Research on Automated Pricing and Replenishment Decision for Vegetable Products based on Optimization Model

Tong Yang[a], Shengjie Xu[b], Jinyu Lu[c], Zhiyun Yu[d], Xinfeng Liu[e]*

Shandong Jianzhu University Jinan, China

[a]2487107665@qq.com, [b]438151971@qq.com, [c]1143016031@qq.com, [d]13789161@qq.com, [e]liuxinfeng18@sdjzu.edu.cn

Abstract. In the current retail environment, efficient and accurate pricing and replenishment strategies are particularly important for vegetable products, which typically have a short shelf life and are highly volatile in market demand. In this work, we analyzed the sales data of vegetable categories and established a decision-making model for automatic pricing and replenishment of supermarkets to maximize the sales revenue of vegetables. Specifically, we established a random forest prediction model based on historical wholesale price data, which can accurately predict the wholesale price of vegetables in the coming week. Subsequently, we utilized the XGBoost regression model based on the unit price and wholesale price to evaluate the predicted wholesale prices of 33 vegetable categories. From our extensive experiments, we can observe that the model performed well with an average R2 value of more than 0.64. This indicates that the model can explain 64% of the variance in this type of data, indicating that the model has relatively good predictive power. Then, a target planning model for maximizing the revenue of supermarkets is established. By using the particle swarm optimization algorithm to solve, we successfully obtained that the total income of 33 single products on July 1 was 3001.897512 yuan under the premise of satisfying the constraints.

Keywords: Pricing and replenishment decision, Vegetable products, Random forest prediction, XGBoost regression.

1 Introduction

At present, supermarkets need to make replenishment and pricing decisions based on the shelf life and quality changes of vegetables. In response to the issue of shelf life, supermarkets can record the shelf life of each vegetable category and use data analysis methods to predict the sales speed of each category, so as to reasonably determine the replenishment time [1]. Regarding the change of appearance, supermarkets can use image recognition technology to evaluate and monitor the quality of vegetables, and through the correlation analysis with sales data, it can determine when to sell goods with poor quality and quality at a discount.

K. Subramaniyam et al. (eds.), *Proceedings of the 3rd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2024)*, Atlantis Highlights in Computer Sciences 21,
https://doi.org/10.2991/978-94-6463-419-8_29

In the modern retail industry, especially in the fast-changing vegetable market, having an effective pricing and replenishment strategy in place is critical to staying competitive and meeting consumer demand. Vegetable commodities present unique challenges for retailers due to their perishability and seasonal fluctuations in demand [2]. Traditional pricing and replenishment methods often rely on empirical judgment and simplified rules, which are not only time-consuming and labor-intensive, but also difficult to adapt to the rapid changes and complexities of the market. Therefore, it is important to develop a method that can automate these decision-making processes to increase efficiency, reduce waste, and maximize profits.

With advancements in data science and machine learning techniques in recent years, retailers now have the opportunity to leverage these tools to optimize their pricing and replenishment strategies. By analyzing historical sales data, consumer behavior, market trends, and other relevant factors, predictive models can be developed that can predict future changes in demand and make more accurate pricing and replenishment decisions [3]. In addition, considering that vegetable prices and demand are affected by a variety of factors, such as weather conditions, holidays, and competitor behavior, a comprehensive optimization model can provide retailers with a framework to dynamically adjust their strategies in response to these complex market conditions.

Optimization models play a central role in the fields of decision science and operations research, providing systematic and quantitative solutions to a wide range of complex problems. These models can help researchers and practitioners find the best decision-making scheme under the given constraints through mathematical methods [4]. With the rapid development of computing technology, optimization models have been widely used in many fields such as industrial engineering, finance, logistics, health management, and retail, helping to solve a series of complex problems from resource allocation to production scheduling to product pricing and inventory management.

