Research On Pricing and Replenishment of Vegetable Products Based on Time Series Prediction

Xu Ma [1, *, #], Jian Ran [2, #], Feng Xiong [1, #]

[1] College of Electrical Engineering, Southwest Minzu University, Chengdu, China, 610225

[2] School of Mathematics, Southwest Minzu University, Chengdu, China, 610225

* Corresponding Author Email: xma_swun@163.com

[#]These authors contributed equally.

Abstract. Based on the commodity information, sales data, and recent loss rate of wholesale price commodities of vegetable categories provided by the superstore, this paper focuses on the relationship between vegetable categories and individual products and the prediction of sales volume in the next seven days. Firstly, the distribution pattern and interrelationship of the sales volume of each category and single product of vegetables were analyzed using data analysis and visualization techniques, and the specific size of the correlation of each category and single product was obtained. Secondly, the ARIMA model was used to predict the sales volume of vegetable categories, and the corresponding replenishment volume was obtained. Finally, through the above analysis, a corresponding replenishment strategy can be provided for the superstore to achieve the subsequent optimal revenue, and the superstore revenue planning model is also discussed and outlooked at the end of the paper.

Keywords: Data processing and visualization, Time series prediction, Nonlinear programming.

1. Introduction

Many ingredients in local fresh supermarkets have a relatively short shelf life, and vegetables that are not sold out within a day cannot be resold. Due to the fact that the vegetables sold by supermarkets are restocked in the early morning, with a wide variety of categories and different origins, supermarkets must accurately restock vegetables based on the historical sales and demand of each product, without knowing the specific individual items and purchase prices[1]. The paper now request to provide replenishment and pricing decisions based on market demand analysis.Due to the possible relationships between different vegetable categories and individual products, this article mainly analyzes the distribution patterns and interrelationships of sales volume of each category and individual product. At the same time, only considering the category as a unit to make replenishment plans, the relationship between the selling price of each vegetable category and the daily total sales quality is analyzed, and combined with historical sales records, a prediction is made for the daily replenishment total and pricing strategy of each vegetable category from July 1 to 7, 2023, Maximizing profits for supermarkets within a week.then analyze the relationship between sales volume, time, and selling price, and explore the periodicity and volatility. Secondly, if there are too many individual products, select the one with high sales volume for analysis. The Spearman correlation coefficient can be used to analyze the correlation between individual products or categories(Data sources:https://cumcm.cnki.net).

2. Analysis of the distribution pattern of sales volume of various categories of vegetables

Firstly, analyze the sales volume of various categories of vegetables, and separately calculate the total sales volume of each type of vegetable from July 2020 to June 2023 and the daily sales volume of each type of vegetable. Use visual methods to intuitively understand the relationship between the total sales volume of each type and observe the trend of daily sales volume, as shown in Fig 1.

