

Replenishment And Automatic Pricing Strategy for Vegetable Products Based on Genetic Algorithm

Yicheng Sheng^{*, 1}, Zhiwen Zhang¹, Zhenru Jin²

¹ Jilin University Economics School, Jilin University, Changchun, China, 130012

² College of Automotive Engineering, Jilin University, Changchun, China, 130025

* Corresponding Author Email: shengyc0621@mails.jlu.edu.cn

Abstract. In order to improve the accuracy of replenishment and pricing of vegetable commodities sold in supermarkets, and to maximise the benefits of vegetable commodities sold in supermarkets, this paper comprehensively establishes a strategic planning model for replenishment and automatic pricing of supermarkets. Firstly, we carry out descriptive statistical analysis and non-linear fitting on the historical sales data of supermarkets to find the distribution pattern of the sales of each single product of vegetables, and then find out the relationship between each single product of vegetables through correlation analysis. and then find out the relationship between each vegetable item through correlation analysis. After that, we built an LSTM time series forecasting model based on the cost data to predict the replenishment volume of the six vegetable categories in the coming week. Finally, in order to maximise the revenue of the superstore, we constructed a pricing planning model for the combination of vegetable categories, and solved the optimal purchase quantity and pricing strategy based on genetic algorithm.

Keywords: Vegetable Commodities, Automation, Pricing Planning, Spearman Correlation Coefficient, LSTM Model, Genetic Algorithm.

1. Introduction

With the continuous improvement of people's living standards, people's demand for fresh food is also getting higher and higher, especially for vegetable commodities [1]. Due to the short freshness period, easy to change, inconvenient transport and other characteristics of vegetables, superstores can only choose the strategy of selling vegetables on the same day of purchase [2]. In order to ensure the quality and freshness of the vegetables sold on the same day, most superstores choose to purchase and trade vegetables from 3:00-4:00 a.m. However, there are many types of vegetables and different types of vegetables with different origins, so merchants decide to replenish the vegetables on the same day without knowing the exact number of items and the price of purchasing [3]. Therefore, an accurate replenishment and automatic pricing strategy is very necessary for supermarkets[4]. Ji Shoufeng established a non-linear planning and in order to improve the accuracy of the replenishment and pricing of vegetable products sold in superstores, this paper comprehensively established a replenishment and automatic pricing strategy planning model, and based on the genetic algorithm solved the model, to obtain the replenishment and automatic pricing strategy to maximise the revenue of superstores [5-6] .

2. Data exploration of vegetable sales and unit price data

2.1. Descriptive statistics on unit prices and sales of vegetables

According to the website www.mcm.edu.cn, we obtained the item code and classification name of each category of vegetables in a large supermarket, as well as information on sales volume, sales unit price, and sales type. As there are a certain number of outliers in the data, we carried out a series of data cleaning and integration, and after obtaining the sales data on a monthly basis, we carried out descriptive statistical analysis, and obtained box plots of the sales volume and unit price of each category of vegetables, which are shown in Figure 1. and Figure 2. below.

