

A study on vegetable replenishment and pricing based on binomial function and deep learning models

Lintao Tang^{#,*}, Huifei Chen[#]

College of information and intelligence, Hunan Agricultural University, Changsha, China, 410128

*Corresponding author: 17580310662@163.com

[#]These authors contributed equally

Abstract. Perishable vegetable commodities for supermarkets, the development of its sales strategy is critical. The purpose of this paper is to develop an integrated model to optimize the daily replenishment and pricing strategy of vegetables, so as to improve the operational efficiency and profitability of supermarkets. By analyzing historical sales data, this paper reveals the negative correlation between vegetable sales volume and cost-plus price, and uses various mathematical models to fit the analysis. It is found that the binomial function model with a weekly period can most accurately describe the relationship between sales volume and price. This paper further utilizes the LSTM long- and short- term memory model for the sliding prediction of replenishment volume and combines it with the ARIMA model to predict wholesale prices. Based on the prediction results, this paper establishes an optimization model with profit maximization as the objective and performs parameter optimization by simulated annealing algorithm. The results of the study show that the maximum profit of ¥9,708.282 can be achieved by using the strategy proposed in this paper in the study of the situation during the week of July 1 to 7, 2023 This study provides a scientific decision-making tool for superstores, which helps to maximize profits while ensuring the freshness of goods.

Keywords: Binomial function, Vegetable replenishment, LSTM, ARIMA.

1. Introduction

Vegetable commodities as a class of perishable products, its sales situation is inevitably affected by its freshness and quality, and most of the vegetable varieties in the super-sales have the characteristics of daily replenishment of daily sales. As a result, without exactly knowing the specific vegetable product and purchase prices, supermarkets usually make replenishment and pricing strategies for each vegetable category based on the historical sales and demand of each commodity. Zhao Ling[1] considered the impact of customer returns and fixed costs on replenishment pricing strategies in her research, and pointed out that in the face of the short-term shelf life challenges of fresh goods, especially vegetable products, accurate prediction of market demand, reasonable price setting and efficient replenishment are the keys to ensuring the profitability of supermarkets. From the perspective of demand and supply and demand, the sales volume of vegetable commodities is related to time, and its supply and demand are affected by seasonal factors. Therefore, it is particularly important to consider factors such as commodity purchase price, historical sales volume, commodity loss rate, and combination purchase and sales among commodities to make replenishment decisions and pricing decisions for vegetable commodities.

In recent years, research on pricing and replenishment strategies for perishable fresh products, especially vegetable commodities, has received much attention. Jinlong Zhang [2] and others constructed a joint pricing and replenishment decision model containing product diffusion, repeat purchase behavior and product deterioration characteristics, emphasized the importance of dynamic batch replenishment, and described the demand evolution process by introducing the repeat purchase rate into the Bass model containing the price effect. In his master's thesis, Xin Chanyun [3] focused on the application of multi-stage pricing and replenishment strategies in perishable fresh products (e.g., vegetables), with particular attention to how the pricing and replenishment of products at different stages can be adapted to changes in market demand. The study analyzes the correlation of sales volume among vegetable categories as well as among individual products through Pearson's

correlation coefficients, and applies ARIMA model to forecast future sales volume. Further with the help of a combination of nonlinear programming and greedy algorithm, the optimal replenishment volume and pricing strategy under the objectives of satisfying market demand and profit maximization were determined. Nie Yuxuan [4] selected different mathematical function fitting according to the sales volume and pricing characteristics of different vegetable categories, and combined with the actual sales space constraints and the minimum display volume requirement of a single product, the greedy algorithm was used to determine the most profitable replenishment and pricing scheme of a single product in the limited display space. Qiao Xue [5] discussed in detail the joint replenishment pricing strategy of fresh products under the scenario of sales loss, emphasized the importance of considering the decay of quantity and quality of goods to the decision-making of retailers, and successfully improved the ability to cope with uncertain demand and maximize profits by using robust replenishment strategy and inventory probability distribution strategy.

In order to formulate the optimal daily replenishment quantity and commodity price, this paper first observes the general trend and correlation by plotting the time series of sales volume and cost-plus price. After that, the specific functional relationship is further explored by fitting the data using statistical methods and machine learning algorithms, and selecting a suitable mathematical model to describe the relationship between the two. Finally, long-term prediction models are used to predict the sales volume and price of each vegetable category, such as LSTM (Long Short-Term Memory Network) and ARIMA (Autoregressive Sliding Average Model). Based on the prediction results, an objective optimization model is built to determine the best pricing strategy, using optimization algorithms such as genetic algorithms or simulated annealing algorithms, which iteratively search the parameter space in order to achieve the goal of maximizing profits or maximizing benefits.

