Fusion Of SSA Mahjong Search and LSTM Model for Vegetable Sales and Pricing Research

Bin Cheng *

College of Software, Jilin University, Changchun, China, 130012

* Corresponding Author Email: cblyh670@outlook.com

Abstract. In supermarkets, vegetable products have a short shelf life, and their quality deteriorates over time, making it challenging to sell them the next day. To formulate rational sales and pricing decisions, this article first analyzes the relationship between sales and prices. Due to the unique nature of vegetable products, pricing may also be influenced by historical sales and pricing information, which can be analyzed through regression data.The study begins with data preprocessing using EMD decomposition and utilizes an LSTM model to predict various vegetable categories. Additionally, an intelligent optimization algorithm, the SSA Sparrow Search Algorithm, is employed to optimize sales forecasts for the upcoming week. Combining historical sales data and pricing information, a multivariate time series prediction is performed, again using the LSTM model, but leveraging Bayesian optimization to determine pricing for various vegetable categories in the upcoming week.This approach provides a pricing strategy that maximizes the benefits for supermarkets.

Keywords: LSTM Model, Replenishment and Pricing Decision, SSA Sparrow Search Algorithm.

1. Introduction

For vegetable sales in supermarkets, due to the perishable nature of vegetables, the freshness of vegetables decreases when they are sold the next day. Therefore, every supermarket needs to replenish and price their vegetables. To formulate reasonable sales and pricing decisions, a regression analysis was conducted on the sales volume and prices of the data to determine the relationship between the two.Taking into consideration the uniqueness of vegetable products, pricing may not necessarily be related to the sales on the same day but could also be influenced by sales in the preceding days. Therefore, this problem should be treated as a multi-input multi-output issue. Predicting the optimal replenishment quantity for the upcoming week, which essentially forecasts the total daily sales for the upcoming week based on historical sales data, is the initial step.Subsequently, utilizing a combination of multivariate time series forecasting and Bayesian optimization with Long Short-Term Memory (LSTM) networks, the study derives both the total sales and pricing strategy for the upcoming week.

2. Establishment and Solution of Replenishment Models

Establishing Six Relationship Charts Using Historical Sales and Pricing Data for Various Vegetable Categories,As shown in Figure 1,When viewed individually, there isn't a strong relationship; there is a slight tendency for higher sales volume with higher prices[1].