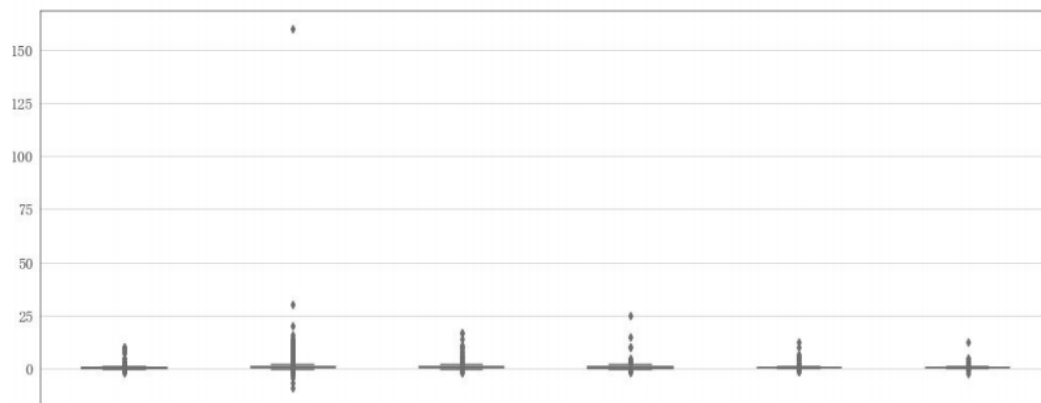


Figure 1. Box plots of sales by category

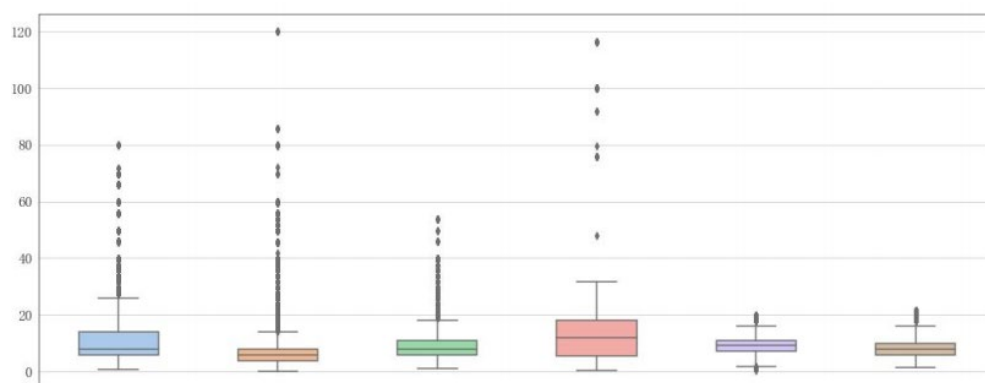


Figure 2. Box plots of unit prices of vegetables by category

2.2. Distribution pattern of vegetable sales volume by category based on time trend analysis

2.2.1 Overall analysis of the vegetable category

Analyse the three years of sales of vegetables in this superstore, with a monthly unit, respectively, the change of six categories of goods in each month, through the visual processing more convenient to study the sales distribution pattern of vegetable categories. As shown in Figure 3.

