BOHR

BOHR International Journal of Operations Management Research
and Practices
2024, Vol. 3, No. 1, pp. 43–53
DOI: 10.54646/bijomrp.2024.28
www.bohrpub.com

CASE STUDY

Vegetable category replenishment and pricing decision model in fresh food supermarkets

Kai Gao* **and Yihao Yin**

School of Management, Tianjin University of Technology, Tianjin, China

***Correspondence:**
Kai Gao,
kai_gao2021@163.com

Received: 14 February 2024; **Accepted:** 15 June 2024; **Published:** 22 October 2024

The vegetable commodities in fresh food supermarkets generally have the characteristics of a wide variety of categories, different origins, and a short shelf life. Moreover, the product quality continuously deteriorates as the sales time increases, and the corresponding price continuously decreases as the sales time increases. At the same time, the purchase and trading time of vegetables is usually from 3:00 to 4:00 in the morning, and the daily sales volume of dishes is unknown. Therefore, merchants must make replenishment decisions for various vegetable categories on the same day without exactly knowing the pricing and purchase quantity of specific dishes. Thus, reliable market demand analysis is particularly important for replenishment and pricing decisions. Based on the historical sales data of six vegetable categories distributed by a supermarket, this paper first uses the Auto Regressive Integrated Moving Average (ARIMA) model to predict the replenishment quantity of each category in the next seven days. Then, through the Back Propagation (in neural networks) (BP), it fits the relationship between sales volume, cost-plus rate, and wholesale price. Combined with the "cost-plus pricing" method, it predicts the pricing of each category of dishes in the next seven days, providing a certain reference for the replenishment and pricing decisions of supermarkets.

Keywords: ARIMA, BP neural network, prediction, vegetables, replenishment

1. Introduction

In today's competitive market environment, fresh food superstores, as an important part of the retail industry, not only face a rich variety of commodity categories and differences in origin, but are also affected by the challenges of the short freshness period of vegetable commodities, easy to change in appearance, and price fluctuations. Especially in the face of vegetables which are fresh commodities, with the dynamic changes of quality and prices, the replenishment needs of supermarkets, the pricing decisions become challenging. To cope with this challenge more effectively, this paper aims to combine Auto Regressive Integrated Moving Average (ARIMA) and Back Propagation (in neural networks) (BP) models to predict the future sales and pricing of various categories of vegetables, so as to provide a certain reference for the replenishment and pricing decisions of supermarkets.

1.1. Literature review

A lot of research has been carried out by scholars at home and abroad in the field of fresh food sales forecasting and pricing decision-making (1, 2). Zhang Yanliang and Dai Peipei explored the customer-perceived product information contained in online review data to assess the accuracy of fresh produce demand prediction, constructed a multivariate Support Vector Regression (SVR) demand prediction model, and found that the extraction of customer-perceived factors in online reviews can effectively improve the accuracy of fresh produce demand prediction (3). Lin and Hu (1) mentioned that review characteristics, transaction characteristics, shop characteristics, service characteristics, and product characteristics have different degrees of influence on the sales of fresh food e-commerce products, among which the influence of review characteristics is the most significant (4, 5). Lu Chao and Xing Miao constructed a pricing decision model for fresh produce supply chain through online reviews and found that when the cost sharing coefficient is high, the optimal level of preservation effort is higher than that in the centralized decision-making model under the decentralized decision-making model with cost sharing and revenue sharing (6, 7). Bi and Zhou (2) combined the characteristics of fresh products to model the problem and define it as a Markov decision process, and then designed a joint inventory control and dynamic pricing algorithm for fresh products based on deep reinforcement learning (2, 8). Tian Zhongwei and Dong Ming developed and solved a two-stage model for the optimal pricing and ordering decisions of retailers for fresh products with different quality levels due to deterioration in the sales process (9, 10). This paper is based on the historical sales data of six vegetable categories distributed by a supermarket to carry out an in-depth study. Firstly, an ARIMA model was used to predict the replenishment volume of each category in the next 7 days to better understand the potential trend of sales. Secondly, through the BP neural network, we try to fit the complex relationship between sales volume, cost-plus rate, and wholesale price, combined with the "cost-plus pricing" strategy, to provide a novel approach to set reasonable pricing for products to adapt to the fast-paced and dynamic changes in the market environment and to provide support and reference.

2. Data description

2.1. Basic description

The data in this paper come from the historical sales data of vegetable commodities in a fresh food supermarket, including 6 categories of vegetables, commodity information of 251 single vegetable products, sales flow detail data, wholesale prices of vegetable commodities, and other data. After data collation, the daily sales data, wholesale price data and cost-plus rate data of various types of vegetables are obtained. After collation, the data table has a total of 1095 rows from July 1, 2020 to June 30, 2023, and a total of 18 columns for the sales volume, wholesale price, and cost-plus rate of each category of vegetables.

2.2. Distribution pattern

Overall, vegetable commodities are divided into six categories, 251 kinds of individual products, in which the sales of cauliflower and leafy category are the highest, followed by chili and edible mushrooms; nightshades have the lowest sales, the sales of cauliflower and leafy category are concentrated in the range of 0–500 kg; the sales fluctuation range is large; cauliflower, aquaticrhizomes, nightshades, chili, and edible mushrooms sales have a relatively small range of fluctuation. The specific results are shown in **Figure 1**; each category of vegetables is subjected to the null value checking, and these null values are filled to a very small value of 0.01 to ensure the integrity of the data.

2.3. Correlation

Spearman's rank correlation coefficient is a non-parametric statistical method that does not require the data to satisfy the assumption of normal distribution and is used to measure the correlation between two continuous variables and is calculated by the formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1}$$

where n is the number of samples, and d represents the rank difference between data x and y. The correlation coefficient matrix was calculated by recoding these six categories and further plotted as a heat map for each vegetable category.

