Research on The Replenishment Pricing Model of Data-oriented Vegetable Products

Zhongyang Han [1, *, #], Jiaxuan Dai [2, #], Jiayu Chen [2, #]

[1] School of Management and Engineering, Capital University of Economics and Business Beijing, China, 100070

[2] School of Statistics, Capital University of Economics and Business, Beijing, China, 100070

* Corresponding Author Email: h15161362141@sina.com

#These authors contributed equally.

Abstract. With the rapid development of economic life and the rapid improvement of people's living standards,Consumer demand for vegetable products has changed from quantity to quality.In view of the increasing requirements of consumers for the freshness, appearance, color and other qualities of vegetable products when purchasing.Based on massive data, the relationship between various categories and single products of vegetables is speculated, find out how their sales are distributed.At the problem of large data volume and relatively unclear structure,the data will be pre-processed,choose Pearson Correlation Analysis to look at overall differences and associations.Pandas library, Numpy library and matplotlib library based on Python,perform a more detailed descriptive statistical analysis of the cleaned data,and make relevant charts to show the visualization results.For complex data relationships.social science research needs to comprehensively consider the actual situation of society.The Apripri correlation analysis algorithm was used to further analyze the correlation between categories and unit sales.Based on the available informationThe relationship between the total sales volume of vegetables and the cost-plus pricing method of each category was analyzed.In view of the problems that the purchase volume of goods is difficult to predict, the difference in consumer consumption is obvious, and the seasonal changes are present,this article uses the LSTM time series and sliding time window to predict the daily sales volume of each category in the coming week.Based on this, the purchase cost is calculated, and the regression model is substituted to predict the pricing.Based on multi-objective optimization thinking and genetic algorithm, the goal decision of "revenue-cost-loss" is used as the objective function to seek the goal decision of benefit maximization.Finally, through the application of the time series prediction model, the relationship between the sales volume of each vegetable category and the cost-plus pricing is inferred, and the daily replenishment volume and pricing strategy are formulated to maximize the benefits of supermarkets.

Keywords: Pearson correlation analysis, Apriori correlation analysis, LSTM sliding time window.

1. Research on the sales volume of vegetable categories and single products

1.1. Data preprocessing

Because this data is unstructured data, it is necessary to clean all the data to determine whether there are missing values. Based on the cleaned data, the overall distribution characteristics of the data are observed through descriptive statistics, and the sales distribution rules of different categories and single products and their correlation are obtained [1]. The daily, monthly and quarterly periodic sales visualization results of the supermarket are shown in Figure 1-3:

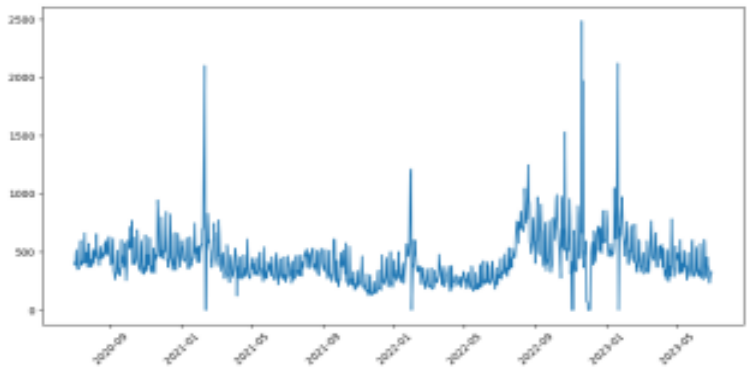


Fig 1. Daily sales statistics of a supermarket from early July 2020 to the end of June 2023.

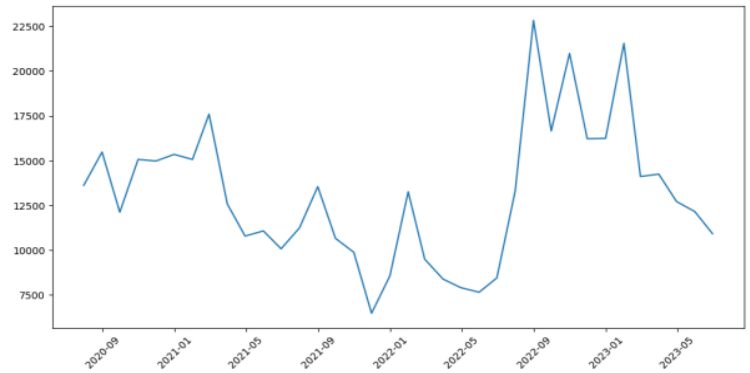


Fig 2. Monthly sales statistics of a supermarket from early July 2020 to the end of June 2023.

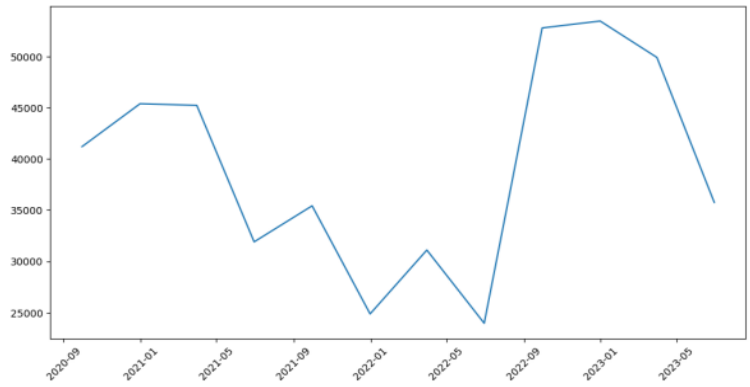


Fig 3. Quarterly sales statistics of a supermarket from early July 2020 to the end of June 2023.