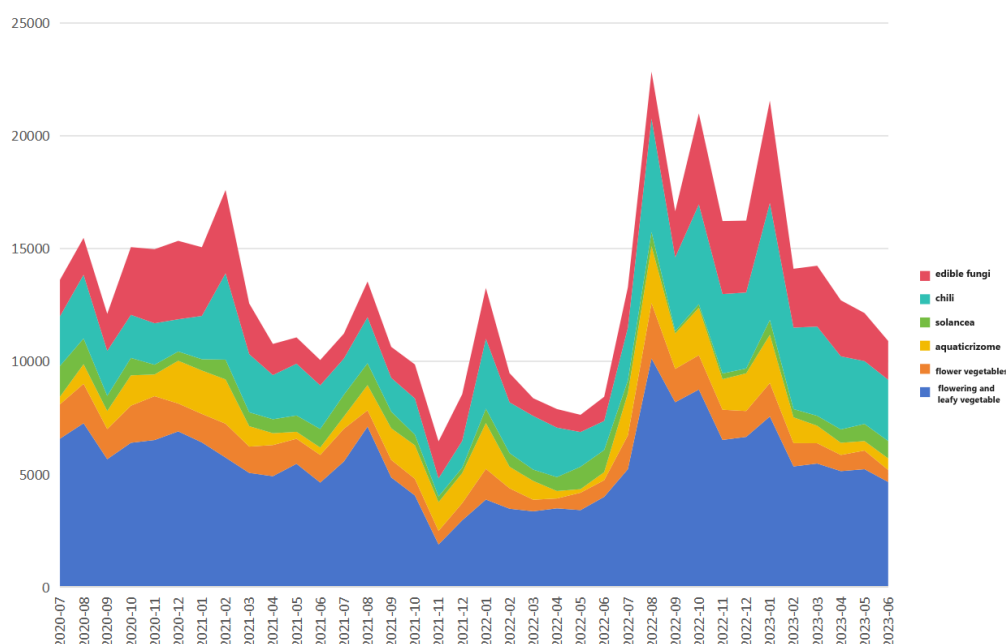


Figure 3. Monthly stacking chart of vegetable category

As can be seen from the above figure, from the category sales volume: flower and leaf data in the six vegetable species species with the highest sales volume, eggplant data with the smallest sales volume; as a whole, the data fluctuations and changes are relatively large, in March 2021 2022 during the period of June data overall sales volume is relatively low, may be affected by the epidemic and other factors; after June 2022, the data show a trend of growth; every year from September to the next year's January for the The peak season for sales is from September to January, and sales are generally low from March to July each year.

2.2.2 Time Trend and Distribution Pattern of Sales Volume of Vegetables by Category

On the basis of the above analysis, we analysed the overall trend of change in the sales volume of vegetables in each of the six major categories, observing the overall trend of change in the vegetable category over time through non-linear fitting on a month-to-month basis, judging the growth trend of the change based on the positive and negative situations of the coefficients of the highest terms of the fitted non-linear regression equations, and judging the magnitude value of the change based on the size of the absolute value of the coefficients. The results of the fitting are shown in Figure 4. and Figure 5.

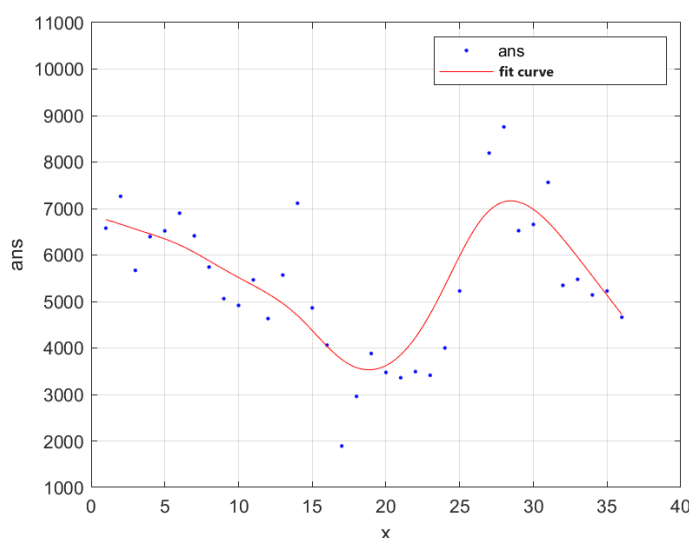


Figure 4. time trend fit of cauliflower vegetable sales

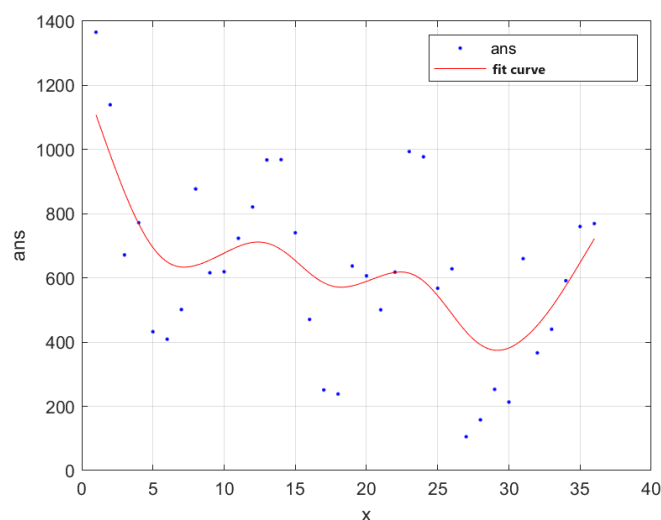


Figure 5. Time trend fit of sales volume of eggplant vegetables

Due to space limitation, we only show the fitting results for cauliflower vegetables and eggplant vegetables. From the fitting results: the sales volume of cauliflower, eggplant and leafy vegetables

showed a decreasing trend, the sales volume of chilli and edible mushroom vegetables showed a growing trend, and the sales volume of aquatic root vegetables showed a flat growth trend. From the sales volume time trend: the sales volume of each category has its own peaks and troughs in a year, showing a clear seasonality. The sales volume of the category of foliage and flowers peaked at specific times of the year, probably related to certain festivals or seasonal events.