By recoding these six categories, the matrix of correlation coefficients was calculated, and the correlation heat map of each vegetable category was further plotted, which is shown in **Figure 2**. In this heat map, the color of each grid indicates the correlation between the corresponding categories, and the redder the color the higher the gene expression, and the bluer the color the lower the gene expression. By observing the correlation heat map, the following results can be obtained: the correlation between the leafy category, cauliflower, aquatic rhizomes, chili, and edible mushrooms is strong, with correlation coefficients greater than 0.5, while the correlation between night shades and other categories is weak, with correlation coefficients less than 0.5.

3. ARIMA model

The ARIMA model is a commonly used time series analysis method that uses difference methods to convert non-stationary time series data into stationary time series data. Based on this data for future data forecasting and modeling, the model combines the characteristics of the autoregressive (AR) model, the difference (I) operation, and the sliding average (MA) model, and its three main parameters are expressed as the autoregressive order (p), the difference number (d), and the sliding average order (q). The general form of the ARIMA (p, d, and q) model is as follows:

$$\begin{aligned} W_\mathrm{t} &= \delta + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} \\ &\quad + \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q} \end{aligned} \tag{2}$$

t represents the moment, W_t represents the smooth sequence obtained after d-order differencing, μ_t represents the random disturbance sequence, δ is the constant term, p is the autoregressive partial order, q is the moving average partial

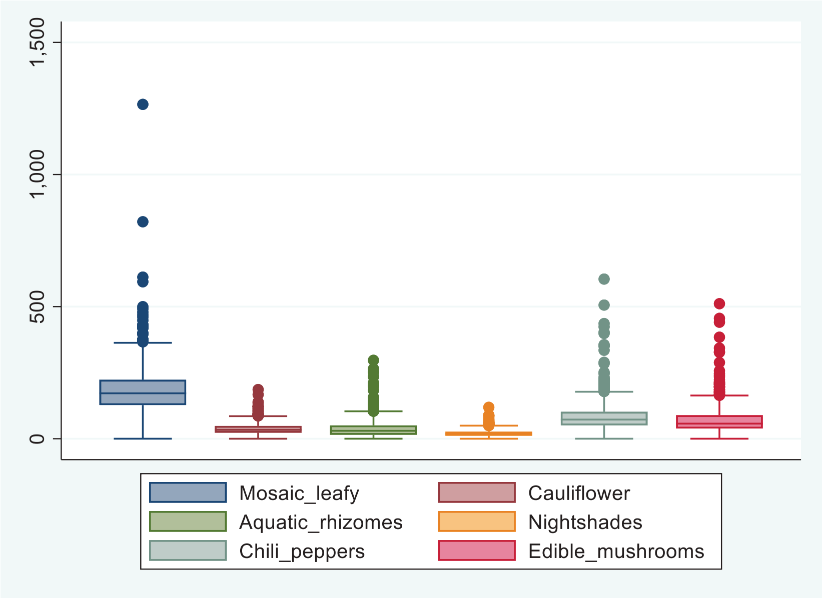

FIGURE 1 | Box line diagram for each category.

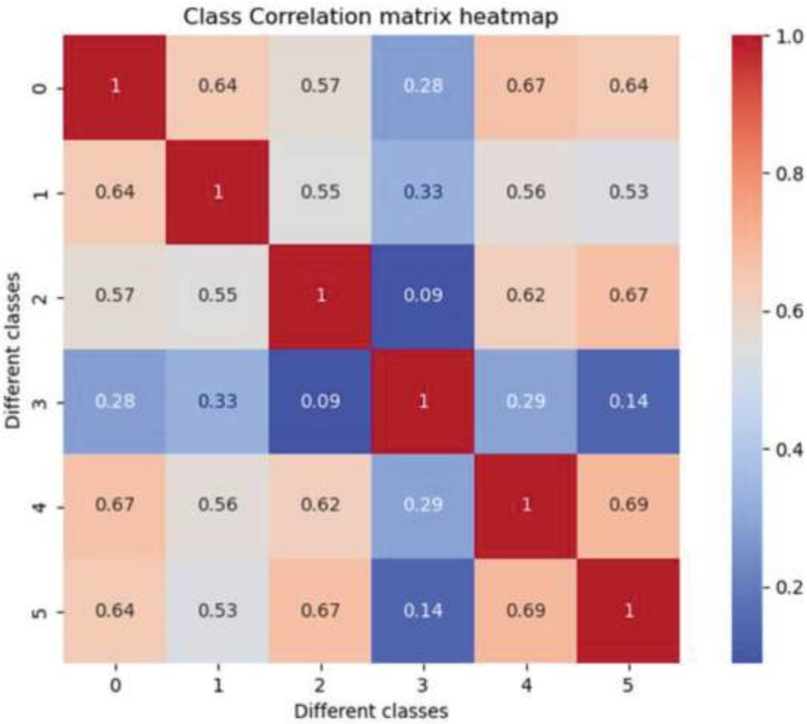

FIGURE 2 | Heat map for each correlation category.

order, ϕ_p is the coefficient of the autoregressive term of order p, and θ_q is the coefficient of the moving average term of order q (11).

3.1. Observation of time series chart

To study the trend of the sales volume of superstores over time, we choose days as the unit of measurement and draw the time-series plot of each category of vegetables over time, and it can be seen that the sales volume of each category shows an oscillating trend of change, and basically smooth. The specific results are shown in **Figure 3**; Combined with the box-and-line diagram in Section 2.1 and the time-series plot below, it can be easily found that there are some values in each category far more than the normal range of fluctuation values of this type of dish in each category, and these values can be considered as outliers, and these

