Research on Vegetable Pricing and Replenishment Based on LSTM Time Series Prediction Method

Xinyue Feng[#, *], Qihang Xia[#]

Normal School of Vocational Techniques, Hubei University of Technology, Wuhan China, 430000

*Corresponding author: 18071131033@163.com

#These authors contributed equally.

Abstract. In the operation of fresh supermarkets, vegetable products exhibit significant price volatility compared to other commodities due to their perishability, freshness, and susceptibility to losses. This study utilizes historical sales and pricing data from supermarkets and employs the LSTM neural network forecasting model to conduct predictive analysis on the replenishment decisions and pricing strategies in fresh supermarkets. The results demonstrate that the proposed forecasting model achieves favorable predictive accuracy, revealing a certain correlation and regularity between the replenishment decisions in fresh supermarkets and historical time series data. This research offers positive economic benefits in guiding the variations of prices and replenishment quantities.

Keywords: Fresh vegetables; Vegetable pricing and replenishment; Dynamic pricing; LSTM time series.

1. Introduction

The supermarket industry occupies a pivotal position in modern retail. The replenishment and pricing strategies of supermarket goods have significant impact on profitability, market competitiveness, and consumer behavior. A wise set of replenishment and pricing strategies helps to increase the supermarket's sales and profits, thereby enhancing market competitiveness. Therefore, discussing the replenishment and pricing of supermarket goods is of great significance.

In fresh supermarkets, businesses provide different types of vegetables, most of which have characteristics such as perishability, freshness, and loss[1-2]. Their quality and appearance will decline over time, and most unsold varieties cannot be sold the next day[3-4]. Many scholars have also endeavored to use different models to predict vegetables during the retail stage. For example, the ARIMA model is used to predict sales volume and price[5]. Studies have shown that due to the price sensitivity of vegetable products and the impact of freshness throughout the day, the fluctuation amplitude and frequency of prices are greater. Some businesses will deliberately lower or raise the retail prices of products during specific time periods[6]. Furthermore, some scholars have pointed out that the pricing of vegetables generally adopts the "cost plus pricing" method[7-8].

In this paper, based on the historical sales volume and historical sales pricing of various vegetables, we plan to use the LSTM time series prediction method to predict the sales volume with periodicity and time-varying changes[9-10]. Based on the predicted sales volume, in order to maximize profits, factors such as wholesale prices, average loss rates, and pricing are considered to determine the markup rate based on the expected sales volume, and finally, determine the prices of various categories. The accuracy of the model is verified by comparing it with historical transaction data for the same period using stacked line charts. This paper has important practical significance in solving the challenges faced by fresh supermarkets in replenishment and pricing and can bring certain economic benefits.

2. Model Building and Solving

2.1. Data Analysis and Model Building

2.1.1 Cost Markup Pricing

According to the requirements, it is necessary to analyze the relationship between the total sales volume of various vegetable categories and the cost markup pricing. Therefore, it is necessary to calculate the cost pricing first. The calculation formula is as follows:

$$Q = D \times (1 + \alpha) \tag{1}$$

$$D = \frac{G + B}{H} \tag{2}$$

$$\alpha = \frac{W - C}{C} \tag{3}$$

represents the vegetable price, represents the unit cost, presents the markup rate, represents the total fixed cost, represents the total variable cost, presents the sales volume, represents the selling price, and presents the purchase price. Based on the distribution characteristics of cost markup pricing and sales volume data, choose to construct a smooth spline curve fitting model.

2.2. Smooth spline curve fitting model

The smooth spline curve fitting divides the data interval into segments for fitting. Between each data point, a local polynomial is constructed, and the coefficients of this polynomial are determined through optimization problems to minimize the trade-off between fitting errors and smoothness. In order to ensure the smoothness of the fitting curve, a penalty term or control of the polynomial order is typically introduced. The goal of smoothness constraint is to achieve smooth transition of the fitting curve at adjacent data points and reduce curve oscillation.