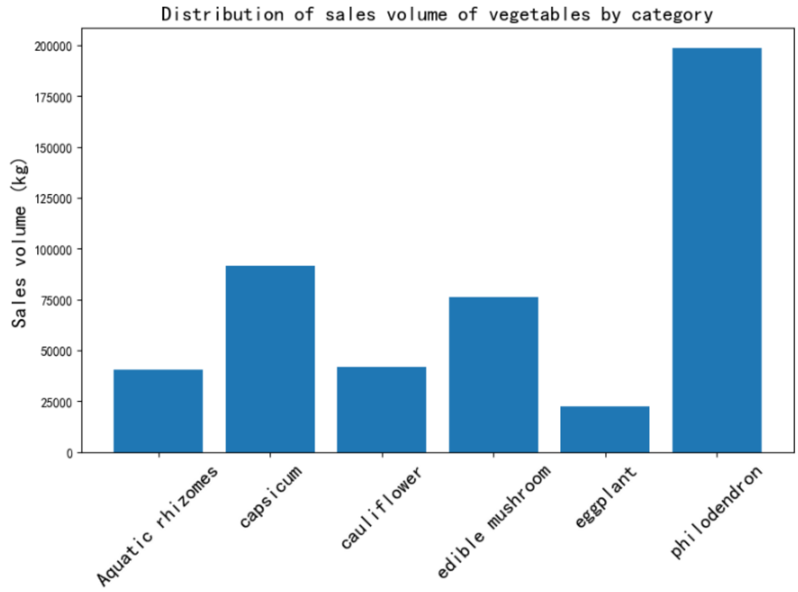

Fig 1. Sales statistics of vegetable categories

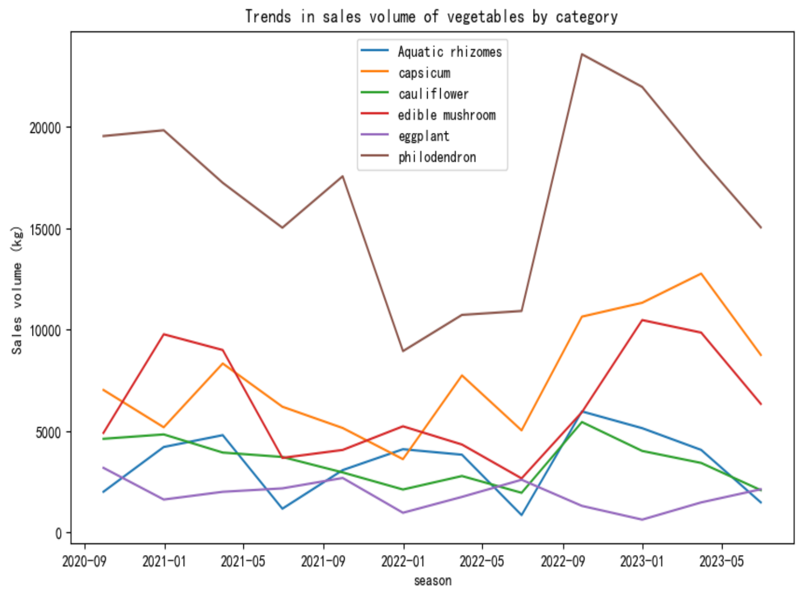

Fig 2. Seasonal sales statistics of various vegetable categories

In order to explore the correlation between the seasonality of various vegetable categories, this article adopts a quarterly statistical analysis, which visualizes the sales volume, average pricing, and total profit from three aspects. The corresponding time to end encoding is (2020-9 to 2023-5), and the corresponding trend is shown in Fig 2.

From the above analysis, it can be concluded that the sales volume of vegetable categories is influenced by time. Taking flower and leaf categories as an example, their sales volume showed a downward trend in the third quarter. The peak of flower and leaf categories appeared in the third and fourth quarters (i.e. winter), while products such as chili peppers appeared in the first and second quarters (spring and summer). There is a correlation between the sales volume of vegetables and the quarter. The sales volume fluctuates greatly before and after each peak, indicating a cyclical supply of some individual products.