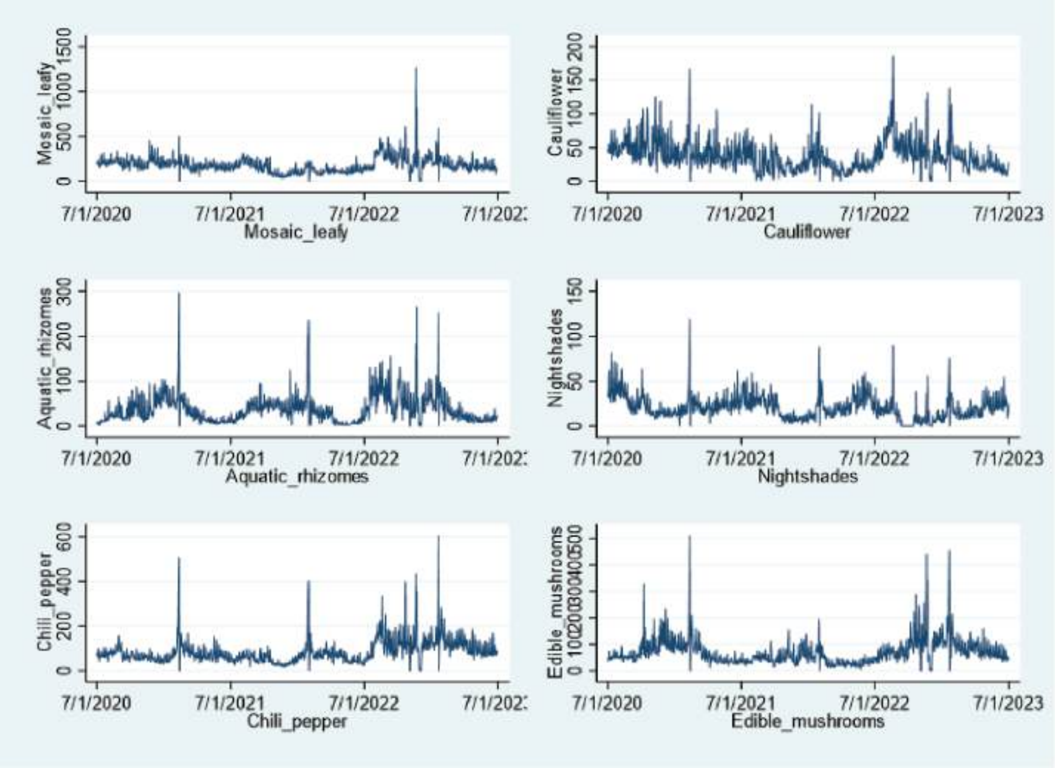

FIGURE 3 | Timing diagram for each category.

values will be replaced by the average value in the subsequent modeling.

3.2. Smoothness and white noise test

The p-value of smoothness test of the daily sales volume of the leafy category = 6.60e-04 < 0.05, which rejects the original hypothesis that there is no unit root, indicating that the time series is smooth; the p-value of stochasticity test of the daily sales volume of the leafy category = 4.30e-08 < 0.05, which indicates that the time series is non-white noise. The p-value of the remaining types of smoothness and white noise test is as follows (**Table 1**).

From **Table 1**, it can be seen that all classes pass the white noise test and smoothness test.

TABLE 1 | Smoothness and white noise test p-values for the remaining five categories.

Remaining five categories	White noise test p-value	Smoothness test p-value
Cauliflower	4.25e-08	7.78e-03
Aquatic_rhizomes	1.18e-08	3.28e-02
Nightshades	3.87e-08	5.27e-04
Chili	1.65e-08	1.94e-02
Mushrooms	5.02e-120	4.55e-02

3.3. Determination of p, d, and q parameters

First of all, we draw the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) diagrams for each category; the specific results are shown in **Figures 4–6**. Take the leafy category as an example. It can be seen from the diagram that the ACF and PACF are trailing, and the original data of the leafy category are smooth and do not need to be differentiated. The order of p and q in the general ARIMA model will not be more than 3 (12); the value of p and q will lead to the overfitting of the model, and there will be a large error in the prediction of the future data. Therefore, the preliminary ARIMA model parameters for the leafy category are (3,0,3). However, the model obtained at this time is not necessarily optimal; the next use of the aic, bic criterion to more accurately select the final order of the model, through the selection of the aic, bic value, is found to be the smallest when the parameter is (2,0,1), and so the parameter is selected as the leafy category ARIMA model parameters. The rest of the ARIMA model parameters are shown in **Table 2**.

3.4. Replenishment prediction

Firstly, it is assumed that the market supply and demand are equal, that is, the replenishment volume is equal

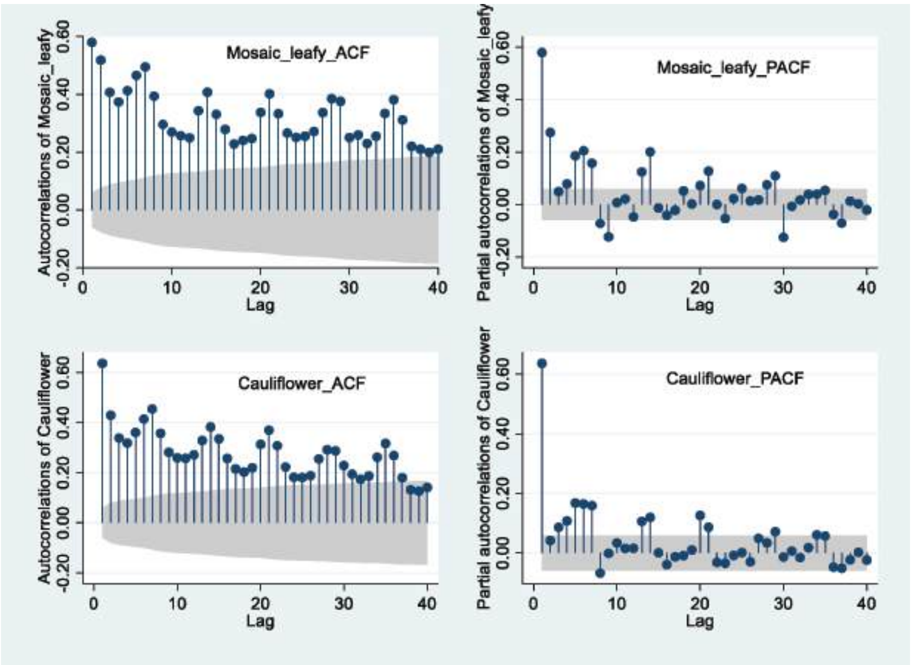

FIGURE 4 | ACF and PACF charts for leafy categories and cauliflowers.

