A study of vegetable replenishment and pricing based on ARIMA modeling

Yixin Yang*, #, Kai Zhang#, Zehan Zhou

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, China, 233030

* Corresponding Author Email: yixincc@aliyun.com

#These authors contributed equally.

Abstract. With the increasing diversification of consumer demand, price fluctuation and supply of vegetables have become key factors affecting the operation of fresh food supermarkets, but the lack of effective formulation strategies for fresh food supermarkets has led to the loss of revenue and efficiency. In this paper, the ARIMA prediction model is constructed by using the optimization of grid search theory, and combined with a large number of actual sales data for verification. The prediction results of the model are compared and analyzed with the actual data, and it is found that the two show significant consistency, thus highlighting the excellent performance of the model in the prediction of vegetable replenishment volume and sales price. This study will contribute significantly to the economic effectiveness of fresh food retailing.

Keywords: vegetable restocking volume, grid search theory, ARIMA prediction model.

1. Introduction.

Vegetables form a fundamental element of the daily diet[1] , however, fluctuations in vegetable prices significantly affect consumers' purchasing decisions. When the price of a particular vegetable increases, consumers turn to other less expensive alternatives. Given that vegetables continue to be perishable from the moment they are picked [2], unsold vegetables not only lead to a waste of natural resources, but also cause economic losses to retailers[3]. Therefore, pricing strategies and inventory management of vegetables are particularly important in the operation of fresh produce supermarkets. The volatility of prices and the perishability of vegetables require retailers to make accurate market demand forecasts in order to minimize losses due to inaccurate demand forecasts.

This problem has attracted extensive attention from related scholars, and in related fields, scholars have used LSTM network model to analyze the time series of demand in fresh fish market [4], and compared its prediction results with empirical plain model, and found that the LSTM model shows significant advantages in several key performance indexes such as root-mean-square error, average absolute error, and average percentage error, but its complexity will lead to its limited ability to generalize under different data i.e. or market environment; in order to improve the accuracy of vegetable price prediction, Li et al. established TPCDR model based on network public opinion, which integrates the main economic factors [5] to identify the key factors affecting vegetable prices and provide theoretical support for predicting future price trends. An innovative STL-ELM hybrid model was also proposed by combining seasonal trend decomposition and extreme learning machine [6], which proved to be effective in predicting short- to long-term trends in seasonal vegetable prices.

In view of the current state of research, this study introduces the ARIMA model, which optimizes parameters through grid search to improve its accuracy in pricing and demand forecasting for vegetable items. The aim of this study is to provide an effective decision support tool for fresh produce supermarkets to help them set selling prices and inventory periods more accurately, thereby achieving the goal of profit maximization and promoting the economic efficiency of the agricultural industry.

2. ARIMA model construction

In this paper, time series data on the sales unit price, sales quantity, and wastage rate of hundreds of vegetable items in six categories, including cauliflower, chili, eggplant, edible fungi, and aquatic roots and tubers, were collected from 2020 to 2023, indicating a time dependence.The ARIMA [7] model, as a common method of time series analysis, has an autoregressive (AR) part that captures the time series data characteristics and the moving average (MA) part takes into account the random errors of the time series.

The ARIMA [8]model has three main parameters, which are the AR parameter p, the difference (I) parameter d, and the MA parameter q. p denotes the number of past observations considered in the model, d denotes the number of differences required to smooth the time series data, and q denotes the number of past prediction errors considered in the model.

The use of ARIMA model is effective in making predictions related to vegetable replenishment and pricing.

2.1. ADF test

Since the ARIMA model can only predict smooth data, so for the non-smooth data, it is necessary to differentiate them, so that the number of differences required to make the data smooth, that is, the value of the parameter d. The use of ADF test can detect whether the data has a unit root, so as to determine whether the data is smooth. Its specific formula as formula (1)

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \delta_2 \Delta Y_{t-2} + \ldots + \delta_p \Delta Y_{t-p} + \varepsilon_t \tag{1}$$

where ΔY_t is the first order difference of Y_t. The calculated test results are shown in Table 1.

Table 1. ADF test results

Vegetable category	*ADF statistic*	*P-value*
Cauliflower	-2.9977	0.0351
Foliage	-2.9682	0.0380
Pepper	-3.8879	0.0021
Aquatic roots and tubers	-3.5502	0.0068
mushrooms	-2.8785	0.0479
eggplant	-3.9007	0.0020

In Table 1, the p-value for each category of vegetables is less than 0.05, which rejects the original hypothesis, indicating that the data passes the smoothness test and the data is smooth with a parameter d of zero.