2.2.3 Distribution pattern of sales volume of vegetables by individual product based on descriptive statistical analysis

We analyse the descriptive statistics of the data over time and analyse descriptive statistics such as mean, maximum, minimum, standard deviation, coefficient of variation, skewness coefficient, kurtosis coefficient and other descriptive statistics, which are used to study the overall situation of the data of the single product of vegetables. The calculated results are shown in Table.1.

Table.1. Table of descriptive statistical values for individual items in the vegetable category

Vegetable singles	average value	maximum values	minimum value	(statistics) standard deviation	skewness factor	kurtosis (chemistry)	coefficient of variation
Green Pepper	782.34	1632.92	0	489.62	-0.21	-0.67	0.63
broccoli	764.92	1656.34	422.35	268.96	1.43	2.79	0.35
Brassica pekinensis	532.98	3706.82	0	885.59	2.27	5.04	1.66
Yunnan lettuce	441.96	1287.41	0	355.65	0.49	-0.62	0.80
purple aubergine	433.22	703.36	67.51	150.99	0.01	-0.43	0.35

Due to space constraints, we show only five of the vegetables. We can find that the average sales volume of individual vegetable categories has increased over time, and the median sales volume has increased substantially from before. In terms of sales volume by item category, turnip peppers, broccoli, and net root are the top three products in terms of sales volume.

According to the analysis of the above descriptive statistics results, it can be seen that the trend of distribution in the data set is to the left, showing a non-normal distribution trend, which does not conform to the normal distribution, and the outliers can not be eliminated. Therefore, we took the six major vegetable categories as reference and analysed the Spearman correlation coefficients of different individual categories in each major category to observe the trend. The results are shown in Figure 6.

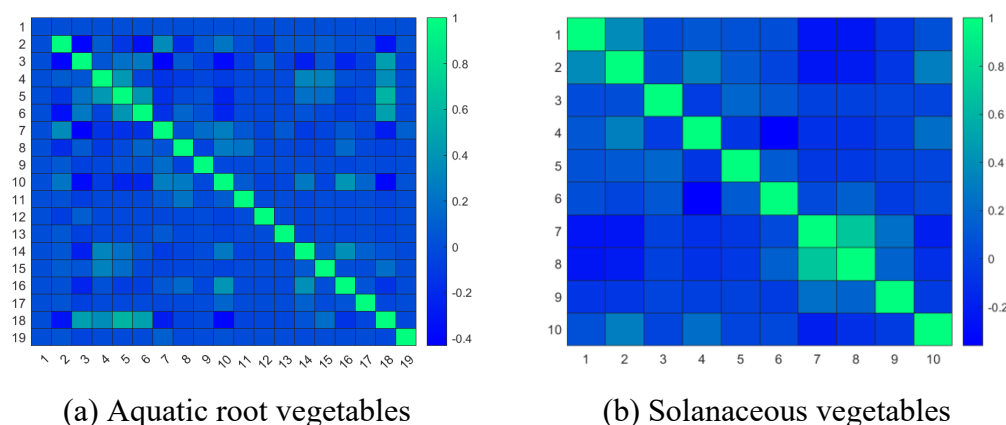


Figure 6. Spearman's correlation coefficient plot between vegetable-like items

From the data in the above table and the above figure, we can conclude that in the category, the rule of change is as follows: leafy vegetables, the strongest correlation coefficient is: $r = 0.932146$, corresponding to a single category of vegetables for: Yunnan lettuce (copies)-Yunnan oleander (copies); cauliflower vegetables, the strongest correlation coefficient is: $r = 0.409635$, corresponding to a single category of vegetables for: green peduncle scattered flowers -Zhijiang green peduncle scattered flowers; In aquatic root vegetables, the strongest correlation coefficient is: $r = 0.579647$, corresponding to a single category of vegetables for: rhododendron - Honghu Lotus Roots; in eggplant vegetables, the strongest correlation coefficient is $r = 0.692479$, corresponding to a single category of vegetables for: green aubergine - purple aubergine; in chilli vegetables, the strongest correlation coefficient is: $r = 0.878787$, corresponding to a single category of vegetables for: small peppercorns (copies) - small wrinkled peppers (copies) - small wrinkled peppers (copies) - small wrinkled peppers (copies); among edible mushroom vegetables, the strongest correlation coefficient is $r = 0.894443$, corresponding to the single category of vegetables: enoki mushrooms - apricot mushrooms.