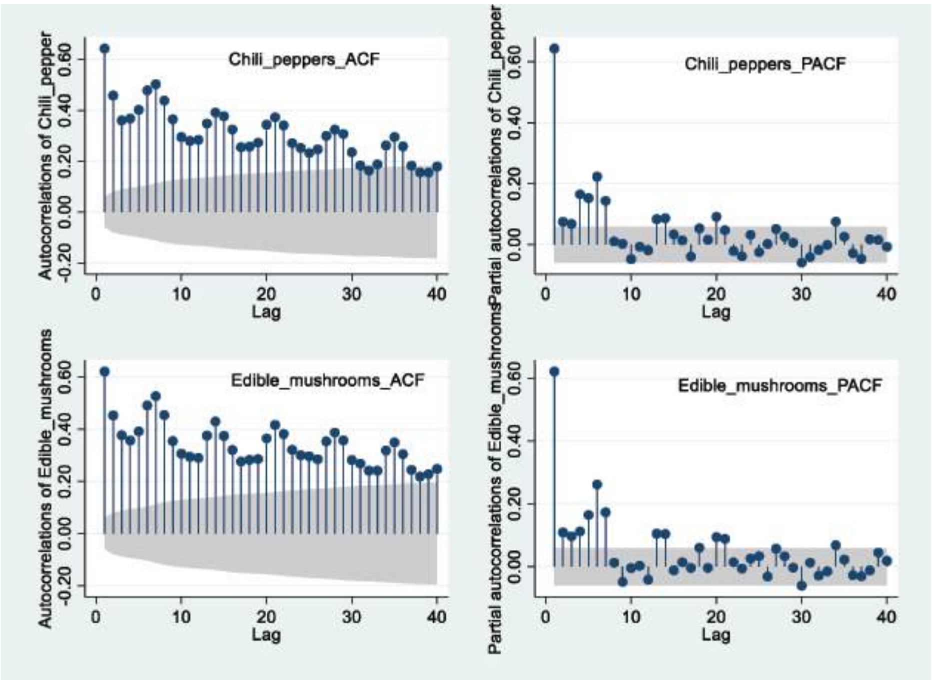

FIGURE 5 | ACF and PACF plots of chili_peppers and edible_mushrooms.

to the sales volume, and then the original sales volume data of various types of dishes are divided, with 90% as the training set and 10% as the test set, and the model test is carried out. Taking the leafy category as an example, the established ARIMA (2,0,1) model was used to predict the replenishment quantity in the coming week, and the specific results were obtained as follows (**Table 3**).

Moreover, the Mean Squared Error (MSE) of the model on the test set is 4.14e-02, which is relatively small, and it can be considered that the model is more reliable, and the average error range on the data scale is 18.62%, which is also small, and the comparison between the actual values and the predicted values on the test set is shown in **Figure 7**.

From **Figure 7**, we can see that the fitting effect is relatively good. The MSE on the rest of the various test

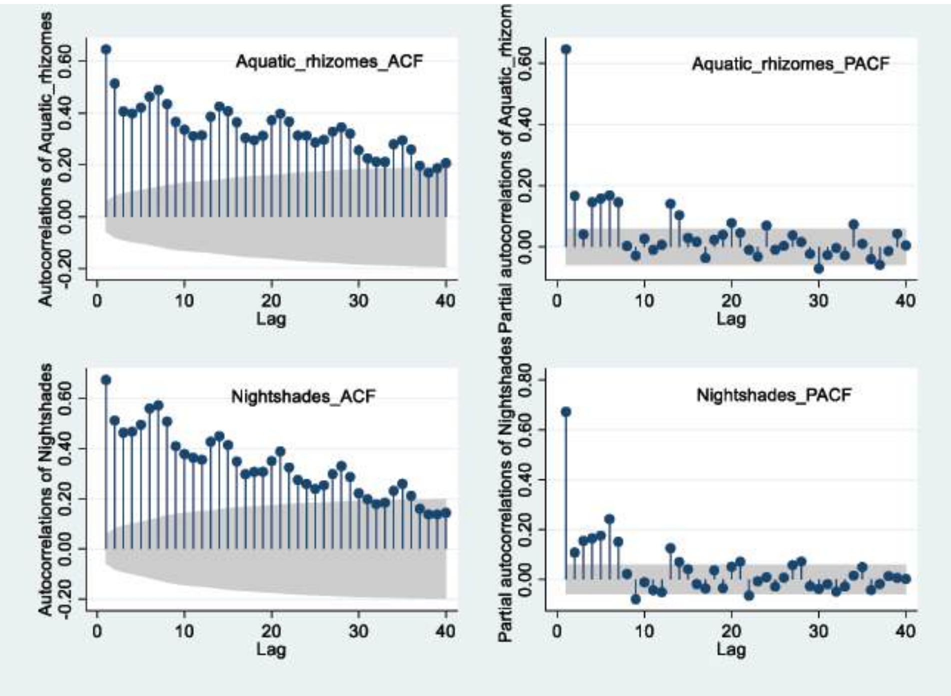

FIGURE 6 | ACF and PACF plots of aquatic_rhizomes and nightshades.

