Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 8, No. 3, 2023

Pricing and Replenishment Decision of Vegetable Products

Le Fang*, Kun Zhang

Stony Brook Institute at Anhui University, Anhui University, Anhui, 230039, China
*Corresponding author: lefang18909664046@163.com

Abstract: Most vegetable products have a short shelf life and need to be restocked according to demand after being sold on the same day. Therefore, it is of great significance to accurately price and make effective replenishment decisions for vegetable products under unknown specific unit and purchase prices. This article analyzes the sales information and wholesale prices of various types of vegetables. Using the system clustering method in cluster analysis, the Euclidean distance and shortest distance are obtained based on the characteristic values between different individual products, and the sales volume correlation is classified according to their strength. Then, the chi square test and Fisher exact probability test are used to obtain the sales volume correlation based on the characteristic values between two different categories. The paper constructs a linear programming model with the goal of minimizing daily transportation costs.

Keywords: Chi-square test, ARIMA model, Logistic regression, Grey prediction, BP neural network.

1. Introduction

In fresh food supermarkets, ordinary vegetable products have a short shelf life and their appearance may deteriorate over time. Most varieties must be sold on the same day, so daily restocking is essential[1]. The variety of vegetable supply is relatively abundant from April to October, and the sales volume of vegetable products is closely related to the time, so reasonable replenishment and pricing decisions are particularly important.

Based on the product information of six vegetable categories sold in a supermarket, this paper establishes a mathematical model for different product sales details, wholesale prices, and recent loss rates from July 1, 2020 to June 30, 2020 and determine the distribution pattern and interrelationship of sales volume between different single products or categories of vegetables. Then we predict the daily replenishment total and pricing strategy for each vegetable category from July 1, 2023 to July 7, 2023, to maximize the revenue of supermarkets. In order to optimize the replenishment and pricing decisions of vegetable products, we propose some opinions and reasons for the relevant data that still needs to be collected by supermarkets.

Section I is the Introduction. Section II gives the preliminary work for the model. Section III and Section IV establish the model and analysis the specific problem. Section V, the Conclusion.

2. Preliminary

2.1. Assumption

1.Neglecting the impact of missing daily sales flow details for some products on the calculation of results.

2.The selected indicators have no significant impact on each other.

3.Neglecting the impact of other unconsidered factors.

4.Specific vegetables are only produced in specific cities.

The general model of artificial neural network consists of four basic elements, which are:

2.2. The determination of the number of network layers

BP neural network is back propagating, mainly composed of three parts: input layer, middle layer and output layer. The number of nodes in the input and output layers is relatively easy to determine, but the determination of the number of nodes in the hidden layer is a very important and complex problem.

Table 1. Symbols Notation

Symbols	Notation
λ	eigenvalue
$d(x_i, x_j)$	Euclidean distance
$D(G_p, G_q)$	Minimum distance
X^2	Chi square statistical value
v	Chi Square Statistical Value Degrees of Freedom
T_{ij}	Normalized data
p_{ij}	The daily selling price for the i-th category on the j-th day
u_{ij}	The i-th category, the daily purchase price on the j-th day
M_{ij}	Total sales of i-th category on day j
$g(z)$	Logistic function
I_{ij}	Profit contribution on day j for the i-th category
$z^{(i)}$	Group (i) operators in GM (1,1)
$p(k)$	Grade ratio deviation value

3. Model Establishment

3.1. Establishment of Principal Component Analysis Models for Each Category

1.Construct a sample matrix y with a size of 4 * n.

$$y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & \vdots & \vdots & \vdots \\ y_{31} & \vdots & \vdots & \vdots \\ y_{41} & y_{42} & \cdots & y_{4n} \end{bmatrix} = (y_1, y_2, \cdots, y_n) \quad (1)$$

2.Standardize the y-matrix. Calculate the mean and standard deviation of the matrix separately by column.

$$\overline{X_j} = \frac{1}{n}\sum_{i=1}^{y} X_{ij} \quad (2)$$

$$S_j = \sqrt{\frac{\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)^2}{n-1}} \quad (3)$$

3.The original sample matrix has been standardized to

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & \vdots & \vdots & \vdots \\ Y_{31} & \vdots & \vdots & \vdots \\ Y_{41} & Y_{42} & \cdots & Y_{4n} \end{bmatrix} = (Y_1, Y_2, \cdots, Y_n) \quad (4)$$

4.Calculate the principal component contribution rate and cumulative contribution rate[2]. Select the representative item of this category to represent the current category, with no more than 3 items in total.

3.2. The solutions

Build a system clustering model and use SPSS software to solve it. By clustering big data, it is found that there is a certain correlation between the sales volume of different vegetable items. After solving the model, it can be concluded that the significant level of correlation between cauliflower and the other four categories is 0.213, which is greater than 0.1. Therefore, cauliflower has a correlation with the other four categories, but the correlation is weak.

This paper uses the entropy weight method[3] to solve the weight vectors of each indicator$(\omega_1, \omega_2, \cdots, \omega_m)$. The results of obtaining weights using the entropy weight method are shown in the Table 1.

