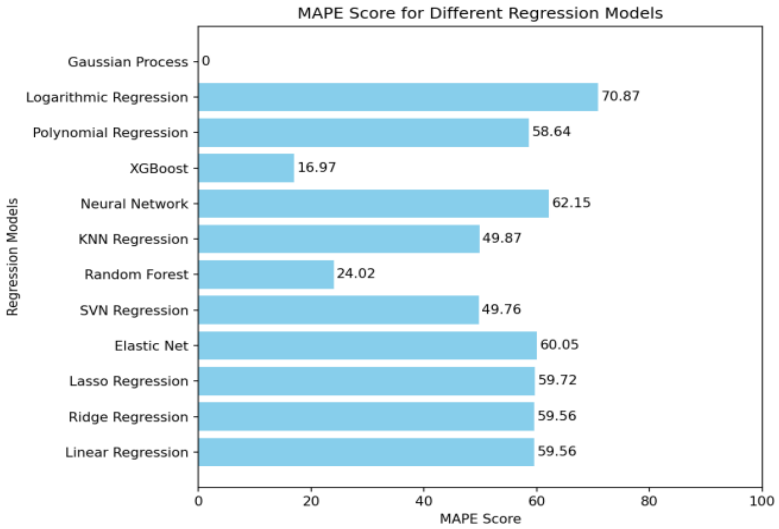


Figure 11. MAPE Metrics.

As shown in Figure 10, the R^2 value obtained using the XGBoost model is 0.939, ranking second and being close to 1. As illustrated in Figure 11, the MAPE score derived from the XGBoost model

is 0.16, which is close to 0. Therefore, choosing the XGBoost model can achieve better predictive results. Although the Gaussian process model exhibits excellent evaluation metrics and outstanding performance, it may suffer from overfitting, which necessitates additional evaluation. The Random Forest algorithm also performed well.

Based on this, the purchase price and selling price of products can be used to predict product sales (XGBoost model). To accurately forecast and formulate pricing strategies, a time series model was chosen, where data from the previous days is used to predict the following day's data, with the process iterating accordingly. The original time series data is transformed into a dataset suitable for supervised learning problems, where each row contains the observation at a given time step as the target and the observations from the previous look_back time steps as features. This generated dataset, as shown in TABLE 2, facilitates subsequent training and calculations.

Table 2. Dataset Generation.

	Y	shiftX_0		Y	shiftX_0
0	7.871938	7.314917	1078	7.596154	7.570375
1	8.120671	7.871938	1079	7.294035	7.596154
2	8.500234	8.120671	1080	7.402807	7.294035
3	7.707923	8.500234	1081	6.268194	7.402807
4	8.557883	7.707923	1082	6.511515	6.268194

Through iterative prediction, the wholesale price forecast for the upcoming week was obtained. The Random Forest regression exhibited favorable characteristics, showing a better fit compared to the Gaussian process. The fitting graph of the Random Forest algorithm is shown in Figure 12, while the fitting graph of the Gaussian process algorithm is presented in Figure 13.

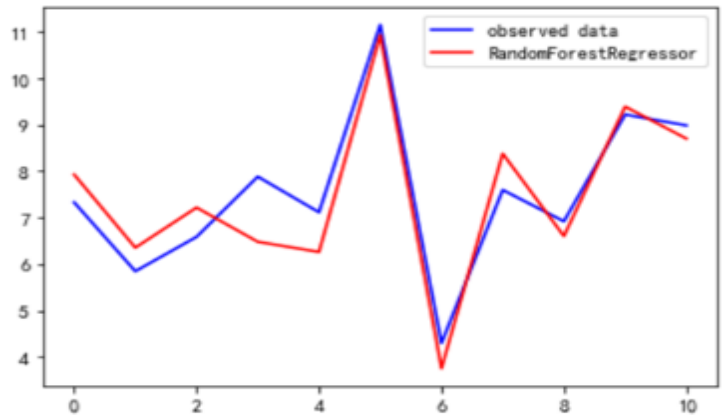


Figure 12. Random Forest Algorithm Fitting.