3. Replenishment and Pricing Strategies for Vegetable Commodities Solved

3.1. Analysis of the relationship between sales volume and cost-plus pricing of vegetable categories based on regression analysis

3.1.1 Calculation of variables

(1) Daily sales volume by category

The daily sales of the six categories are denoted as S_i , by summing the total daily sales of each individual item under that vegetable category.

$$S_i = \sum_{i=1}^n s_i \quad (1)$$

Where s_i denotes the total daily sales of each individual item under the vegetable category.

(2) Daily wholesale prices by category

The daily wholesale price C_i for the six categories is determined by the wholesale price of individual items, weighted by the percentage of total sales of each individual item.

$$C_i = \frac{s_i}{\sum_{i=1}^n s_i} \quad (2)$$

Where C_i is the daily wholesale price of each individual item.

(3) Loss rate per category

The wastage rate L_i for the six categories is determined by the wastage rate of individual items, weighted by the percentage of total sales of each individual item.

$$L_i = \frac{s_i}{\sum_{i=1}^n s_i} \times l_i \quad (3)$$

Where l_i is the attrition rate for each individual item.

(4) Cost price per category

Express the actual cost price of the superstore based on the wastage rate of each individual item in Annex IV

$$\dot{C}_i = \frac{c_i}{1-L_i} \quad (4)$$

3.1.2 Modelling and Solving

Based on the trend analysis of the sample observations, the following model is proposed.

$$P = \alpha S_i^2 + \beta S_i + \mu \quad (5)$$

Parameter estimation was carried out using least squares estimation and the preliminary fitted regression equations for the six vegetable categories were calculated by MATLAB as follows:

$$P_1 = -0.0001S_i^2 - 0.0417S_i + 134.1269 \quad (6)$$

$$P_2 = -0.0006S_i^2 + 0.1058S_i + 12.9922 \quad (7)$$

$$P_3 = 0.0014S_i^2 - 0.3419S_i + 42.4140 \quad (8)$$

$$P_4 = -0.0041S_i^2 + 0.4140S_i + 21.1211 \quad (9)$$

$$P_5 = 0.0006S_i^2 - 0.2387S_i + 127.9863 \quad (10)$$

$$P_6 = -0.0005S_i^2 + 0.3075S_i + 60.3900 \quad (11)$$

3.2. Vegetable category mix pricing model for revenue maximisation in hypermarkets based on genetic algorithm

3.2.1 Cost Forecasting Analysis of Vegetables Category Based on LSMT Model

Long short-term memory (LSTM)^[7] is a special variant of Recurrent neural network (RNN). LSTM has a "gate" structure, through the logical control of the gate unit to decide whether to update the data or choose to discard, overcoming the shortcomings of RNN, such as excessive influence of weights, easy to produce gradient disappearance, so that the network can be better and faster convergence, and can effectively improve the prediction accuracy. The LSTM has three gates, the forgetting gate, the input gate, and the output gate, which determine how much information is remembered and forgotten at each moment. The input gate determines how much new information will be added to the cell, the forgetting gate controls whether the information will be forgotten at each moment, and the output gate determines whether the information will be output at each moment.

The univariate prediction was carried out using the foliage data as an example, setting the moving step to 7 for scrolling, and using the fitted LSTM model to predict the future values of the six vegetable categories respectively, and the results are shown in Table .2.