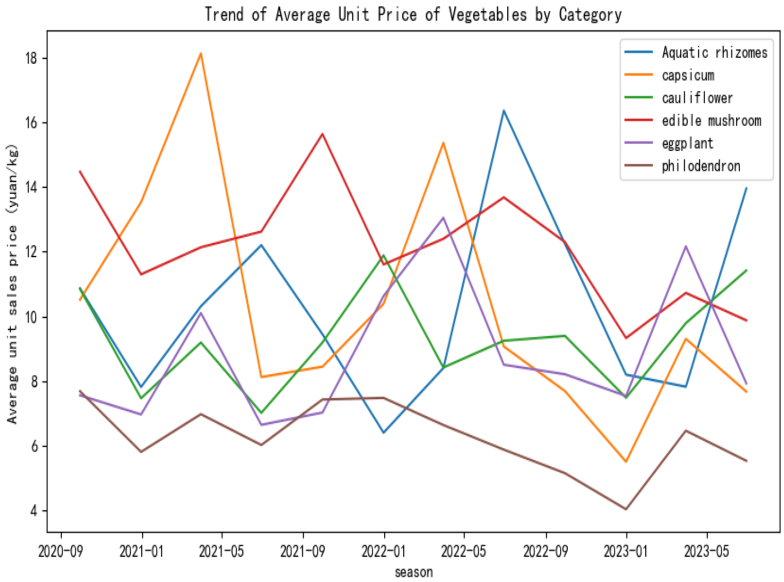

Fig 3. Seasonal average pricing statistics for various vegetable categories

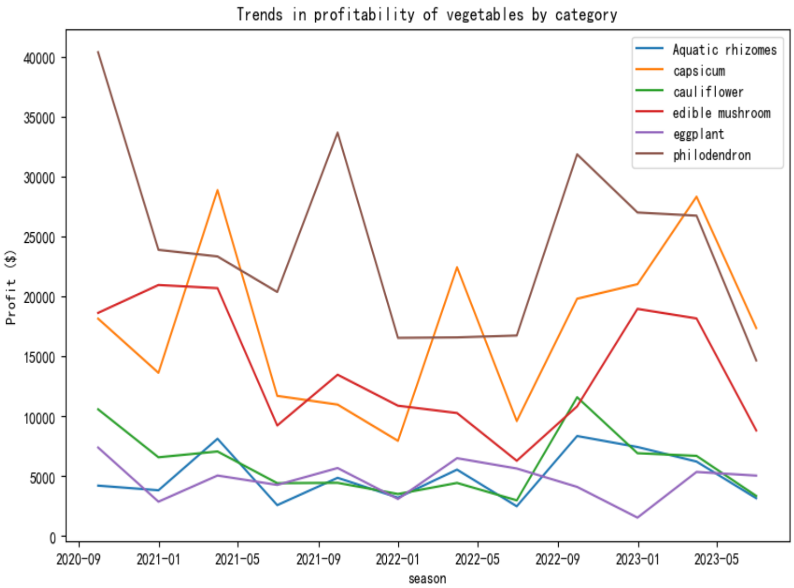

Fig 4. Seasonal profit trends of various vegetable categories

From Fig 3 and 4, it can be seen that the pricing of the product category also satisfies the saying "rarity is more expensive". For example, in Fig 4, the yield of chili peppers in winter is relatively low, so the season with higher prices appears between the fourth and first quarters; The production of eggplants was relatively high in the third quarter, so the season with lower prices appeared in the third quarter. Although the sales unit price of flowers and leaves is the lowest in every quarter, their total profit is relatively objective, that is, the total sales amount is relatively high. This indicates a certain relationship between sales volume and sales pricing, which is speculated to be in line with general consumer psychology: sales price is inversely proportional to purchase intention.

From the perspective of profit and sales trends, the trend of change is roughly the same. Due to the high sales volume of leafy vegetables, they bring more profits, which is higher than other curves. The profits of eggplants, aquatic rhizomes, and cauliflower are not significantly different. It is worth mentioning that the bonus rate of aquatic rhizomes in April 2021 is only about 0.29%, which means they are often sold at a discount, with a profit of only 76 yuan.

2.1. Analysis of the correlation between vegetable categories and individual products

When studying the correlation between various vegetable categories, it was found that some varieties did not follow a normal distribution, so the Spearman correlation coefficient was used here. Analyze the correlation between different categories by calculating the correlation coefficient [2] r_{ij}.

$$r_{XY}=\frac{Cov(X,Y)}{S_X S_Y} \tag{1}$$

$$\bar{X}=\frac{1}{36}\sum_{k}^{36}x_{ik} \tag{2}$$

$$\bar{Y}=\frac{1}{36}\sum_{k}^{36}x_{jk} \tag{3}$$

$$Cov(X,Y)=\frac{\sum_{i=1}^{36}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{35} \tag{4}$$

Among them, X and Y are sequences subjected to pairwise correlation analysis, r_{XY} is the Spearman correlation coefficient between X_i and X_j, $\bar{X}$ and $\bar{Y}$ are the average values of X_i, Y_j, and x_{ik} and x_{jk} represent the k th data of X_i and X_j, respectively.

