Research on automatic pricing and replenishment decision model based on vegetable products

Yang Liu*, [1], Xi Hu[1], Jinyang Wang[2]

[1]School of Computer Science and Engineering, Anhui University Of Science & Technology, Huainan, China, 232001

[2]School of Mechanics and Optoelectronic Physics, Anhui University Of Science & Technology, Huainan, China, 232001

*Corresponding author: ly1026yws@163.com

Abstract. This paper studies and analyzes the pricing and replenishment decision-making problems in vegetable sales management, and discusses the relationship between the sales volume of various categories and single products of vegetables. After Morlet wavelet analysis, it is found that the sales volume changes in a seasonal cycle. This paper chooses Spearman correlation coefficient for correlation analysis. After using K-Means clustering, Apriori algorithm is used to mine association rules to complete the interpretation of the distribution law and mutual relationship of various categories and single product sales of vegetables. In view of the relationship between the total sales of vegetables and the cost-plus pricing, this paper finds that the distribution relationship between the total sales and the cost-plus pricing is exponential distribution. In addition, a variety of regression analysis algorithms are used for fitting, and the random forest model is selected as the most suitable model. At the same time, based on the seasonal ARIMA algorithm, the demand and cost price in the next week are predicted. In this paper, the genetic algorithm is used to solve the model, and finally the daily replenishment amount and pricing strategy of vegetable category in the next week are obtained.

Keywords: Wavelet analysis; Association rules mining; ARIMA algorithm; Genetic algorithm.

1. Introduction

The scale of China 's fresh food business super industry has a trillion-level market volume. As one of China 's basic consumer goods, with the increase of per capita disposable income and consumption expenditure in recent years, China's fresh retail market has maintained steady growth, and its important position in the national economy has become increasingly prominent[1]. The market size has increased from 3.67 trillion yuan in 2017 to 5.64 trillion yuan in 2021. In the fresh supermarket, due to the variety of vegetables sold and the large number of places of origin, the purchase transaction time of vegetables is generally 3 to 4 o 'clock in the morning[2]. Therefore, merchants must make the replenishment decision of vegetable categories on the same day according to the historical sales and demand of commodities without knowing the specific items and purchase prices. Because the fresh-keeping period of vegetable products is relatively short, the product phase becomes worse with the increase of sales time, and it cannot be sold again every other day. Therefore, the replenishment decision and pricing decision of the supermarket are very important for commodity sales[3]. Only by formulating a reasonable daily vegetable replenishment quantity and sales pricing strategy, can the sales profit be maximized as much as possible.

Most supermarkets usually make replenishment every day according to the sales history and demand of each commodity. Therefore, based on the three-year sales flow data of Shangchao given in question C of the 2023 Mathematical Modeling National Competition[4], this paper analyzes the possible correlation between different categories and different items in vegetable products, as well as the distribution law and mutual relationship of sales volume. Through the data analysis of the total sales volume and cost pricing of vegetable categories, based on the seasonal ARIMA algorithm, the demand and cost price in the next week are predicted, and then the daily replenishment volume and pricing decision of vegetable categories from July 1 to 7,2023 are given, so that the profit of

Shangchao in the next week is the largest. The research results of this paper provide important decision support for vegetable sales management, which is of great significance for improving the sales efficiency and profit of supermarkets[5].