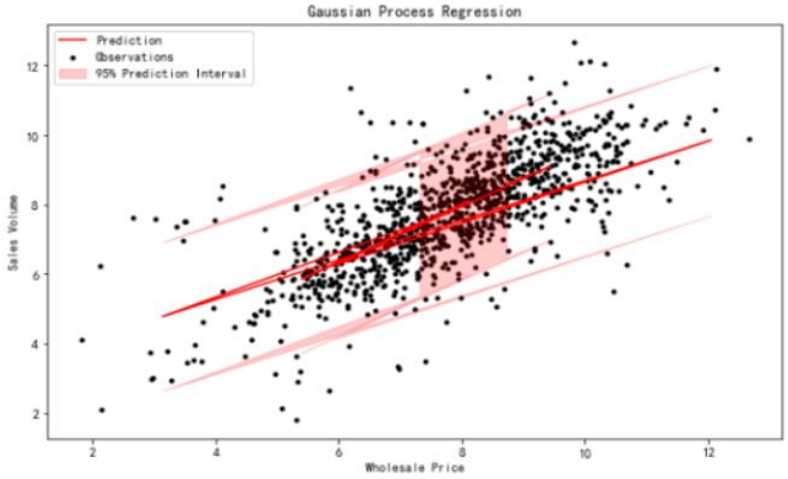


Figure 13. Gaussian Process Fitting.

The accuracy of the data obtained solely through model predictions still needs improvement, so optimization algorithms were selected to enhance the data to achieve maximum profitability. An optimization function was written, and the DEAP genetic algorithm was used to optimize the wholesale price through multiple iterations. For each iteration, a chart was plotted using Matplotlib to show the convergence of the PSO algorithm. The chart's x-axis represents the generation (iteration count), and the y-axis represents the maximum fitness value.

By examining this chart, one can observe whether the algorithm is gradually converging to an optimal solution and how the maximum fitness value changes in each generation. Observing the changes in the y-axis provides insight into the improvement of individual fitness across generations and whether the algorithm is converging to an optimal solution. Generally, as the number of generations increases, the fitness value should gradually increase or stabilize, eventually approaching a maximum value, indicating that the algorithm has found a relatively optimal solution.

The iteration chart for the maximum fitness on the 7th day is shown in Figure 14.

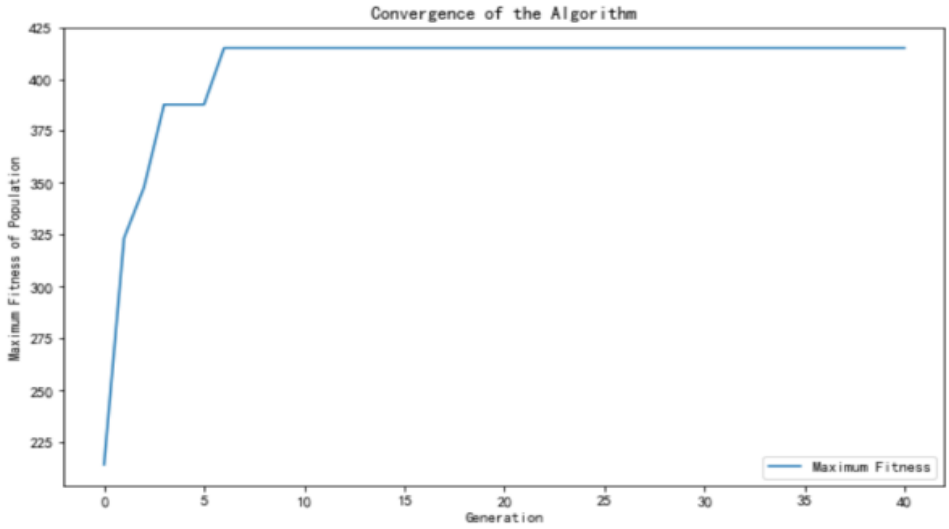


Figure 14. Iteration Chart of Maximum Fitness on Day 7.

Finally, the results were obtained using the profit calculation formula (7):

$$Pf_i = \widetilde{S_i} \times \left(1 - \frac{l_i}{100}\right) \times \left(\widetilde{P_i} - \widetilde{f_i}\right) \tag{7}$$

Partial Prediction Results are shown in Table 3.

Table 3. Predicted Results for Vegetable Types.

Future Day	Selling Price	Sales Volume	Commercial Profit	Category
First day	7.747651	197.4117	354.2348	Chili Pepper
First day	29.45392	278.0027	6305.862	Leafy Vegetables
First day	19.11957	20.00851	136.8952	Aquatic Root Vegetables
First day	12.42675	237.8097	871.1934	Edible Fungi
First day	13.55863	61.61338	300.2304	Cauliflower
First day	7.929897	56.12144	183.5352	Eggplants

4. Model Prediction Under Constraints

First, a regression model needs to be selected to predict sales volume based on the selling price and wholesale price. Unlike the previous operations, this section focuses on the procurement patterns of individual products, and the predicted data should be based on individual products, with both the selling price and wholesale price reflecting the prices of these products.