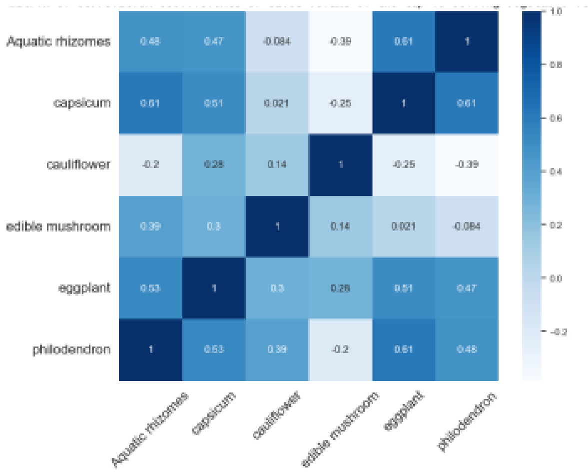

Fig 5. Heat map of profit correlation coefficients for various vegetable types

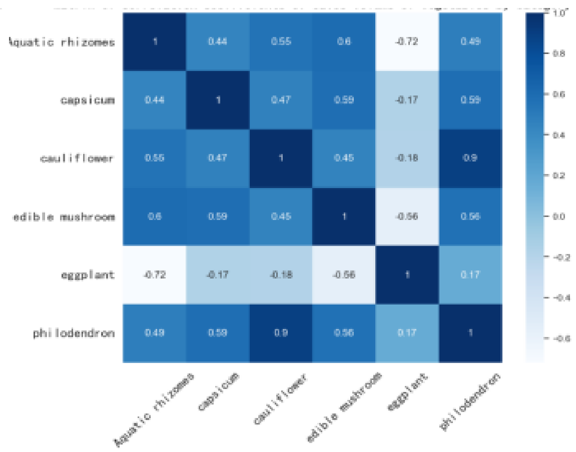

Fig 6. Heat map of correlation coefficients between sales volume of various vegetable types

From Figs 5 and 6, it can be concluded that the correlation coefficients between cauliflower and edible fungi are 0.615 and 0.42, with a strong positive correlation; The correlation coefficients between aquatic rhizomes and edible fungi are 0.615 and 0.625, indicating a strong positive correlation; The correlation coefficient between aquatic rhizomes and eggplants is -0.385, with a negative correlation of -0.715. The two positively correlated categories will exhibit positive sales trends such as bundled sales. Negative correlation can lead to negative sales trends such as mutual competition.due to the large correlation coefficient matrix of all 252 individual products being studied and not having significant practical significance. Therefore, this article selects the top 15 products in terms of sales volume for correlation analysis [3].