Based on the known conclusions above, the first five dates of sales volume of vegetable products and their specific sales volume were calculated respectively, and a visual chart was made, as shown in Figure 4-8:

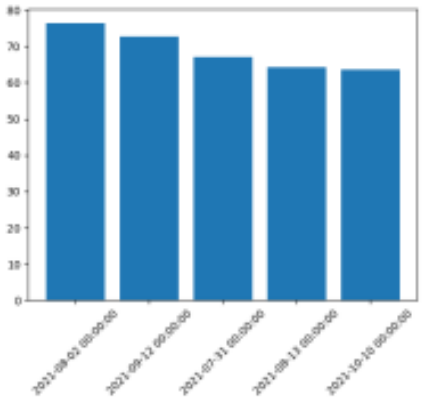


Fig 4. Yunnan lettuce.

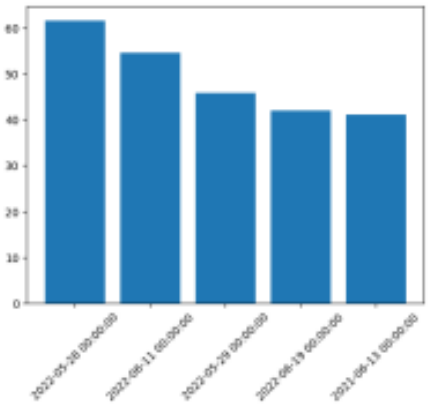


Fig 5. Shanghai Green.

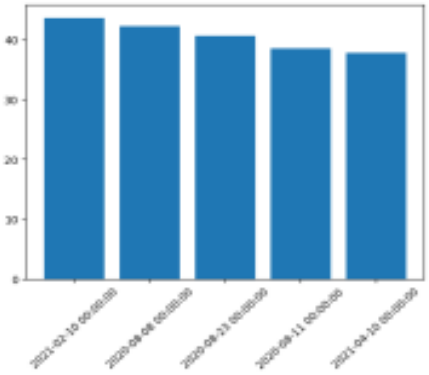


Fig 6. Bamboo leaf vegetables.

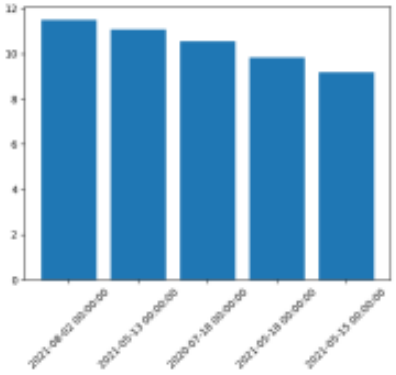


Fig 7. Little cabbage.

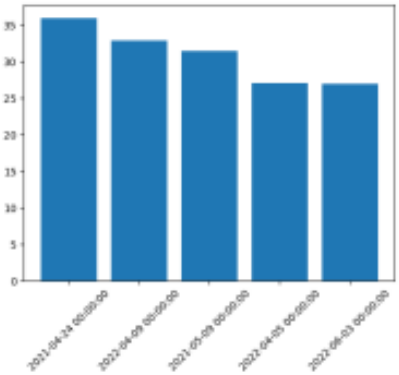


Fig 8. Amaranth.

The scatter plot of daily/monthly/quarterly sales data for each item/category is shown in Figure 9-14:

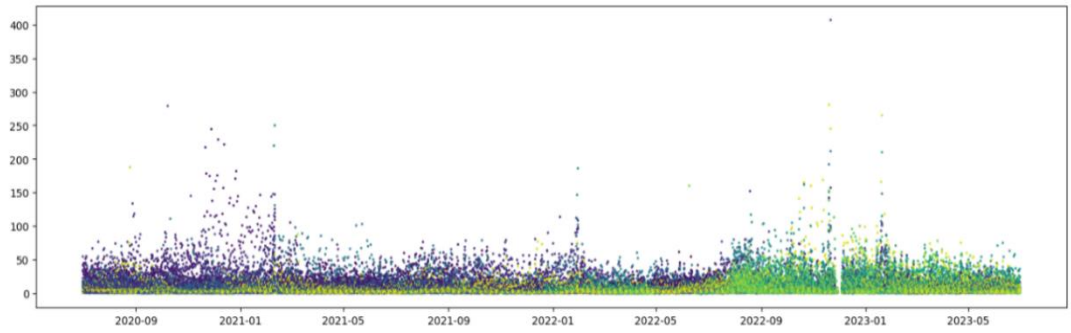


Fig 9. Scatter plot of daily sales data of single product.

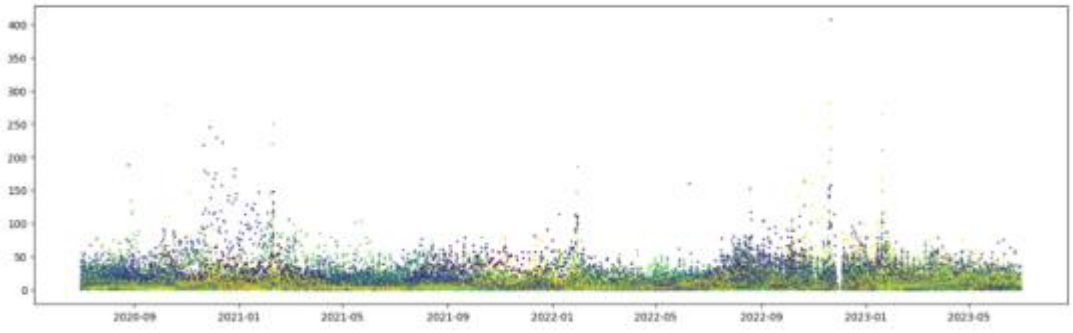


Fig 10. Scatter plot of daily sales data of each category.