By using MAPE and R^2 as evaluation metrics, as shown in Table 3, it was found that only the Random Forest and XGBoost models produced values within the normal range and had relatively

high accuracy. Other models, such as the linear regression model, yielded a MAPE greater than 1 and an R^2 close to 0, making them unsuitable for predicting this data, as shown in Table 4.

Table 4. Models and Metrics**.**

Model	MAPE	R^2
Random Forest	0.52558	0.8533329343630891
XGBoost	0.34208	0.9353862995993835
Linear Regression	1.37466	0.05725894987524305
…	…	…

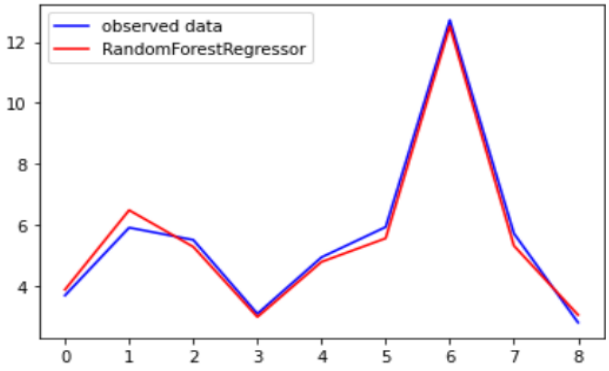
XGBoost continues to be selected as the primary regression model. After obtaining the predicted sales volume for individual products, the wholesale price data is transformed into a dataset suitable for supervised learning. The Random Forest regression model is used for data fitting, and a comparison chart of observed values and predicted values from the model is plotted.

The next time point's sales volume is predicted using a combination of time series models and genetic algorithms. Outliers are calculated using the first quartile (Q1) and third quartile (Q3) and are then removed. The merchant's profits are then calculated using a specific formula and stored in a list for visualization.

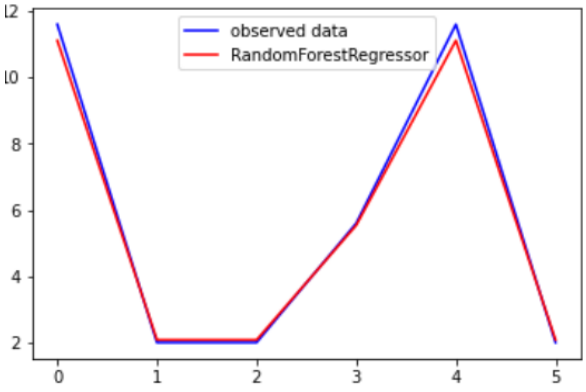
The PSO algorithm's objective function and constraints are set to find the optimal selling price using the PSO algorithm. The optimal selling price, sales quantity, and supermarket profit are output, and finally, a convergence chart for the PSO algorithm is plotted.

Based on previous data, 49 individual products were listed in descending order by their frequency of appearance, and products ranked from 27th to 33rd were selected for model analysis. It was ensured that the sales volume of each product was no less than 2.5 kg.

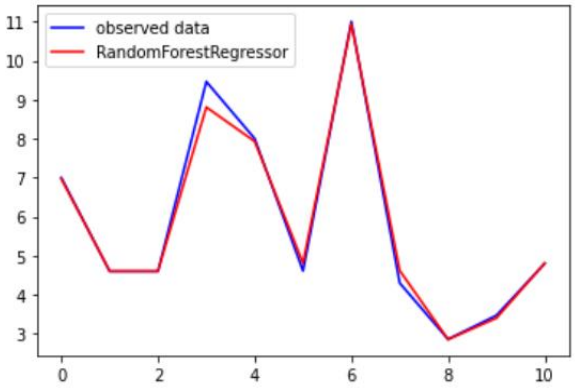
As shown in Figure 15, the results were highly accurate when using Random Forest to fit the test set.



Partial Fitting Chart for Random Forest (a).