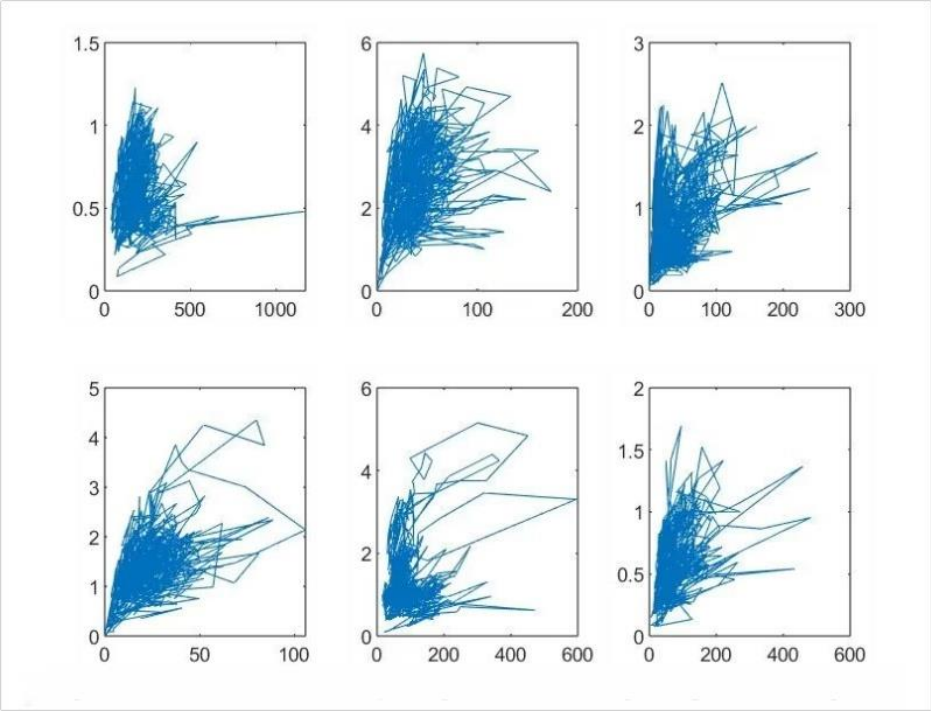


Figure 1. Historical sales and pricing data for six major vegetable categories

Due to the unique nature of vegetable products, pricing may not necessarily be related to the sales on the same day but could also be influenced by sales in the preceding days. **These warrants** conducting a regression analysis to study the relationship between the prices on a given day and the sales on that day, as well as the sales from the previous days, making it a multi-input multi-output problem.

In predicting the pricing and replenishment quantities for various vegetable categories for the upcoming 7 days, our model utilizes a combination of Empirical Mode Decomposition (EMD), multivariate time series forecasting (accounting for unknown future features), Long Short-Term Memory (LSTM)[2] networks, and Bayesian optimization. This combination offers multiple advantages, providing significant benefits across various aspects of the problem.

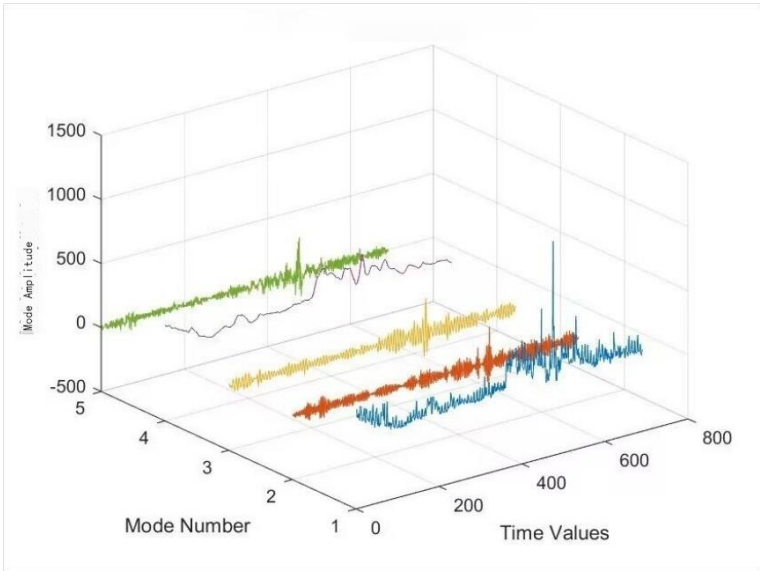


Figure 2. EMD Decomposition

To address the issue of sales forecasting, we first employed the **EMD [**2] decomposition method to simplify the problem. The EMD decomposition is illustrated in Figure 2, and its steps are as follows:

Initialization: We begin by selecting the original signal as the starting point for the first decomposition, defining it as the current Intrinsic Mode Function (IMF)[3].

Decomposition Process: For the current IMF, we iterate through the following steps:

a) Calculate the local maxima and minima of the data sequence, which represent the turning points of the signal.

b) Fit a spline curve to these turning points to obtain upper and lower envelopes.

c) Calculate the current IMF by averaging the upper and lower envelopes. If the IMF meets the termination criteria, the iteration stops. Otherwise, the current IMF becomes the new input signal for further decomposition.

Termination Criteria: The iteration is stopped based on different termination criteria. Common termination criteria include:

(1) Checking if the number of extrema points of the current IMF reaches a predefined threshold.

(2) Evaluating if the difference between the current IMF function and its upper and lower envelopes is sufficiently small.

Extracting the Last IMF: When the termination criteria are met, the remaining signal is considered the last IMF.

Next, we capture the long-term dependencies between replenishment quantity, pricing, and time for various product categories using a specialized recurrent neural network called Long Short-Term Memory (LSTM**) [**4]. LSTM effectively analyzes the input based on time series data, allowing us to efficiently convey and express information from long time sequences without overlooking valuable information from the distant past. The specific iterative calculation process is depicted in Figure 3 and Figure **4. This** significantly improves prediction accuracy compared to traditional linear models.

Furthermore, we utilize the SSA Sparrow Search **Algorithm [**5] to optimize the sales forecasts for the upcoming week.