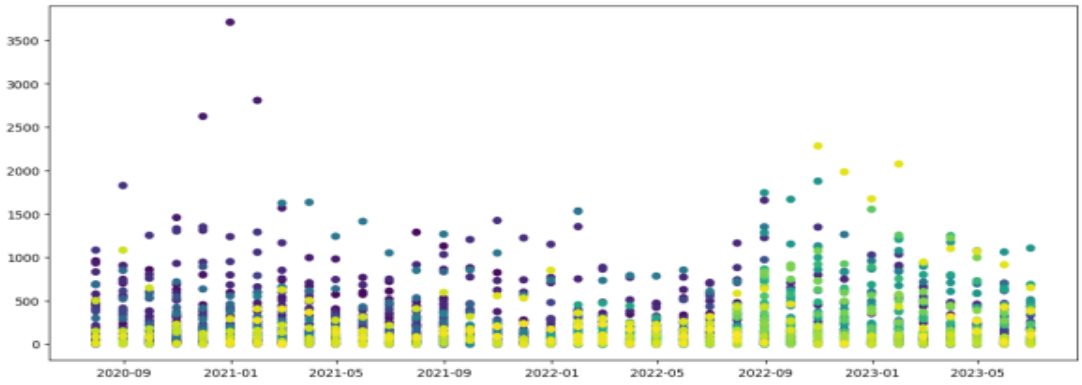


Fig 11. Data chart of monthly sales of single product

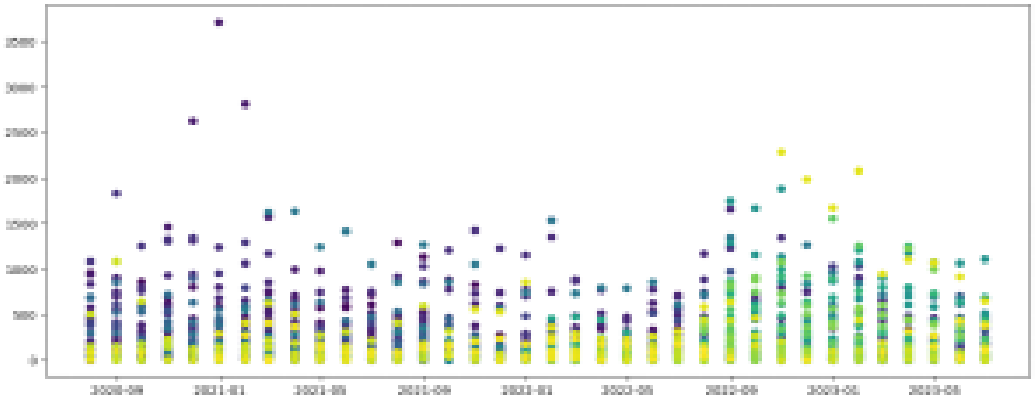


Fig 12. Monthly sales data chart of each category.



Fig 13. Data chart of quarterly sales of single product.

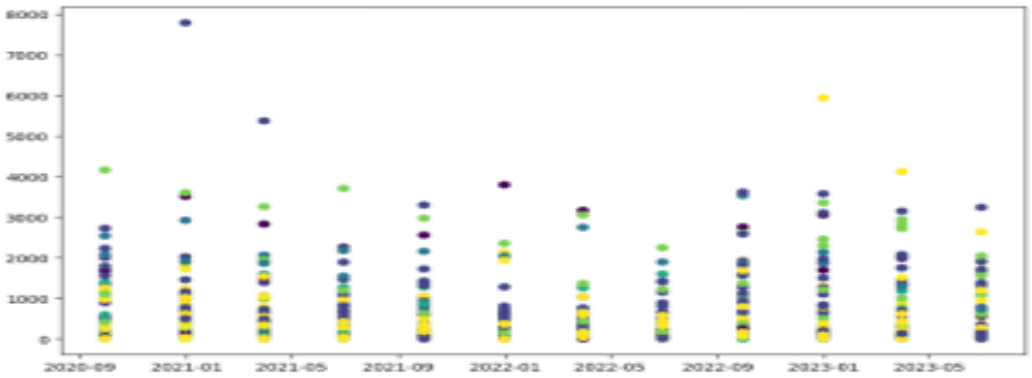


Fig 14. Quarterly sales data chart of each category.

Then, the data was verified by drawing the QQ chart of daily total sales (as shown in Figure 15), and the analysis found that it did not conform to the normal distribution.

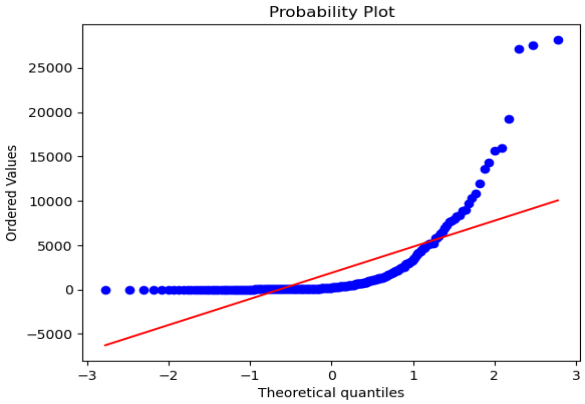


Fig 15. QQ chart of total sales volume.