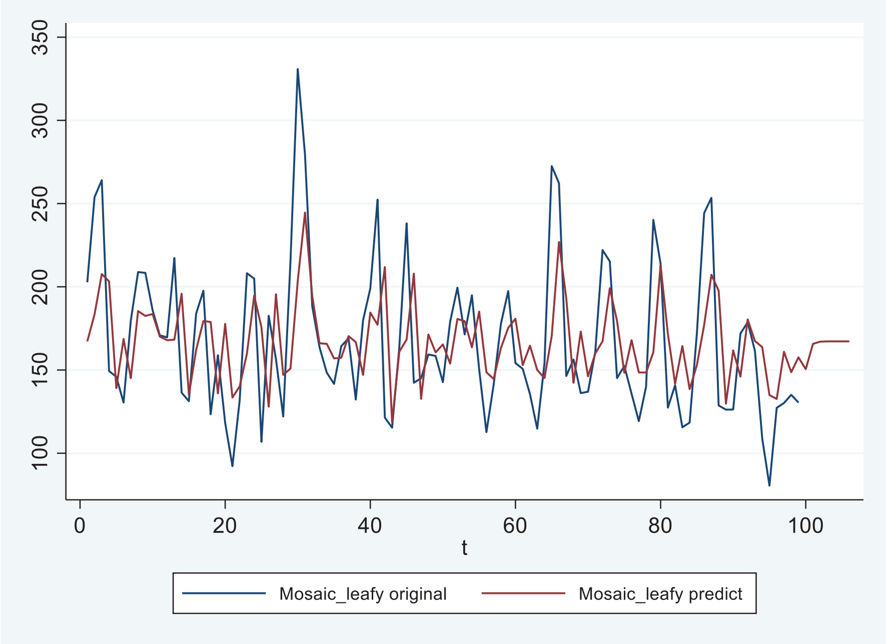

FIGURE 7 | Plot of actual versus predicted values on the test set for the leafy category.

sets, the average error ranges are shown in Table 4, the predicted values of the replenishment volume for the next 7 days are shown in Table 5, and the comparison of the actual and predicted values on the test sets is shown in Figure 8.

As can be seen from Figure 8, the remaining five categories also fit better on the test set, and the model is relatively reliable.

4. BP neural network modeling for predictive pricing

4.1. Model construction and fitting

For fresh commodities, the core of the pricing strategy is to use low gross profit to stimulate sales and maintain the freshness and turnover of commodities. Cost-plus pricing is

TABLE 2 | ARIMA parameters for the remaining five categories.

Remaining five categories	p	q
Cauliflower	1	2
Aquatic_rhizomes	2	1
Nightshades	2	2
Chili	1	1
Edible_mushrooms	1	1

TABLE 3 | Forecast for restocking of the leafy category in the coming week.

Day	1	2	3	4	5	6	7
Replenishment	159.570	177.712	176.517	168.268	163.813	164.564	166.864

a method of setting product prices based on the unit cost of the product plus a certain percentage of profit, which is calculated by the formula:

$$P = C(1 + w) \tag{3}$$

Where, P denotes the price, C denotes the average cost, and w denotes the cost-plus rate, w = (selling price - purchase price)/selling price × × 100%, which reflects the profit margin of the enterprise in selling the product or service (13).

To analyze the relationship between the replenishment quantity and cost-plus pricing of each category of vegetables, firstly, we consider the replenishment quantity and merchant's wholesale price and cost-plus rate and carry out linear regression and multivariate binomial regression and find that the fitting coefficients obtained under linear regression $R^2 = 0.0536$ and $R^2 = 0.0467$ under multivariate binomial regression, the fitting effect is poor; therefore, we use the BP neural network to build a model for prediction BP neural network model structure, including input layer, intermediate layer (hidden layer), and output layer. The training algorithm of the neural network is to continuously adjust the connection weights through the back-propagation algorithm, trainlm, to improve the training speed and fitting accuracy to achieve the goal of minimizing the training error so that the prediction effect of the whole network is better (14). According to the data set obtained from the collation, the neural network input neuron cell number is 1, indicating the replenishment of each category; the hidden layer has 10 cells; the output layer neuron cell number is 2, indicating the wholesale price and the cost of markup coefficients; and its model structure is shown in the following Figure 9.

At the same time, the BP neural network model uses the activation function to add nonlinear factors to the network to solve the problems that cannot be solved by the linear model, and by limiting the range of the output result of the activation function, it helps to make the gradient decrease smoothly (15). The activation function added in this paper is as follows:

$$\varphi\ (x) = \frac{1}{1 + e^{-x}} \tag{4}$$

The loss function is also added to the BP neural network to evaluate and validate the model and evaluate the model fit. In this paper, the performance function MSE is set to 1e-1, below which the iteration stops, and its specific formula is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - t_i)^2 \tag{5}$$

where t denotes the actual value and y denotes the predicted value. Then 90% of the historical sales volume data of various types of food in supermarkets are used as the training set and 10% as the test set, so that the replenishment volume data in the input layer are obtained, and then the historical wholesale price data and historical cost markup rate data are brought into the output layer to fit the BP neural network. Taking the leafy category as an example, its mean squared loss function image is shown in Figure 10, and the model fitting is shown in Figure 11.

From Figure 11, it can be seen that the leafy category reaches the set value of the mean square loss function after 20 iterations in the test set, and the fit coefficient of the model in the test set reaches 0.85, which is a good fit of the model. The number of iterations and the fitting coefficients on the test set for the remaining classes are shown in Table 6.

From Table 6, it can be seen that except for chili peppers, the fitting coefficients of the other categories are high, and it can be considered that the model is effective.

TABLE 4 | MSE and mean error ranges on the remaining five category test sets.

	MSE	Average range of error
Cauliflower	4.78E-02	23.28%
Aquatic_rhizomes	4.36E-02	31.44%
Nightshades	7.74E-02	22.02%
Chili	4.89E-02	21.47%
Edible_mushrooms	3.77E-02	24.56%

TABLE 5 | Restocking of the remaining five categories for the next 7 days.