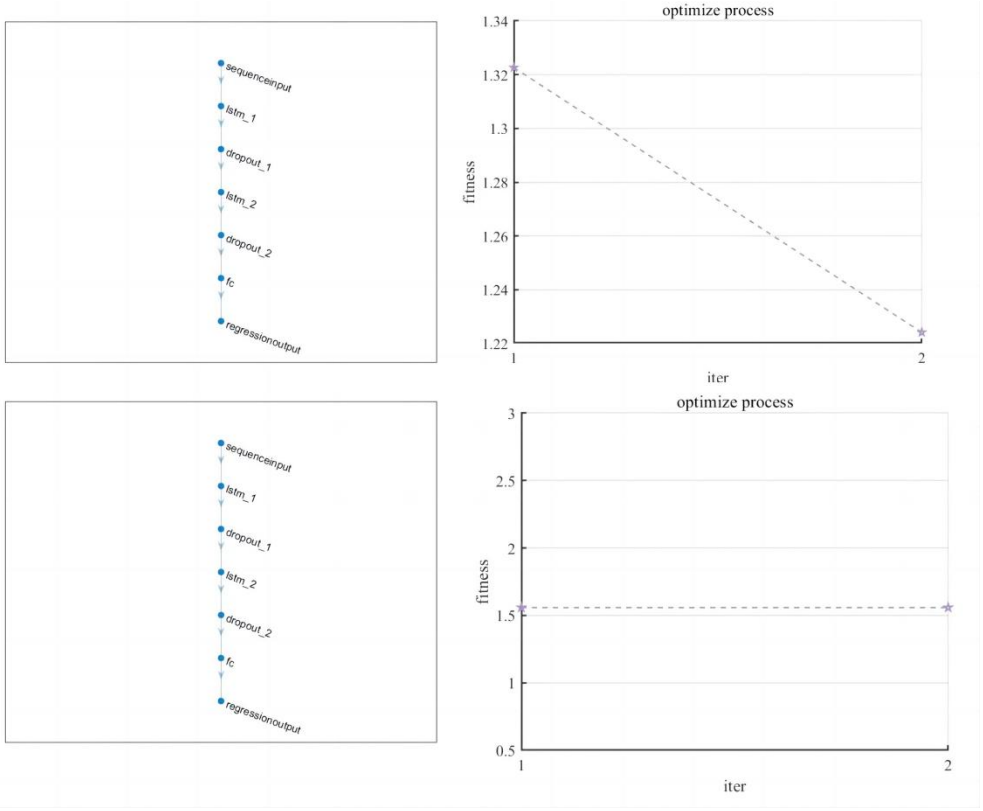


Figure 3. SSA Sparrow Search Algorithm Optimization Process

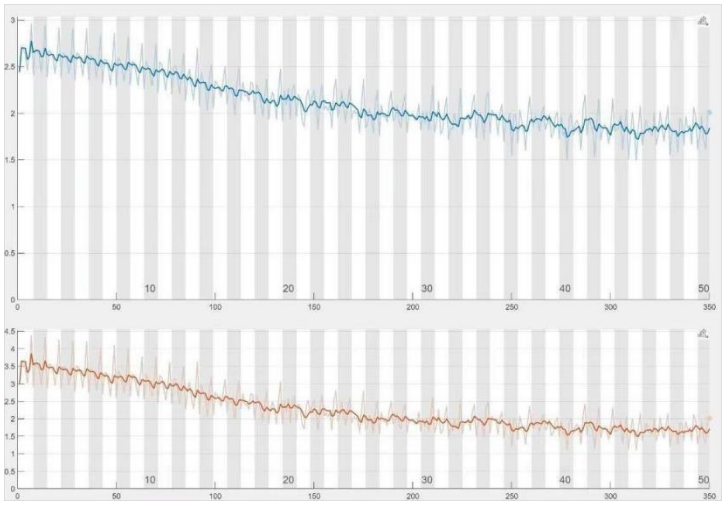


Figure 4. Training Set RMSE

The final results for the replenishment quantities are presented in Table 1 and Figure 5:

Table.1. Daily Replenishment Quantities for Each Product

Date	7.1	7.2	7.3	7.4	7.5	7.6	7.7
Leafy Vegetables	167.376	157.575	133.97	114.893	131.227	161.163	155.351
Cauliflower	24.5659	12.6243	17.6965	12.1343	16.0519	25.9752	22.1605
Aquatic Roots and Stems	18.8808	19.2994	21.137	14.9728	15.3791	16.0461	16.5513
Tomatoes	20.8162	17.4682	15.047	15.5667	14.065	15.1216	17.7004
Chili Peppers	85.4898	87.6675	77.9533	65.3657	75.1782	93.3268	96.4188
Edible Mushrooms	52.7508	53.2699	43.3262	37.6278	46.2431	54.9066	54.7252

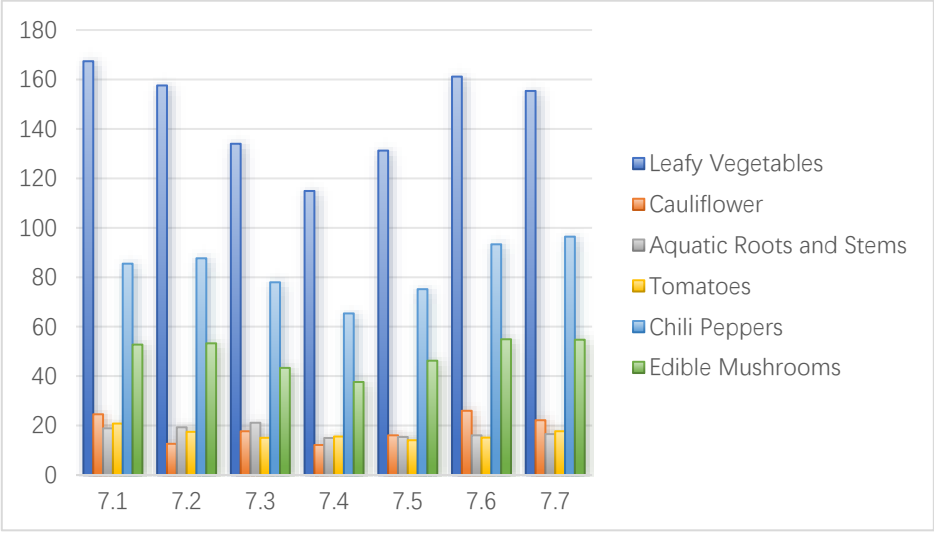


Figure 5. Daily Replenishment Quantities for Each Product

Subsequently, in order to predict pricing for the next seven days in a more scientific manner, considering multiple random variables, we accounted for potential interdependencies and interactions among multiple time series. Multivariate Time Series Forecasting[6] was employed to model complex dynamic behaviors among various system components or different feature variables. In the context of considering "unknown future features" – for example, when predicting for the seventh day, the information from the previous six days is considered "unknown future" – this model can only use current and past observed values to make predictions. This approach ensures the practicality and operability of the model[7].In addition to this, we continued to employ the LSTM algorithm to derive

pricing strategies for various vegetable categories for the upcoming week[8], and The results for the training set, validation set, and test set are displayed in Figure 6, Figure 7, and Figure 8, respectively。