2.2.1 The penalty formula is:

$$RSS(f, \lambda) = \sum_{i=1}^{N} \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt \tag{4}$$

2.2.2 The free spline formula is:

$$f(x) = \sum_{j=1}^{N} N_j(x) \theta_j \tag{5}$$

2.3. Model Solution

Based on the statistics of product information and sales revenue details with wholesale prices for 6 vegetable categories in a local supermarket from July 1, 2020 to June 30, 2023, smooth spline curve fitting graphs for cost markup pricing and sales volume for each vegetable category were plotted using Matlab and the results are shown in Figures 1, 2, 3, 4, 5, 6.

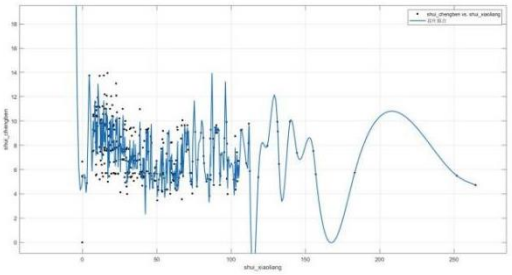


Figure 1 Leafy vegetable fitting graph

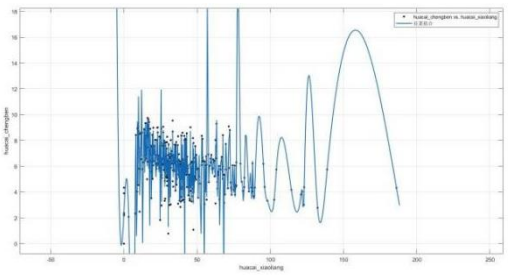


Figure 2 Cauliflower fitting graph

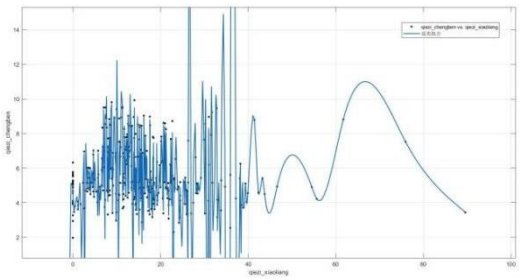


Figure 3 Leafy vegetable fitting graph

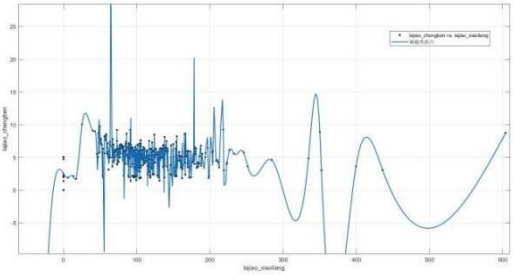


Figure 4 Cauliflower fitting graph

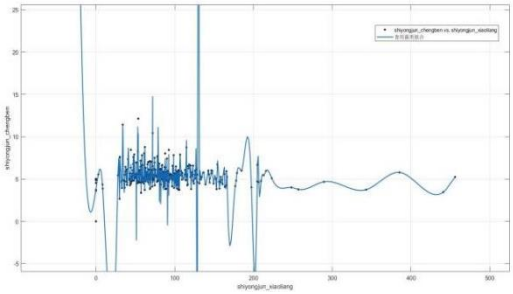


Figure 5 Edible fungus fitting graph

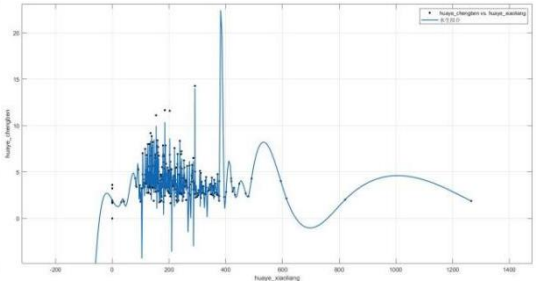


Figure 6 Aquatic root and stem fitting graph

From the analysis of the six graphs above, it can be seen that the sales volume of each category and the cost mark-up pricing is greater than 0.6, indicating a good fitting effect. It can also be observed that the sales volume of each category and the cost mark-up pricing exhibit a certain linear relationship. Additionally, based on the trend changes in the fitting graph, it can be further determined that there is a negative correlation between the sales volume of each category and the cost mark-up pricing, but the negative correlation effect is very weak.

According to the preprocessing of the survey data, the total sales volume of the six major vegetables exhibits a cyclical trend. It is necessary to predict the replenishment quantity of each category, therefore, for sales volume with periodic and time-varying changes, an LSTM time series forecasting method is chosen for solving.