	1	2	3	4	5	6	7
Cauliflower	22.226	20.375	19.562	19.738	20.904	21.060	20.207
Aquatic rhizomes	16.815	15.932	16.089	15.454	16.658	16.691	15.655
Nightshades	30.795	26.049	19.060	15.411	17.743	23.872	28.791
Chilli	88.132	96.539	98.403	97.816	98.908	98.928	97.933
Edible mushrooms	58.433	70.238	71.378	70.488	71.499	71.500	70.500

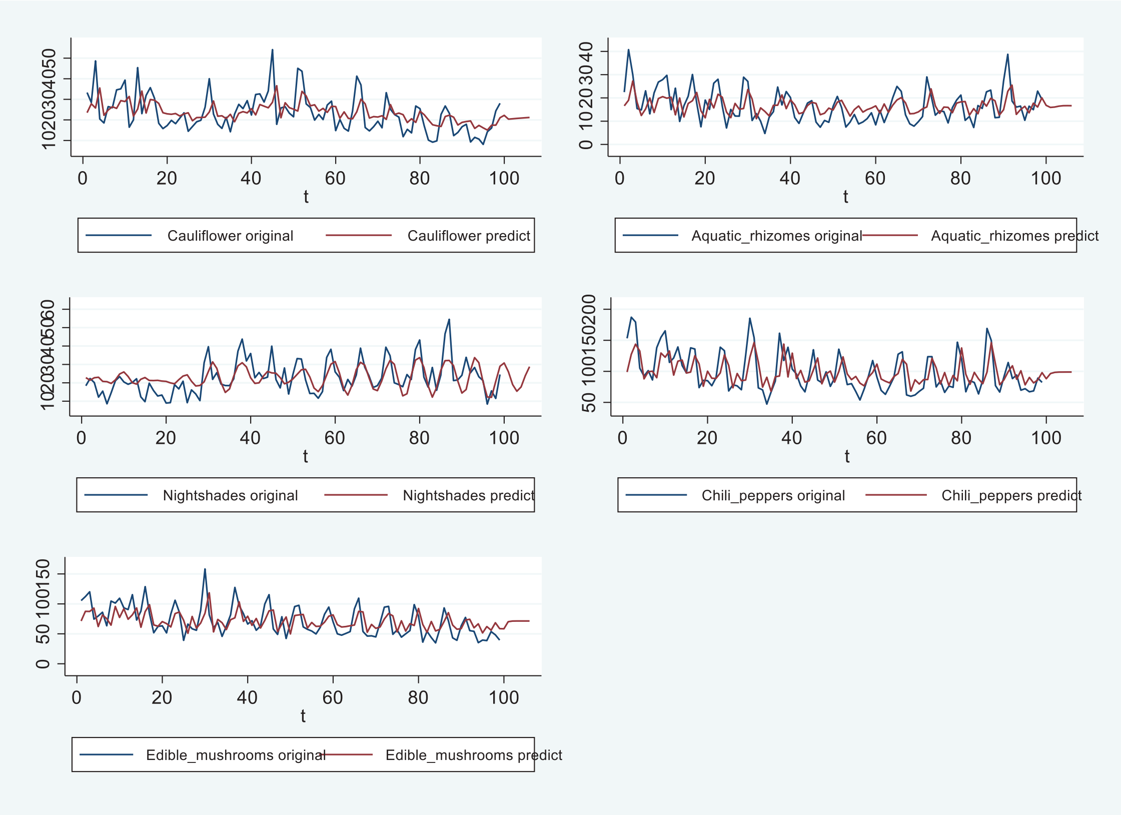

FIGURE 8 | Comparison of actual and predicted values on the remaining five types of test sets.

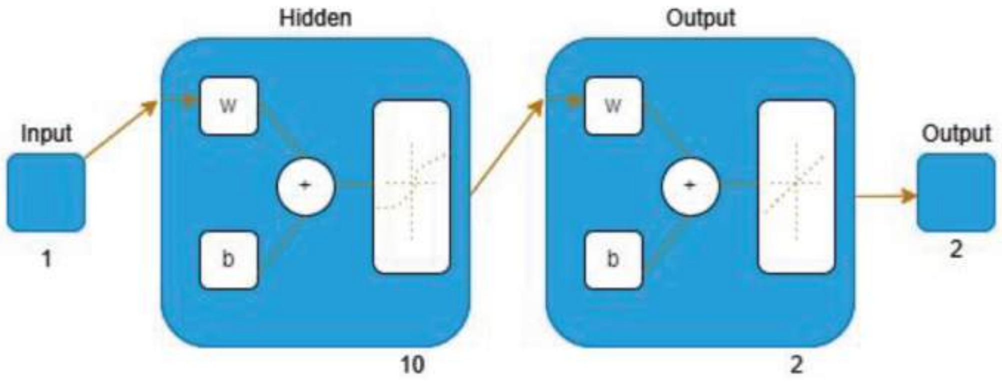

FIGURE 9 | Structure of the constructed BP neural network.

4.2 Predictive pricing

The content in Section 4.1 indicates that the BP neural network model successfully fits the relationship between replenishment quantity, wholesale price, and cost markup rate. Building on this, combined with the replenishment quantities forecasted for the next seven days by the ARIMA model in Section 3, it is possible to calculate the wholesale prices and cost markup rates for each category over the next seven days, thereby determining the pricing strategy. Taking leafy vegetables as an example, the pricing results are shown in Table 7, while the pricing strategies for other categories are detailed in Table 8.

The remaining types of pricing strategies are shown in Table 8.

