# LR-GA Algorithm Based Study on Vegetable Replenishment and Pricing Decision Making

Zixuan Liu*, Xiaoyu Yang

North Alabama International College of Engineering and Technology, Guizhou University, Guizhou, 550025, China

* Corresponding Author Email: xuan001166@163.com

**Abstract.** The sales volume and sales price of vegetable products fluctuate greatly due to factors such as origin, variety and freshness period. Therefore, rational replenishment and pricing decisions are particularly important for supermarkets. Therefore, this paper proposes a replenishment and pricing strategy model for vegetable products based on linear regression-genetic algorithm. In this paper, all vegetable commodities are first classified into four categories using K-mean cluster analysis, and the sales performance of the four categories is observed to understand their sales behaviors and provide references for sales strategies. Then, the cost, price and profit margin of each vegetable commodity are calculated using the weighted method, and the linear regression equation between the sales volume of vegetable commodities and the weighted sales price is given. Finally, using the total profit formula considering the loss rate of each item as the objective function and the linear regression equation between the total sales volume of each vegetable item and the weighted sales price as the constraints, the optimization search is carried out by using linear regression and Genetic Algorithm (LR-GA) to find out the sales price and the sales volume under the maximum profit so as to realize the sales strategy.

**Keywords:** Vegetable Pricing, K-means, Linear Regression, Genetic Algorithm.

## 1. Introduction

Fresh vegetables have become an indispensable business item in the development of supermarkets, taking residents' daily consumption as the starting point [1]. Fresh vegetables have become an indispensable business item in the process of supermarket development Fresh vegetables have become an important area for domestic and foreign supermarkets to compete for the market. Fresh vegetable business has become an important field for domestic and foreign supermarkets to compete for the market [2-3]. Although the fresh vegetable business of supermarkets has great potential and significance, fresh vegetables are characterized by perishability, easy deterioration and constant loss of value. However, fresh vegetables are characterized by perishability and constant loss of value [4]. Ensure to improve the quality and safety of fresh vegetables and maintain the reasonable price of fresh vegetables in supermarkets. Maintaining the reasonable price of fresh vegetables in supermarkets and utilizing the "customer gathering power" of fresh vegetables while increasing the profitability of fresh vegetable business will help promote the modernization of the retail industry as well as the development of the retail industry. Only by doing so will it help to promote the modernization of the retail industry and enhance the competitiveness of the city. Therefore, supermarkets at home and abroad are committed to solving the various conflicts faced by supermarkets in the operation of fresh vegetables [5-6].

# 2. Research Methods and Results

## 2.1. Data pre-processing

The data in this paper comes from http://www.mcm.edu.cn/html_cn/node/c74d72127066f510a5723a94b5323a26.html.The following table 1 shows the classification of vegetable items into category data, which contains the date of sale, item code, sales volume, sales unit price and category information.

**Table 1.** Data preprocessing for single product categorization into categories

| Date | product code | Sales volume (kg) | Sales unit price (¥/kg) | category |
|---|---|---|---|---|
| 2020-07-01 | 102900005117056 | 0.396 | 7.60 | capsicum |
| 2020-07-01 | 102900005115960 | 0.849 | 3.20 | philodendron |
| 2020-07-01 | 102900005117056 | 0.409 | 7.60 | capsicum |
| ... | ... | ... | ... | ... |
| 2023-06-30 | 102900005115250 | 0.125 | 24.00 | mushrooms |

Also below is the order quantity and pricing strategy for July 1 based on the 49 vegetable items from June 24-30, 2023 for the vegetable items. The data contains the cost, price, sales volume, and profit margin for each vegetable item. In order to more accurately reflect the economic performance of each category, the idea of weighting is used to recalculate the price, cost and profit margin of each category. The calculation steps are as follows: first, get the proportion of each item's daily sales volume to the total sales volume, i.e., the "normalization factor", and the results are shown in Table 2.

