Research on Vegetable Sales Strategies Based on Multi-objective

Yu Zhang[1, #], Yitong Ding[2, #], Manqing Shi[3], Mingda Wang[4], Sumei Li[5], Xiangyu Ren[6, *]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Weihai, China, 264209

[2]School of Economics and Management, Harbin Institute of Technology, Weihai, China, 264209

3Faculty of Science, Harbin Institute of Technology, Weihai, China, 264209

[4]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai China, 200433

[5]Faculty of Ocean Engineering, Harbin Institute of Technology, Weihai, China, 264209

[6]College of Automotive Engineering, Jilin University, Changchun, China, 130022

*Corresponding author: 18888229612@163.com

#These authors contributed equally.

Abstract. This paper analyzes and predicts the sales and pricing strategies of fresh produce through the Prophet time series prediction model, difference equation model, and multi-objective optimization model. Observations of the data reveal that sales exhibit intra-week and intra-year cyclical fluctuations, particularly significant changes during weekends and between March and June, with a peak in August. The analysis of six major categories of fresh produce using the Pearson and Spearman coefficients shows a strong positive correlation, reflecting the complementarity and substitutability of market demands. Integrating the results from these models, this study constructs an optimization model aimed at maximizing profits, proposing optimal replenishment volumes and pricing strategies, thereby validating the predictive accuracy and practical value of the models. Moreover, the model analysis sheds light on market trends and shifts in consumer preferences, offering new insights into supply chain management and product positioning within the fresh produce industry.

Keywords: Prophet Model, Difference Equation Model, Multi-Objective Optimization Model.

1. Introduction

Based on the insights gleaned from time series forecasting results and the dynamic relationship between sales volume and pricing revealed by the difference equation model, this study constructs an optimization model aimed at maximizing profits. This model not only offers optimal solutions for replenishment volumes for the upcoming week but also provides recommendations for pricing strategies, striving to balance cost, demand, and market competition factors. Through a comprehensive evaluation of the model, this study validates its predictive accuracy and practical value, offering scientific decision support for the sales strategies of fresh produce. In the field of machine learning, numerous studies have addressed this issue. Some researchers have integrated the LightGBM model and LSTM (Long Short-Term Memory) network model, employing methods such as the arithmetic mean method, entropy weight method, variance inverse method, and ensemble learning stacking approach for model fusion [1]. Other scholars have utilized econometric methods such as the Probit model, Logit model, propensity score matching, instrumental variable method, mediation effect model, and moderation effect model[2]. Additionally, many have employed machine learning techniques like the LSTM and other predictive models for forecasting.

2. The basic funamental of BP neural network Sales Volume Distribution Model Based on Prophet

2.1. Application of the Prophet Model

2.1.1 Basic Components of the Prophet Model

Considering the sales volume has both a long-term stable underlying trend and noticeable seasonal fluctuations, and taking into account sudden social factors (such as pandemics, sudden weather changes, etc.), the model can be established from four aspects: Trend, Seasonality, Holidays, and Error Term. The expression of the Prophet algorithm is represented by the following formula:

$$y(t) = g(t) + s(t) + h(t) + \epsilon \tag{1}$$

Where: y(t) represents the actual sales volume or forecasted sales volume of a certain category of vegetables, g(t) represents the non-periodic basic trend changes in the sales volume of a certain category of vegetables, s(t) represents the periodic trend changes in the sales volume of a certain category of vegetables, analyzed from weekly and yearly dimensions, h(t) represents the extreme changes in the sales volume of a certain category of vegetables, which are not periodic, and ε represents the error variation, dealing with abnormal changes in sales volume caused by sudden factors.

2.1.2 Non-Periodic Trend Component of a Certain Category of Vegetable Sales

The non-periodic underlying trend changes can be analyzed and established from two directions, namely the logistic growth model based on logistic regression and the piecewise linear model. This trend excludes the impact of periodic factors and extreme changes, making a linear fit in the non-periodic part while avoiding the impact of missing data in conventional models. [3]

The change in the underlying trend plays a crucial foundational role in the Prophet model, allowing for reasonable trend analysis and prediction based on the existing categories of vegetables and their corresponding sales volume changes.