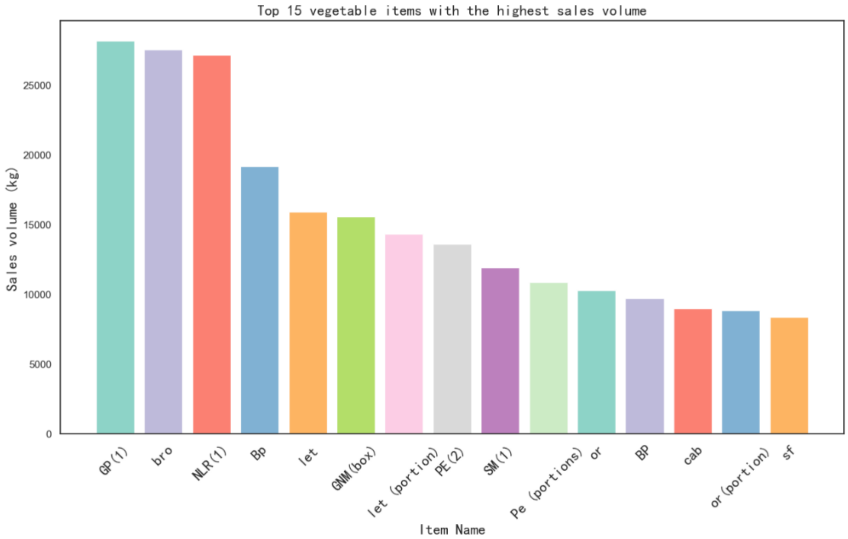

Fig 7. Top 15 Single Product Sales

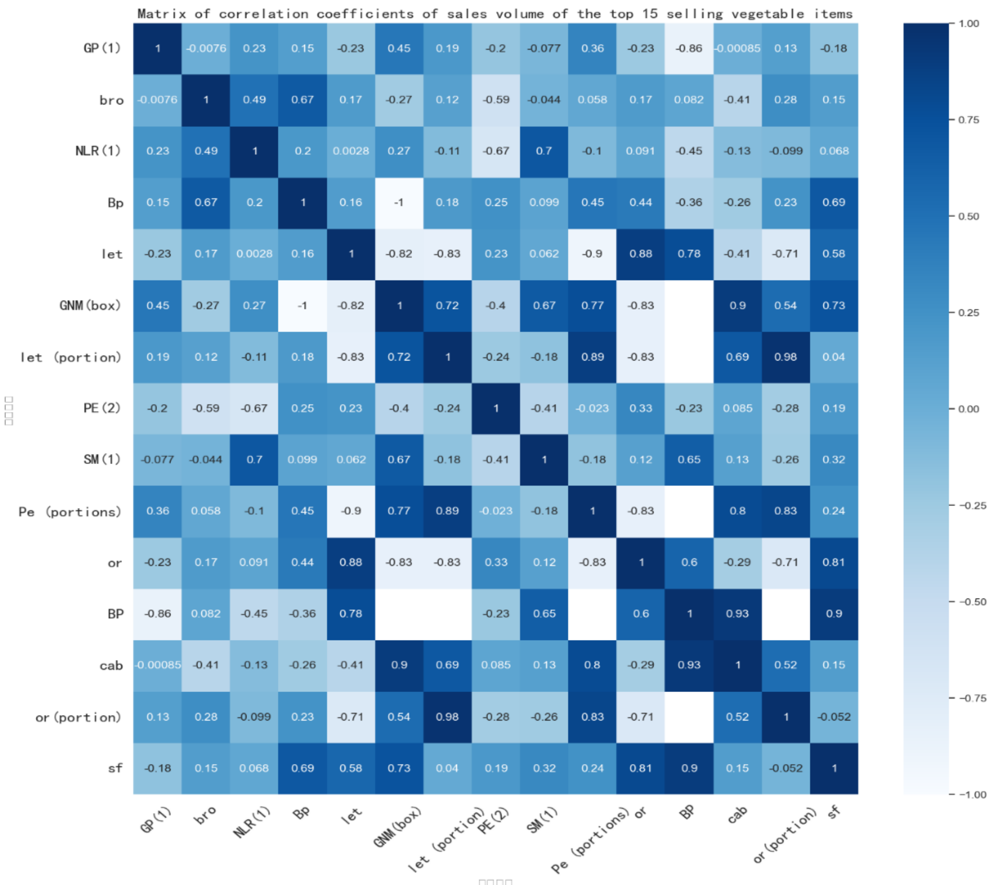

Fig 8. Correlation coefficients of typical individual products

The full names of the abbreviations in Fig 7 and Fig 8 are: Green Pepper (1); broccoli; Net Lotus Root (1); Brassica pekinensis; lettuce; Golden Needle Mushroom(box); lettuce (portion); Shiitake Mushroom (1); Purple Eggplant (2); oilseed rape; cabbage; Bubble Pepper Peppers (portions); scattered flowers.

Studying the main single items in terms of sales duration and sales quantity in various vegetable categories is conducive to a deeper exploration of bundled and mutually exclusive single items. With the analysis of the aforementioned vegetable categories, calculating the Spearman correlation coefficient can yield the following results.

According to Fig 8, the negative correlation coefficient between Wuhu green pepper (1) and Paopao pepper (premium) is -0.86, and the correlation coefficient between Jin needle mushroom (box) and Chinese cabbage is -1; The correlation coefficient between Xiaomi pepper and Yunnan lettuce is -0.9. The positive correlation coefficient is 0.98 for Yunnan lettuce (portion) and Yunnan oilseed rape (portion); The correlation coefficient between Qinggan Sanhua and Paopao Jiao (premium) is 0.9; The correlation coefficient between Bubble Pepper (Premium) and Baby Cabbage is 0.93. A negatively correlated product, also known as "you increase, and I decrease," belongs to competitive products, while a positively correlated product is suitable for bundled sales and retail, that is, "same increase and same decrease"[4].

2.2. Analysis of the Relationship between Total Sales and Cost-Plus Pricing

Regression is a commonly used method to explore the interrelationships between variables, but the data in this article is time series data [5]. Most of the variables composed of time series data are non-stationary. If non-stationary variables are directly used for regression analysis, especially in the case of tens of thousands of big data like this question, it will lead to incorrect conclusions that all variables have correlation relationships, using actually unrelated non-stationary variables for regression analysis is a false regression (pseudoregression). To avoid such problems, the first step is to conduct stationarity tests on each time series [6]. Using SPSS for ADF unit root test, all six categories of data rejected the null hypothesis at the 0.1 level, indicating that the sequence was stationary.

In order to explore the correlation between total category sales and cost-plus pricing, the Spearman correlation coefficient was used for analysis, and the following results were obtained, as shown in Table 1:

Table 1. Correlation coefficient of each category

category	spearman correlation coefficient
florescent vegetables	-0.236
Chili peppers	-0.312
Flower leaf class	-0.16
Solanaceae	0.798
Edible fungi	-0.461
Aquatic rhizomes	-0.442

There is a significant positive correlation between the total sales volume of eggplants and their pricing, with a Spearman correlation coefficient of 0.798; The other five categories all showed a negative correlation, showing a trend of "rise and fall", with edible fungi and aquatic rhizomes being more prominent [7].