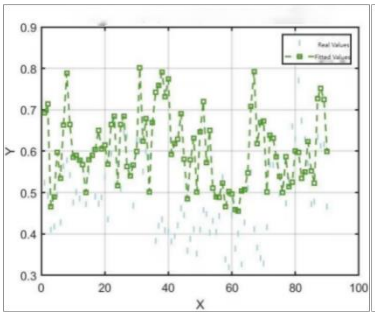


Figure 6. Validation Set Results

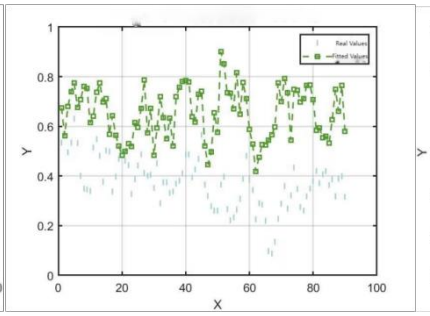


Figure 7. Test Set Results

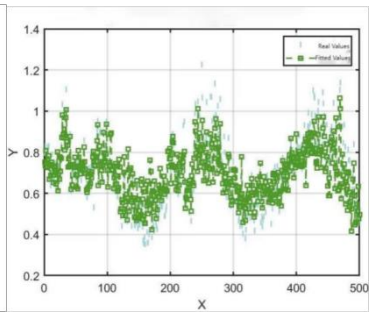


Figure 8. train Set Results

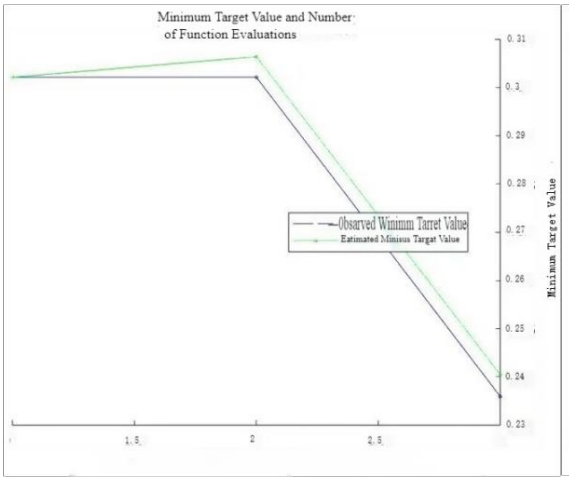


Figure 9. Function Evaluations (1)

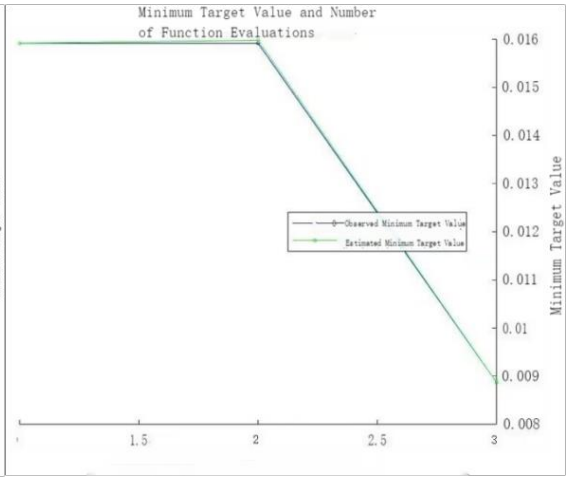


Figure 10. Function Evaluations (2)

I also employed Bayesian optimization [9], which allowed me to efficiently search for the optimal model parameters, reducing the time and computational resources required for model tuning [10].The optimization process is illustrated in Figure 9 and Figure 10.

The final pricing for each product category from July 1st to July 7th is presented in Table 2 and Figure 11:

Table.2. Daily Pricing for Each Product Categort

Date	7.1	7.2	7.3	7.4	7.5	7.6	7.7
Leafy Vegetables	4.9131	4.3312	4.3345	5.0872	4.8333	4.2874	4.6901
Cauliflower	12.4754	12.4245	12.3898	12.356	12.3481	12.3472	12.3598
Aquatic Roots and Stems	11.3852	11.6938	11.9582	12.2311	12.5006	12.7738	13.0397
Tomatoes	3.4598	3.4265	3.3679	3.3142	3.2703	3.2549	3.2234
Chili Peppers	20.5649	20.7345	20.9149	21.1301	21.3186	21.4179	21.561
Edible Mushrooms	52.7978	54.6339	55.4842	56.6173	57.3082	57.5904	57.7892