2. Modeling and Solving for Vegetable Replenishment and Pricing

2.1. Relationship between vegetable category sales and cost-plus pricing

In economic theory, there is a negative correlation between the sales volume of a certain commodity and its own price, i.e., an increase in the price of a commodity leads to a decrease in the sales volume of the commodity, and a decrease in the price leads to an increase in the sales volume of the commodity. The traditional form of the sales volume function is divided into two categories: one is the linear sales volume function and the other is the inverse proportional sales volume function. In the study, vegetables as perishable goods, their price and sales volume are prone to fluctuation over time, and it may not be desirable to use the traditional price-sales function to assess the relationship between the price and sales volume of vegetable commodities, so we analyze the existing dataset to explore the price-sales model of vegetable commodities.

Because vegetables are usually sold using cost-plus pricing, we assume that the historical unit prices of the vegetable items given in the dataset are cost-plus pricing. First, we find the monthly sales volume and average unit price for each category of vegetable price and sales data for each day from July 1, 2020 to June 30, 2023, whose monthly sales volume and average unit price are summarized in Table 1

Table.1. Summary of Monthly Sales and Average Unit Price for Vegetable Category

Timing	Philodendron		Cauliflower (Brassica oleracea var. botrytis)	
	Monthly sales (kg)	Average monthly unit price	Monthly sales (kg)	Average monthly unit price
June 2020	6577.08	7.65	2198.05	12.28
July 2020	7261.52	7.80	2822.34	9.25
August 2020	5668.90	7.59	1992.41	10.18
September 2020	6395.31	7.78	1908.94	11.79
...	...	...	...	...
March 2023	5476.18	6.48	3953.09	8.73
April 2023	5140.51	5.39	3240.12	7.58
May 2023	5225.56	5.74	2783.96	8.32
June 2023	4660.20	5.46	2714.93	7.07

Note: Data source: http://www.mcm.edu.cn/html_cn/node/c74d72127066f510a5723a94b5323a26.html

According to the above data table to make the average monthly sales and average unit price of flowers and foliage category of the trend is shown in Figure 1:



Figure 1. Trend of average monthly sales and average unit price of foliage category

In terms of overall changes, there is a certain dependence between the average monthly price and the average monthly sales volume, and it is manifested in the fact that when the average price is higher, the monthly sales volume is lower, and when the average price is lower, the monthly sales volume is higher.

To further analyze the statistical relationship between the average price and sales volume of each category of vegetables between July 2020 and July 2023, we described the correlation between the two variables by fitting a univariate linear regression equation on the average price and sales volume of each category of vegetables on a weekly, monthly, and quarterly cycle, respectively, in order to provide a basis for forecasting the price and sales volume of each category of vegetables.

Let the monthly (weekly and quarterly) sales volume of a vegetable category be S_i , $i=1,2,3...6$ denoting leafy flowers, cauliflower, chili peppers, eggplant, edible mushrooms, and aquatic roots and tubers, respectively, and the average price of the monthly (weekly and quarterly) price be P_i , and in order to determine the optimal regression model, the establishment of the corresponding evaluation model expression is considered, and the commonly used evaluation of the goodness-of-fit is adopted, with the evaluation criterion of the greater the R^2 , then the better the fit of the fitted model.

The following 3 function fits were established:

(1) Linear relationship

$$S_i = \alpha_i + \beta_i p_i, i = 1, 2 \dots 6 \quad (1)$$

(2) Exponential relationship

$$S_i = \alpha_i e^{\beta_i p_i}, i = 1, 2 \dots 6 \quad (2)$$

(3) Binomial relations

$$S_i = \alpha_i + \beta_{i1} p_i + \beta_{i2} p_i^2, i = 1, 2 \dots 6 \quad (3)$$

A comparative analysis of the goodness of fit of the fitted models reveals that the binomial model has the highest goodness of fit, followed by the linear model and the exponential model. We selected the binomial model to fit to get six regression functions for each of monthly, weekly and quarterly, and by comparing their goodness of fit R^2 , we can find that the regression effect of weekly sales and weekly average price of vegetable categories is the best, and the regression functions of sales and price of six vegetable categories are shown in Table 2:

Table.2. Binomial fit function of weekly sales and weekly average price for six vegetable categories

Binomial fitting function	R ²	P-value
$s_1 = 63707.39 - 14092.42p_1 + 1020.26p_1^2$	0.892	0.07***
$s_2 = 467 + 960.31p_2 - 66.68p_2^2$	0.768	0.00***
$s_3 = 23049.45 - 2718.33p_3 + 106.55p_3^2$	0.847	0.002***
$s_4 = 4897.34 - 560.96p_4 + 23.5p_4^2$	0.911	0.00***
$s_5 = 36075.6 - 3848.2p_5 + 113.15p_5^2$	0.868	0.005***
$s_6 = -3900.9 + 0.22p_6 - 1.22p_6^2$	0.790	0.0038

Note: ***, **, * represent 1%, 5% and 10% significance levels, respectively.

2.2. Predicting Vegetable Category Sales Based on LSTM Modeling

For vegetables such as perishable goods, their future sales volume is usually closely related to the historical sales volume, merchants in the case of not knowing exactly the specific single product and the price of goods, usually based on the historical sales volume of the same period in the past for replenishment. So to predict the total daily replenishment of each vegetable category in the coming week (July 1-7, 2023), we consider a weekly cycle, based on the sales volume in the same period in the past to predict the future sales volume, for example, using the sales volume of June 24, 2023 to predict the sales volume of July 1, 2023, using June 25, 2023 to predict the sales volume of July 3, 2023, and so on, to carry out cycle sliding. And so on for period sliding prediction. In this regard, we selected the LSTM model [6-7] to predict the sales volume of each category of vegetables for the coming week.

Based on the principle of LSTM model, using the sales data of each category of vegetables, the predicted results of the sales volume of six vegetable categories in the coming week (July 1-7, 2023) were obtained using LSTM model prediction are shown in Table 3:

Table.3. Sales Volume Forecasts for 6 Vegetable Categories for July 1-7, 2023

Timing	Philodendron	Cauliflower (Brassica oleracea var. botrytis)	Capsicum	Eggplant	Edible mushroom	Aquatic rhizomes
July 1,2023	150.640	24.116	82.879	23.245	48.566	20.691
July 2,2023	133.368	22.044	83.867	21.378	52.083	19.847
July 3,2023	112.243	21.720	84.235	20.920	53.517	20.811
July 4,2023	102.698	21.995	84.289	20.996	54.158	20.574
July 5,2023	110.425	22.349	84.269	21.090	54.497	21.047
July 6,2023	127.892	22.670	84.250	21.130	54.720	21.074
July 7,2023	140.927	22.962	84.240	21.150	54.899	21.369

2.3. Predicting Costs of Vegetable Categories Based on ARIMA Models

Through the wholesale prices of different vegetable individual items given in the dataset, we can find that the wholesale prices are different every day and have some volatility in time, i.e., they have time series characteristics. So we consider forecasting the wholesale price of vegetables in the coming week by ARIMA model [8]. Therefore, we establish the ARIMA prediction model based on the original data of six vegetable categories, and the sequence plots drawn from the data of flowering and leafy vegetables (see Fig. 2) show that the sequence is more volatile, and then we analyze the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the sequences, which are trailing, indicating that the sequences are non-stationary.