2.2. ACF function

The parameter q of the model can be determined using the ACF function. ACF indicates the correlation between the time series and its own past values, while q represents the effect of the past values of the time series itself on the current values, if ACF truncates, the lag number at the first truncation is q. If ACF drags the tail, it indicates that the model may be a MA process.

Taking leafy and flowering vegetables as an example, the q-values of the ACF functions calculated for the sales volume and sales price data are shown in Figures 2 and 3.

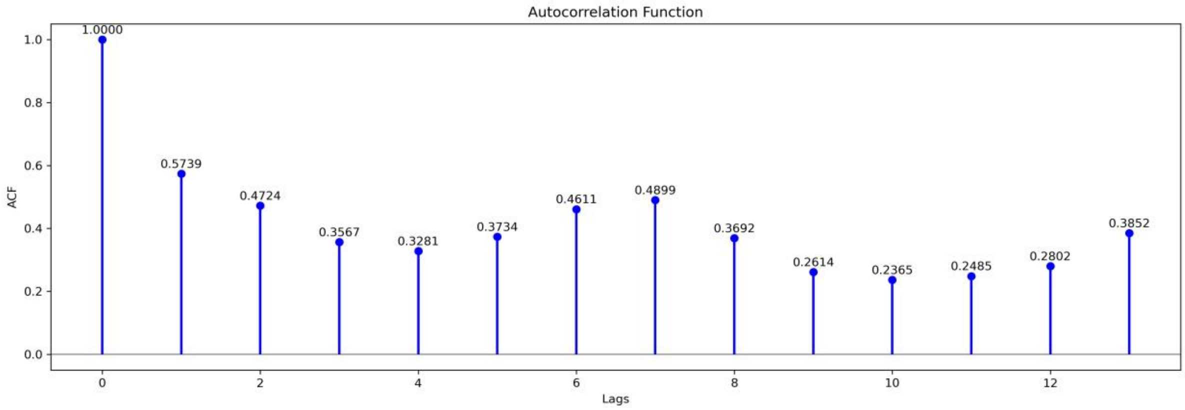


Figure 1. Sales parameters of foliage category

The data in Figure 1 illustrate that the data point at moment t has a correlation of 0.5739 with the data point at moment t-1, and a correlation of 0.4724 with the data point at moment t-2. Since the ACF shows a decreasing fluctuation trend, but does not quickly truncate the tail, and the value at Lag6 picks up again, 2 is conservatively chosen as the parameter of model q. Afterwards, adjustments will be made according to the test results of the model.

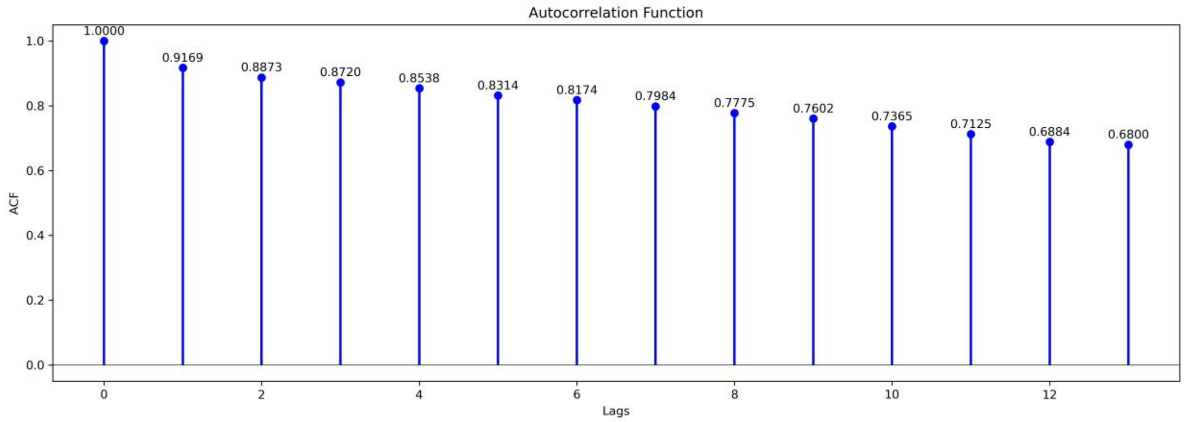


Figure 2. Parameters of sales price of foliage category

The ACF value in Figure 2 gradually decreases, showing an obvious trailing phenomenon, indicating that the model q value may be low, so the tentative q parameter value is 1.