At the heart of the optimization problem is the process of finding the process of maximizing or minimizing the value of an objective function while satisfying a set of constraints. This objective function can be cost, benefit, time, efficiency, or any other quantified performance metric. Constraints define the boundaries of the problem, including, but not limited to, resource constraints, technical parameters, or market rules [5]. The design and solution of optimization models involves complex mathematical theories and algorithms, including linear programming, nonlinear programming, integer programming, dynamic programming, and more recently, heuristic and meta-heuristic algorithms.

At present, with the advancement of big data and artificial intelligence technology, the application scope and capabilities of optimization models are constantly expanding. By incorporating machine learning methods, optimized models are not only able to handle more complex and dynamic problems, but also learn from historical data and predict future trends to provide more accurate and effective decision support [6]. For example, in supply chain management, optimization models can help companies anticipate future fluctuations in demand and optimize inventory levels and replenishment schedules accordingly to reduce costs and improve service levels. In the field of energy management, optimization models can be used to balance supply and demand, optimize generation and transmission plans, and improve energy efficiency and reduce environmental impact [7].

Although optimization models provide powerful tools, they still face a number of challenges in real-world applications, including the complexity of the model, the need for computational resources, and the uncertainty and dynamics of real-world data [8]. Therefore, the research and development of optimization models requires not only in-depth theoretical foundation, but also interdisciplinary knowledge and technology, as well as a deep understanding of practical problems and innovative solutions. As research deepens and technology advances, optimization models will continue to play an essential role in helping people better understand and improve our world.

2 Notions

We summarize the primary used parameters in following Table 1.

Table 1. Primary notions description

Notions	Definition
$L(\theta)$	Loss function
$\Omega(f)$	Regularization terms
yi	True value
y^i	Predicted value
w	Wholesale price on day i
x_i	Unit price of the sale on the i-th day
X_t	Time is a eigenvalue of t
Loss_rate	Attrition rate
P_w^{t+1}	Wholesale price
q	Samples map to a tree structure of leaf nodes
T	Number of leaf nodes
y^T	L1 regex

3 METHODOLOGIES

3.1 Sales volume forecasts

Following items describe the detail procedure for proposed prediction model.

- Filter vegetable categories: Select a specific vegetable category from all vegetables for analysis.
- Build a price prediction model: Use historical wholesale price data to build a random forest model that can predict the wholesale price of the vegetable category in the coming week.
- Build a sales volume prediction model: Based on the characteristics of sales volume, sales unit price and wholesale price, build an XGBoost model to predict the sales volume of the vegetable category on the same day.

- Build a revenue maximization model: According to factors such as sales volume and sales unit price, build a revenue maximization model for supermarkets. By setting the constraint function and parameters and using the heuristic algorithm to solve the problem, the sales volume and sales unit price strategy that maximizes the revenue every day are obtained.
- Derive the total replenishment volume and pricing strategy: Repeat above steps obtain the daily replenishment volume and pricing strategy for the vegetable category for the coming week (January 1 to 7, 2023).

XGBoost is the abbreviation of extreme gradient ascension, and XGBoost algorithm [9] is a kind of synthesis algorithm that combines basis functions and weights to form a good fit to data. Due to the advantages of the XGBoost model such as strong generalization ability, high scalability, and fast operation speed, the XGBoost model can be expressed as Equation 1 for datasets containing n m-dimensions.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F(i = 1,2, \dots, n)(1)$$

Additionally, q is the tree structure of the samples mapped to the leaf nodes, T is the number of leaf nodes, and w is the real fraction of the leaf nodes. When constructing an XGBoost model, it is necessary to find the optimal parameters according to the principle of objective function minimization to establish the optimal model. The objective function of the XGBoost model can be divided into the error function term L and the model complexity function term Ω. The objective function can be written as Equation 2, where γT is an L1 regex and $\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$ is an L2 regex term. When using training data to optimize the model, you need to keep the original model unchanged and add a new function f to the model to reduce the objective function as much as possible.

$$\begin{gathered} Obj = L + \Omega \\ L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (2) \\ \Omega = \gamma \mathrm{T} + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \end{gathered}$$