2. Analysis of vegetable commodity replenishment and pricing strategy

2.1. Association rule mining of sales price of vegetable products

In the sales data set, there are a total of six categories and two hundred and fifty-one items from July 1,2020 to June 30,2023, more than eight hundred thousand consumer flow details, a huge amount of data. From the supply side and the demand side, there is a certain correlation between the sales of vegetable commodities and the sales time, and the supermarket is also very important for the combined sales of vegetables. For the association rule mining and analysis of different vegetable products and sales prices, this paper uses continuous wavelet transform time-frequency diagram and histogram to find the periodicity and frequency law of data[6]. This paper analyzes the distribution law of vegetable commodities on a single day and the distribution law of vegetable commodities on each quarter, and considers the influence of holidays, Saturdays and Sundays on the distribution law. There are six categories of vegetables analyzed in this paper : flowers and leaves, peppers, edible fungi, cauliflower, aquatic rhizomes, and eggplants. Firstly, the data visualization is used to obtain the curve of vegetable sales with time and the quarterly distribution of vegetable sales with date, as shown in Fig1 and Fig2. It can be seen from Fig1 that for different categories of vegetables, their daily sales show similar distribution patterns. Each category of vegetables will basically reach the first peak around 10 : 00 daily, and reach the second peak around 18 : 00 daily. Of course, there are also differences : for example, flowers and leaves will reach the peak of daily sales around 10 : 00 every day ; for peppers and edible fungi, the daily sales around 10 : 00 and 18 : 00 were similar. From Fig 2 can see the distribution of vegetable sales with the date quarter.

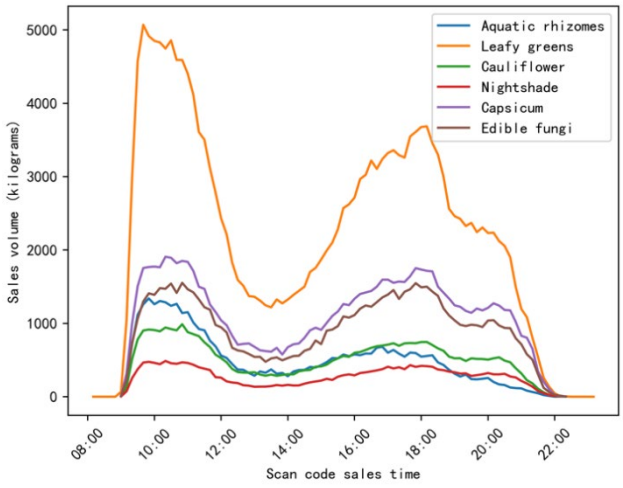

Fig1 Sales over time change chart

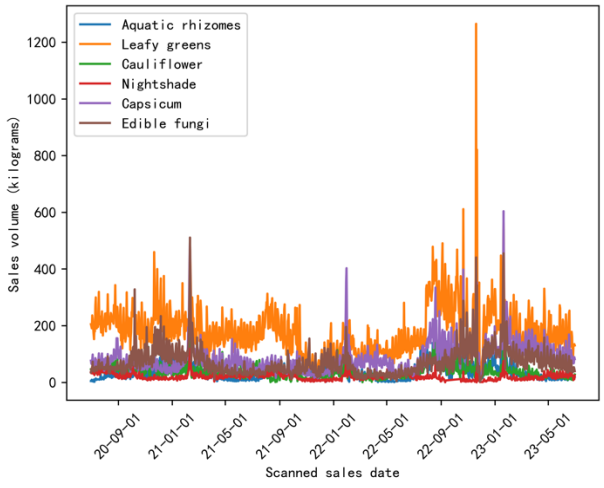

Fig2 Quarterly distribution of sales

Through the following graph analysis, the sales volume of vegetable categories changes periodically with the date. In order to further explore the periodic distribution law of various categories of vegetables, the wavelet analysis is used to analyze various vegetable categories. The results are shown in Fig3-8. It can be concluded that from the sales time of vegetable commodities, the distribution of vegetable categories with the quarter is cyclical[7].