1.2. Pearson correation analysis

This paper quantifies the association between different categories or single items, so Pearson correlation coefficient is used to quantify the linear relationship between random variables [2]. According to the definition, the Pearson correlation coefficient of the sample is expressed as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{Y})^2}} \ , (r \in [-1,1]) \quad (1)$$

Its specific calculation results are more complex and occupy too much space, see Annex I for details. Table 1 shows the correlation between vegetable categories:

Table 1. Pearson correlation coefficient of each category of vegetables.

	Cauliflower	Chili pepper	Edible fungus	Aquatic root vegetables	Nightshade species	Mosaic leaf
Cauliflower	1	0.544261824	0.519611425	0.541716793	0.347625264	0.622311125
Chili pepper	0.544261824	1	0.686001571	0.610251832	0.317525239	0.652635366
Edible fungus	0.519611425	0.686001571	1	0.670316829	0.140319009	0.630379719
Aquatic root vegetables	0.541716793	0.610251832	0.670316829	1	0.105562452	0.555334032
Nightshade species	0.347625264	0.317525239	0.140319009	0.105562452	1	0.322177101
Mosaic leaf	0.622311125	0.652635366	0.630379719	0.555334032	0.322177101	1

Based on the above table, the top four single product codes and their sales volume of daily vegetable products are screened out, and cluster analysis is carried out on them. The top ten products in sales volume were filtered out, and the top ten products in sales volume were extracted by category for visual analysis, as shown in Figure 16-17:

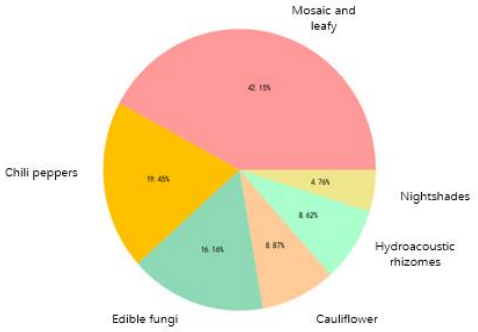


Fig 16. The top ten products by item code.

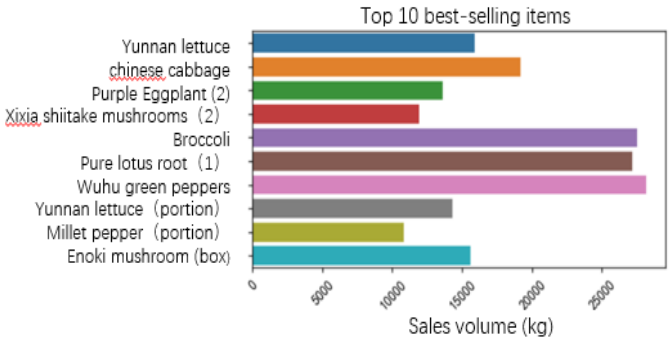


Fig 17. Top ten vegetable categories by sales volume.

1.3. Association analysis of Apriori algorithm

This paper studies the correlation problem of each category and its single product sales of vegetables. Because its sample space satisfies the corresponding analysis conditions of Apriori algorithm, it is selected as the analysis algorithm to mine its internal association rules [3]. The frequent sets of bool association rules are found by an iterative method of searching layer by layer.

The data set is traversed, and the support, the antecedents and the consequent are obtained respectively. The antecedents and the consequent are spliced into the form of association rules, and then the confidence and lift are extracted and the results are output.

2. Sales and pricing strategies of various categories and categories of vegetables

2.1. Data preprocessing

In this paper, the basic pandas and numpy libraries are used for data preprocessing. The wholesale price, purchase quantity, purchase cost and sales revenue were extracted as the required new data set. Based on the time series correlation algorithm, combined with the monthly and seasonal periodicity, the sales and pricing of various types of vegetables are predicted respectively [4]. The result is shown in Figure 18-23:

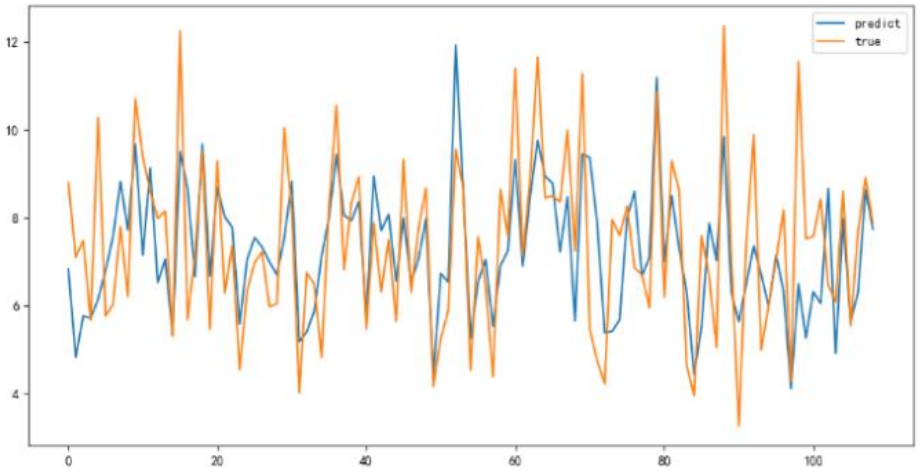


Fig 18. Mosaic sales and pricing forecast.

