Information Systems and Economics (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2024.050308
ISSN 2523-6407 Vol. 5 Num. 3

A study on vegetable replenishment and pricing decisions based on polynomial regression and neural network prediction

Ke Xu[1], Qingyu Zhang[2], Chenyou Guo[2], Gong Zhang[2,*]

[1]College of Arts, Changchun University of Science and Technology, Changchun, 130022, China
[2]School of Optoelectronic Engineering, Changchun University of Science and Technology, Changchun, 130022, China
[*]Corresponding author: gongzhang@cust.edu.cn

Keywords: Polynomial regression, neural network time series forecasting, pricing strategy

Abstract: Fresh produce superstores sell a large variety of vegetables with short freshness periods, and their quality deteriorates with the increase in selling time. Therefore, fresh produce superstores' replenishment and pricing decisions are particularly important. In order to maximize the revenue of supermarkets, this paper establishes and solves Pearson's correlation coefficient model for the average selling price of vegetables and the daily sales volume and combines the polynomial regression model and linear regression to obtain the relationship between the sales volume of each vegetable category and the cost-plus pricing. In order to better predict the daily replenishment and pricing strategy of each vegetable category in the coming week, this paper establishes a time series model under the BP neural network by combining the daily sales data of each vegetable category of the supermarket in the past three years. It solves the replenishment and pricing decision of the vegetables when meeting the market demand and maximizing the supermarket's revenue through validation. The model can accurately predict and optimize the replenishment and pricing decisions of the hypermarket and provide a decision basis for the operation and revenue enhancement of the hypermarket.

1. Introduction

With the continuous development of the social economy and improving people's quality of life, people increasingly favor high-quality fresh commodities. However, even if the fresh food superstore takes preservation measures to reduce the rate of decay and deterioration of fruits and vegetables, most of the vegetable varieties in the next day will still be poor quality and can not be sold normally.[1] Supermarkets sell a variety of vegetables with different origins, usually between 3:00 am and 4:00 am, so merchants need to make decisions without knowing exactly the specific vegetable varieties and purchase prices. Therefore, it is very important to analyze and model the vegetable sales data of superstores in recent years to obtain replenishment decisions and pricing decisions.

In recent years, scholars have conducted studies to predict vegetable replenishment and pricing in superstores. In 2017, Afshin Oroojlooyjadid et al. proposed a machine learning algorithm based on deep Q-networks. They combined it with a migration learning algorithm to optimize the replenishment decision at a given stage. [2] In 2018, Xiong T et al. used a hybrid method based on Seasonal-Trend Loess and Extreme Learning Machines (STL-ELM) for short-, medium-, and long-term forecasting of seasonal vegetable prices, which greatly expanded the application of vegetable price forecasting scope. [3] In 2020, Ghosh S et al. implemented forecasting of vegetable prices using the ARIMA model with input series and seasonal autoregressive integrated moving average model. [4] In 2021, Linsheng Chen et al. used MEEMD to process the fluctuation characteristics of vegetable price time series signals and combined it with the KELM model to propose a short-term forecasting method for vegetable prices based on MEEMD-KELM, which has high accuracy and stability. [5] In 2023, Hu Yanjun et al. proposed a Gated Recurrent Unit (GRU) neural network model incorporating an Attention mechanism (Attention) based on time-series price and stock data, which greatly improved the vegetable pricing prediction accuracy in terms of mean-square error. [6]

In recent years, scholars in the research of predicting the direction of vegetable replenishment and pricing in superstores, mainly using traditional econometric methods and mathematical and statistical methods [7], including time series analysis, neural network model, machine learning algorithms or a combination of both to optimize the algorithm and other methods, through the processing and analysis of the vegetable sales data, a series of research results have been achieved, which provide important theoretical support and practical guidance for the management of superstore vegetables and operation provides important theoretical support and practical guidance. This paper takes the sales data of a fresh food supermarket in the past three years as the research object and finds out the replenishment quantity and pricing strategy of each single product every day that can maximize the supermarket's revenue. The Pearson correlation coefficient model is introduced to derive the correlation between the daily average selling price and the sales distribution of different single products in different categories of vegetables. On this basis, a time series prediction model under the BP neural network algorithm is introduced to find the optimal replenishment quantity and pricing for 1-7 July 2023 that can maximize the revenue of the hypermarket by using the vegetable sales volume of each category in the last three years of historical sales data as the training set.