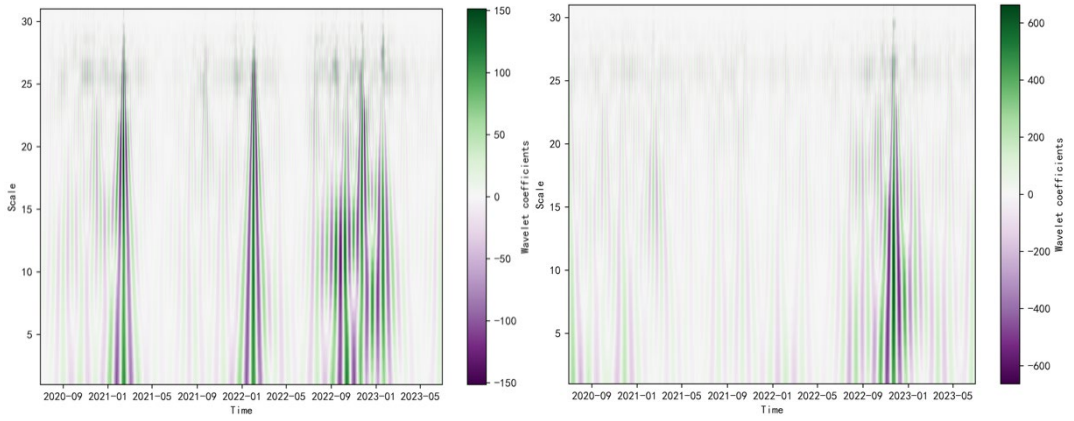

Fig3 Aquatic rhizomes **Fig4** Leafy greens

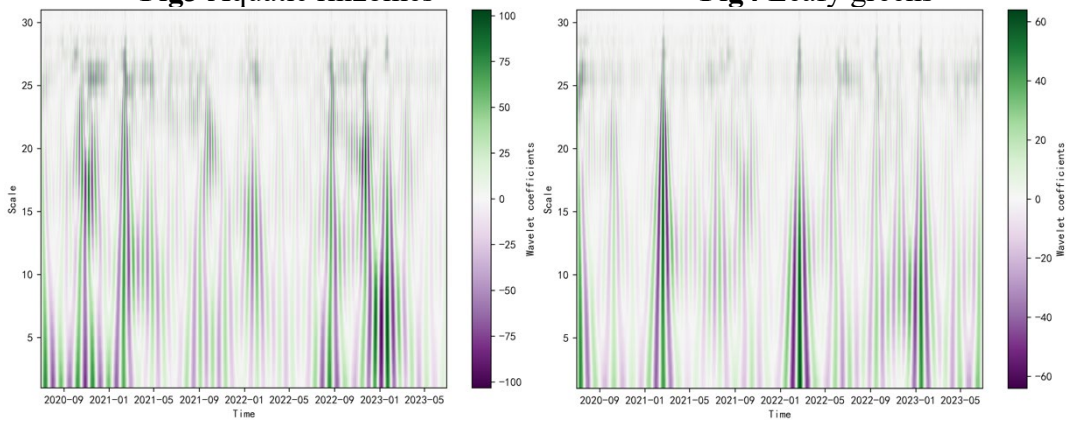

Fig5 Flower vegetables **Fig6** Eggplant

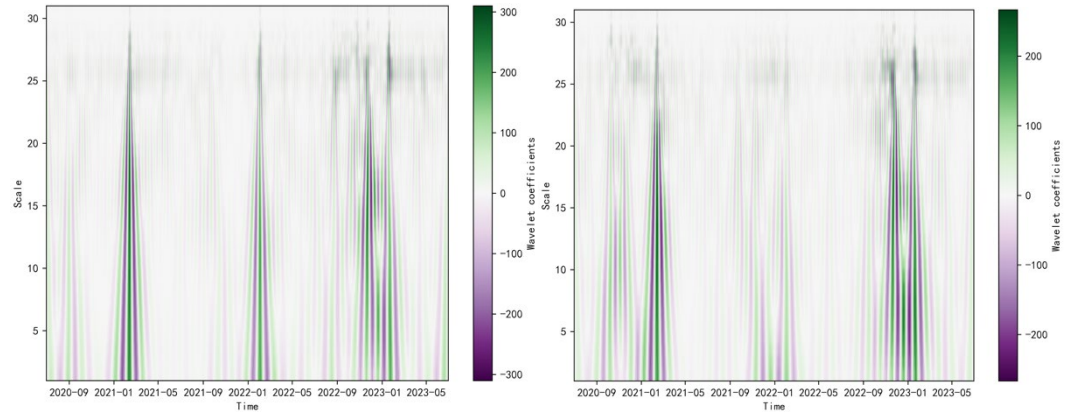

Fig7 Peppers **Fig8** Edible fungi

In order to further explore the possible correlation between different vegetable categories, this paper uses Apriori association rule mining algorithm to analyze the correlation between vegetable categories and single product sales. Apriori algorithm is an association rule mining algorithm. Association rules are rules that describe the potential relationship between data items in the database. The form is as follows[8]:

$$A_1 < A_2 < \ldots < A_m \rightarrow B_1 < B_2 < \ldots < B_n \tag{1}$$

Confidence is the standard for dividing strong association rules / weak association rules, and the specific abstract represents the ratio of the probability of A and B appearing at the same time to the probability of A appearing. In this paper, Apriori algorithm is used to calculate the confidence of different vegetable categories, so as to determine the association rules and interrelationships between them.

After K-Means clustering analysis of the data set, Shapiro-Wilk normality test was performed. However, because the data did not pass the test, it meant that the data did not conform to the hypothesis of normal distribution. Therefore, Spearman correlation coefficient is used for correlation analysis and significance test. The Spearman correlation coefficient formula is [9]:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{2}$$

Confidence intervals for Spearman correlation coefficients are calculated and heat maps for different types of vegetables are obtained, as shown in Fig9 :

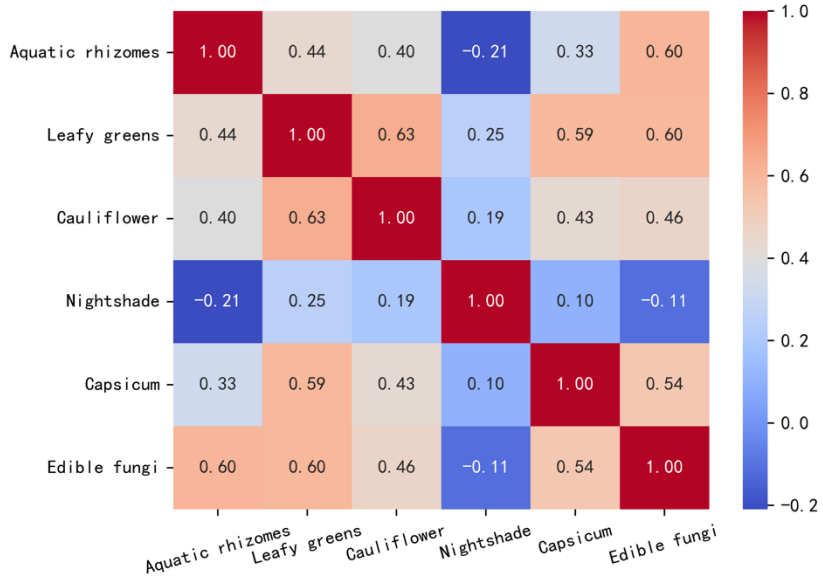

Fig9 Spearman correlation coefficient heatmap

2.2. The model establishment and solution of replenishment and pricing strategy

Based on the knowledge of cost-plus pricing method, this paper analyzes the relationship between sales volume and cost-plus pricing. After data normalization, a variety of different methods are used to fit the distribution of markup and sales volume, and different root mean square errors are obtained. By establishing ARIMA time series model, the total daily replenishment in the next week is determined, and finally a reasonable pricing strategy is given by genetic algorithm.

The basic idea of the cost markup pricing method is the target price of the product = cost base + number.

The formula used is as follows : $Cost_{unit}$ which represents the unit cost, $Cost_{rugular}$ represents the total fixed cost, and $Cost_{change}$ represents the total variable cost[10].

$$Cost_{unit} = \frac{Cost_{regular} + Cost_{change}}{Sale} = Cost_{unit-regular} + Cost_{unit-change} \tag{3}$$

Considering that the cost markup pricing is affected by the discount sales of goods with loss of transportation and poor phase change, this paper calculates the cost pricing and markup rate of a single day by weighted average of the average sales price of a single day according to the sales volume. From this, the corresponding data of the total sales volume and cost markup pricing of each vegetable category can be obtained.