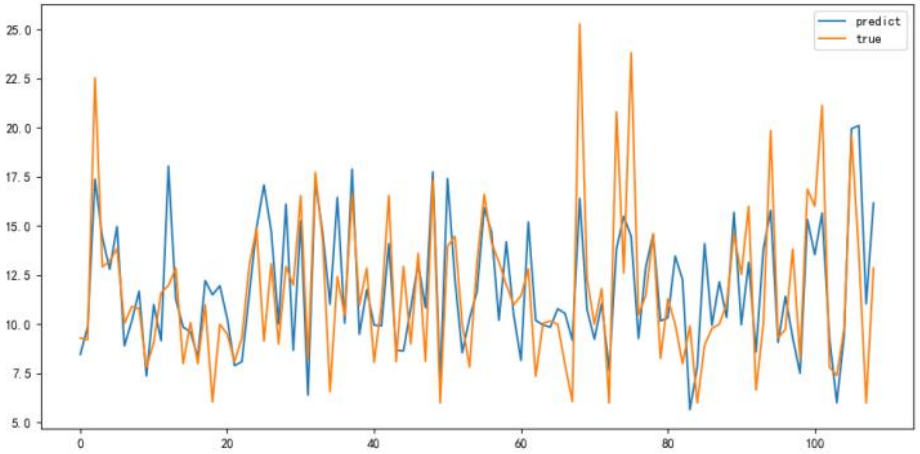


Fig 19. Sales and pricing forecast of aquatic rhizomes.

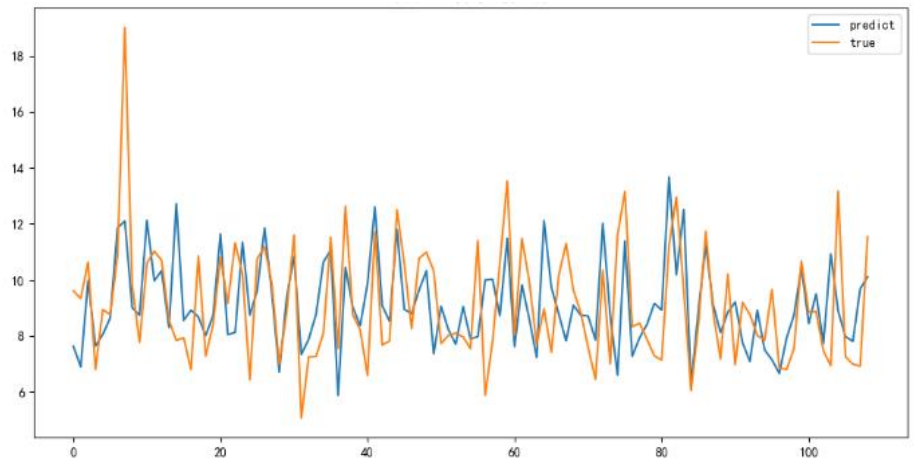


Fig 20. Edible mushroom sales and pricing forecast.

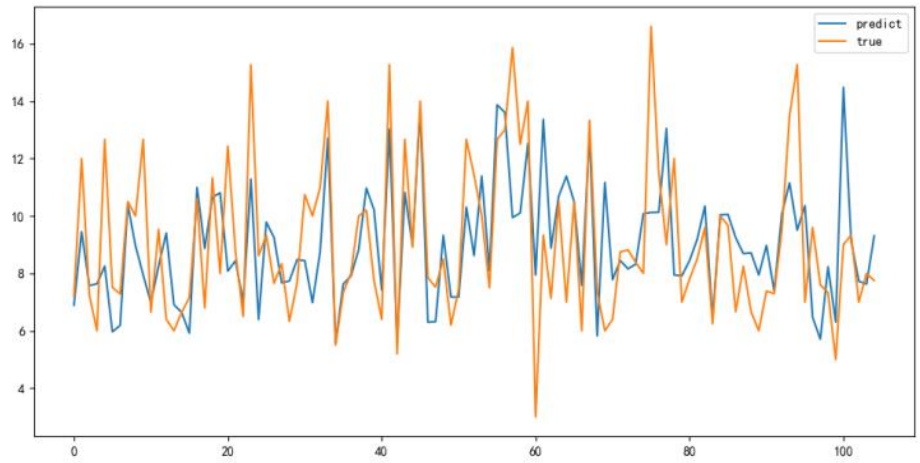


Fig 21. Nightshade sales and pricing forecast.

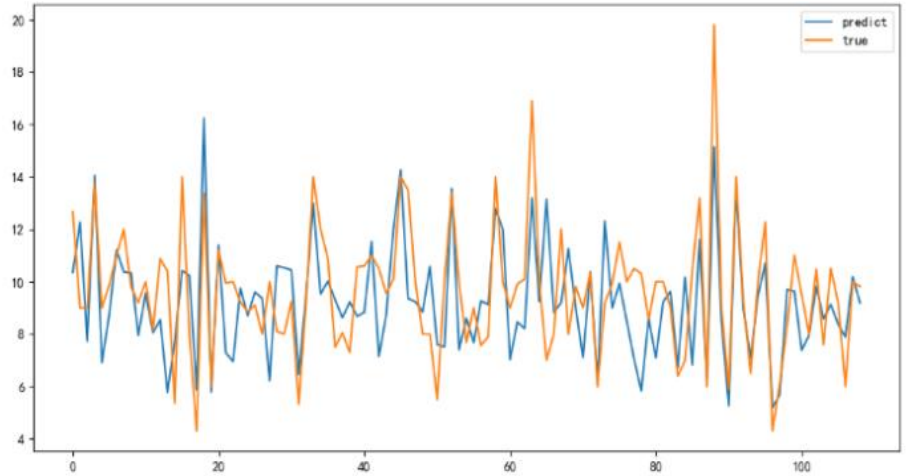


Fig 22. Cauliflower sales and pricing forecast.