2.3. Prediction models for various statistical indicators

After clarifying the interrelationships between each category, this article adopted two methods to analyze the future 7-day prediction data of six categories. The 7-day prediction results of the VAR model [8].

Based on a time series model [9], predictions were made for each category from July 1st to July 7th. The expert modeler in SPSS can effectively help us find the most suitable ARIMA (p, d, q) model. The projected sales values for each vegetable category are shown in Table 2.[10]

Table 2. Sales forecasts by vegetable category

Time	florescent vegetables	Flower leaf class	Chili peppers	Solanaceae	Edible fungi	Aquatic rhizomes
July 1	23.386	132.5	79.59	23.08	43.14	21.43
July 2	21.851	126.53	80.16	22.17	46.19	19.99
July 3	20.82	126.65	80.47	18.85	49.56	20.95
July 4	21.896	139.84	82.15	20.15	53.95	20.62
July 5	22.895	148.07	81.78	19.47	53.38	20.99
July 6	23.64	140.83	84.41	22.06	52.73	20.97
July7	24.816	142.86	86.31	21.8	50.84	21.16

3. Conclusions

This paper processed the data and divided it into 6 vegetable categories. Relevant data analysis software was used to study the correlation and distribution rules among each vegetable category and typical single product according to the overall, seasonal trend and Spearman correlation coefficient, and the results were visually analyzed. The results show that the sales volume of vegetable products is affected by the time of receipt, the sales volume is inversely proportional to the sales price, and the profit trend is roughly the same as the sales volume. The two categories showing positive correlation in vegetable categories will show positive sales trends such as bundling and the two categories showing negative correlation will show negative sales trends such as competition! For example, the correlation coefficients between the profits and sales of cauliflower and edible fungi were 0.65 and 0.42, and the correlation coefficients between the profits and sales of aquatic rhizoma and eggplant were -0.385 and 0.715. There is a negative correlation of "you increase and I decrease" and a positive correlation of "same increase and decrease" among typical products. For example, the correlation coefficient between Flammulina mushroom (box) and Chinese cabbage is -1, which is a competitive product, and the correlation coefficient between Yunnan lettuce (part) and Yunnan romaine lettuce (part) is 0.98, which is suitable for bundled sales. The above results can make the supermarket more clear on the choice of seasonal products, choose the more competitive vegetables, avoid the vegetables that do not produce the expected profit for the supermarket, and purchase the vegetables suitable for bundling in a proper proportion to achieve greater profits.

In the following, to make the supermarket more clear about the quantity of vegetable replenishment and pricing strategy, and maximize the profit. We use the provided data and time series model to make predictions and use nonlinear programming to search the total replenishment and pricing strategies under different constraints, taking categories and parts as units. Finally, we give the data and reasons that are conducive to the supermarket to make better decisions.

References

[1] Zeng Minmin. Research on dynamic pricing strategy of fresh community supermarket based on time context A [D]. Southwest University of Finance and Economics,2022.

[2] Wang Mingle, Hou Qingfeng. Fluctuations in Vegetable Prices and Their Influencing Factors in Baiyin City, Gansu Province: Empirical Analysis Based on VAR Model [J]. Land and Natural Resources Research, 2021, (04): 82-86.

[3] Fan Taotao, Kou Yanting, Liu Chen, Yan Hongcan. Research on Determining the Stability of Data in Time Series Analysis [J]. Modern Electronic Technology, 2013,36 (04): 66-68+72

[4] Wei Wenlong A Study on Optimization of Company A's Replenishment Strategy Based on CPFR [D]. Xiamen University, 2014.

[5] Mao Lisha. Research on Pricing Strategies and Production and Sales Models of Vegetable Wholesale Market from the Perspective of Supply Chain [D]. Central South University of Forestry and Technology, 2022

[6] Shou kui Si, Xi ting Sun. Mathematical Modeling Algorithms and Applications [M]. National Defense Industry Press,2011.

[7] Lee D. Review of correlation studies [J]. 2021.

[8] Feng Bao. Research and application of solving algorithms for nonlinear programming problems based on inequality constraints [D]. Nanjing University of Finance and Economics,2022.

[9] Liu Bin, Liu Zengjie, Liu Yu, et al. Review of data visualization. Journal of Hebei University of Science & Technology, 2021, 42(6).

[10] Yang Haimin, Pan Zhisong, Bai Wei. Review of time series prediction methods [J]. Computer Science,2019(1):21-28.