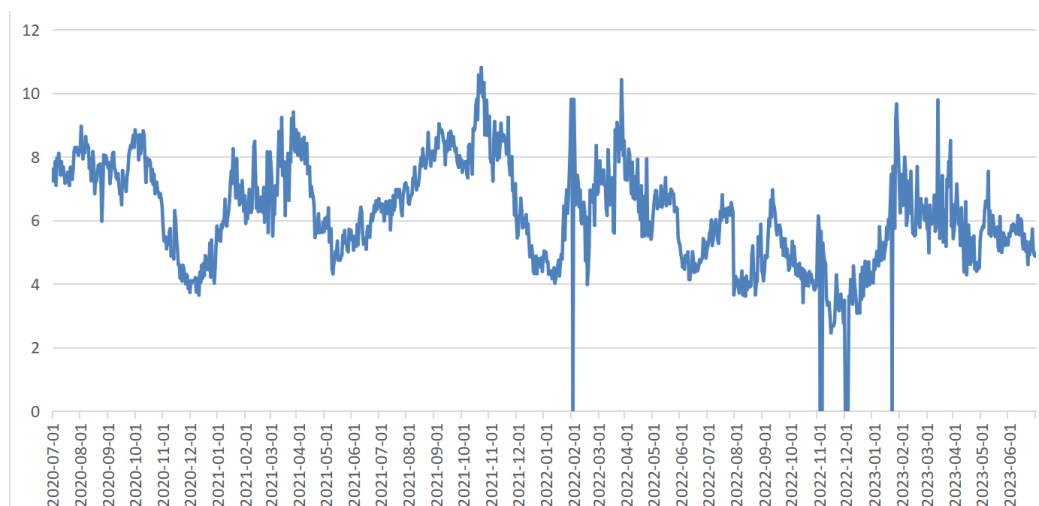


Figure 2. Time series of average daily wholesale prices of flowering and leafy vegetables

The sequence is then differenced, and theoretically, a sufficient number of differencing operations can fully extract the non-smooth deterministic information in the original time series, but it is not the case that the more orders of differencing operations, the better. Differencing is the process of extracting and processing information, and there is a loss of information in each differencing, so the order of differencing needs to be appropriate to avoid over-differencing. Through multiple difference operations on the sequence, the best difference of the eggplant vegetable data was obtained as the first order, so the ARIMA (p,1,q) model was established, and after repeated tests on the model, the model was determined to be ARIMA (3,1,1):

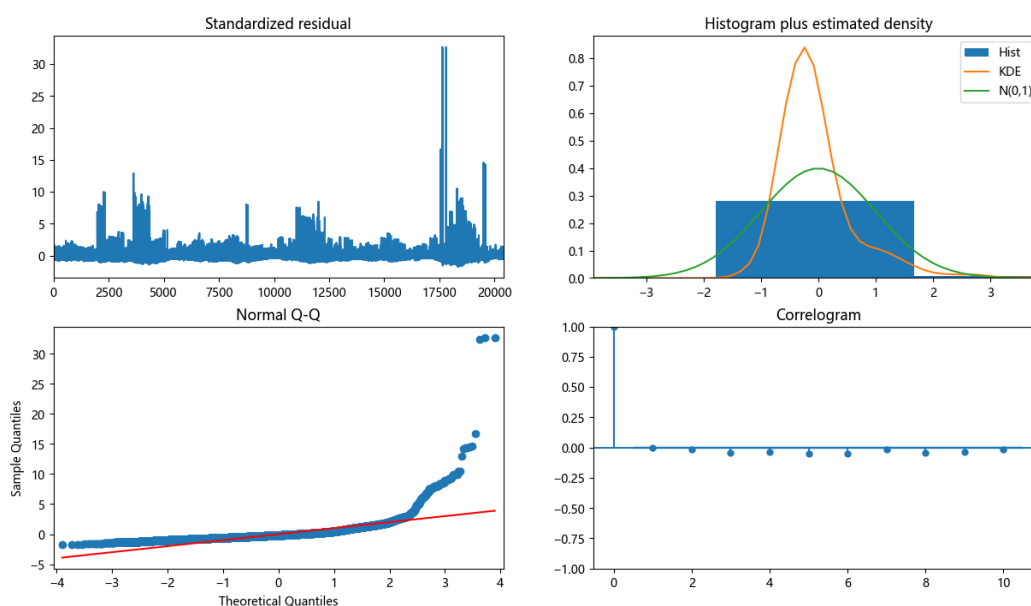


Figure 3. Test of ARIMA prediction model for eggplant vegetables

The ACF and PACF tests of the residuals revealed that the p-values were greater than 0.05, indicating that the residuals were partially uncorrelated and the model fit was good, so the ARIMA (3,1,1) model could be used to predict the future shipments of the daily average wholesale price of the route eggplant vegetables.

Applying a similar procedure, the prediction model can be obtained as ARIMA (1,0,1) for leafy vegetables, ARIMA (3,0,3) for cauliflower vegetables, ARIMA (3,0,3) for chili vegetables, ARIMA (3,0,3) for edible vegetables, ARIMA (3,0,1) for aquatic root vegetables, and ARIMA (3, 0,3). The model parameters were obtained by fitting the data based on the results of the ARIMA prediction models for the six vegetable categories, and all passed the significance test, resulting in the average wholesale prices for the six vegetable categories from July 1-7, 2023, as shown in the Table 4:

Table.4. Projected Average Wholesale Prices for 6 Vegetable Categories, July 1-7, 2023