(1)Piecewise Logistic Regression Model

The sales volume of vegetables exhibits non-periodic changes. These foundational changes can be represented by a piecewise logistic regression model, as shown in Equation 2:

$$g(t) = \frac{C(t)}{1+\exp\left(-(k+\alpha(t)^t\delta)\cdot(t-(m+\alpha(t)^T\gamma))\right)} \tag{2}$$

In this equation, $\alpha_j(t) = \begin{cases} 1, & t \geqslant s_j \\ 0, & \text{others} \end{cases}$, Moreover, t represents time, C(t) signifies the carrying capacity, k denotes the growth rate of vegetable sales, and m is the offset parameter, with the growth rate of sales being $k + \alpha^T\delta$.

(2)Piecewise Linear Model

Additionally, this paper can also employ a piecewise linear model for fitting to observe its degree of fit, thereby choosing the better-performing model between the two. This model can be represented by Equation 3.

$$g(t) = [k(t) + a(t)^T\delta]t + [m(t) + a(t)^T]\gamma \tag{3}$$

In the formula, it is stipulated that $\gamma = -s_t \times \delta_t$, ensuring the function is continuous and related to the inflection points associated with extreme changes in sales volume;

k(t) represents the growth rate of vegetable sales over time t, m represents the offset parameter, δ represents the change in the efficiency growth rate of vegetable sales, and a(t) represents the frequency of change in inflection points prior to the moment t when the model predicts extreme values in sales volume.

2.1.3 Seasonal Component of a Certain Category of Vegetable Sales

Based on everyday experience, it can be inferred that the seasonality of vegetable sales fluctuations primarily manifests in two periodic dimensions: one is a weekly cycle measured in days, and the other is an annual cycle measured in months[4]

In the Prophet model, to construct the seasonal model, the inclusion of a Fourier series is chosen to fit the changes in the sales volume time series, which can be expressed as Equation 4.

$$s(t) = \sum_{n=1}^{N}\left(an\cos\frac{2n\pi t}{P} + bn\frac{2n\pi t}{P}\right) \tag{4}$$

In Equation 4, P represents the length of the cycle for the daily sales volume time series, a_n and b_n represent Fourier coefficients, and N denotes the order of the self-adjusting Fourier series, which acts similarly to a low-pass filter.

When selecting the order N for self-adjustment in the Prophet model, different choices yield varying effects, thus necessitating continuous experimentation to determine the optimal value of N. Research has shown that the best fitting results are achieved when N=10 (for annual cycles) and N=3 (for weekly cycles).

For the seasonal component s(t), this paper can further express it using Equation 5:

$$s(t) = X(t)\beta\ ,\ \beta \sim Normal(0, \sigma^2) \tag{5}$$

$$X(t) = [\cos(\frac{2\pi(1)t}{P}), \sin(\frac{2\pi(1)t}{P}), \cdots, \cos(\frac{2\pi(N)t}{P}), \sin(\frac{2\pi(N)t}{P})] \tag{6}$$

2.1.4 The Impact of Extreme Events on Vegetable Sales Volume Variations

Taking real-life circumstances into account, and considering sudden social factors such as pandemics, abrupt weather changes, or sporadic promotional events held by shopping centers, the sales volume predicted by the Prophet model may deviate from the actual sales volume. Therefore, it is necessary to allocate a specific allowance for these sudden extreme impacts to enhance the accuracy of the fit. This extreme term is represented by Equation 7.

$$h(t) = Z(t)k = \sum_{i=1}^{L} k_i \times 1\ (t \in D_i), k \sim Normal(0, \upsilon^2) \tag{7}$$

$$Z(t) = [1(t \in D_1), 1(t \in D_2) \cdots, 1(t \in D_L)] \tag{8}$$

L represents the set of sudden extreme events, containing i elements, where k_i denotes the impact factor of sudden extreme events. D_i represents the set of time points t within the time variation window, as expressed in Equation 8. Z(t) symbolizes the influence of a single extreme event, and v represents the coefficient of impact of extreme events.