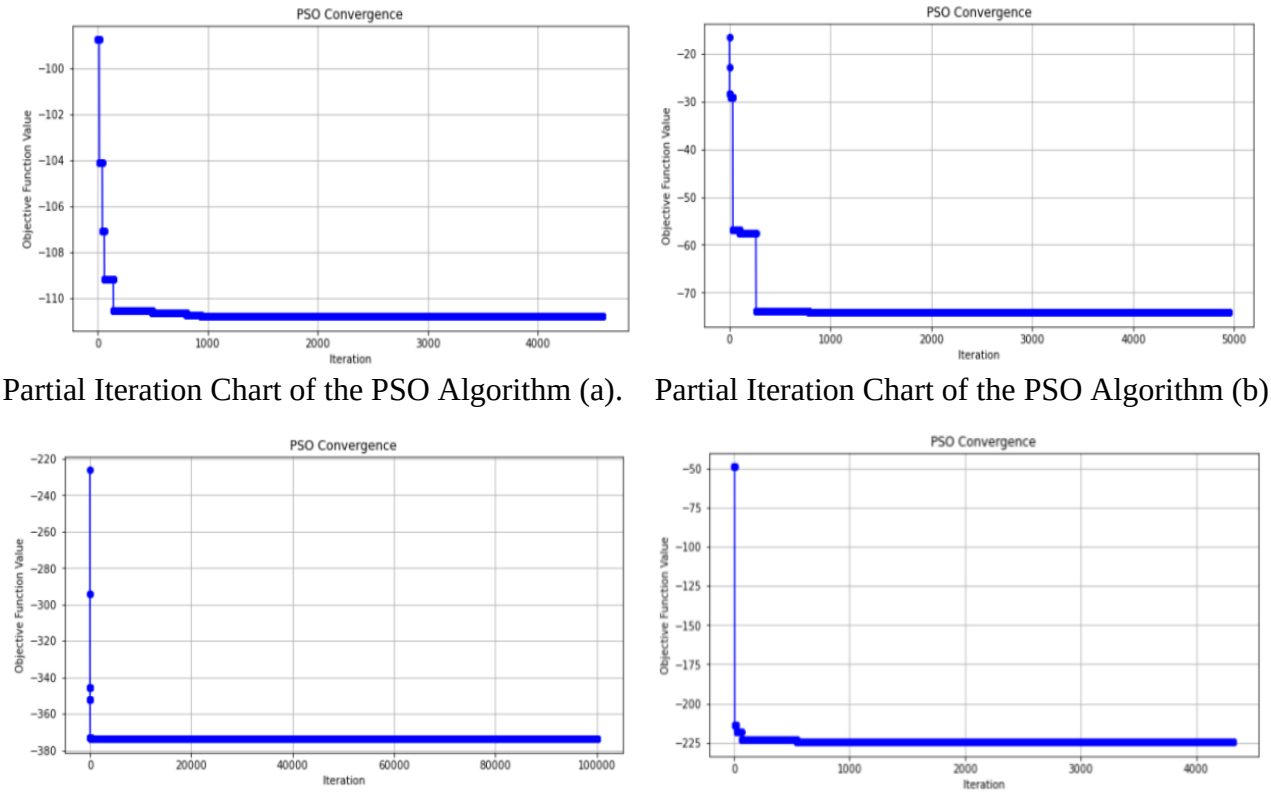
Partial Fitting Chart for Random Forest (b).



Partial Fitting Chart for Random Forest (c)

Figure 15. Fitting Chart for Random Forest.

Partial Results from Data Optimization Using the PSO Algorithm are shown in Figure 16.



Partial Iteration Chart of the PSO Algorithm (a). Partial Iteration Chart of the PSO Algorithm (b).

Partial Iteration Chart of the PSO Algorithm (c). Partial Iteration Chart of the PSO Algorithm (d).

Figure 16. Iteration Chart of the PSO Algorithm.

As shown in Figure 17, after a large number of iterations, the estimated profit of individual products tends to stabilize. Consequently, this paper calculates the total merchant profit for different numbers of individual products and compares the results, as illustrated in Figure 18.

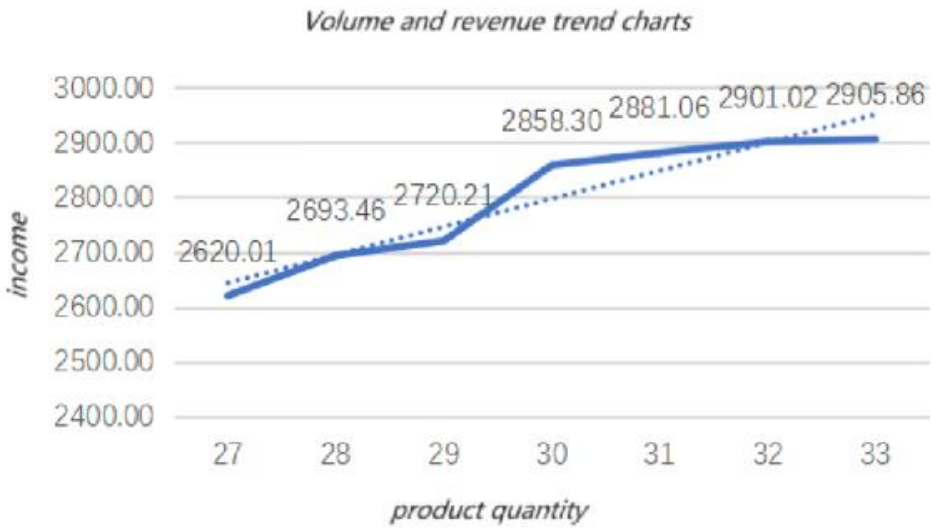


Figure 17. Quantity vs. Profit Trend Chart.