2.3. PACF function

The parameter p of the model can be determined using the PACF function.PACF describes the direct relationship between the time series and its lagged values.If the PACF is truncated, then the current value is directly related to all previous values, and the lag corresponding to the time just truncated is the appropriate p value for the model.

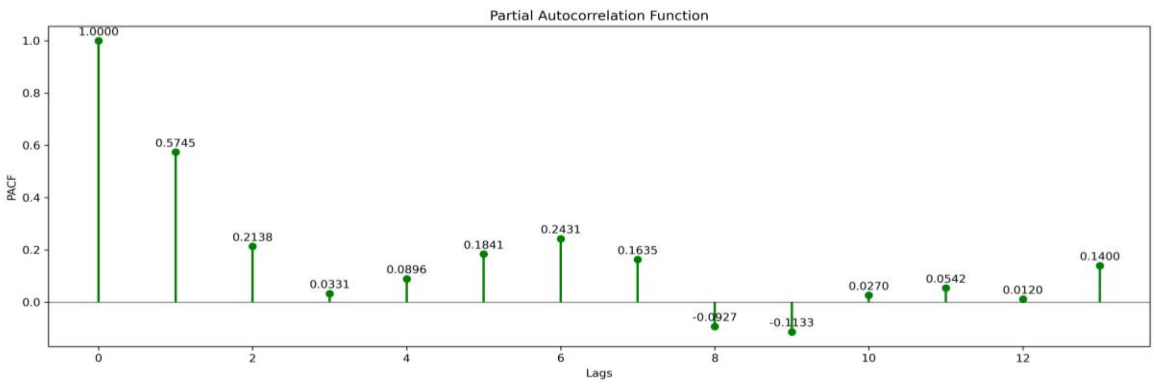


Figure 3. Parameters of sales of flower and leafy vegetables

The value of Lag1 in Figure 3 is 0.5745, while the value of Lag2 decreases to 0.2138, after which it gets closer and closer to 0, indicating that the value of parameter p in the model of sales of flower and leafy vegetables is 1.

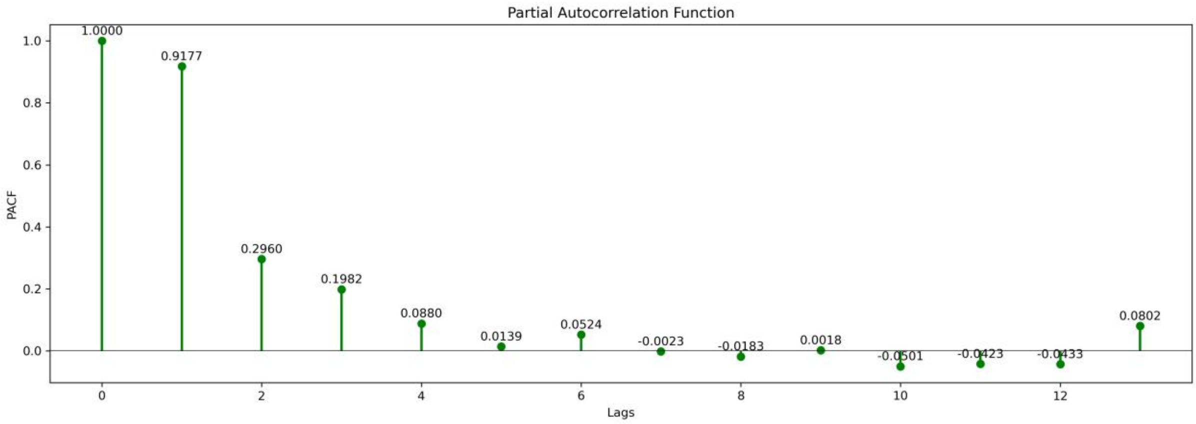


Figure 4. Parameters of sales price of foliage category

The value of lag1 of the PACF data shown in Figure 4 is 0.9177, which shows strong autoregressivity, and at lag2, it rapidly decreases to 0.2960 and gradually converges to 0 afterward, i.e., the PACF is truncated, and the model parameter p-value is 1.

3. Results and Discussion

3.1. Analysis of results

The values of total daily sales and sales price of each category of vegetables from July 1 to 7, 2023 predicted based on ARIMA model using Python code are shown in Figures 6 and 7.