2.1.5 Solving the Prophet Model

By selecting a training set and a validation set, a Prophet time series prediction model is constructed. The sales volume predicted by the Prophet model can be refined through adjustment points to enhance the fitting degree or to adjust the trend of overfitting. There are four parameters related to change points, among which, the larger the value of changepoint_prior_scale, the greater the flexibility of the fitting curve. After adjusting the coefficients, the prediction values derived from the training set are compared with the validation set's COR (Correlation Coefficient) results and adjusted until the fitting precision meets the standard, at which point the COR value can converge steadily. Predictions and validations are carried out as required, followed by result analysis.

2.1.6 Result Analysis

To investigate the distribution patterns of sales volumes across different categories of vegetables, this study utilized the time series of sales volumes for six types of vegetables from 2020 to 2023.

Based on the Prophet model, time series decomposition graphs were drawn (Figs 1 and 2). Fig 1 displays the time series decomposition graph for the cauliflower category, which has distinctive features, and uses this as a case study to explore the distribution pattern of sales volumes. The x-axes of the three subgraphs in this set represent the time gradient, while the y-axes have different meanings.

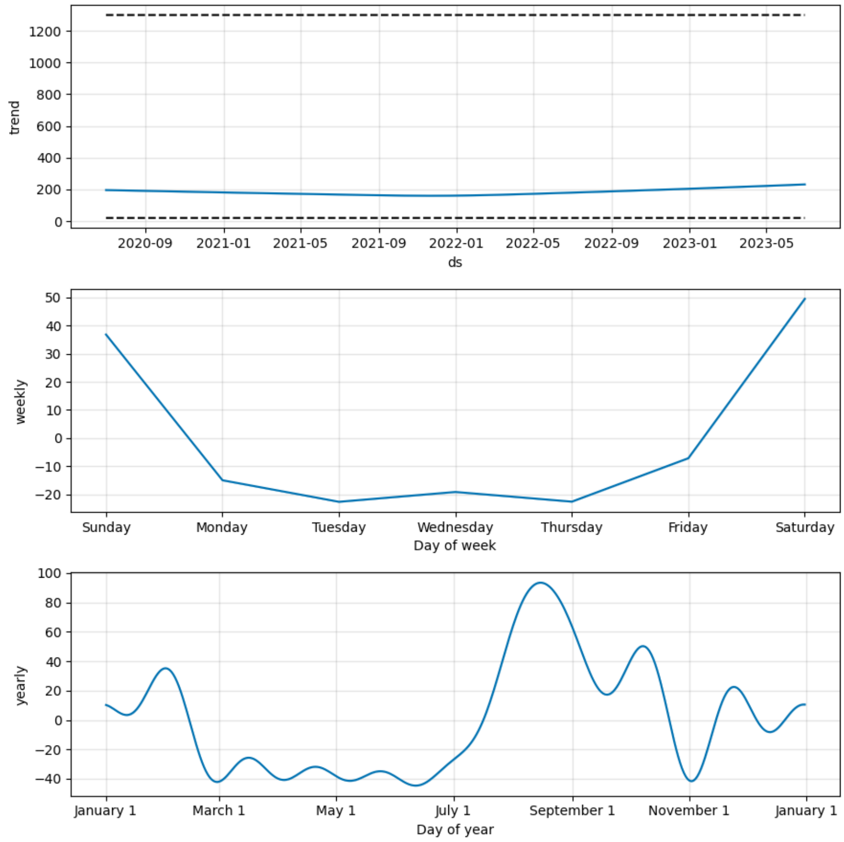

Fig 1. Time Series Decomposition Graph of Sales Volume Based on the Prophet Model (Cauliflower Category)