2.4. LSTM Time Series Forecasting

When processing information, LSTM achieves fine control of the state C through the coordinated action of the forget gate, input gate, and output gate. At each time point, LSTM can precisely determine how much old information should be retained, how much new information should be absorbed, and how much content should be output based on the current input and the previous time step's state. This mechanism ensures that important historical information is not forgotten, and at the same time, it avoids excessive attention to recent non-critical information, thus achieving precise control of information processing.Its processing is shown in Figure 7:

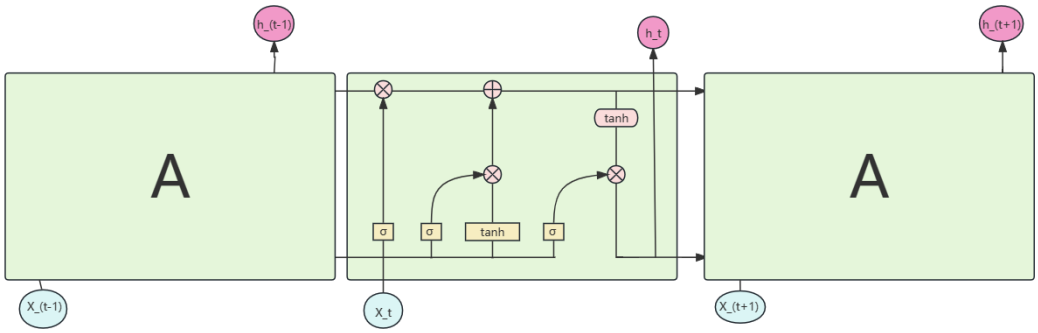


Figure 7 LSTM time flowchart

2.4.1 LSTM Related Formulas

2.4.1.1 The formula related to the forget gate is:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{6}$$

2.4.1.2 The formula related to the input gate is:

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{7}$$

$$C_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{8}$$

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{9}$$

2.4.1.3 The formula related to the output gate is:

$$k_t = \sigma\left(W_k \left[h_{t-1}, x_t\right] + b_k\right) \tag{10}$$

$$h_t = k_t * \tanh\left(C_t\right) \tag{11}$$

2.4.2 Data Input

LSTM does not have a strong requirement on the length of the input time series. The longer the time series length, the longer the learning opportunity and time, and the more accurate the prediction will be. Therefore, the total sales volume data of 1095 for each category is input into the LSTM time series forecasting model. 80% of the data is used as the training set, and 20% of the data is used as the test set. Finally, 1095 data points are used to predict the total sales volume for the last week.

2.4.3 Model Solution

By inputting the training set data into the LSTM time series forecasting model and running it using Matlab, the daily sales volume prediction graphs and the predicted seven-day sales volume curve marked with a red box for various types of vegetables are obtained, as shown in Figure 8, 9, 10, 11, 12, and 13:

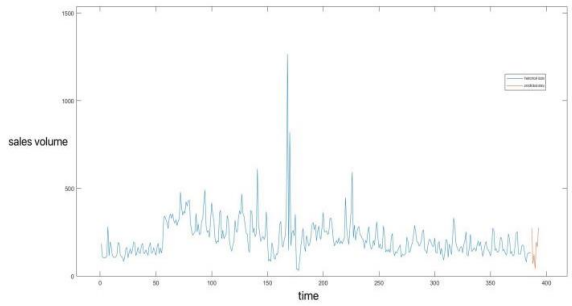


Figure 8 Floral leaf

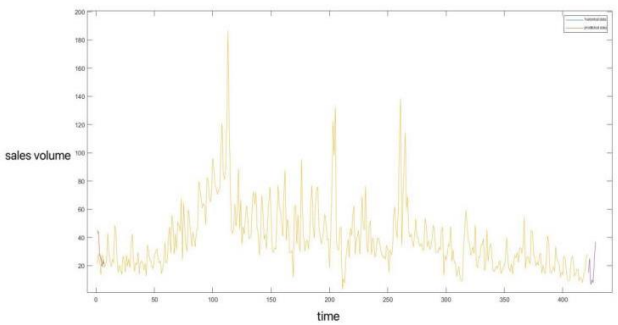


Figure 9 Cauliflower

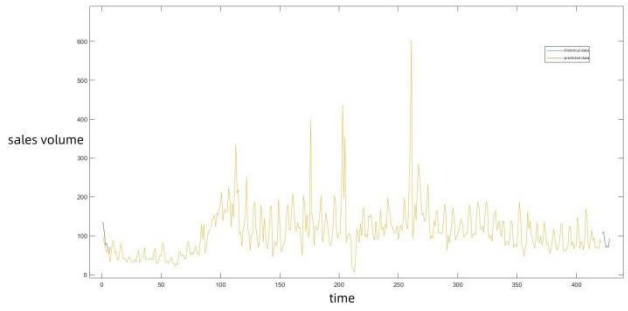


Figure 10 Pepper

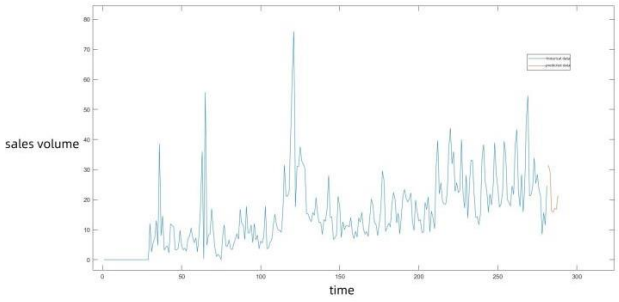


Figure 11 Eggplant

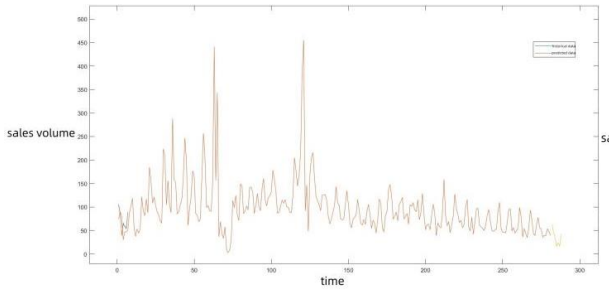


Figure 12 Edible mushrooms

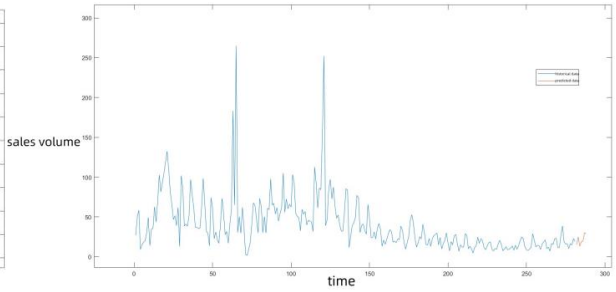


Figure 13 Aquatic rhizome

From the daily sales volume prediction graphs of various types of vegetables, the replenishment quantity table for each category of vegetables can be obtained, as shown in the following table 1:

Table 1 Forecasted Replenishment Quantity for Various Categories of Vegetables

date	floral leaf	cauliflower	pepper	eggplant	edible mushrooms	Aquatic rhizome
2023-07-01	222.74	15.31	105.38	31.46	62.90	15.54
2023-07-02	161.07	25.03	110.66	29.29	46.84	24.76
2023-07-03	47.55	6.56	86.09	16.23	35.20	13.37
2023-07-04	73.25	10.07	69.06	15.86	16.07	19.20
2023-07-05	130.61	7.95	74.75	17.06	23.41	19.35
2023-07-06	128.11	22.55	69.85	16.71	16.87	30.31
2023-07-07	151.63	36.91	92.86	21.29	42.90	28.77

Draw a line graph for the forecasted replenishment quantity for various categories of vegetables, as shown in Figure 14:

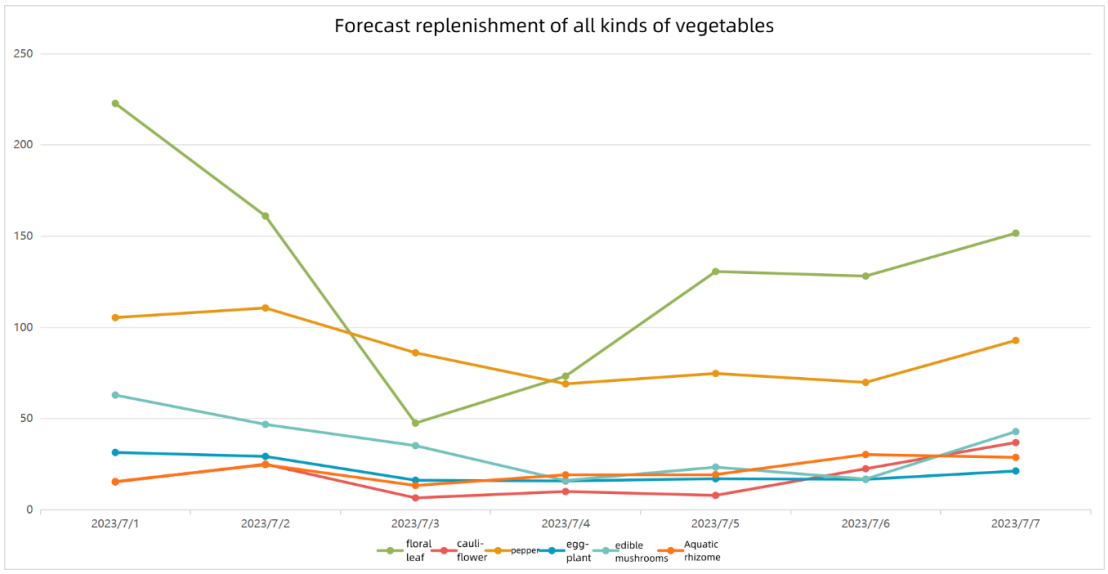


Figure 14 Forecast replenishment of all kinds of vegetables

From Figure 14, it can be observed that, in the upcoming week, the replenishment quantity of Floral leaf will decrease from the 1st to the 3rd, then rise from the 4th to the 7th, sometimes with a longer rising period, showing a trend of increase followed by decrease; the replenishment quantity of peppers will slightly increase from the 1st, but then exhibit a decreasing trend or relatively stable change from the 2nd to the 6th, although there is a slight increase in replenishment quantity on the 7th, the magnitude is not large; the replenishment quantity of edible mushrooms will continue to decrease from the 1st to the 4th, remain relatively stable from the 4th to the 6th, and show some increase by the 7th, although the magnitude of change is still not significant; the categories such as cauliflower, eggplant, and aquatic rhizome show minimal fluctuation, remaining relatively stable.

3. Results and Discussion

3.1. Forecast Results

Based on the forecasted sales volume, in order to maximize profits, we need to consider wholesale prices, average loss rates, and pricing, with the specific steps as follows:

Step 1) Calculate the cost for each category, the calculation formula is as follows:

$$D = C \times \left(1 + \frac{\theta}{100}\right) \tag{12}$$

Where, D is the cost, C is the wholesale price, and θ is the loss rate.Table 2 shows the cost list for each category.

Table 2 Cost Price Table for Each Category

Vegetable category	Cost price(￥)
Aquatic rhizome	12.13
Floral leaf	5.34
Cauliflower	8.47
Eggplant	6.17
Pepper	7.41
Edible mushrooms	7.30

Step 2) Determine the markup rate based on expected sales volume

To explore the distribution patterns of various vegetable categories, this study first calculates the total sales volume of the six major vegetable categories and then draws the corresponding histogram based on the calculation results, as shown in Figure 15:

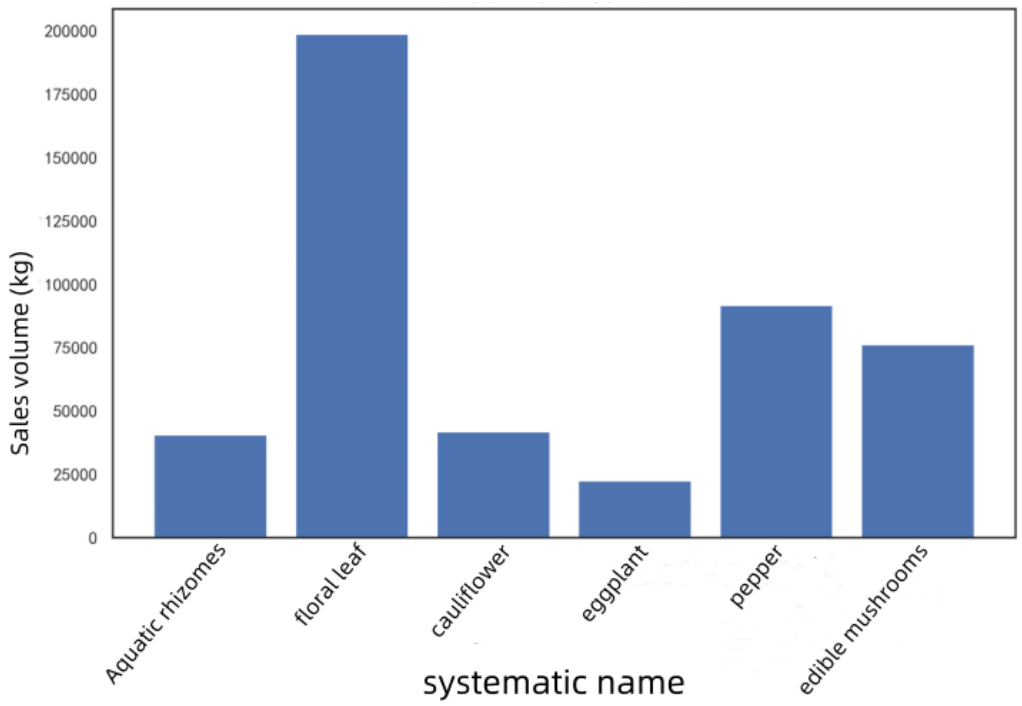


Figure 15 Distribution of Sales Volume for Various Vegetable Categories

From the above figure, the ranking of sales volume for various vegetable categories is as follows:Floral leaf>Pepper>Edible mushrooms>Cauliflower≈Aquatic rhizome>Eggplant, with the sales volume of floral leaf significantly higher than other vegetable categories; the sales volume of pepper and edible mushrooms is relatively close; the sales volume of cauliflower and aquatic rhizome is approximately equal; and the sales volume of eggplant is roughly in a 1:2 proportion with cauliflower.

Next, further proceed by performing the following operations on SPSS, as illustrated in the process flow in Figure 16:

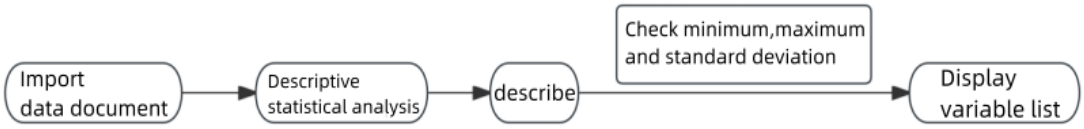


Figure 16 SPSS operation flow chart

The minimum, maximum, median, mean, and standard deviation of the sales volume for the six major vegetable categories can be separately calculated, resulting in the following results as shown in Table 3.

Table 3 Descriptive Statistics Table for Six Major Vegetable Categories

	Minimum statistics	Maximum value statistics	The median	Average statistics	Standard deviation statistics
Floral leaf	31.30	1265.47	173.192	182.97	86.20
Cauliflower	0.63	186.16	34.0725	38.53	22.68
Aquatic rhizome	0.93	296.79	31.831	37.40	31.36
Edible mushrooms	3.01	511.14	57.535	70.13	48.49
Eggplant	0.25	118.93	18.757	21.36	13.16
Pepper	6.07	604.23	72.925	84.41	53.44

By analyzing the sales volume data, we can use the minimum value to understand the lower limit of sales volume, which helps to understand the worst sales situation; the maximum value can be used to understand the upper limit of sales volume. Floral leaf have the highest overall sales volume, with the highest maximum value statistic, the highest sales volume upper limit, and the minimum value statistic is the largest among the minimum values of the six major vegetable categories, so its sales volume lower limit is the highest among the six major vegetable categories, but the standard deviation value is the largest, so the overall sales volume data has the greatest fluctuation and may have a greater degree of uncertainty; while the overall sales volume of eggplant is the lowest, with the lower and upper limits of sales volume being the lowest, and its standard deviation is the smallest, so the overall sales volume data has the smallest fluctuation, and its sales volume is relatively stable.