**Table 2.** Normalization coefficients for vegetable commodities

| | Date | product code | normalization factor |
|---|---|---|---|
| 1 | 2023-06-24' | 102900005115250 | 0.190923414 |
| 2 | 2023-06-25' | 102900005115250 | 0.131947233 |
| 3 | 2023-06-26' | 102900005115250 | 0.110661435 |
| ... | ... | ... | ... |
| 281 | 2023-06-28' | 106971533450003 | 0.096774194 |
| 282 | 2023-06-29' | 106971533450003 | 0.177419355 |
| 283 | 2023-06-30' | 106971533450003 | 0.096774194 |

Next, the coefficients were used to calculate the weighted wholesale price, weighted selling price, and weighted cost margin for a simplified treatment of the data. Total sales, total profit, weighted wholesale price, weighted selling price, and weighted cost margin are then calculated for each group.

## 2.2. K-means cluster analysis

In addition to using Pearson's correlation coefficient, this paper also performs K-means cluster analysis based on the sales volume data of each single product in 2022, selecting the annual sales volume of a single product and the highest monthly sales volume of a single product as features.K-means cluster analysis is an unsupervised learning algorithm that is used to group the input data, so that the similarity of the data points within the same group to each other is maximised and the similarity of the data points between different groups is minimised.K-means cluster analysis can help us to perform data exploration and data compression[7-8]. This similarity is usually measured by calculating the distance between the data points.K-means clustering helps us in data exploration, data preprocessing, data compression etc.

The formula (1) for the clustering effect evaluation index SSE for K-means cluster analysis is as follows:

$$\begin{cases} SSE = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2 \\ \mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \end{cases} \tag{1}$$

where SSE denotes the sum of all sample points to the cardinality distances of the categories they belong to, with smaller values indicating better aggregation, x is the current sample point, $C_i$ is the set of all sample points in the ith category, and $\mu_i$ is the centre of mass of the ith category.

Prior to the analysis, all feature data were normalised to remove the effect of data range differences on the results. In the first table, each row represents the highest sales volume of an item in a particular month, and this volume was normalised to obtain the results shown in Table 2 (only partially shown for space reasons).

Prior to analysis, all characteristic data were normalized to eliminate the effect of data range differences on the results. This normalization makes it easier to compare sales of different items and analyze the sales characteristics of each item, such as seasonality and trends. These analyses are then combined with forecasting models, such as ARIMA and Random Forest, to predict the future sales volume of each item and develop replenishment levels and pricing strategies accordingly.

After normalisation, the normalised data was subjected to K-means cluster analysis and the results are shown in **Figure 1.**