Timing	Philodendron	Cauliflower (Brassica oleracea var. botrytis)	Capsicum	Eggplant	Edible mushroom	Aquatic rhizomes
July 1,2023	3.762	7.705	6.057	23.245	5.323	5.554
July 2,2023	3.762	7.799	6.040	21.378	5.074	9.397
July 3,2023	3.762	7.792	6.201	20.920	4.633	12.947
July 4,2023	3.762	7.701	5.943	20.996	4.690	13.246
July 5,2023	3.762	7.732	6.259	21.090	4.746	11.814
July 6,2023	3.761	7.635	5.915	21.130	4.735	10.821
July 7,2023	3.761	7.667	6.262	21.150	4.728	10.842

2.4. Maximum benefit pricing model

Cost-plus pricing method [8] is a method of setting the price of a commodity based on the unit cost of the commodity plus a certain percentage of the profit set by the profit. In most commodity pricing conversions, the expression is as follows:

$$P = C(1 + r) \quad (4)$$

where P denotes the unit selling price of a good, C denotes the total cost of the good, and r denotes the target profit margin. For the target utilization rate, it can be obtained from the following expression:

$$\gamma = \frac{P - c}{p} \times 100\% \quad (5)$$

Vegetables pass through four distribution links from the source to the consumers - procurement link, transportation link, wholesale link and retail link, while the short life cycle and perishable character of vegetables cause value loss throughout the distribution link, which leads to an increase in cost. So the cost of vegetable wholesale and sales of superstores includes the following aspects: purchasing cost, which refers to the cost that superstores need to pay for purchasing vegetables from suppliers, which is determined by the wholesale price given in the dataset; transportation cost, which means that superstores need to transport the purchased vegetables to the point of sale, and the transportation cost includes the cost of the transportation vehicle, the cost of fuel, as well as other logistic costs that may be incurred, and so on. In the sales process, with the changes in sales time, its quality is bound to deteriorate, which leads to the loss of vegetables, that is, the actual sales volume of goods is not equal to the amount of goods. At the same time, with the deterioration of the quality, consumers will reduce the desire to buy the goods, for the deterioration of the quality and the day for the sale of vegetables, supermarkets will usually take the discount promotion. In addition, supermarkets incur costs such as labor when operating.

To sum up, the costs of wholesale and sales of vegetables in supermarkets include various aspects such as purchasing costs, transportation costs, wholesale costs, wastage costs, discount promotion costs and labor costs. The increase of these costs will lead to an increase in the total cost of vegetables, and ultimately reflected in the price of goods. As a result, we establish the following vegetable sellingprofit function:

$$W = \sum S_{ij} (1 - \gamma_i) (p_{ij} - C_{ij} - C_0) \quad (6)$$

Where W denotes the total profit earned by the superstore vegetables from July1-7, 2023; S_{ij} denotes the predicted sales volume of the vegetable category on day; γ_i denotes the average wastage rate of the vegetable category, and we average the wastage rate of a single product based on the category in Data set as the wastage rate of the vegetable category, and assume the wastage rate of thevegetable category to remain unchanged. P_{ij} denotes the pricing of the vegetable category on day C_{ij} denotes the P denotes the pricing of vegetable category0 on day ; C denotes the average wholesale price of vegetable category on day ; and C_0 denotes the fixed cost.

In order to find the maximum profit from July 1-7, 2023, we take the vegetable category selling profit function as the optimization objective function. And for the decision variable P_{ij} , we use the minimum selling price and the maximum selling price of each vegetable category three-year category as its constraints, and finally get the following pricing optimization model:

$$\begin{aligned} \max W = & \sum_{i=1}^6 \sum_{j=1}^7 S_{ij} (1 - \gamma_i) (p_{ij} - C_{ij} - C_0) \\ \text{s.t.} \left\{ \begin{array}{l} 0.1 \leq p_{2j} \leq 85.6 \\ 0.9 \leq p_{3j} \leq 79.8 \\ 1.5 \leq p_{4j} \leq 21.6 \\ 0.5 \leq p_{5j} \leq 116 \\ 1.9 \leq p_{6j} \leq 53.8 \end{array} \right. \end{aligned} \quad (7)$$

2.5. Introduction of heuristic algorithm to simulate annealing algorithm for parameter optimization

(1) Initial solution determination

When applying the simulated annealing algorithm [9-10], the initial solution of the algorithm X_0 is set to any value between July 1, 2020 and June 30, 2023 when the sales price of vegetables in each category is at its maximum value