2. Interrelationship between total sales volume and cost-plus pricing in the vegetable category

2.1 Cost-plus pricing

Cost-Plus Pricing (CPP) is a common pricing strategy and a cost-based pricing theory. [8] It's based on the cost of producing a product plus an expected profit to determine the final sales price. Its formula is:

$$\text{Cost-plus pricing (per item/category)} = \text{unit cost } (1 + \text{cost margin}) \tag{1}$$

2.2 Pearson correlation discrimination

The Pearson correlation coefficient (PCC) is a statistical measure of the strength and direction of a linear relationship between two variables. It is commonly used to explore the relationship between two continuous variables. The Pearson correlation coefficient ranges from -1 to 1, where -1 means a perfect negative correlation, 0 means no correlation, and 1 means a perfect positive correlation. The following formula calculates Pearson's correlation coefficient:

$$\mathrm{r} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{2}$$

The daily sales and average selling price of each vegetable category were solved using Pearson's correlation coefficient model, and the results are shown in Table 1:

Table 1: Pearson's correlation coefficients for the six vegetable categories

	flower-leaf class	florescent vegetable class	eggplant class	aquatic rhizomes class	edible fungi class	chili class
correlation coefficient	-0.288	-0.223	-0.182	-0.305	-0.351	-0.114

From the data in Table 1, the Pearson correlation coefficients for the flower-leaf class, florescent vegetable class, eggplant class, aquatic rhizomes class, edible fungi class, and chili class are all less than 0, which can be verified that there is a negative correlation between the sales volume of each vegetable category and cost-plus pricing.

2.3 Building and solving polynomial and linear regression

(1) Polynomial regression

A polynomial regression model is a regression analysis method used to model non-linear relationships. Unlike a simple linear regression model, a polynomial regression model fits the data by adding higher terms to the independent variables, which allows for better adaptation to complex data patterns.

The general form of a polynomial regression model can be expressed as:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \varepsilon \tag{3}$$

Where y is the dependent variable, x is the independent variable, $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the parameters of the model representing the coefficients in each polynomial, n is the order of the polynomials, and ε is the error term indicating the factors that cannot be fully explained in the model. [9]

(2) Linear regression

A linear programming model is a mathematical planning model that solves for variables by means of an objective function and constraints. [10] It is widely used in various fields for predicting and explaining relationships between variables. The general form of linear regression model can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{4}$$

Where y is the dependent variable, $x_1, x_2, \ldots, x_n$ are the independent variables, $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients of the model and ε is the error term. The goal of a linear regression model is to find a set of coefficients that minimise the sum of squares of the errors between the predicted values of the model and the actual observed values.

In order to arrive at the relationship that exists between total sales and cost-plus pricing for each vegetable category, different models, such as polynomial regression and linear regression, were attempted to determine the equations to which total sales and cost-plus pricing for each vegetable category confirmed to observe the degree of fit and to select the model that fits best among them in order to arrive at specific conclusions. In this case, a linear fit was used for the chili and eggplant classes due to uneven distribution of daily sales and a need for more data on high sales. The fitted graph and the final selected model and expression are shown in Figure 1 and Table 2:

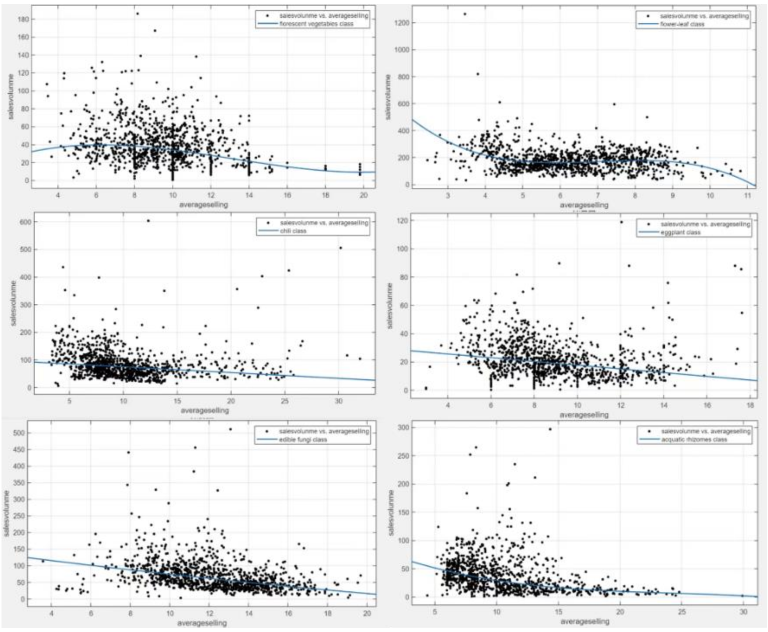

Figure 1: Fitted Curve of Gross Vegetable Sales and Cost Plus Pricing

Table 2: Gross vegetable sales and cost-plus pricing results

Kind	Model name	Polynomial equation results	R^2	RMSE
flower-leaf class	polynomial regression	$f(x) = 0.3606x^3 - 1.573x^2 - 5.964x + 35.18$	0.9778	3.38
florescent vegetable class	polynomial regression	$f(x) = -10.68x^3 + 15.75x^2 - 9.581x + 161.8$	0.9806	12.04
aquatic rhizomes class	polynomial regression	$f(x) = -0.234x^3 + 2.602x^2 - 12.19x + 27.88$	0.8891	10.46
edible fungi class	polynomial regression	$f(x) = 0.4651x^2 - 16.36x + 60.41$	0.9698	8.432
chili class	linear regression	$f(x) =$ -9.277x +74.04	0.9704	9.195
eggplant class	linear regression	$f(x) = -3.28x + 19.05$	0.8643	4.85

3. Vegetable replenishment and pricing optimization for each category

3.1 Time series forecasting models

BP neural network-based time series forecasting model is a method of using BP neural networks for time series data forecasting; this model learns the patterns and trends of time series data by training the neural network to forecast future values. Compared to general neural network models, the model used in this paper does rolling forecasts on a weekly basis, with each group using the first six days of data to predict the seventh day of data, resulting in the optimal replenishment scenario and pricing strategy for the 1-7 July 2023 period.

3.2 Time series forecasting model building and solution

Since the vegetables sold by the superstore contain seasonal vegetables, such as hollow cabbage, rape, Shanghai green, etc., and the vegetable price fluctuation with the season is large, this paper only selects the daily sales volume of each vegetable category in April-August in the last three years of a fresh food supermarket as input and divides the samples into three categories according to a certain proportion to be processed, which are the training set, the test set, and the test set. The number of hidden neurons is chosen as 5, and the delay compensation is 6. Predictions are made across one time point, and rolling predictions are made on a weekly basis by predicting the last 1 data from the first 6 data. The goodness of fit for the floral and foliage class is shown in Fig. 2, and the coefficients of determination for each type of vegetable are shown in Table 3:

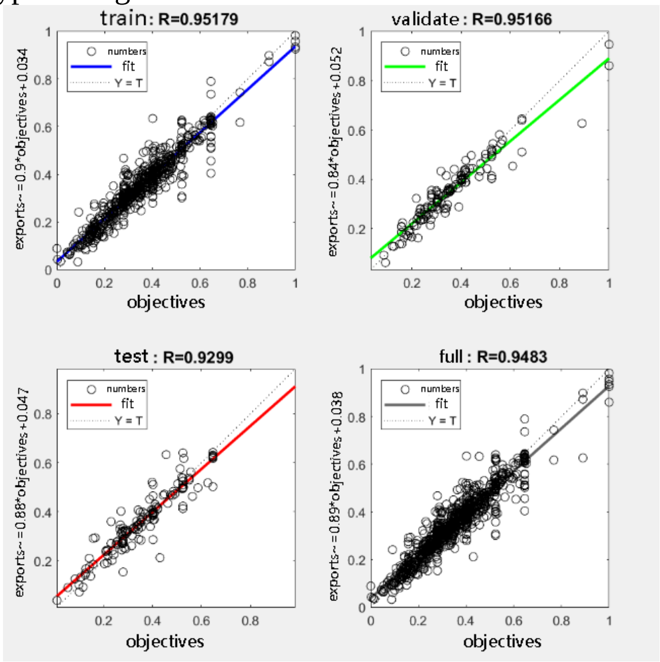

Figure 2: BP neural network goodness of fit (in the case of foliar species)

Table 3: Coefficients of judgment for each vegetable category

kind	flower-leaf class	florescent vegetable class	aquatic rhizomes class	edible fungi class	chill class	eggplant class
R^2	0.8793	0.8993	0.7609	0.6851	0.9422	0.8991

As can be seen from Table 3, the values obtained from the fitting of the six categories are closer to 1.R^2values are closer to 1, so it can be considered that the model fits better with higher goodness of fit. However R^2, it does not indicate whether the model adequately fits the data, i.e., it is not possible to determine whether the predicted values generated by the model have deviations from the true values, so this paper uses residual plots to compare the results of the true values with the predicted values. Figure 3 shows the results of the real value and the predicted value of the flower-leaf class as an example; according to the results and data, we can find that the sales of each vegetable and the seasonal correlation are very large, and its predicted value has a small deviation from the real value, the root-mean-square error RMSE is more accurate, so the results of this time-series model after training can be accepted.

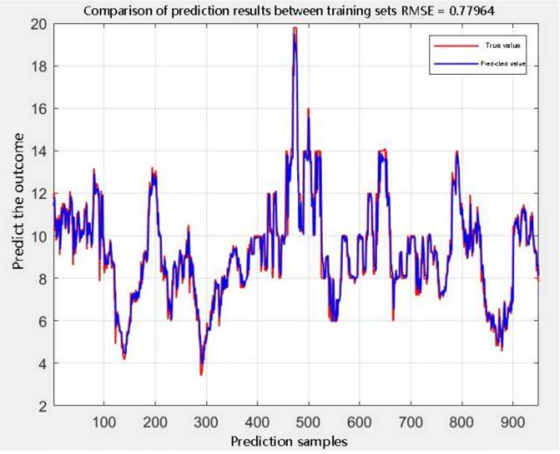

Figure 3: Comparison of results between true and predicted values (in the case of flower-leaf class)

In order to effectively maximize the profits of the vegetable superstore, it is also necessary to use the above model on the daily replenishment of vegetable categories and pricing strategy for the next week of a single-day vegetable replenishment, pricing strategy prediction, predictive value results as shown in Table 4:

Table 4: Projected values of daily vegetable replenishment and pricing strategies for the coming week

Dates	flower-leaf class		florescent vegetable class		aquatic rhizomes class		edible fungi class		chill class	
	daily replenishment	Pricing strategy	daily replenishment	Pricing strategy	daily replenishment	Pricing strategy	daily replenishment	Pricing strategy	daily replenishment	Pricing strategy
7.1	144.539	5.02115	22.1892	11.3349	58.888	14.6129	58.888	11.2191	81.3187	7.40971
7.2	148.93	5.10841	24.7018	11.11673	50.6829	16.0013	50.6829	10.2762	79.3563	7.39899
7.3	128.856	5.00095	20.522	10.85	47.7368	14.0581	47.7368	10.1167	70.4618	7.28359
7.4	145.15	5.22206	21.019	11.0156	30.5643	15.0619	30.5643	11.3568	81.3832	7.40485
7.5	129.712	5.04666	21.065	10.6157	56.5835	15.5641	56.5835	11.2477	78.4695	7.35934
7.6	150.636	5.15707	23.7226	10.7154	59.5151	14.293	59.5151	11.3697	81.1264	7.25155
7.7	137.778	5.10141	24.3155	10.541	52.7456	14.3417	52.7456	10.9032	69.3557	7.40845