Table .2. 2023-7-1 to 2023-7-7 Cost Projections

	7-1	7-2	7-3	7-4	7-5	7-6	7-7
philodendron	67.8439	67.7840	72.8709	71.8678	72.3778	72.9050	74.4463
cauliflower	20.1382	17.9367	17.8940	15.9680	16.6626	16.0103	14.9568
eggplant	20.9335	21.1342	21.5248	21.5148	22.1163	21.2545	21.0076

3.2.2 Modelling and Solving

Based on the information provided, we can model the problem as an optimisation problem, i.e. finding the maximum profit gain, given the daily replenishment and pricing strategy. We can use linear programming to solve this problem and use genetic algorithm to solve it. In order to maximise the profit gain of the superstore, we analyse the sales volume and profit per item of the superstore to establish the objective function as follows:

$$Q = [C_1 - (1 + L) \times C_2] \times S_1 \quad (12)$$

Where C_1 denotes the cost, C_2 denotes the cost after prediction by LSTM neural network, S_1 denotes the sales volume, and L denotes the attrition rate. With the restriction that the wastage rate is not less than zero, we construct a portfolio pricing model to maximise the revenue of the superstore, and use genetic algorithm^[8-9] to solve the model, the specific process is shown in Figure 7.

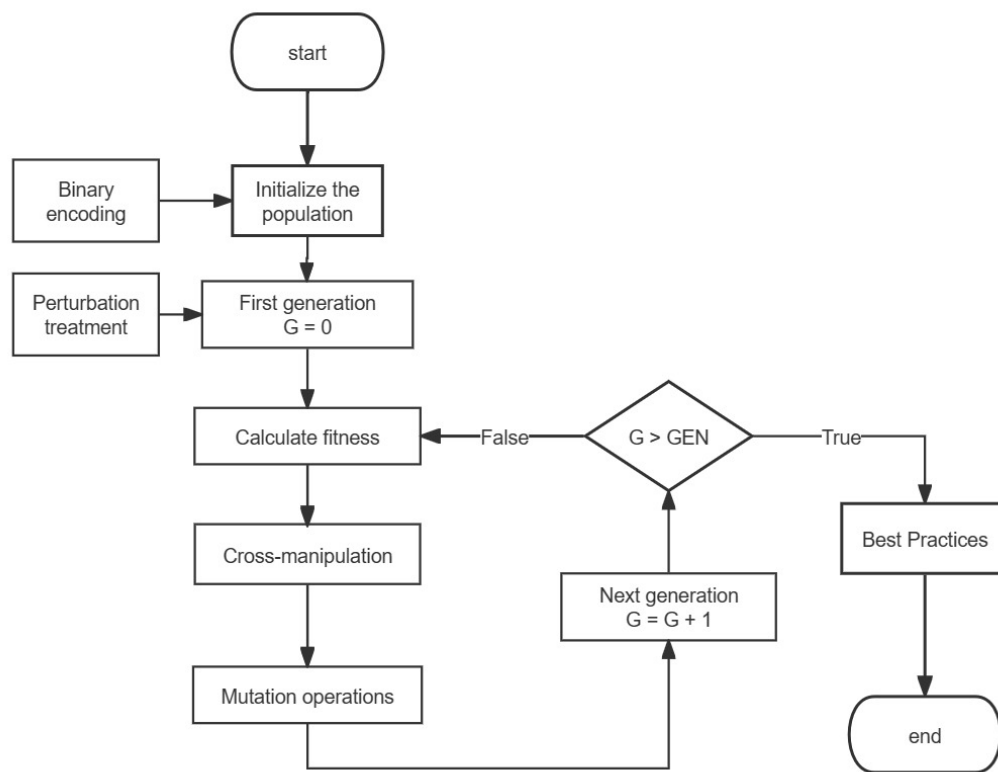


Figure 7. Flowchart of genetic algorithm

MATLAB was used to program the solution and the final resultant pricing strategy was brought into the model to solve the shipping case. The final results are shown in Table .3. and Table .4.