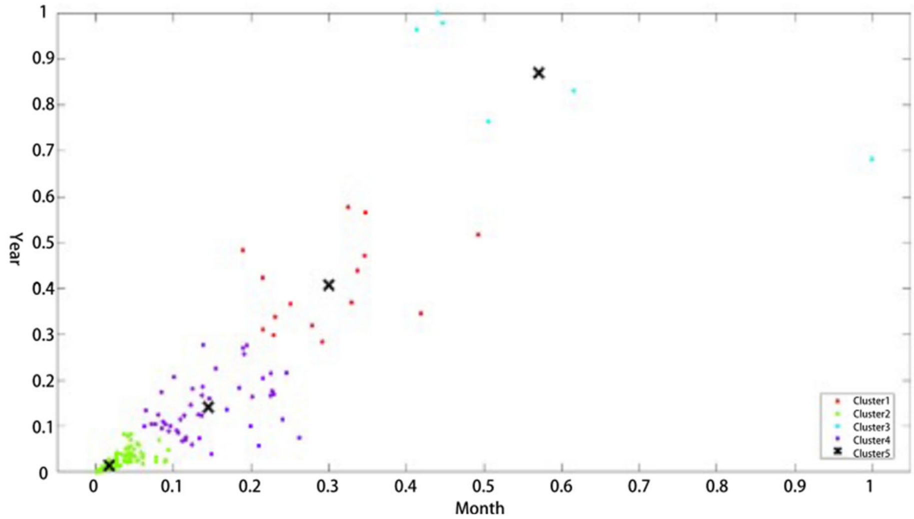


**Fig. 1** Cluster analysis of annual sales volume of a single product and its highest monthly sales value

Looking at Figure 1, it can be observed that the highest monthly and annual sales volumes for the single item category numbered 1 form a major cluster in the lower left portion of the chart. This implies that the highest monthly sales volume and annual sales volume of these individual items show similarity, and that the range of fluctuation of these sales volumes is small and the values are relatively small, which may indicate that the annual sales of these individual items are underperforming. In contrast, the single-item category numbered 3 has a higher maximum monthly sales volume and annual sales volume, forming a smaller cluster. However, this cluster is generally farther away from the centre, possibly reflecting the fact that these individual items have special sales characteristics, for example, they may be affected by seasonal factors and therefore have higher than average sales in some months and lower than average sales in other months, resulting in their annual sales being below the average, or that the sales of the item are relatively stable in every month of the year, resulting in the highest monthly sales being low, but the annual sales at a high level. Finally, the single item category numbered 4 has both annual sales and highest monthly sales well above the average, forming another cluster, which may indicate that this type of single item is performing well. As well as from the charts, it can be analysed that most of the individual products' highest monthly sales and annual sales are clustered at lower values, as well as being more affected by the seasons, and only a few of the individual products' highest monthly sales and annual sales are at higher values, and are also less affected by the seasons. In summary, through K-means cluster analysis, the different performance and characteristics of each individual product in terms of sales volume can be observed, and at the same time, the influence of different vegetable individual products on the highest monthly sales volume and annual sales volume is reflected.

# 3. Vegetable commodity pricing and replenishment strategies

## 3.1. Regression Equation for Vegetable Unit Sales Price and Volume of Sales

Through the June 24-30, 2023, only 49 kinds of saleable varieties, the establishment of 49 kinds of single product respectively linear regression model[9-10], in order to approximate the solution of the relationship between the single product sales price and sales volume, the process of reference to solve the category sales price, the results are shown in Table 3.

**Table 3.** Relationship between selling price and sales volume of different individual products corresponding to individual products

| | Name of Product | Linear Regression Equation |
|---|---|---|
| 1 | Yunnan oilseed rape (Brassica campestris L.) | sales = -1.21*price + 21.55 |
| 2 | Yunnan oilseed rape (portion) | sales = -0.57*price + 19.23 |
| 3 | Yunnan lettuce | sales = -0.94*price + 26.70 |
| ... | ... | ... |
| 47 | Zijiang Qingdian Scattered Flowers | sales = -0.80*price + 22.25 |
| 48 | Honghu Lotus Roots | sales = -0.15*price + 7.45 |
| 49 | Seafood Mushroom(Pack) | sales = -0.74*price + 11.29 |

## 3.2. Optimal price and replenishment volume selection

The objective function for solving the maximum return is established as in equation (2), where$W_2$ denotes the profit of a single product, and$S_j$ denotes the sales volume of a single product, and$P_j$ denotes the selling price of a single product, and$B_j$ denotes the wholesale price of a single product , and$L_j$ denotes the loss rate of a single product.

$$W_2 = \sum_{j=1}^{\alpha}\left(S_j P_j - S_j \frac{B_j}{1-L_j}\right)$$
$$\text{St.} \quad 27 \le \alpha \le 33, \quad S_j = kP_j + b \tag{2}$$

Through genetic algorithm, the expected return at the corresponding price is obtained, i.e., demand multiplied by (selling price - cost). Finally, the price with the highest expected return is selected as the optimal price, and the corresponding demand is used to find the optimal replenishment quantity, let the optimal replenishment quantity of each individual product is$C_k$ The sales volume is$S_k$ and the corresponding wastage rate is$L_k$ Solve the equation as follows:

$$\begin{cases} C_k = \frac{S_k}{1-L_k} \\ C_k > 2.5 \end{cases} \tag{3}$$

With this approach, the optimal replenishment volume and pricing strategy can be successfully found for each product, thus maximizing the expected revenue.