According to Table 3, the descriptive statistics table of the six major vegetable categories, it can be seen that the median values of floral leaf, cauliflower, aquatic rhizome, edible mushrooms, eggplant, and peppers are 173.192, 34.0725, 31.831, 57.535, 18.757, and 72.925, respectively. Combined with the predicted values, the sales volume is divided into three intervals. In order to maximize profits under the condition of determining the predicted sales volume, it is necessary to sell low-sales products at high prices to increase the unit price, increase the price of medium-priced products relatively low, and sell high-sales products at low prices to increase sales volume, with the lowest price increase rate greater than the historical average profit rate. The detailed premium rates for various categories of vegetables are shown in Table 4:

Table 4 Price Increase Rate Table for Each Vegetable Category Based on Expected Sales Volume Determination

Vegetable category	Expected sales（kg)	Price increase rate
Floral leaf	h< 100	140%
	100< h<150	130%
	h>150	110%
Cauliflower	h<10	140%
	10<h<20	130%
	h>20	110%
Pepper	h<70	140%
	70<h<90	130%
	h>90	110%
Eggplant	h<20	140%
	20<h<30	130%
	h>30	110%
Edible mushrooms	h<20	140%
	20<h<40	130%
	h>40	110%
Aquatic rhizome	h<20	140%
	20<h<30	130%
	h>30	110%

3.2. Determination of Prices for Each Category

The price calculation formula is:

$$W = D \cdot \beta \tag{13}$$

Where W is the selling price, D is the cost, and β is the markup rate.

The pricing for each category is calculated according to the formula and shown in Table 5：

Table 5 Pricing for Each Category

date	floral leaf	cauliflower	pepper	eggplant	edible mushrooms	Aquatic rhizome
2023-07-01	6.55	9.86	9.65	6.52	8.54	16.53
2023-07-02	6.32	9.88	9.84	6.47	9.56	15.24
2023-07-03	5.69	8.65	8.52	5.62	10.56	15.36
2023-07-04	4.58	7.44	10.65	5.45	12.56	14.51
2023-07-05	6.54	10.56	11.25	6.98	10.45	14.23
2023-07-06	5.23	9.64	9.45	6.52	12.35	13.25
2023-07-07	6.68	11.23	8.23	6.45	8.35	15.66

3.3. Validate the accuracy of the forecast

To validate the accuracy of the predictions, we compared the sales volume data for the past three years and this year using stacked line charts. The results are shown in Figures 17, 18, 19, 20, 21, and 22：