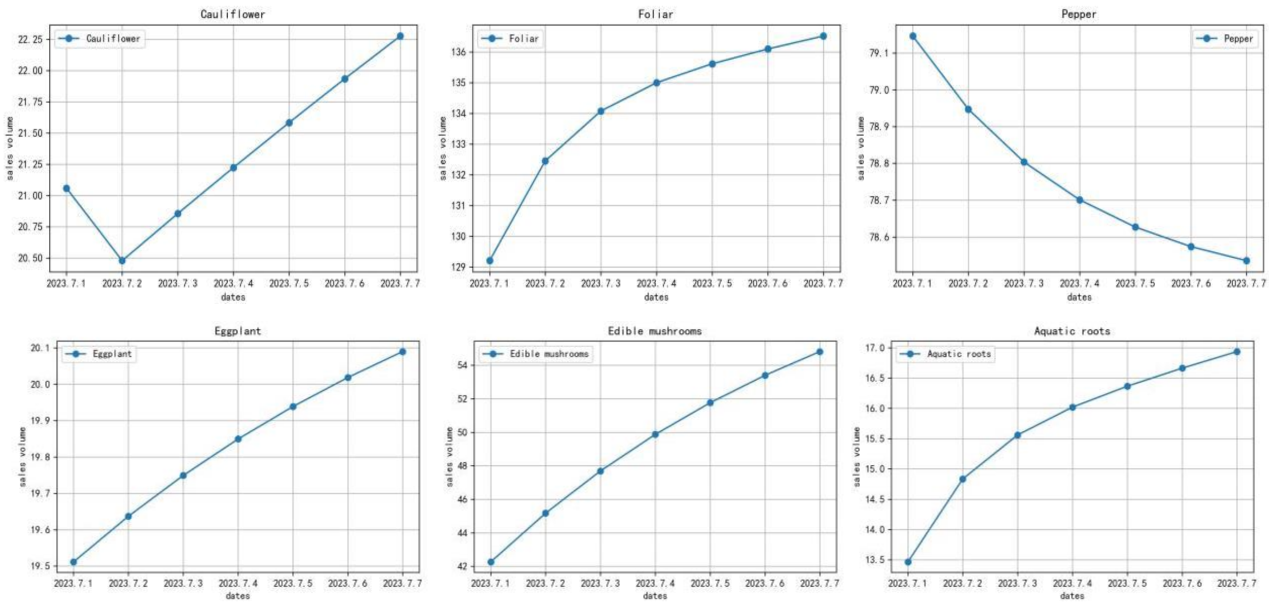


Figure 5. Total daily sales forecast table

Figure 5 shows the results of the seven-day sales forecast for each category of vegetables. It can be seen that the total sales volume of the four categories of vegetables, namely, foliage, eggplant, edible mushrooms and aquatic root vegetables, will increase with the increase of time, while chili peppers, on the contrary, show a decreasing trend in their sales volume with the time, and cauliflower vegetables' sales volume reached a minimum of 20.48 on July 2, and then continued to rise.

Comparing these six types of vegetables, cauliflower vegetables have the highest sales, chili peppers rank second, and aquatic root vegetables are the least popular among consumers.

Since commodities have a wastage rate, the prediction of the total daily replenishment can be derived from the daily sales volume / (1 - wastage rate).