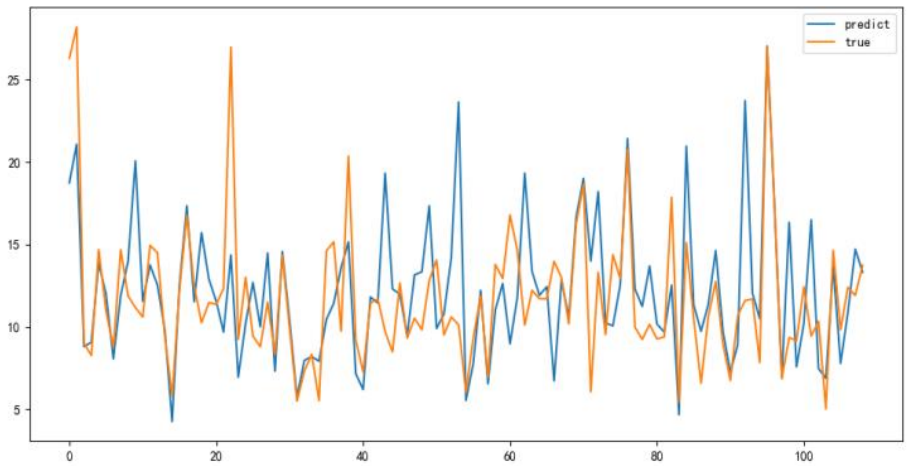


Fig 23. Sales and pricing forecast of pepper category.

2.2. Linear regression models

Based on the existing data set, a linear regression model is established, and the model is trained to obtain the characteristic coefficients [5]. The calculation equation of the functional relationship between the sales of various types of vegetables is calculated as follows:

Equation of sales volume and pricing function of Mosaic leaf:

$$y_1 = (-0.08 * Sales\ \ volume(kg)) + \left(0.07 * Purchase\ \ quantity(kg)\right) + 7.63 \quad (2)$$

Relationship between sales volume and pricing function of aquatic rhizomes:

$$y_2 = (-1.91 * Sales\ \ volume(kg)) + \left(1.72 * Purchase\ \ quantity(kg)\right) + 12.36 \quad (3)$$

Equation of sales and pricing function of edible fungi:

$$y_3 = (-0.23 * Sales\ \ volume(kg)) + \left(0.22 * Purchase\ \ quantity(kg)\right) + 9.25 \quad (4)$$

Solanum solanum sales and pricing function equation:

$$y_4 = (1.68 * Sales\ \ volume(kg)) + \left(-1.60 * Purchase\ \ quantity(kg)\right) + 9.89 \quad (5)$$

Equation of sales volume and pricing function of cauliflower:

$$y_5 = (0.10 * Sales\ \ volume(kg)) + \left(-0.12 * Purchase\ \ quantity(kg)\right) + 10.56 \quad (6)$$

The functional relationship between sales volume and pricing of peppers:

$$y_6 = (1.64 * Sales\ \ volume(kg)) + \left(-1.51 * Purchase\ \ quantity(kg)\right) + 13.09 \quad (7)$$

The above linear regression relationship analysis was used to fit the linear relationship between total sales and cost markup and visualize it as shown in Fig.24:

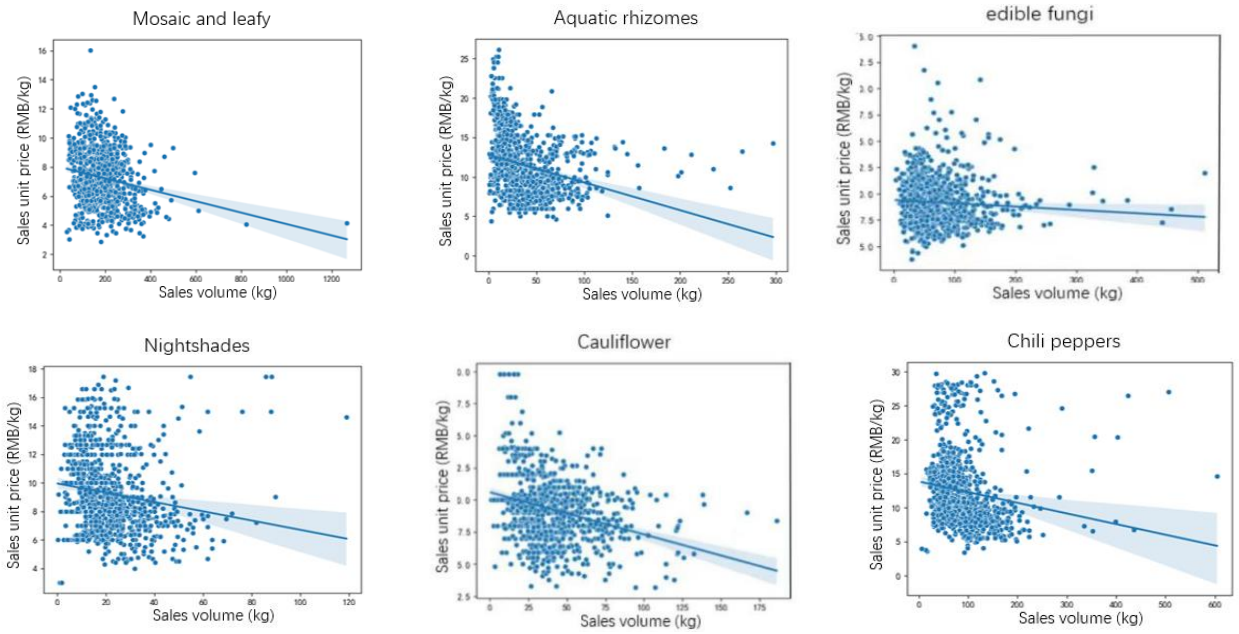


Fig 24. Relationship between total sales volume and pricing of various vegetable commodities.