In the XGBoost algorithm, in order to quickly find the parameters that minimize the objective function, the second-order Taylor expansion of the objective function is carried out to obtain the approximate objective function, and when the constant term is removed, it can be seen that the objective function is only related to the first-order and second-order derivatives of the error function. In this case, the objective function is expressed as Equation 3.

$$\begin{aligned} Obj^{(t)} &= \sum_{i=1}^{n} [g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2] + \gamma \mathrm{T} + \frac{1}{2} \sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T} [\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma \mathrm{T} \ (3) \end{aligned}$$

Obj is a scoring function that can be used as an evaluation model, and the smaller the Obj value, the better the model performance. By recursively calling the above tree building method, a large number of regression tree structures can be obtained, and the optimal XGBoost model can be established by using OBJ to search for the optimal tree structure and put it into the existing model. Following Figure 1 describes the principle of XGBoost model.

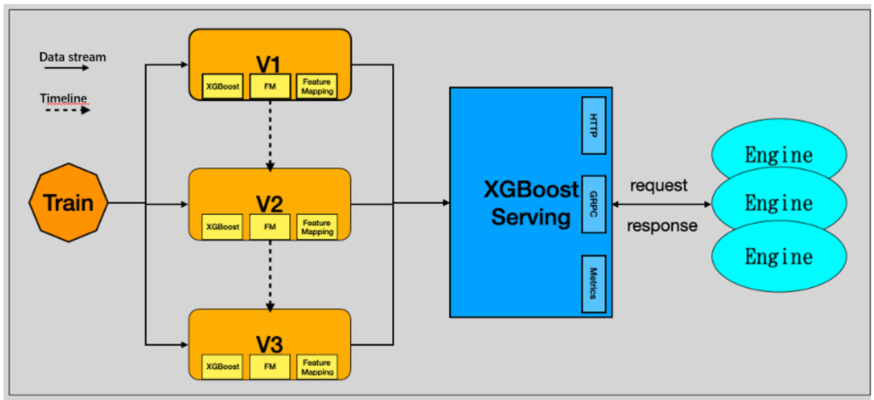

Fig. 1. XGBoost schematic illustration.

3.2 Particle swarm optimization

In order to solve the above planning model, the PSO algorithm can be used. Particle Swarm Optimization (PSO) is a heuristic optimization algorithm that simulates social behavior, inspired by the foraging behavior of swarms and schools of fish. In the PSO, each solution is seen as a "particle" in search space. All particles have a fitness value, which is evaluated by the objective function of the optimization problem. The specific steps are as following items. We need to find the optimal selling price x in such a way that the objective function Π(x,w) is maximized, while satisfying the given constraints.

- Initialization: The selling price of each vegetable item is taken as a set of initial values for the model, and an initial velocity is assigned to this set of values.
- Evaluation: The fitness values of each vegetable category were evaluated using the objective function Π(x,w).
- Update: Update the velocity and position of particles based on the historical best position of the item and category, as well as other parameters such as learning factor and inertia weight.
- Check constraints: After each update of the location, make sure that the new location meets the constraints. If not, appropriate adjustments are required.
- Termination: Stop the algorithm when a preset number of iterations is reached or when the value of the objective function changes very little over several successive iterations.

Following Figure 2 describes the general procedures of optimization process.

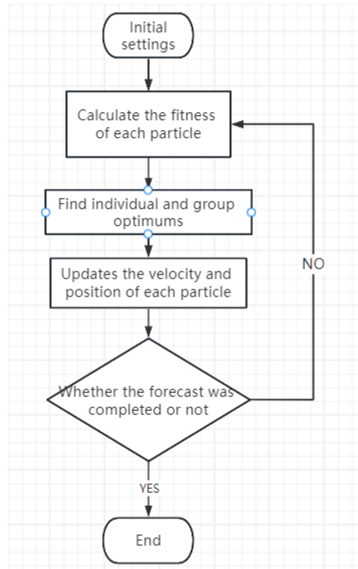

Fig. 2. PSO algorithm flowchart.