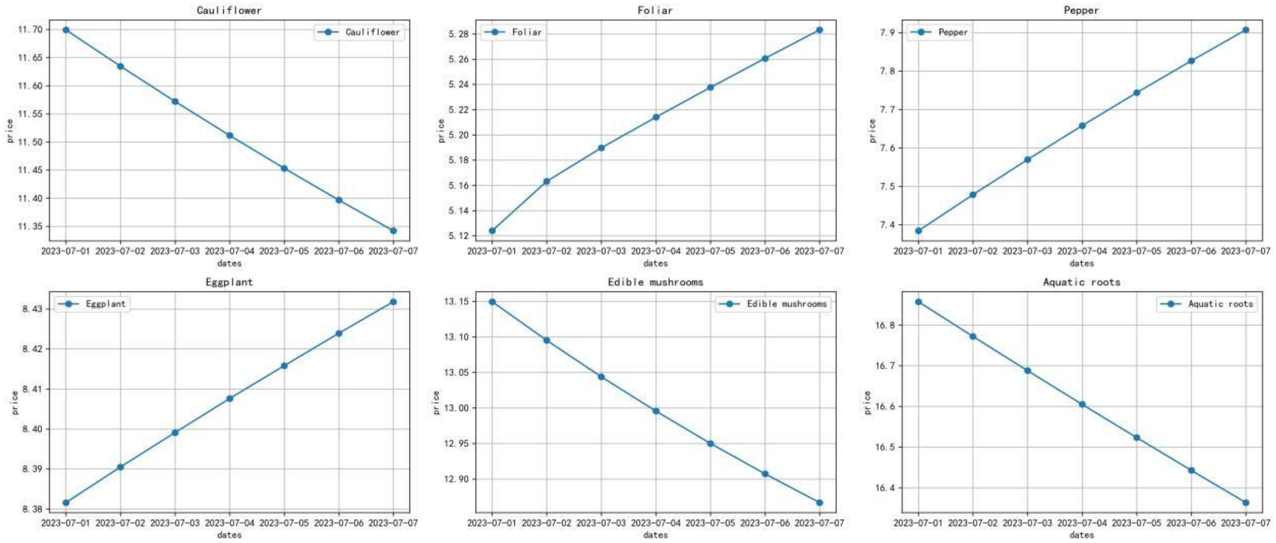


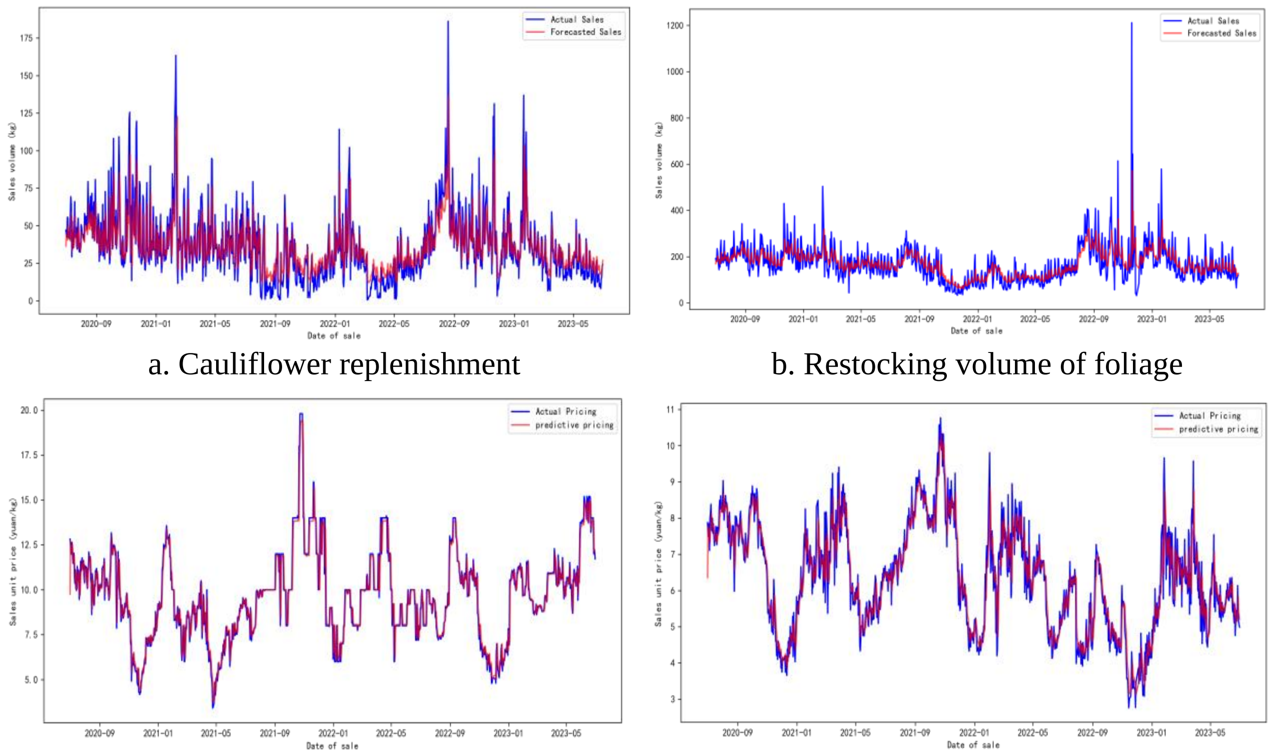
Figure 6. Sales Price Forecast Table

Figure 6 shows the results of the sales price forecast for each category of vegetables in the cauliflower and aquatic roots, edible fungi vegetables predicted a week of sales prices gradually decreasing, in the case of cauliflower, for example, its price fell from 11.699 to 11.3422, foliage and eggplant, chili pepper vegetables prices are the opposite of the price, such as foliage price rose from 5.124 on the 1st to 5.2832 on the 7th. comparing the six types of vegetables. The average price of aquatic root vegetables is the highest, over 16 yuan, and the price of foliage vegetables is the lowest, only about 5 yuan.