2.3. LSTM time series prediction model

LSTM time series model is a model for learning and predicting long time series data. Based on the data characteristics of known data, this model is selected to predict the sales data of various types of vegetables relatively accurately in the next week [6-7].

Due to the mismatch of the existing data form, the existing data is converted into the form of supervised learning set based on python language. Data preprocessing was carried out by data set splicing, deleting missing rows where missing values were located, and normalization processing. Based on the processed data, it is divided into training set and test set. With a period of one year, the LSTM model is trained to initialize, and the prediction is made based on this [8]. After the inverse scaling of the predicted data and the real data is completed, the real data and the predicted data are visualized, as shown in Figure 25-30:

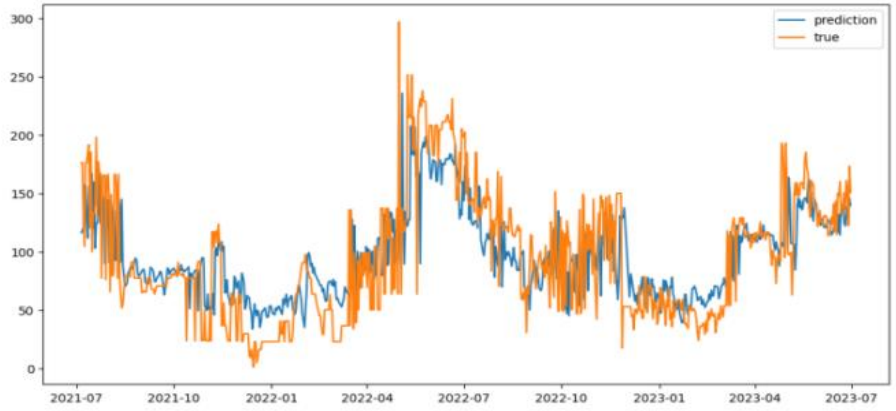


Fig 25. Sales data forecast chart of aquatic rhizomes in the next 7 days

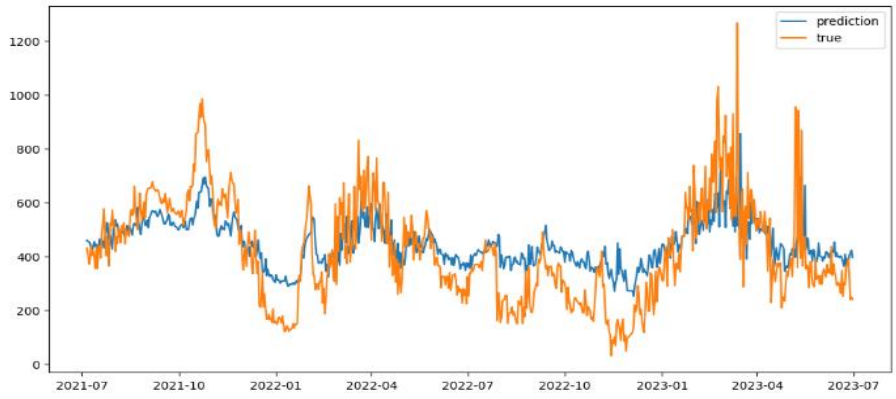


Fig 26. Sales data prediction chart of Mosaic category in the next 7 days.

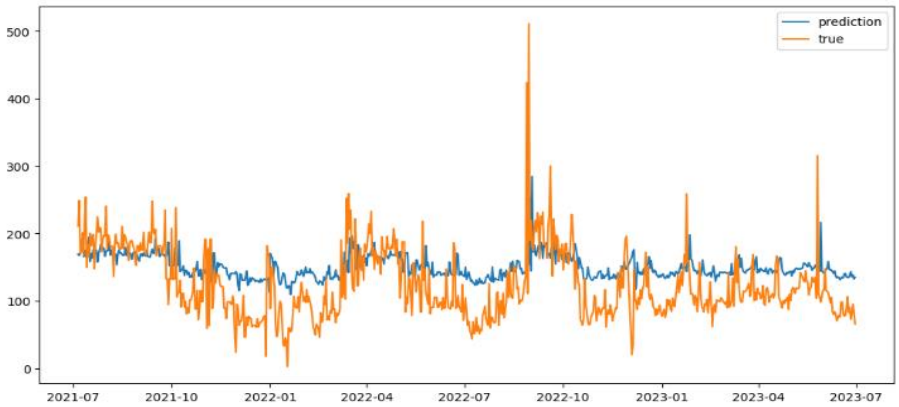


Fig 27. Prediction chart of sales data of edible fungi in the next 7 days.



Fig 28. Sales data forecast chart of nightshade in the next 7 days.