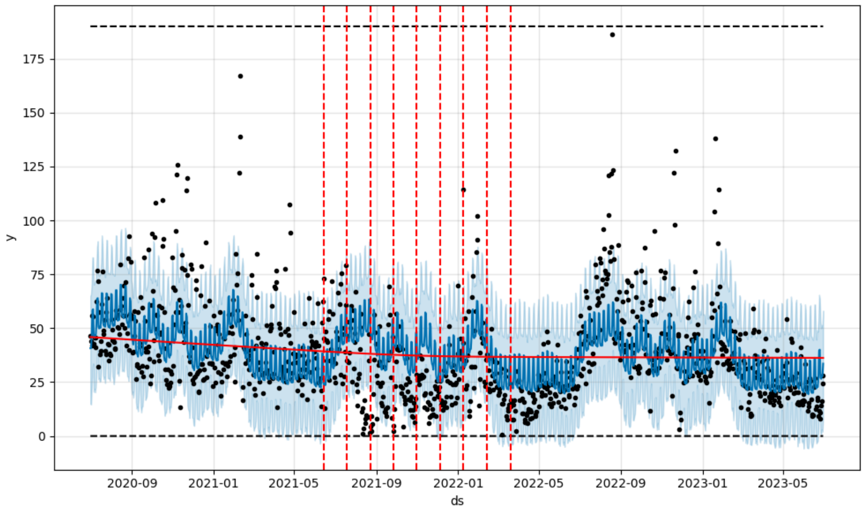

Fig 2. Time Series Decomposition Graph of Sales Volume Based on the Prophet Model (Cauliflower Category)

Figs 1 and 2 illustrate the seasonal variation trends in vegetable sales volumes through time series analysis. The Weekly subplot in Fig 1 shows that the sales volumes of various vegetable categories tend to decline from Monday to Thursday compared to the baseline trend, begin to recover slightly on Friday, and significantly increase over the weekend. This reflects the customer preference for making bulk purchases during weekends. The Yearly subplot in Fig 2 reveals a steady decline in vegetable sales volumes relative to the annual baseline trend from March to June. This phenomenon could be attributed to the increased variety of vegetables available in the market and intensified price competition during this period, leading to customers spreading their purchases across different stalls, thereby affecting the sales volumes of individual vegetable categories. Together, these charts reveal the impact of weekly and annual seasonality on vegetable sales volumes, aligning with actual sales data.

2.2. Distribution Pattern of High Sales Days

2.2.1 Model Selection and Establishment

This paper defines a high sales day—if the total sales volume of a certain category of vegetables on a given day exceeds 90% of the total sales volume for all time periods, that day is considered a high sales day. Statistical analysis was conducted using Excel.

2.2.2 Result Analysis

The high sales days of different categories, when tallied by month, display varied outcomes, which may be attributed to customer demand and the seasonal nature of vegetable harvesting. Looking for commonalities among these individual patterns, it's still not difficult to observe that, with the exception of edible mushrooms, the number of high sales days in August is the highest or second-highest for all categories, with the count of high sales days within the month ranging between 18 to 30 days, exceeding half the number of days in the month. Moreover, for most categories, the low points of high sales days occur in spring (March to May), and there may even be no high sales days at all during this period. As can be seen from the right chart, the high sales days of the six vegetable categories show a high degree of similarity in the dimension of daily statistics within a week. The number of high sales days on Saturday and Sunday, the weekend days, are the highest and second-highest, respectively, indicating that weekends stimulate customer demand, resulting in a high concentration of high sales days on these two days.

2.3. Interrelationships Among Various Vegetable Categories

2.3.1 Result Analysis

To explore the correlations among the six categories of vegetables, Scatterstats graphs were created based on the processed daily sales data of the six types of vegetables, illustrating the degree of association between two categories of vegetables.

Fig 3 show cases four scatter plots with the highest Pearson correlation coefficients, where black dots represent the daily sales volumes of two categories of vegetables, and the blue line is the trend line obtained from fitting these points. The histograms on the top and right sides respectively represent the distribution of daily sales volumes for the two categories of vegetables.