As shown in Figure 18, when the number of individual products reaches 33, the overall predicted profit is the highest. The general trend increases with the number of products, though the growth rate slows down. The results for 33 individual products are partly illustrated in Table 5.

Table 5. Prediction Results for 33 Individual Products**.**

Item ID	Unit Price	Quantity Sold	Revenue
1.03E+14	6.3	41.19	110.80
1.03E+14	13.88	17.17	154.50
1.03E+14	10.05	65.09	373.96
1.03E+14	4.97	30.75	74.19
1.03E+14	13.84	37.08	200.78
1.03E+14	16.99	6.49	58.70
1.03E+14	21.17	45.22	224.84
1.03E+14	5.82	25.51	64.66
1.03E+14	9.93	27.08	189.36
1.03E+14	6.88	54.13	198.86
1.03E+14	44.18	2.77	78.19
1.03E+14	6.99	16.87	51.63
1.03E+14	11.89	24.77	74.77
1.03E+14	23.86	7.03	92.13
1.03E+14	2.85	98.29	136.91
1.03E+14	12.57	9.59	73.44
1.03E+14	8.87	50.08	138.09
1.03E+14	14.94	22.91	94.02
1.03E+14	4.79	41.27	88.21
1.03E+14	6.42	30.36	130.28
Total Revenue			2905.86

5. Conclusion

This paper identified distribution patterns and interrelationships among different categories and individual products using clustering algorithms. It was found that clusters with a larger number of members generally included the six major vegetable categories, indicating customers' tendency to purchase across various categories. However, in clusters with fewer members, such as fresh rice dumpling leaves and black porcini mushrooms, a strong negative correlation was observed, suggesting that black porcini mushrooms appeared in a different cluster. The analysis also examined the relationship between existing sales volumes and pricing by category. Using regression models, this paper predicted replenishment and pricing strategies for the upcoming week, with results summarized in tables and figures. These strategies were formulated based on available products from the previous week, considering constraints on individual product ordering.

The paper employed a multi-model comparison approach, using regression models like XGBoost, Random Forest, Gaussian Process, and clustering, selecting the most suitable model for the problem based on practical needs. This method significantly improved the model's accuracy and was further supported by a comprehensive assessment using multiple evaluation metrics. Additionally, data processing techniques, such as merging data, extracting useful information, and handling outliers using quartiles and IQR, enhanced the model's robustness and predictive capability. The use of various graphical visualization methods, including data comparison charts and algorithm convergence charts, made the model results more intuitive and easier to understand, improving both the model's interpretability and its innovativeness.

References

[1] Lu Y. Research on Dynamic Pricing of High-Quality Fresh Vegetables in Supermarkets in China. Beijing: Beijing Jiaotong University, 2010.

[2] Lu M, Zhang W, Xu T. Optimizable replenishment strategy based on dynamic matrix model. J Comput Eng Appl, 2021, 57(7): 263–268.

[3] Xuan Y, Wan Y, Chen J. LSTM time series classification based on multi-scale convolution and attention mechanism. Comput Appl, 2022, 42(8): 2343–2352.

[4] Wang W, Luo Y, Yang X. Automated pricing and replenishment strategies for vegetable products based on SVR and simulated annealing. Int J Glob Econ Manag, 2024, 2(3): 1–10.

[5] Han J, Gan S. Review of cost-plus pricing method. Finance & Accounting Monthly, 2012, (22): 74–75.

[6] Automatic pricing forecast and replenishment decision of vegetable products based on LSTM model. Acad J Sci Technol, 2023, 58(9): 85–92.

[7] Chen L, Liu F, Wang W. Replenishment and pricing strategy analysis for vegetable products based on time series model. J Econ Manag, 2023, 47(5): 131–137.

[8] Paul R K, Yeasin M, Kumar P. Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. PLoS ONE, 2022, 17(7): e0270553.

[9] Kumari P, Goswami V, Harshith N, Pundir R S. Recurrent neural network architecture for forecasting banana prices in Gujarat, India. PLoS ONE, 2023, 18(6): e0275702.

[10] Hasan M R, Mashud A H M, Daryanto Y, Wee H M. A non-instantaneous inventory model of agricultural products considering deteriorating impacts and pricing policies. Kybernetes, 2021, 50(8): 2264–2288.