4. Conclusion

In this paper, we aim to address the challenges faced by fresh produce superstores in replenishment and pricing decisions for vegetable items. Considering the characteristics of vegetables, such as short freshness period and variable character, we propose a comprehensive evaluation model that aims to help superstores maximize the profitability of vegetable items. First, we established and solved the Pearson correlation coefficient model for the distribution of average selling price and sales volume of vegetables and combined the polynomial regression model and the linear regression model to obtain the relationship between the total sales volume of vegetables in a single day, the total sales volume of each vegetable category, and the cost-plus pricing. Secondly, in order to better predict the total daily replenishment and pricing strategy of each vegetable category in the coming week, we combined the external and temporal influences and made rolling forecasts on a weekly basis, established a time-series model under the BP neural network, and through validation, solved the vegetable replenishment and pricing decisions when satisfying market demand and maximizing the revenue of the superstore.

Through this study, we have established a reliable decision support system, which provides an effective basis for superstores to make replenishment and pricing decisions and is expected to improve the operational efficiency and profitability of superstores. Our method not only provides practical operational guidance for superstores but also provides important theoretical support for the stable development of the fresh food market. In the future, we will continue to refine the model and improve the prediction accuracy to cope with the constant changes in the market and make greater contributions to the sustainable development of hypermarkets.

References

[1] Pan Xiaofei, Zhang Tao. Optimization of freshness efforts and pricing in fresh produce supply chain considering loss aversion [J]. Highway Traffic Science and Technology. 2023, 40(05).

[2] Afshin Oroojlooyjadid; Mohammad Reza Nazari; Lawrence Snyder; Martin Takáč; "A Deep Q-Network For The Beer Game: a Deep Reinforcement Learning Algorithm To Solve Inventory Optimization Problems", ARXIV-CS.LG, 2017.(IF: 3).

[3] Xiong T, Li C, Nbao Y. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China[J]. Neurocomputing, 2018(275): 2831-2844.

[4] Ghosh S, Singh K N, Thangasamy A, et al. Forecasting of onion (Allium cepa) price and volatility movements using ARIMAX-GARCH and DCC models[J]. Indian Journal of Agricultural Sciences, 2020, 90(5): 169-173.

[5] Chen Linsheng, Sun Lijun, Ma Jia. A comparative study of short-term forecasting models for vegetable prices--Take the price of green vegetables in Shanghai as an example[J]. Price Theory and Practice. 2020(09).

[6] Hu Yanjun, Zhang Pingchuan, Shang Zheng, Wang Huimin, Qiao Yongfeng. Research on garlic price prediction based on deep learning[J]. Journal of Henan Institute of Science and Technology (Natural Science Edition). 2023, 51(03).

[7] Yu Weige,Wu Huarui,Peng Cheng.Short-Term Price Forecast of Vegetables Based on Combination Model of Lasso Regression Method and BP Neural Network [J]. Smart Agricul ture,2020,2(3):108-117. doi:10.12133/j.smartag. 2020. 2.3. 202008-SA003.

[8] Jue Wang. A post-Keynesian model of corporate pricing - the cost-plus pricing principle[N]. Journal of Lanzhou University. 2003(03).

[9] Wang Wannian, Zhu Xu, Chen Zhanxing, Zhou Minxu, Xing Qiwei, Wang Xiaohong, Ma Tengfei. Multiple linear regression model for empirical parameters to predict the yield strength of high-entropy alloys[J]. Special casting and non-ferrous alloys 2024(03).

[10] Huang Zhengpeng, Ma Xin, Chen Xue, Liu Na. Analysis and application of colorful Guizhou tourism data based on linear regression algorithm[J]. Software Engineering. 2024, 27(03).