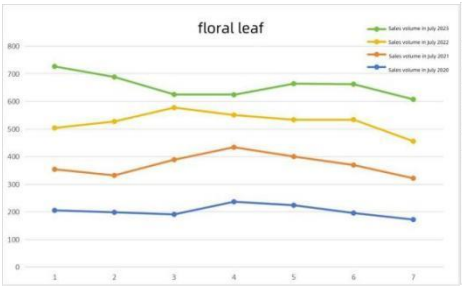


Figure 17 Floral leaf

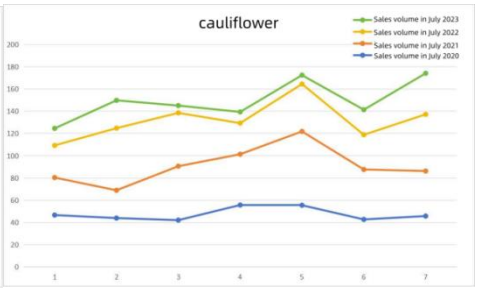


Figure 18 Cauliflower

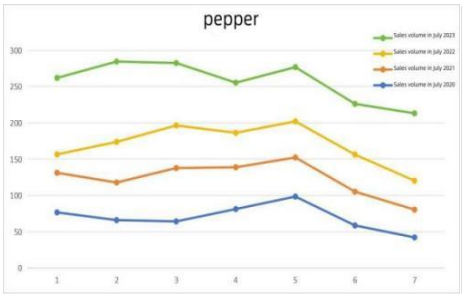


Figure 19 Pepper

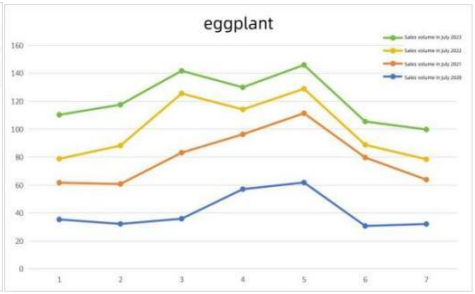


Figure 20 Eggplant

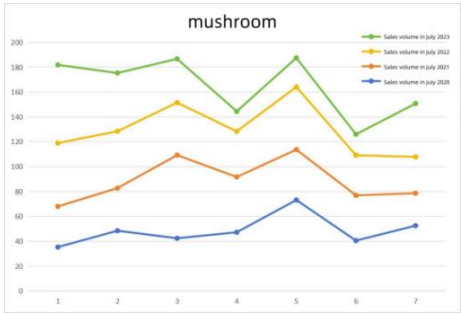


Figure 21 Mushroom

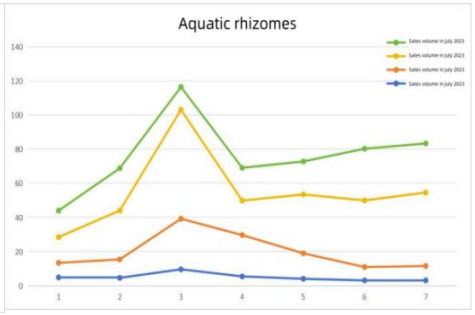


Figure 22 Aquatic rhizomes

After a detailed analysis of the stacked line graphs of the six categories of vegetables, it was found that during the same time period, although the total sales amounts for each year were different, the overall trend remained consistent. This regular trend indicates that the peaks and valleys of the total sales amount showed a high degree of similarity. Based on this consistency, we inferred that the fitting effect is good and the prediction accuracy is relatively high.

4. Conclusion

This paper uses LSTM time series prediction model to forecast replenishment and pricing of vegetable categories in supermarkets. The results demonstrate that there is a certain correlation and regularity between the replenishment decisions of the fresh supermarket and historical time series data. At the same time, there is a positive correlation between pricing and sales volume, meaning that the pricing of vegetable varieties with higher sales volume will correspondingly increase. The comparison of past data also verified the accuracy of the LSTM model in predicting sales trends, providing important basis for forecasting and optimizing replenishment strategies. This paper has important practical significance for addressing the challenges faced by fresh supermarkets in replenishment and pricing, and can bring about certain economic benefits.

References

[1] Pan Lin,Xu Xiajing,Zhou Rongting.Research on Dual Channel Dynamic Pricing of Community Fresh Food Supply Chain from the Perspective of Competition[J/OL].Chinese Journal of Management Science,2021,1506:1-12.

[2] Yu Z .Vegetable income insurance pricing: Based on semi-parametric Copula method[J].Research of Agricultural Modernization, 2019,40(2): 308-315.

[3] McLaughlin E W. The dynamics of fresh fruit and vegetable pricing in the supermarket channel[J]. Preventive Medicine, 2004, 39: 81-87

[4] Qin Y, Wang J, Wei C. Joint pricing and inventory control for fresh produce and foods with quality and physical quantity deteriorating simultaneously[J]. International Journal of Production Economics, 2014, 152: 42-48.

[5] Shukla M , Jharkharia S .Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation[J].International Journal of Information Systems and Supply Chain Management (IJISSCM), 2013, 6(3):105-119.

[6] Bi Wenjie,Zhou Yubing.Research on inventory control and dynamic princing of fresh produce based on deep reinforcement learning[J].Application Research of Computers, 2022,39(09): 2660-2664.

[7] Kartika E , Bakhtiar M R .Penentuan Cost of Good Sold dan Penerapan Cost Plus Pricing Method dalam Menentukan Harga Jual: Study Penggilingan Padi UD Budi Luhur[J]. 2020,10(2):67-73.

[8] Han Junhua,Gan shengdao.Review of Cost-Plus Pricing Method[J].Finance and Accounting Monthly, 2012, 22(No.626):76-77.

[9] Wang X , Wu J , Liu C ,et al.Exploring LSTM based recurrent neural network for failure time series prediction[J].Beijing Hangkong Hangtian Daxue Xuebao/Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4):772-784.

[10] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, 2019, 31(7): 1235- 1270.