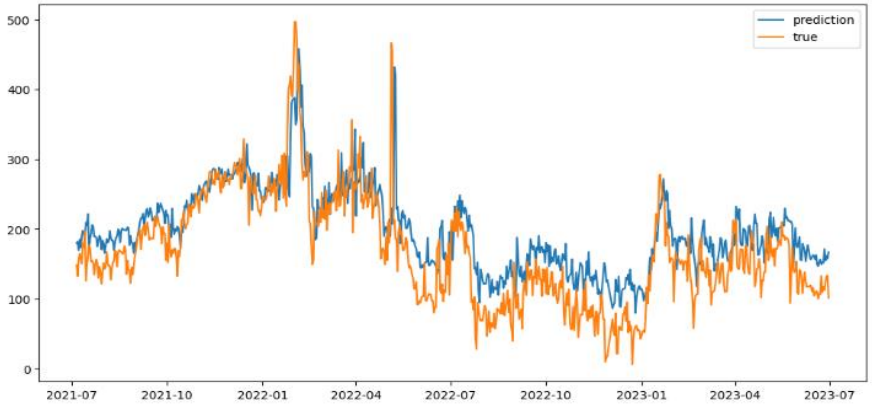


Fig 29. Sales data prediction chart of chili in the next 7 days.

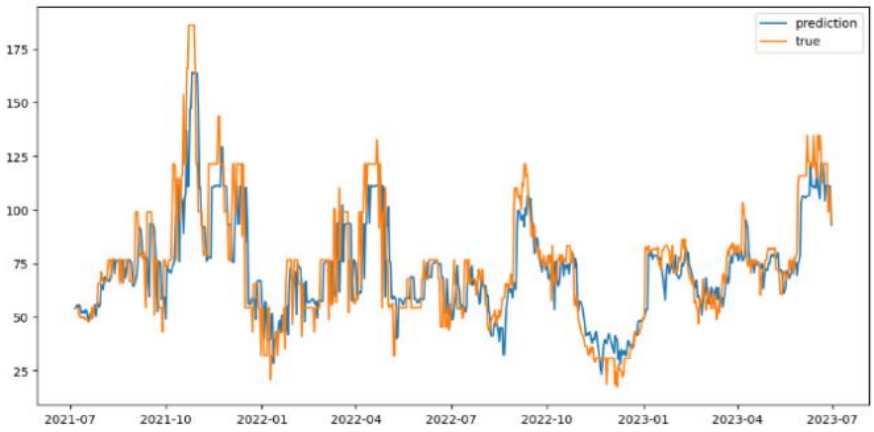


Fig 30. Sales data of cauliflower category forecast for the next 7 days.

Then, the multi-objective optimization method was used to solve the possible sales volume of each category in the next seven days, the amount of replenishment needed under the algorithm optimization strategy, the maximum revenue and pricing. Finally, the replenishment pricing strategy in the next seven days was obtained, and the supermarket profit was determined to be the maximum under this strategy [9-12].

3. Conclusions

Supermarkets have a long history of selling vegetables and have accumulated massive data. However, due to the limitation of technology, the formulation of pricing and replenishment strategies is mostly based on the seller's own experience, with large individual differences and low accuracy and scientificity. In this paper, Pearson correlation analysis, Apriori algorithm correlation analysis, and LSTM time series prediction model are used to construct a producer auxiliary model under the support of relevant economic theory. The model is based on the reference and analysis of a large number of literature, which is highly scientific. And use a variety of models to reference each other, the results are relatively accurate, intensive reading is high. Due to the high correlation of a large number of food commodities and their derivatives, the formulation of pricing and replying strategies is similar, and their applications are wide. With a solid mathematical and computer foundation, persuasive, credible, and innovative, relatively novel. And after the model passes the stationarity test, it is found that the fitting degree is good, the data fluctuation is small, and it has high application value.

References

[1] Tao Shancheng. Research on spatio-temporal data analysis, prediction and visualization system of grain temperature [D]. Nanjing University of Finance and Economics, 2024.

[2] Gong H ,Li Y ,Zhang J , et al.A new filter feature selection algorithm for classification task by ensembling pearson correlation coefficient and mutual information[J].Engineering Applications of Artificial Intelligence,2024,131107865-.

[3] Mujianto H A ,Mashuri C ,Andriani A , et al.Consumer Customs Analysis Using the Association Rule and Apriori Algorithm for Determining Sales Strategies in Retail Central[J].E3S Web of Conferences,2019,12523003.

[4] [Chen J, Kang S. Joint decision-making of agricultural product pricing and inventory replenishment for dual-channel sales [J]. Industrial Engineering, 2023, 26(03):39-46.

[5] Shen X .Analysis of Commodity Housing Price in Shanghai Based on Multiple Linear Regression Model[J].Academic Journal of Business Management,2022,4.0(3.0):

[6] Elmaghraby W, Keskinocak P. Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions [J]. Management Science, 2003, 49(10): 1287-1309.

[7] Xu Zexin, Yang Lei, Li Kangshun. Prediction model of short long series time series [J/OL]. Computer Application :1-10[2024-01-12]

[8] Shen Chen, Mu Yueying. Analysis of Time series Change of Vegetable Price in Our Country [J]. Statistics and Decision, 2011, (16):78-80.

[9] LI Shuai. Research on commodity pricing and ordering strategy considering strategic purchase behavior under E-commerce environment [D]. Yanshan University, 2019.

[10] Cachon G P, Feldman P. Price Commitments with Strategic Consumers: Why it can Be Optimal to Discount more frequently than Optimal [J]. Manufacturing & Service Operations Management, 2015, 17(3): 399-410.

[11] Shum S, Tong S, Xiao T. On the Impact of Uncertain Cost Reduction When Selling to Strategic Customers [J]. Management Science, 2016, 63(3): 843-860.

[12] Savva N. Dynamic Pricing in the Presence of Social Learning and Strategic Consumers [J]. History, 2017, 63(4): 919-939.