Table .3. 2023-7-1 to 2023-7-7 replenishment strategy

	7-1	7-2	7-3	7-4	7-5	7-6	7-7
philodendron	159.7414	159.9975	160.0649	159.9614	160.1486	159.9048	160.2002
cauliflower	2.07802	22.1002	22.08577	22.0472	22.08922	21.90044	22.09229
Aquatic rhizomes	11.05334	10.94375	10.82813	10.80803	10.79247	10.78517	10.77935
eggplant	17.07341	17.08081	17.06499	17.06977	17.0663	17.06447	17.06293
chilli	76.95735	76.7876	77.96742	76.81373	76.98788	76.8376	77.00889
edible mushroom	54.24866	52.53689	54.98256	53.4243	51.83144	53.38439	52.05502

Table .4. 2023-7-1 to 2023-7-7 Pricing strategy table

	7-1	7-2	7-3	7-4	7-5	7-6	7-7
philodendron	123.23	121.89	122.36	121.98	122.03	122.35	122.42
cauliflower	11.86	11.86	11.86	11.86	11.86	11.86	11.86
Aquatic rhizomes	81.98	81.98	81.98	81.98	81.98	81.98	81.98
eggplant	28.15	28.15	28.16	28.16	28.16	28.17	28.17
chilli	44.13	44.13	44.13	44.13	44.13	44.13	44.13
edible mushroom	38.99	38.99	38.99	38.99	38.99	38.99	38.99

4. Conclusion

In this paper, an in-depth data mining has been carried out in terms of data preprocessing, according to different time dimensions for days, months, and,quarterly and yearly data are summarised and classified for discussion, which helps to derive the rule of change from different time

dimensions; when analysing the distribution law of changes in vegetable categories and individual products, in-depth data exploration is carried out from different perspectives such as the trend of changes in vegetables, the frequency of changes and the periodicity of changes in vegetables, which helps to observe the change situation from different levels. When analysing the interrelationships between vegetable categories as well as individual products, the analysis was carried out through Spearman's correlation coefficient, and the magnitude of the correlation coefficient was used to determine the strength of the interrelationship between the two. The LSTM model and genetic algorithm were also used to predict and calculate the stocking strategy and pricing strategy of various vegetables in the superstore for the coming week, and it resulted in a 10% increase in the average weekly profit of the superstore.

However, in terms of solving, we used heuristic algorithms to obtain approximate solutions, which cannot obtain accurate solutions; in future research, the use of heuristic algorithms can be avoided when solving optimisation problems. Secondly, the robustness[10] of the model is not strong; in future research, methods such as Label Smoothing can be used to enhance the robustness of the model.

References

- [1] Huang Guanyu. Design and Implementation of Supermarket Fresh Food Purchasing and Distribution System [D]. Xiamen University,2019.
- [2] Mao Lisha. Research on Pricing Strategy and Production and Marketing Mode of Vegetable Wholesale Market under the Perspective of Supply Chain [D]. Central South Forestry University of Science and Technology,2023.
- [3] Zhou Yaying. Research on fresh food game [D]. Shanghai University of Finance and Economics,2022.
- [4] Wang Junxia. A review of research on joint replenishment based on perishables grouping strategy[J]. China Storage and Transportation,2023(08):97-98.
- [5] Chen Jinxiao,Chen Jian. Profit game and equilibrium pricing in supply chains[J]. China Management Science,2022,30(09):128-139.
- [6] Miao Jun. Research on dynamic pricing strategy of fresh food [D]. Nanjing University of Finance and Economics,2020.
- [7] WANG Tao, WANG Xu, XU Ye et al. Study on LSTM photovoltaic power prediction model taking similar days into account [J]. Journal of Solar Energy,2023,44(08):316-323.
- [8] Qingsong Feng, Long Wang, Kui Sun et al. multi-objective optimization of ballastless track structure of city express line based on NSGA-II genetic algorithm [J/OL]. Railway Standard Design:1-8.
- [9] Hou Rongfeng. Optimisation analysis of fresh food distribution path based on genetic algorithm [J]. Railway Purchasing and Logistics,2023,18(02):56-59.
- [10] Baofeng Tian,Jingwen Zhang,Zhiyao Shi. Robust bi-objective critical chain project scheduling optimization for eliminating secondary resource conflicts [J/OL]. Journal of Management Engineering,2024(02):1-14.