Table 1. The weights of each vegetables

Category name	ω_1	ω_2	ω_3
Florifolias	0.34	0.33	0.33
florescent vegetables	0.501	0.499	-
Aquatic rhizomes	0.4963	0.5037	-
Solanaceae	0.48	0.52	-
Edible fungi	0.51	0.49	-
Chili peppers	0.433	0.567	-

Establish a logistic regression mathematical model, and the specific solving steps are as follows.

$$\theta_0 + \theta_1 m_1 + \cdots + \theta_n m_n = \sum_{t=0}^{n} \theta m_i = \theta^T m \quad (5)$$

To control the output between 0 and 1, select the Logistic function.

$$g(z) = \frac{1}{1+e^{-2}} \quad (6)$$

The Loss Function of Logistic Regression[4] is

$$J(\theta) = -\frac{1}{s}\left[\sum_{i=1}^{s} u_i \log(h_\theta(m_i)) + (1 - u_i)\log(1 - h_\theta(m_i))\right] \quad (7)$$

3.3. White noise processing

By substituting the data again, the predicted results of the model increase steadily, and the residual is also white noise. All other classes pass the white noise test, that is, the model passes the white noise test. Perform ARIMA model[5] predictions for different vegetable categories, and the predicted results are shown in the following figure.

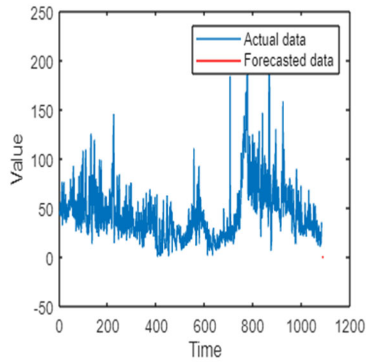

Figure 1. Forecast results(a)

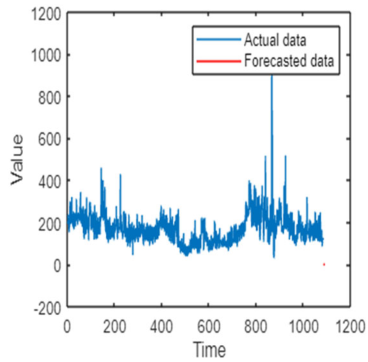

Figure 2. Forecast results(b)

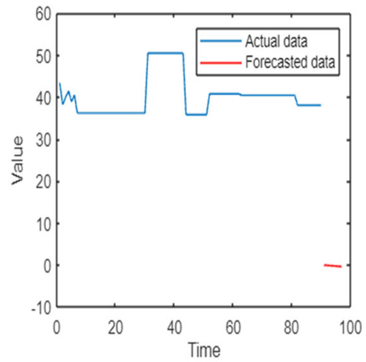

Figure 3. Forecast results(c)

By comparing the predicted value with the actual value, it was calculated that=0.892, indicating that the model has good predictive accuracy and can be used for prediction. Using the ARIMA model, predict the total daily replenishment volume and pricing strategy from July 1 to July 7, 2023.

3.4. Predicting the prices

To predict the replenishment volume and pricing strategy of individual products on July 1, 2023, grey GM (1,1) is used due to the limited sample data. To improve the accuracy of problem prediction, a BP neural network model needs to be combined. Therefore, this article constructs a combined mathematical model based on grey GM (1,1)[6] and BP neural network[7] for prediction.

4. Discussions

1. The logistic regression model is a non parametric model that does not require assumptions about the distribution of data and has strong adaptability and flexibility.

2. The time prediction series ARIMA model is relatively flexible and can achieve different types of predictions and analysis by adjusting the parameters in ARIMA (p, d, q).

3. The combination of grey GM (1,1) prediction model and BP neural network can comprehensively utilize their respective strengths to improve the accuracy and robustness of prediction.

4. The logistic regression model is sensitive to the presence of multicollinearity between features and requires measures such as feature selection and dimensionality reduction.

5. The grey GM (1,1) prediction model and BP neural network combined model are more complex than the single model, which requires more computing resources and time to train and adjust model parameters.

5. Conclusion

This paper analysis the prediction of variable vegetables' price based on the history data through PCA, BP neural network,etc. Based on the results, the model provides some useful strategy for businesses to make the profit biggest. We discuss the change of the prices and productions in different angles in the specific times, which is helpful for the managers to make up more economical plans. In addition to time series prediction analysis, ARIMA models can also be used in fields such as financial data analysis, meteorological prediction, and signal processing. The combination of grey GM (1,1) prediction model and BP neural network model can also be used in fields such as sales prediction, traffic flow prediction, and production planning.

Acknowledgment

The authors gratefully acknowledge the financial support from xxx funds.

References

[1] Tanatar M .EU HRC market activity slow as distributors avoid restocking[J].SBB daily eBriefing, 2023.

[2] Riibsamen R , Kopp C , Doerrscheidt G J ,et al.Principal component analysis applied to action potentials reveals neuronal interaction in auditory brainstem nuclei[J]. 2022.

[3] Wang L , Hu J , Hu Y .Investigation of the global stock trading based on visibility graph and entropy weight method [J]. Fluctuation and Noise Letters, 2023.

[4] Baffour B .Modelling Census Under-Enumeration A logistic regression perspective[J]. 2022.

[5] Lee H, Zhang S , Lin Z ,et al.Time Series Analysis of Hotel Room Sales in Busan Using Seasonal ARIMA Model [J]. International Journal of Tourism Management and Sciences, 2023.

[6] Zhang Y , Wu Z , Xu J ,et al.Remaining lifespan prediction of cross-linked polyethylene material based on GM(1, N) grey models[J].IET generation, transmission & distribution, 2022 (2):16.

[7] Chen G , Huang W X , Ronch A D ,et al.BP neural Network-Kalman filter fusion method for unmanned aerial vehicle target tracking:[J].Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2023, 237(18):4203-4212.