5. Summary and prospect

Through the research in this paper, the sales data of the vegetable category is deeply analyzed, and a restocking and pricing decision-making framework based on the ARIMA model and BP neural network is proposed. By predicting the replenishment and pricing in the next 7 days, the framework allows supermarkets to make more targeted purchases and pricing, thus better meeting market demand

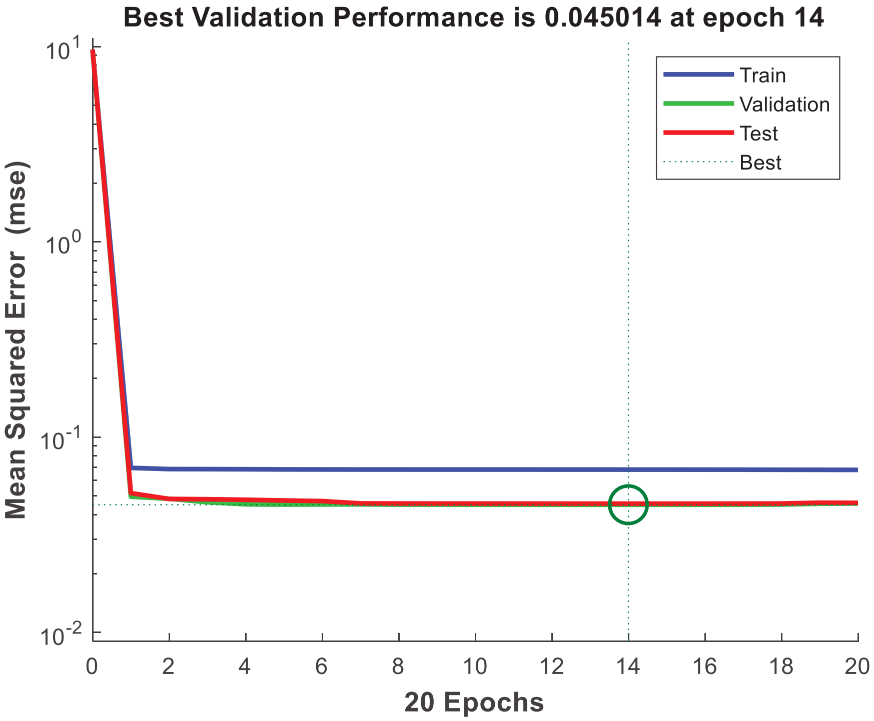

FIGURE 10 | Image of the mean-variance loss function for the leafy category.

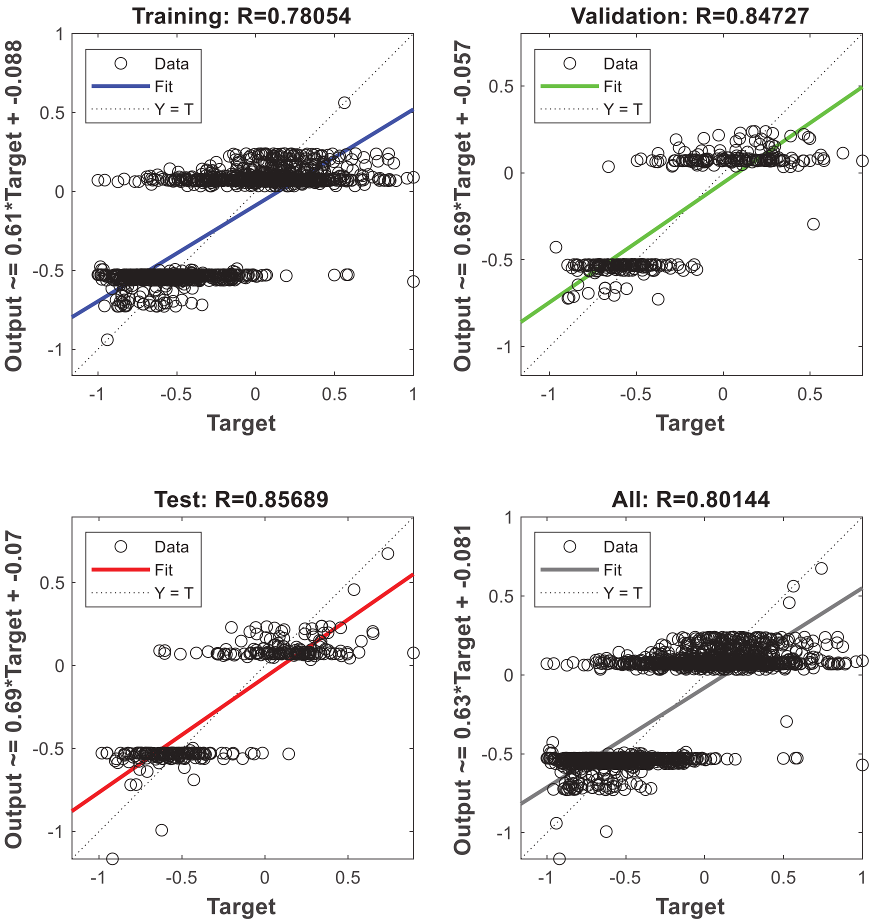

FIGURE 11 | Fitting of the leafy category.

TABLE 6 | Number of iterations and fit coefficients for the remaining categories.

Category	Number of iterations	R^2
Cauliflower	4	0.73757
Aquatic_rhizomes	6	0.46318
Nightshades	25	0.48066
Chilli	1	0.31615
Edible_mushrooms	1	0.65866

TABLE 7 | Pricing strategy for the leafy category.

Sales	Wholesale prices	Cost-plus ratio	Set a price
159.57024	4.58497487	0.418270388	6.5027
177.7123	4.58874695	0.419013799	6.5115
176.51704	4.47508296	0.451777889	6.4968
168.26784	4.48452522	0.437810039	6.4479
163.81308	4.49677659	0.427539234	6.4193
164.56416	4.48874378	0.433607457	6.4351
166.86438	4.54826066	0.415045533	6.4360

TABLE 8 | Pricing strategies for the remaining categories.