At the same time, linear fitting, Ridge regression, Lasso regression, SVR regression, random forest regression, K-nearest neighbor regression, gradient boosting regression and other methods are used for fitting. The final root mean square error is shown in Table1 :

Table1 The root mean square error of each regression equation algorithm

serial number	name	root mean square error
1	linear regression	0.055765902228400464
2	Ridge regression	0.055765902228400464
3	Lasso regression	0.05747654250354727
4	SVR regression	0.0737563613970582
5	random forest regression	0.024881671337439917
6	k-nearest regression	0.04944392621273388

It can be seen from Table.1 that the random forest regression is the smallest root mean square error in this process, and the random forest analysis can be used as the relationship model between vegetable category sales and cost pricing.

Based on the results of wavelet analysis, it can be seen that the sales volume of different vegetables and single products changes periodically with the seasons, and it can be seen from Table2 that the ADF test of sales volume data shows that the sequence is stable. Therefore, this paper chooses to use the seasonal ARIMA time series model to predict the sales volume in the next seven days.

Table2 ADF detection of various categories of vegetables

serial number	trade price	numerical value
1	Wholesale price of aquatic rhizomes	0.03328
2	Wholesale price of flowers and leaves	0.01634
3	Wholesale price of cauliflower	0.00886
4	Wholesale price of eggplants	0.00000
5	Wholesale price of peppers	0.00014
6	Wholesale price of edible fungi	0.03176

The sales volume and wholesale volume of the next week predicted by ARIMA time series model are shown in Fig10 and Fig11 :

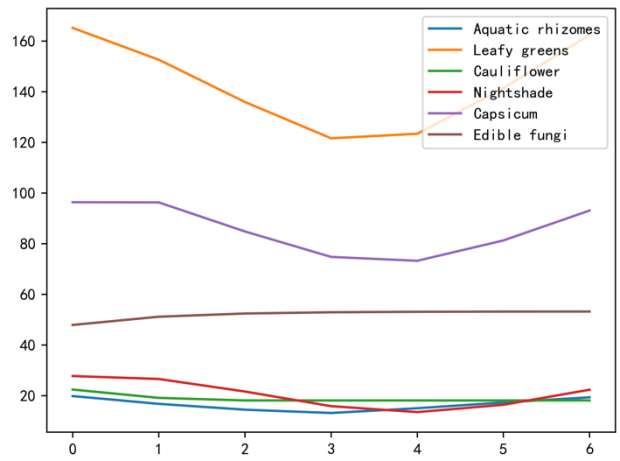

Fig10 7-day sales forecast

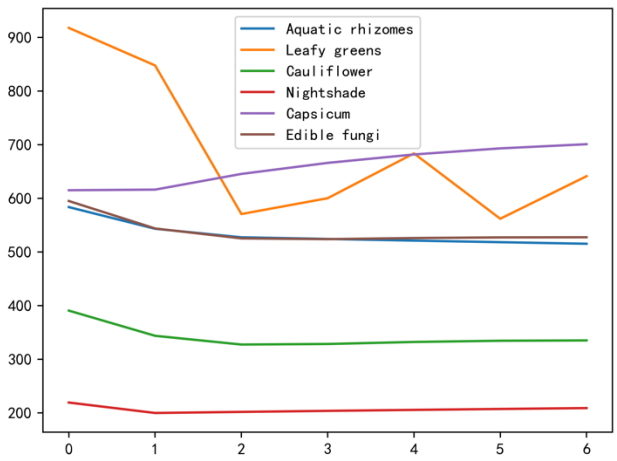

Fig11 7-day wholesale price forecast

After the optimization of genetic algorithm, the interpretation fitness curve of genetic algorithm is obtained, as shown in Fig12: the curve tends to be stable and does not change significantly. It can be considered that the genetic algorithm has found a better solution[11].