### 3.3. Analysis of results

The finalized effective replenishment plan and pricing strategy is shown in Table 4:

**Table 4.** Optimal commodities and optimal replenishment and pricing strategies

| | category | weighted wholesale price(¥) | Replenishment(kg) | Price(¥) | projected revenue(¥) |
|---|---|---|---|---|---|
| 1 | capsicum | 7.80 | 33.74 | 14.5 | 226.09 |
| 2 | philodendron | 3.50 | 54.56 | 6.95 | 188.33 |
| 3 | mushrooms | 6.46 | 39.11 | 10.49 | 157.65 |
| ... | ... | ... | ... | ... | ... |
| 31 | lettuce | 4.10 | 6.43 | 8.15 | 26.09 |
| 32 | eggplant | 2.44 | 14.08 | 4.26 | 25.64 |
| 33 | cucumbers | 1.96 | 12.64 | 3.84 | 23.77 |

As can be seen from the table, the model makes detailed predictions and optimizations for each item. For example, for capsicum species, the model predicts that its optimal replenishment volume on July 1, 2023 is 33.74kg, its optimal pricing is ¥14.5, and its projected revenue is ¥226.09. These data provide firms with a clear guideline on what replenishment volume and selling price they should adopt to maximize revenue. With the pricing strategies in Table 4, the projected returns for each individual item on July 1 are summed to obtain a maximum projected return of ¥2205.3 for the superstore on July 1st.

## 4. Conclusion

In this paper, all vegetable commodities are first classified into four categories using K-mean cluster analysis, and the sales performance of the four categories is observed to understand their sales behaviors and provide references for sales strategies. Then, the cost, price and profit margin of each vegetable commodity are calculated using the weighted method, and the linear regression equation between the sales volume of vegetable commodities and the weighted sales price is given. Finally, using the total profit formula considering the loss rate of each item as the objective function and the linear regression equation between the total sales volume of each vegetable item and the weighted sales price as the constraints, the optimization search is carried out by using linear regression and Genetic Algorithm (LR-GA) to find out the sales price and the sales volume under the maximum profit so as to realize the sales strategy.

In this paper, the effectiveness of vegetable replenishment and pricing strategies may be affected by many factors, such as changes in market demand, competitors' strategies and so on. In order to optimize the replenishment and pricing strategy, the next step can also collect historical sales data, consumer behavior data, price sensitivity data, etc. for analysis.

## References

[1] Songheng Ji, Qinhong Zhang. A study on joint decision making of presale inventory and pricing in fresh food e-commerce considering consumers' freshness preference[J]. Logistics Engineering and Management, 2021, 043(008):P.116-125.

[2] WANG Shuzhong, LIU Yongxia, WANG Yizhong, et al. 40 years of vegetable industry development in reform and opening up Beijing[J]. 2022(3).

[3] Wei Danlin. Research on vegetable income insurance risk and pricing in Shandong[J]. Rural Economy and Technology, 2021.

[4] Tan Lang Chu Futang Xiang Hongmei Yi Youguang. Current situation and suggestions for the development of vegetable advantageous single product in Bijiang District[J]. Agricultural Technology Service, 2022, 39(5):99-101.

[5] Zhang Weihong. Development status of alpine vegetables in Chongqing [J]. Southern Agriculture, 2022, 16(7):26-28.

[6] WANG Haoyu, GAO Peichao, XIE Yiru, et al. Land use optimization based on genetic algorithms: a comparative study of NSGA-II and NSGA-III[J]. Journal of Ecology, 2023, 43(2):639-649.

[7] Ren Jianchao,Feng Yanjun,Lv Yang,et al. Spatial-temporal analysis of microseismic energy clusters based on K-Means fuzzy clustering for rock overburden transport in quarries[J]. Energy and Environmental Protection, 2023, 45(3):308-312.

[8] Chen Ming, Li Junxiang, Qu Deqiang, et al. Planning of urban emergency logistics facilities based on K-means clustering algorithm - An example of vegetable delivery data at a certain stage in Changchun City[J]. Logistics Science and Technology, 2023, 46(17):57-60.

[9] Cai Chengwang, Pan Zhiying, Cai Xiaoqing, et al. Research on coke quality prediction model based on linear regression[J]. Contemporary Chemical Research, 2023(1):1-4.

[10] Shuang Hongchun, Dai Yue, Anna, et al. A linear regression-based longitudinal protection method for flexible DC networks[J]. Journal of Electrotechnology, 2022, 37(13):15.