4 ExperimentS

4.1 Experimental setups

The collected vegetable sales data are merged, and the date and single product code are stringed firstly, combined as the main key, and the wholesale price is obtained based on the merger and merge the merged data.

4.2 Experimental analysis

During the iterative process, the algorithm will continuously select individuals with higher fitness to cross and mutate to generate new populations. Ideally, as the number of iterations increases, the maximum fitness in the population gradually increases until a satisfactory solution or maximum number of iterations is reached. Through the iterative graph, the optimization efficiency and convergence speed of the algorithm can be visually observed, so as to evaluate the performance of the algorithm, as well as the parameters or strategies that may need to be adjusted to improve the search ability of the algorithm and the quality of the solution. Following Figure 3 shows the iterations of used optimization model.

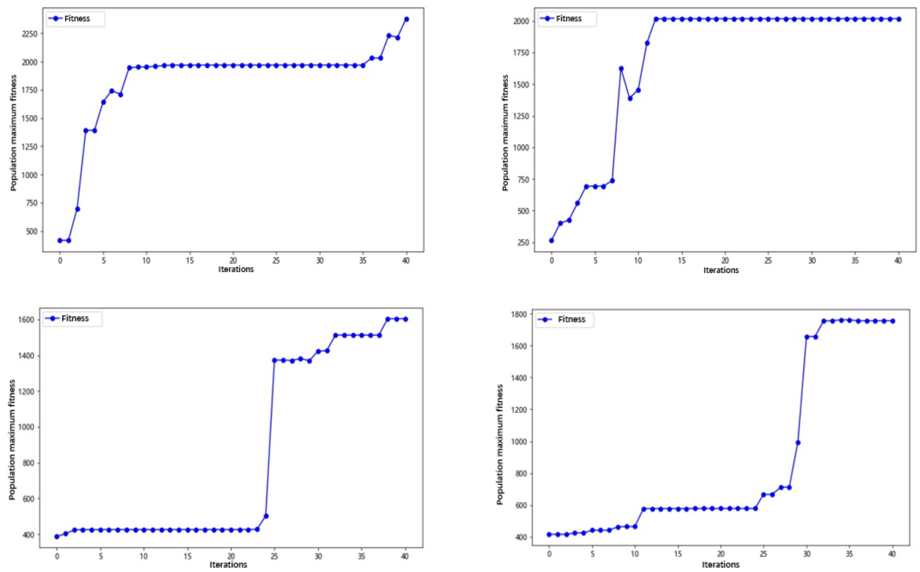

Fig. 3. Iterative graphs of the 0th, 2nd, 4th, and 6th days of the category mosaic and leaf category.

By solving the above steps, we can get the next 7 days for each category. The sales unit price and sales quantity corresponding to the largest revenue of the supermarket are not given because the data of the question does not give its storage capacity and finally the total revenue in the next seven days is 41398.93 yuan, as shown in Table 2.

Table 2. Prediction results table

Future Days	Unit price	Number of sales	Supermarket benefits	Category
1	8.613115057	115.635437	299.8229247	Chili peppers
2	13.82737698	206.4257202	1594.332769	-
3	17.65087266	302.0381165	3565.839372	-
4	17.64268482	292.8843689	3443.715013	-
5	6.733972923	132.2991028	245.9635908	-
6	21.39543144	396.0528564	5999.128596	-
7	7.747888693	241.2694855	575.7107347	-
1	8.317788223	478.3677673	2017.517613	Mosaic and leafy
2	8.318289895	513.0485229	2254.994827	-
3	14.55631163	290.2860413	2800.098106	-
4	4.854934449	417.6813354	489.180818	-
5	8.317783287	427.2704773	1781.039488	-
6	7.689602439	363.3972168	1290.950237	-
7	20.52037454	312.6834106	4704.377834	-
1	14.15913375	54.91007996	130.5344852	Aquatic rhizomes
2	19.12153541	31.67584229	237.7609618	-
3	14.60112219	35.76191711	132.7154032	-