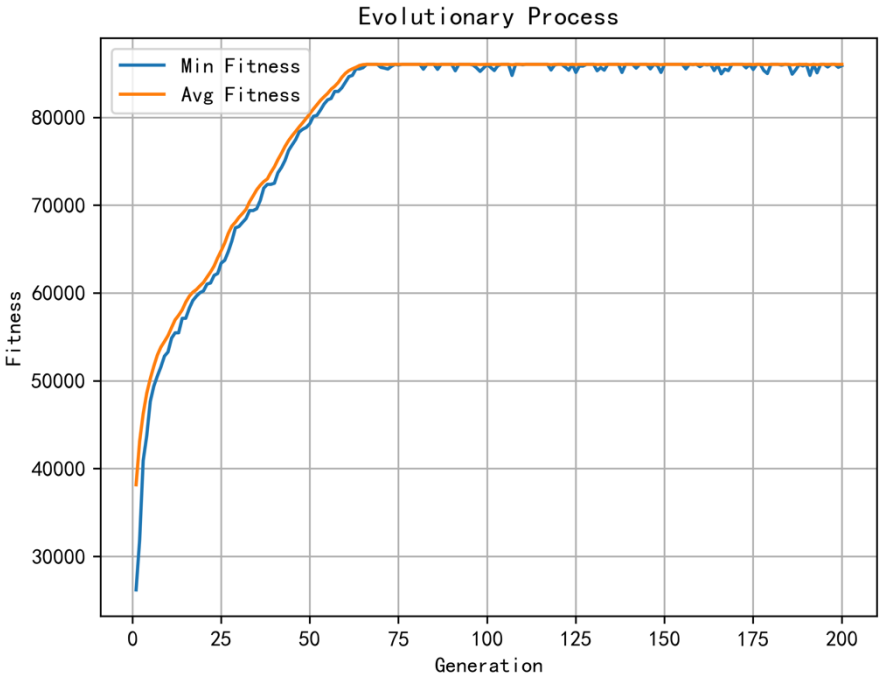

Fig12 Evolutionary process

3. Results

3.1. Association rules of different vegetable categories

This paper first clusters the customer 's shopping data, defines different shopping cart contents according to the time interval, and uses the Aprior algorithm to mine the association rules when analyzing this problem[12], and sets the support degree greater than 0.05 as the frequent itemsets, and the confidence degree greater than 0.3 as the association rules. The final vegetable category association rules are shown in Table3.

Table3 Association rules

antecedents	consequences	confidence
(Aquatic rhizomes)	(Leafy greens)	0.480803
(flower vegetables)	(Leafy greens)	0.474013
(solanaceae)	(Leafy greens)	0.491556
(Leafy greens)	(peppers)	0.312988
(peppers)	(Leafy greens)	0.447471
(edible mushrooms)	(Leafy greens)	0.460453
(flower vegetables)	(peppers)	0.345168
(edible mushrooms)	(peppers)	0.338289
(Leafy greens, edible mushrooms)	(peppers)	0.442875
(Leafy greens, peppers)	(edible mushrooms)	0.338196

3.2. Replenishment and pricing strategy prediction

In this paper, the ARIMA time series model is used to predict the daily replenishment and cost-plus pricing of vegetables in the next week, as shown in Table4 and Table5 :

Table4 Replenishment in the next 7 days

replenishment volume	July 1 to 7						
Aquatic rhizomes	19.909	16.848	14.580	13.221	15.050	17.271	19.327
Leafy greens	165.061	152.543	136.012	121.754	123.530	141.420	162.391
flower vegetables	22.444	19.157	18.141	18.140	18.141	18.141	18.141
solanaceae	28.957	27.705	23.408	17.910	15.682	18.664	24.512
peppers	95.732	95.68	84.721	75.266	73.880	81.497	92.622
edible mushrooms	47.945	51.200	52.465	52.957	53.148	53.223	53.251