(2) Probability of acceptance of a new solution

If the solution of the objective function obtained during one iteration is smaller than the current global optimal solution [10], then the probability of accepting the new solution needs to be calculated, using the Metropolis acceptance criterion, whose probability is calculated as follows:

$$df = MIN_{current} - MIN_{best} \quad (8)$$

$$p_{accept} = e^{-\frac{df}{T}} \quad (9)$$

(3) Setting of important parameters

Setting of important parameters: initial temperature $T_0 = 1000$, minimum temperature $T_m = 800$, maximum number of iterations $T_k = 1000$, temperature decay coefficient $a = 0.98$, number of iterations at each temperature $N_k = 980$.

The maximum profit for each vegetable category on July 17, 2023 can be obtained as \$9,708.282, and its corresponding daily pricing strategy is shown in Table 5:

Table.5. Pricing by Vegetable Category, July 1-7, 2023

Timing	Philodendron	Cauliflower (Brassica oleracea var. botrytis)	Capsicum	Eggplant	Edible mushroom	Aquatic rhizomes
July 1,2023	6.60	9.15	11.51	10.20	12.53	12.00
July 2,2023	6.60	9.15	11.50	10.20	12.50	12.00
July 3,2023	6.61	9.15	11.51	10.20	12.51	12.00
July 4,2023	6.62	9.15	11.51	10.20	12.51	12.01
July 5,2023	6.61	9.15	11.51	10.20	12.52	12.01
July 6,2023	6.61	9.15	11.51	10.21	12.51	12.00
July 7,2023	6.60	9.15	11.51	10.20	12.53	12.00

3. Conclusions

For the problem of pricing replenishment of perishable vegetables, this paper constructs a comprehensive model, and through in-depth analysis of historical sales data, it is found that the binomial function model with a weekly period can more accurately simulate the dynamic relationship

between the sales volume of each category of vegetables and the cost-plus pricing, and binomial function equations for the sales volume of each category of vegetables and the cost-plus price are obtained. Then an LSTM long and short-term memory model was established to slidingly predict the restocking volume of each category of vegetables. Meanwhile, the ARIMA autoregressive sliding average model was used to predict wholesale prices. Based on the prediction results and the average loss rate of each category of vegetables, an optimization model with the objective of maximizing profit was established, and finally, the maximum profit was found using simulated annealing algorithm parameter optimization. The results of the study show that in the week of July 1-7, 2023, a maximum profit of \$9708.282 can be achieved using the strategy proposed in this paper.

References

- [1] Zhao Ling, Liu Zhixue. Joint replenishment and pricing strategy considering customer returns and fixed costs [J]. *Operations Research and Management*, 2022, 31(06): 105-110+124.
- [2] JINLONG ZHANG, XIANG WU, HAOXUAN XU. A joint decision model for pricing and replenishment of perishable new products [J]. *Journal of Systems Engineering*, 2018,33(01): 79-89.
- [3] Xin Chanyun. Research on multi-stage pricing of perishable fresh products [D]. Chongqing University, 2022.
- [4] Nie Yuxuan. Research on automatic pricing and replenishment strategy of fresh commodities based on ARIMA prediction optimization model with vegetable commodities as an example [J]. *Commercial Exhibition Economy*, 2024(05): 19-22.
- [5] Qiao Xue. Joint replenishment pricing strategy of fresh products considering sales loss [D]. Southeast University, 2021.
- [6] Zhang Y, Kong W, Dong Z Y, et al. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network [J]. *IEEE Transactions on Smart Grid*, 2019.
- [7] Adaryani F R, Mousavi S J, Jafari F. Short-term rainfall forecasting using machine learning- based approaches of PSO-SVR, LSTM and CNN [J]. *Hydrology*, 2022(614): 1-8.
- [8] SHI Qingyan, YUE Juzai, HAN Ping, et al. Short-term flight trajectory prediction based on LSTM-ARIMA model [J]. *Signal Processing*, 2019, 35(12): 10-10.
- [9] PENG Laihu, WANG Weihua, WAN Changjiang et al. Energy-saving optimization of flexible flow shop scheduling based on genetic simulated annealing algorithm [J]. *Software Engineering*, 2022, 25(11): 49-55.
- [10] Li C G Z D S. Novel Adaptive Simulated Annealing Algorithm for Constrained Multi-Objective Optimization [J]. *China Communications*, 2022(9): 1-11