Future Days	Unit price	Number of sales	Supermarket benefits	Category
4	12.75020551	66.66902161	212.082205	-
5	13.11492021	116.2756195	362.5873563	-
6	13.13519621	123.8185959	452.5431273	-
7	11.25218598	90.85707092	264.060278	-
1	11.92044831	120.8604965	386.9294046	Edible fungi
2	13.77264158	110.5566254	485.4260481	-
3	12.42502027	234.0227356	685.2815425	-
4	12.72553299	100.3593521	343.7587976	-
5	9.939506564	201.1493835	378.5923746	-
6	11.92297467	164.2913666	674.3232167	-
7	11.95581547	98.17132568	440.0090983	-
1	13.56001968	61.61337662	300.307067	Cauliflower
2	12.51985181	65.0848465	267.6136753	-
3	13.52389177	41.30136108	234.3276926	-
4	11.23388696	59.70246887	212.0377074	-
5	11.0359859	62.86208725	199.1342039	-
6	13.56001974	48.33254242	295.8128887	-
7	13.55330716	48.33254242	295.4046999	-
1	7.900512539	43.3454895	140.5611267	Nightshades
2	7.491109133	38.53446579	100.6028781	-
3	7.914223404	31.78683853	97.5959376	-
4	7.752346655	36.65195847	117.699163	-
5	7.919685834	39.71208572	133.7758369	-
6	7.248416365	46.97365952	129.1979625	-
7	7.570892977	36.00680542	98.24442118	-

5 Conclusion

In conclusion, Through the establishment and solution of the above model, we can see that the model has certain limitations in terms of factor consideration. The model only focuses on factors such as wholesale price, sales unit price, and attrition rate, but ignores other possible influencing factors. However, in real life, vegetable yield, quality, and price are greatly affected by seasonal factors. Therefore, supermarkets can use seasonal data to predict the supply of vegetables in a specific season, and reasonably formulate purchase plans and pricing strategies. To improve the comprehensiveness and accuracy of the model, we can introduce seasonal factor data. For example, the seasonal factor is taken as an influencing factor S and taken into account in the calculation results of the wholesale price in the original model. Through such improvements, we can more fully consider the impact of seasonal factors on the operation of supermarkets.

References

1. Hou, M. (2023). Automated Pricing and Replenishment Decisions for Vegetable Items. Information Systems and Economics, 4(9), 105-111.
2. Zhang, Z., Wu, R., & Cui, H. (2023). Automated Pricing and Replenishment Decisions for Supermarket Fresh Vegetables. Advances in Computer, Signals and Systems, 7(9), 64-72.
3. Liu, Y., Hu, X., & Wang, J. (2024). Research on automatic pricing and replenishment decision model based on vegetable products. Journal of Education, Humanities and Social Sciences, 25, 149-156.
4. Qian, M., & Wang, Y. (2024). Automated Vegetable Pricing and Replenishment Decisions based on PSO. Journal of Education, Humanities and Social Sciences, 25, 41-46.
5. Guo, X., & Zhao, G. (2023). Automatic Pricing and Replenishment Strategy for Vegetable Products Based on XGBoost and LSTM Prediction Models. Highlights in Science, Engineering and Technology, 70, 615-623.
6. Van Bruggen, G. H., Smidts, A., & Wierenga, B. (1998). Improving decision making by means of a marketing decision support system. Management Science, 44(5), 645-658.
7. Almagrabi, A. O., Abdullah, S., Shams, M., Al-Otaibi, Y. D., & Ashraf, S. (2022). A new approach to q-linear Diophantine fuzzy emergency decision support system for COVID19. Journal of Ambient Intelligence and Humanized Computing, 1-27.
8. Pagano, A., Giordano, R., & Vurro, M. (2021). A decision support system based on AHP for ranking strategies to manage emergencies on drinking water supply systems. Water Resources Management, 35(2), 613-628.
9. Cao, L. (2017). Data science: a comprehensive overview. ACM Computing Surveys (CSUR), 50(3), 1-42.