	Sales	Wholesale prices	Cost-plus ratio	Set a price
Cauliflower	22.22606	5.604864171	0.663921181	9.326052212
	20.374576	5.638370903	0.666967332	9.398980099
	19.562089	6.328516579	0.725881666	10.92227074
	19.738409	6.738478773	0.749115354	11.78637668
	20.904206	6.46794876	0.736577945	11.23209716
	21.060107	6.402873045	0.731701808	11.08786683
	20.206702	6.115652505	0.708358728	10.44772833
Aquatic_ rhizomes	16.814954	7.453934592	0.336638312	9.96321455
	15.932203	7.620024034	0.34917092	10.28071484
	16.088553	8.583192492	0.415905964	12.15299344
	15.453662	9.666204898	0.463720736	14.14862455
	16.658085	9.038068981	0.441812524	13.03120105
	16.691415	8.417501461	0.405345368	11.82949669
	15.654502	8.053977246	0.380680379	11.11996836
Nightshades	30.795475	5.428813177	0.619494779	8.791934593
	26.049241	5.181828347	0.620315619	8.396197405
	19.059821	5.624282968	0.588106932	8.93196277
	15.411392	5.95050633	0.562786137	9.299368797
	17.743159	5.801484132	0.56749237	9.093782112
	23.872477	5.733678404	0.571773191	9.012042001
	28.7905	5.622064532	0.588689264	8.931713567
Chilli	88.132054	6.742380682	0.427551109	9.625093021
	96.539308	6.699111186	0.432239793	9.594733618
	98.403058	7.465304575	0.389567775	10.37354667
	97.81622	7.59467949	0.383719187	10.50890373
	98.907812	7.661984357	0.380678411	10.57873639
	98.928116	8.109022781	0.360329256	11.03094093
	97.932617	7.376267606	0.393604857	10.27960236
Edible_ mushrooms	58.433188	5.904714027	1.013131641	11.88696664
	70.237555	6.038854402	1.084428325	12.58755917
	71.377909	6.224934997	1.173842926	13.53203091
	70.488072	6.221549966	1.172594697	13.51690646
	71.498714	6.146095477	1.138476789	13.14328252
	71.499742	6.193997859	1.161073499	13.38568463
	70.499842	6.000164115	1.06438706	12.38666116

and improving operational efficiency. However, it should be noted that any model has its limitations due to the changing market environment. Therefore, it should be flexibly adjusted in the actual situation and combined with other decision-making tools to improve the overall decision-making level, and future research can be more specific to the replenishment volume and pricing decisions for each single product and continuously expand the model to provide more useful references and insights for the efficient and effective development of the fresh food retailing industry.

With the continuous progress of science and technology and the evolution of the business environment, the replenishment and pricing decision-making framework based on the ARIMA model and BP neural network proposed in this paper can provide some references for the operational decision-making of supermarkets to a certain extent. In the future, more complex models and algorithms can be further explored to improve the accuracy of market demand and sensitivity to sales trends. At the same time, by combining big data, artificial intelligence, and other technologies, the decision-making system can be further improved to make it more real-time and adaptive, so as to better adapt to the dynamic changes of the market. In addition, with the development of digitalization and intelligence in the fresh food supply chain, supermarkets can consider introducing more advanced technological means, such as the Internet of Things (IoT), blockchain, etc., to improve the traceability of vegetable commodities and the level of quality assurance. This will not only help consumers better understand the source and quality of goods, but also help supermarkets to realize more efficient management and coordination in the supply chain.

References

1. Lin H, Hu Y. Big data profiling of influencing factors of fresh food e-commerce platform sales–an empirical study based on the support vector machine method. *Guizhou Soc Sci.* (2021) 3:129–38. doi: 10.13713/j.cnki.cssci.2021.03.017
2. Bi W, Zhou Y. Research on joint inventory control and dynamic pricing of fresh products based on deep reinforcement learning. *Comput Appl Res.* (2022) 39:2660. doi: 10.19734/j.issn.1001-3695.2022.01.0056
3. Zhou C, Li H, Zhang L, Ren Y. Optimal recommendation strategies for AI-powered e-commerce platforms: A study of duopoly manufacturers

and market competition. *J Theoretical Appl Electron Commerce Res.* (2023) 18:1086–106.

4. Han L. Research on academic paper evaluation model based on BP neural network. *Modern Intell.* (2024) 44:170–7.
5. Yu J, Zhou C, Leng G. Is it always advantageous to establish self-built logistics for online platforms in a competitive retailing setting? *IEEE Trans Eng Manag.* (2024) 71:1726–43.
6. Zhou C, Li X, Ren Y, Yu J. How do fairness concern and power structure affect competition between e-platforms and third-party sellers? *IEEE Transa Eng Manag.* (2023) 21:2318. doi: 10.1109/TEM.2023.3262318
7. Lu C, Xing Y. Supply chain pricing strategy for fresh products based on online reviews. *Operat Res Manag.* (2022) 31:91–7.
8. Zhang H, Zhang Q, Yu J. Property analysis and improvement of activation function in convolutional neural networks. *Comput Simul.* (2022) 39:328–34.
9. Zhang Y, Dai P. Multivariate SVR demand forecasting for fresh products - customer perception factor extraction based on online reviews. *J China Agric Univers.* (2022) 27:275–82.
10. Tian Z, Dong M. Pricing and ordering strategies for fresh products with different quality levels. *J Shanghai Jiao Tong Univers.* (2014) 48:306–11. doi: 10.16183/ j.cnki.jsjtu.2014.02.025
11. Zhai X, Zhou X, Song S. *Research on mine disaster prediction method based on ARIMA model.* (2023).
12. Sun M, Duan Y, Zheng L. Application of ARIMA, ARIMAX, and NGO-LSTM models in predicting the number of tuberculosis cases in Liaocheng City, Shandong Province. *Chin J NTI Tuberculosis.* (2023) 26:1–16. doi: 10.19982/j.issn.1000-662
13. Cui X, Zhou C, Yu J, Khan AN. Interaction between manufacturer's recycling strategy and e-commerce platform's extended warranty service. *J Clean Prod.* (2023) 399:1–16. doi: 10.13301/j.cnki.ct.2024.06.034
14. Wang F, Liu X. Economic analysis of the reform of China's natural gas price formation mechanism-from "cost-plus" pricing method to "market netback" pricing method. *Nat Gas Ind.* (2014) 34:135–42.
15. Yu J, Zhao J, Zhou C, Ren Y. Strategic business mode choices for e-commerce platforms under brand competition. *J Theoretical Appl Electron Commerce Res.* (2022) 17:1769–90.