Table5 Cost markup rate in the next 7 days

cost-plus rate	July 1 to 7						
Aquatic rhizomes	0 .920	1.695	0.700	0.493	0.608	0.364	0.850
Leafy greens	0.826	1.562	0.462	0.492	0.349	0.402	0.620
flower vegetables	0.932	1.500	0.642	0.571	0.598	0.164	0.362
solanaceae	0.882	1.660	0.732	0.473	0.317	0.728	0.582
peppers	0.897	1.582	0.499	0.388	0.532	0.611	0.686
edible mushrooms	0.826	1.554	0.846	0.982	0.586	0.618	0.758

4. Conclusions

In this paper, Morlet wavelet is used to analyze the periodic variation of time series data and explain the seasonal periodic variation of sales volume. The conclusions are as follows : From the perspective of the sales time of vegetable products, the quarterly distribution of vegetable categories is cyclical. Under the condition of non-normal distribution data, Spearman correlation coefficient is used to measure the correlation between vegetable category and single product sales. K-Means clustering can cluster the shopping basket information and help to find the association rules between vegetable categories, such as the strong correlation between cauliflower and mosaic vegetables.

For the daily replenishment and pricing decision of supermarket vegetables, through the comparison of different models, the random forest model is finally established, and the regression analysis is carried out to describe the relationship between the total sales volume and the cost markup pricing. The ADF stability test helps to determine the stability between the sales volume and the cost of vegetable categories. Based on the ARIMA algorithm, the vegetable demand and cost price in the next week are predicted. On this basis, a multi-objective optimization model is established. Finally, through the solution of genetic algorithm, the daily replenishment quantity and pricing strategy of different vegetable categories in the next week are finally obtained. The results are shown in Table.4 and Table.5 above, so as to realize the problem of maximizing the profit of the supermarket.

References

[1] Feng Bihong,Gu Yang.Study on the influence of weather factors on vegetable prices [J].Agricultural development and equipment,2023(01):89-90.

[2] Kurtulu M,Savaskan R C,Wang C.Drivers and implications of Direct-Store-Delivery in distribution channels[J].Production & Operations Management,2020,(29):2621-2636

[3] Kim G,Moon I.Integrated planning for product selection, shelf-space allocation, and replenishment decision with elasticity and positing effects[J].Journal of Retailing and Consumer Services,2021,(58):102274.

[4] Bianchi A,et al.Retail shelf space planning problems:A comprehensive review and classification framework[J].European Journal of Operational Research,2021,(1):1-16.

[5] Niu B Z,Dai Z P,Chen L.Information leakage in a cross-border logistics supply chain considering demand uncertainty and signal inference[J].Annals of Operations Research,2022,309(2):785-816.

[6] National Bureau of Statistics of the People's Republic of China.China Statistical Yearbook [M].China Statistics Press,2019.

[7] Wang Xiaotong;research on inventory control strategy of fresh fruit and vegetable products in supermarket chain based on DIT[D];southeast University;2020

[8] KAYIKCI Y,DEMIR S,MANGLA S K,et al.Data-driven optimal dynamic pricing strategy for re-ducing perishable food waste at retailers[J].Journal of Cleaner Production,2022,344:131068.

[9] Wang Xiaotong.Research on inventory control of fresh fruit and vegetable products in chain supermarkets considering price discounts [J].Fujian Quality Management ,2020,41(09),110-111.

[10] Duan Yongrui,Yang Jin,Huo Jiazhen.Ordering and transshipment strategies for perishable products with demand dependent on inventory levels in the display area[J].Operation and management,2014,23(04):51-57.

[11] Chen Yijuan.Research on multi-stage dynamic pricing of fresh food [D].Hainan Normal University,2015.

[12] DHAR S,BOSE I.Walking on air or hopping mad?Understanding the impact of emotions, sentiments and reactions on rating in online customer reviews of mobile apps[J].Decision Support System,2022,162:113769.