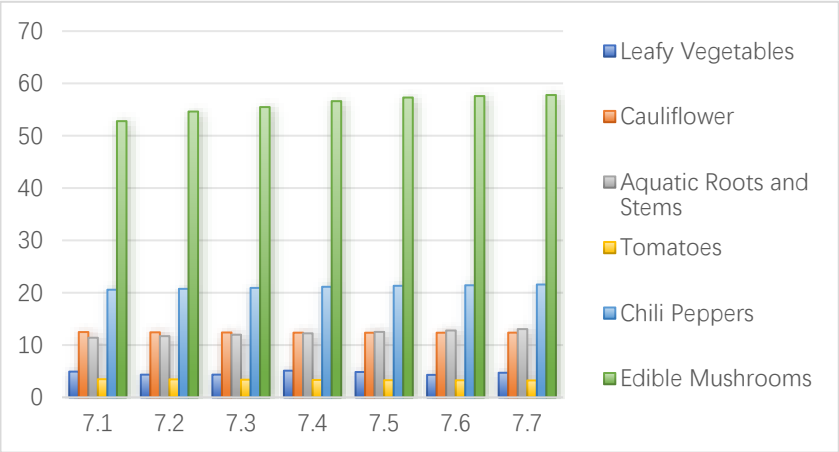


Figure 11. Daily Pricing for Each Product Category

3. Conclusions

In this paper, I conducted an in-depth study on the unique sales and pricing challenges of vegetable products in supermarkets. Due to the short shelf life of vegetables and their appearance deteriorating over time, traditional sales and pricing strategies often struggle to adapt to this highly dynamic market environment. To address this issue, I first analyzed the relationship between sales data and pricing, establishing the dynamic connection between the two. Based on this, I used EMD (Empirical Mode Decomposition) for data preprocessing and combined the LSTM (Long Short-Term Memory) model to predict the sales of various vegetable categories. Furthermore, to optimize the prediction results, I applied the SSA (Sparrow Search Algorithm) intelligent optimization algorithm.

In terms of multivariate time series forecasting, I also relied on the LSTM model and incorporated Bayesian optimization techniques, effectively predicting the pricing of different vegetable categories for the upcoming week by combining historical sales and pricing strategies. With these advanced analysis and prediction methods, I not only improved the accuracy of sales forecasting but also developed a scientific and reasonable pricing strategy for supermarkets to maximize their profitability.In summary, I proposed an innovative data analysis and forecasting framework in the highly volatile environment of vegetable sales. By integrating EMD preprocessing, the LSTM prediction model, and SSA and Bayesian optimization algorithms, I not only enhanced the accuracy of predictions but also provided effective sales and pricing strategies for supermarkets to meet rapidly changing market demands. This research not only helps supermarkets increase profitability but also serves as a valuable reference and inspiration for similar fields.

References

[1] Zhou, J., & Wang, S. (2021). A Carbon Price Prediction Model Based on the Secondary Decomposition Algorithm and Influencing Factors. Energies, 14(5), 1328.

[2] Liu, F., Qin, P., You, J., & Fu, Y. (2022). Sparrow Search Algorithm-Optimized Long Short-Term Memory Model for Stock Trend Prediction. Computational Intelligence and Neuroscience, 2022.

[3] Guo, F., Deng, S., Zheng, W., Wen, A., Du, J., Huang, G., & Wang, R. (2022). Short-Term Electricity Price Forecasting Based on the Two-Layer VMD Decomposition Technique and SSA-LSTM. Energies, 15(22), 8445.

[4] Zhang, C., Fu, S., Ou, B., Liu, Z., & Hu, M. (2022). Prediction of Dam Deformation Using SSA-LSTM Model Based on Empirical Mode Decomposition Method and Wavelet Threshold Noise Reduction. Water, 14(21), 3380.

[5] Guo, J., Liu, P., & An, Z. (2021). Research on Computer Prediction Model Using SSA-BP Neural Network and Sparrow Search Algorithm. Journal of Physics: Conference Series, 2033(1), 012003.

[6] Ouyang, C., Zhu, D., & Wang, F. (2021). A Learning Sparrow Search Algorithm. Computational Intelligence and Neuroscience, 2021.

[7] Zhang, Y., He, D., & Wu, Q. (2023). Forecasting of PM2.5 concentration time series based on SSA-LSTM model. [Details not available].

[8] Wang, Y., & Tuo, J. (2020). Blood Glucose Prediction Based on Empirical Mode Decomposition and SSA-KELM. 2020 Chinese Automation Congress (CAC), 4759-4763.

[9] Zhou, J., & Chen, D. (2021). Carbon Price Forecasting Based on Improved CEEMDAN and Extreme Learning Machine Optimized by Sparrow Search Algorithm. Sustainability, 13(9), 4896.

[10] Zhang, Y., Tao, P., Wu, X., Yang, C., Han, G., Zhou, H., & Hu, Y. (2022). Hourly Electricity Price Prediction for Electricity Market with High Proportion of Wind and Solar Power. Energies, 15(4), 1345.