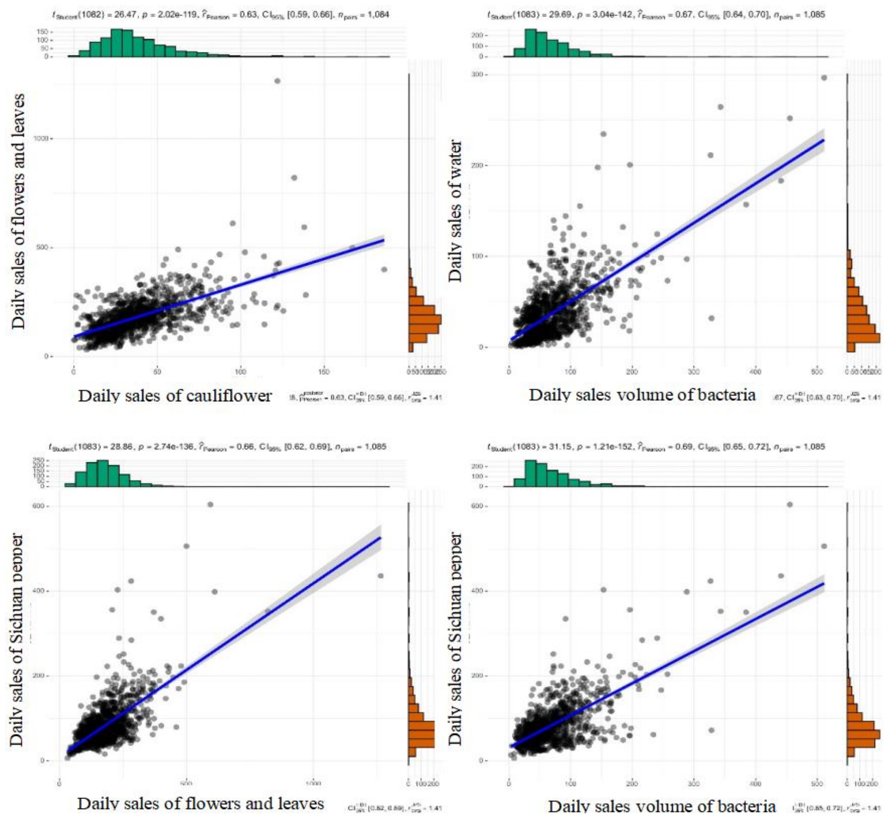

Fig 3: Data Distribution Characteristics of Daily Sales Volumes for Two Categories of Vegetables

Through the application of Pearson and Spearman correlation coefficients analysis, this study found a significant positive correlation between the daily sales volumes of edible mushrooms and chili peppers. Specifically, the Pearson correlation coefficient is 0.69, indicating a strong linear relationship; the Spearman correlation coefficient is 0.535, suggesting a positive monotonic relationship between these two types of vegetables. This means that an increase in sales of edible mushrooms often accompanies an increase in chili pepper sales, reflecting a customer tendency to purchase these two types of vegetables together. The discovery of this relationship holds significant implications for formulating sales and promotional strategies. [5]

3. Sales Volume and Pricing Relationship Model Based on Difference Equations

3.1. Sales Volume Based on Difference Equations

3.1.1 Model Establishment

Let the total daily sales volume of a certain category of vegetables on the ith day be x_i, and the total daily sales volume on the (i-1)th day be $x_{\mathrm{i-1}}$, The difference between the two, denoted as Δx_i, represents the fluctuation in sales volume. The difference in pricing between the ith day and the (i-1)th day is denoted as Δy_{i}, Indicating pricing fluctuations. Here, i=1,2,3,...,n, where n is the total number of days covered in the sales details from 2020 to 2023 as presented in the data source.

The difference equation that represents the relationship between the daily sales volume and the change in daily pricing is: $f(x_i - x_{i-1}) = f(\Delta x_i) = \Delta y_i$. By solving this difference equation, we can determine the magnitude ofΔx_i, which can then be used to infer Δy_i.

Based on the aforementioned Prophet model, the predicted daily sales volume for the ith day can be denoted as Y. Considering the impact of price fluctuations on sales volume as per the difference equation model, the daily sales volume for the ith day can be further expressed as $Y + \Delta y_i$. Utilizing the daily sales volume and daily pricing values derived from this model, along with the sales volume for the current day, reasonable decisions can be made regarding the replenishment volume for the following day. [6]

3.1.2 Solving the Model

(1) For "Leafy Vegetables":

The statistical results of the model fitted using important features for the "Leafy Vegetables" data are as follows: R^2: 0.531, Adjusted R^2: 0.52. All selected features are statistically significant. The final model equation is:

$$\Delta Sales\ Volume = 0.0002 + (0.0780 \times \Delta Price) + (-0.0078 \times \Delta Price^2) + (-0.0690 \times sin(\Delta Price))$$

(2) For cauliflower-like products

For cauliflower-like products, the R^2 is 0.249 and the adjusted R^2 is 0.246. The statistical significance (p-value) of the features is as follows: Δprice $p<0.001$ (significant), sin(Δprice): $p<0.001$ (significant), cos(Δprice): $p=0.011$ (significant). The final model equation is:

$$\Delta sales = -0.0023 + (0.0547\times\Delta price) + (-0.0309\times\sin(\Delta price)) + (0.0076\times\cos(\Delta price))$$