Combining the predicted sales and average pricing of each category of vegetables, it can be found that the flower and leafy vegetables may be due to its low price is loved by consumers, even if its price has a slight increase, does not affect the purchase of vegetables; chili peppers, edible mushrooms, aquatic roots and tubers of three kinds of vegetables, the change in the sales of sales situation are inversely proportional to the pricing, implying that the sales of their sales are more sensitive to the changes in price.

3.2. Model Test

Comparing the gap between the model prediction results and the real value, the model can be tested to verify the reliability of its prediction results[10]. Figure 7 shows the comparison between the predicted and true values of some categories of vegetables.



a. Cauliflower replenishment b. Restocking volume of foliage

c. Cauliflower pricing d. Foliar pricing

Figure 7. Comparison of actual value and predicted value

From the overlap of the two folds in Figure 8, it can be seen that both the forecast of the replenishment quantity and the forecast of the sales price, the predicted value of the model and the real value results are highly similar, indicating that the predicted value of the model is close to the real value, the prediction results are reliable, and the model test passes.

4. Conclusion

Vegetables as a necessity for people's healthy life, its sales situation can reflect the residents' consumption tendency and life situation, reasonable prediction of the replenishment and pricing information of each category of vegetables, not only can provide information support for the sales and purchase of supermarkets, but also can reflect the side of the consumer's demand for different kinds of vegetables and feelings.

This study focuses on the dynamic analysis of the vegetable market, employing the ARIMA model to predict and analyze the replenishment quantities and sales prices of vegetables. The research findings reveal that the ARIMA model exhibits significant predictive capabilities, with its forecast results showing a high degree of consistency with actual market data. The application of this model not only aids fresh food supermarkets in achieving cost control and profit maximization but also enables a more precise capture of consumer demand fluctuations for different vegetable varieties, thereby enhancing the market competitiveness of the fresh retail industry.

References

[1] Bo L, Junqi D, Zhengqing Y, et al. Optimized neural network combined model based on the induced ordered weighted averaging operator for vegetable price forecasting [J]. Expert Systems with Applications, 2020,114232.

[2] Huber J, Gossmann A, Stuckenschmidt H. Cluster-based hierarchical demand forecasting for perishable goods [J]. Expert Systems with Applications, 2017, 76 140-151.

[3] Chetna C ,Amandeep D ,Ul M A , et al. Food loss and waste in food supply chains. A systematic literature review and framework development approach [J]. Journal of Cleaner Production, 2021, 295.

[4] Lucia V M ,André P ,João P , et al. Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning [J]. Journal of Cleaner Production, 2022, 359.

[5] Li Y ,Yao J ,Song J , et al. Investigation of causal public opinion indexes for price fluctuation in vegetable marketing [J]. Computers and Electrical Engineering, 2024, 116 109227.

[6] Xiong T ,Li C ,Bao 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.

[7] Yunxia C ,Xiaopeng L ,Chunmei J , et al.Influencing Factors and Prediction of Risk of Returning to Ecological Poverty in Liupan Mountain Region, China[J].Chinese Geographical Science,2024,34(03):420-435.

[8] Renata Rojas Guerra;Anna Vizziello;Pietro Savazzi;Emanuele Goldoni;Paolo Gamba . Forecasting LoRaWAN RSSI using weather parameters: A comparative study of ARIMA, artificial intelligence and hybrid approaches[J]. Computer Networks,2024,243:110258.

[9] Ashwini Darekar;A Amarender Reddy.Price forecasting of pulses: case of pigeonpea.Journal of Food Legumes[J],2017,30:212-216.

[10] Amir Reza R Niknam;Maryam Sabaghzadeh;Ali Barzkar;Davood Shishebori.Comparing ARIMA and various deep learning models for long-term water quality index forecasting in Dez River, Iran. Environmental science and pollution research international[J],2024.