3.2. Pricing strategy model based on single-objective programming model

Profit maximization is the main consideration in the pricing of this paper, and the pricing strategy problem is transformed into a single-objective optimization problem. [7]

3.2.1 Establishment of the objective function

To make decisions on pricing and replenishment for the next week, it is necessary to construct an objective function based on the goal of profit maximization. The assumption made earlier in this paper is that returned products will not be resold, so the daily profit of the fresh supermarket can be represented by the following formula:

Daily profit = Net daily sales volume × (Profit from discounted part + Profit from non-discounted part) - Return amount

(1) β_j: Daily return probability of category j

First, define the frequency of return behavior per day over all time periods. Within 1085 days, the number of days with return behavior is denoted as, where i represents the sequence number of the days of return, arranged in chronological order.

$$\beta_j = f_i / Number\ of\ days\ within\ the\ statistical\ range \tag{9}$$

Secondly, define within all days of return behavior, based on the daily sales revenue, the average frequency of return behavior for a certain category:

Average return rate on days with returns=[Number of returns on that day / (Number of returns on that day + Number of non-returns on that day)]/ f_i

(2) γ_j : Daily discount probability for category j

Define the discount rate using the same method as the return rate calculation. Firstly, for a certain category of vegetables:

$$\gamma_j = \text{Number of days with discounts} / \text{Total number of days within the statistical range} \tag{10}$$

Secondly, for a certain category of vegetables, the frequency at which the product is purchased at a discount on discount days is:

[Number of discounted sales on that day / (Number of discounted sales on that day + Number of regular price sales on that day)] / Number of days when discount behavior occurs.

(3) t_j : Average sales volume proportion when discounts occur for category j t_j = Discounted sales volume of category j / Total sales volume of category j

3.2.2 Establishment of constraints

The constraints are explained as follows: the relationship between daily sales volume differences $\Delta x_i^{(1)}$ 、 $\Delta x_i^{(2)}$under different sales price statuses and the daily pricing difference Δy_i, where the daily pricing difference Δy_ion day i is defined as the difference in pricing between day i and day i−1. The daily sales volume predicted for day i based on the Prophet model is defined as: the sum of the discounted and regular price sales volumes predicted for day i based on the Prophet model. Based on life experience, to achieve profit, the normal selling price $y_i$on day i must be higher than or equal to the cost price μ_i on that day. Since vegetables may not have been sold before a discount on a certain day, satisfying greater than or equal to 0; however, the daily sales volume cannot be infinitely large, influenced by factors such as daily stock levels, it must be restricted within different ranges daily, thus the predicted daily sales volume satisfies less than or equal to a certain condition. Analyzing the attachment data, it is known that there are situations where some category of goods are not sold at a discount on certain days, therefore, the predicted daily sales volume of discounted sales is restricted to be greater than or equal to 0.[8]

Thus, this paper successfully establishes a single-objective optimization model with the goal of maximizing profit as follows.

Objective function: (In the following equation $i=1,2,\cdots,7$)

$$\begin{aligned}\max\sum_{i=1}^{7}\ &(1-\beta_j)\left[(1-\gamma_j)\left(y_j^{(i)}-\mu_j^{(i)}\right)\left(x_j^{(i)}+\Delta x_j^{(i)}\right)+\gamma_j t_j\left(cy_j^{(i)}-\mu_j^{(i)}\right)\left(x_j^{(i)}+\Delta x_j^{(i)}\right)\right]+\\ &\beta_j\left[(1-g_j)(1-\gamma_j)\left(y_j^{(i)}-\mu_j^{(i)}\right)\left(x_j^{(i)}+\Delta x_j^{(i)}\right)+(1-g_j)\gamma_j(1-t_j)\left(y_j^{(i)}-\mu_j^{(i)}\right)\left(x_j^{(i)}+\Delta x_j^{(i)}\right)\right]\\ &-\beta_j\mu_j^{(i)}g\left(x_j^{(i)}+\Delta x_j^{(i)}\right)\end{aligned} \tag{11}$$

$$\text{constraints：} s.t.\begin{cases}\Delta x_i^{(1)}+\Delta x_i^{(2)}=f(\Delta y_i)\\ \Delta y_i=y_i-y_{i-1}\\ X_i=x_i^{(1)}+x_i^{(2)}\\ y_i\geqslant\mu_i\\ y_0=d\\ x_i^{(1)}\in[0,h]\\ x_i^{(2)}\geqslant 0\end{cases} \tag{12}$$

3.2.3 Solving the model

(1) Data Feature Extraction

In the objective function, discount and return rates are key factors for constructing the function. Analysis and processing of the provided data were conducted, focusing on six categories of vegetables to obtain the frequency of occurrence for return and discount behaviors. [9]

(2) Model Solution

The maximum profit for the six major categories of vegetables can be obtained using genetic algorithms. [10]

4. Conclusions

This study conducted a comprehensive analysis and prediction of fresh vegetable sales and pricing strategies using the Prophet time series forecasting model, difference equation model, and multi-objective optimization model. The results reveal the seasonal patterns of vegetable sales, particularly the significant fluctuations during weekends and the period from March to June, as well as the sales peak in August. Through Pearson and Spearman correlation analyses, a significant positive correlation was found among different categories of vegetables, indicating complementary and substitutable characteristics in market demand.

Combining the time series forecast results with the relationship between sales and pricing revealed by the difference equation model, an optimization model aimed at maximizing profit was constructed. This model provides scientific decision support for replenishment volumes and pricing strategies for the upcoming week, balancing cost, demand, and market competition factors. A comprehensive evaluation of the model confirmed its predictive accuracy and practical value, offering strong guidance for fresh vegetable sales strategies.

Moreover, the model analysis of this study also helped to reveal changes in market trends and consumer preferences, providing new insights for supply chain management and product positioning in the fresh produce industry. Through meticulous data analysis and model construction, this research emphasizes the importance of data-driven decision-making in fresh vegetable sales management and offers effective strategic tools for industry practitioners.

In summary, the study of vegetable sales strategies based on multi-objective planning not only enhances our understanding of the dynamics of vegetable sales but also provides practical methodological guidance for improving sales efficiency, optimizing inventory management, and formulating effective pricing strategies. Future research could further explore sales strategies under different market conditions and the application of emerging technologies in sales forecasting and optimization to continuously enhance the competitiveness of the fresh vegetable industry.

References

[1] Zhang He. Vegetable Sales Forecasting Based on Ensemble Learning [D]. Yunnan Normal University, 2021.

[2] Li Quanhai. The Impact of E-commerce Participation on Green Production Behavior of Facility Vegetable Growers [D]. Shandong Agricultural University, 2024.

[3] Nie Yuxuan. Research on Automatic Pricing and Replenishment Strategies for Fresh Commodities Based on the ARIMA Forecast Optimization Model: A Case Study of Vegetables [J]. Business Show Economy, 2024(05): 19-22.

[4] Li Quanhai. The Impact of E-commerce Participation on Green Production Behavior of Facility Vegetable Growers [D]. Shandong Agricultural University, 2024.

[5] Tian Zhongwei, Dong Ming. Pricing and Ordering Strategies for Fresh Products of Different Quality Grades [J]. Journal of Shanghai Jiao Tong University, 2014, 48(02): 306-311.

[6] Liu Shanshan. Dynamic Pricing Research of Fresh Products Based on Consumer Time Preferences [D]. Hebei University, 2020.

[7] Yang Shuai, Huang Xiangmeng, Wang Junbin. Research on Joint Optimization Strategy of Shelf Allocation and Pricing for Fresh Food [J]. Supply Chain Management, 2022, 3(08): 49-59.

[8] Fan Fengyan. Research on the Pricing Model of Supermarket Fresh Products [D]. Fujian Normal University, 2020.

[9] Hong Junjie. Research on Product Assortment and Ordering Decision of Small and Medium-Sized Cross-Border E-commerce [D]. Jinan University, 2019.

[10] Zeng Zhiqiang, Cai Wenqing. Research on Multi-objective Optimization of Flexible Flow Shops Based on Total Setup Time and Maximum Completion Time [J]. China Pulp & Paper, 2024, 39(01): 82-90.