ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

Research on Pricing and Replenishment of Vegetable Products Based on Optimization Models

Wanru Bai*, Yijing Qiao, Zixi He

International Business College, Dongbei University of Finance and Economics, Dalian 116025, China
**Corresponding author: 17866551624@163.com*

Abstract: *The perishable nature of vegetable commodities requires fresh supermarkets to make scientific sourcing and pricing decisions, which is vital for promoting the welfare of residents and balancing the interests of parties related to the industry. Based on historical sales data, this paper studies the distribution patterns and interrelationships of sales volumes of different vegetable categories, as well as the relationship between sales volumes and cost-plus pricing mainly by establishing Spearman model and conducting regression analysis in which MATLAB fitting toolbox is used to derive the functional relationship. The result reveals that a significant positive correlation exists between the foliage and cauliflower category, aquatic rhizomes and edible fungi category. The sales volume of flowering and leafy vegetables is the highest, while that of the nightshade family is the lowest. Subsequently, with sales volume as the decision variable and supermarket profits as the objective function, an optimization model is constructed and solved, taking into account the predictions of ARIMA and its seasonal improvement model to ensure that a relatively reasonable and optimal daily replenishment and pricing strategy that maximize the profits is ultimately obtained.*

Keywords: *Spearman Model, ARIMA Seasonality Model, Regression Analysis, Optimization Model*

1. Introduction

Vegetables are indispensable in people's daily lives and are one of the most consumed food types. Due to their characteristics of perishability and value loss, many domestic supermarkets still can only judge the freshness of vegetables by visual inspection and feeling, and then adjust their sales prices, lacking professional standard measurement. In addition, they need to decide the daily replenishment amount based on historical sales data and demand. Liang Shuting (2020) has pointed out that in a competitive market, avoiding inventory backlogs or stock-outs of fresh produce is crucial to improving the profitability of community fresh produce retailers and improving the lives of residents [1]. Regarding how to determine reasonable vegetable pricing and replenishment strategies, many scholars have conducted research from multiple perspectives in recent years. Mao Lisha (2022) from the supply chain perspective quantitative analysis of the vegetable pricing strategy and production and marketing model [2], but it is mainly focuses on wholesale markets as the object of study, while the supermarket as a consumer direct consumption behavior of the place also has a strong research significance. Lu Yajie (2010) shows that many merchants have introduced various forms of pricing methods associated with different promotional techniques for frequent price adjustments, and he also quantitatively analyzes the relationship between the selling price of fresh vegetables in supermarkets and the rate of value loss, thus deriving a dynamic pricing model [3].

This paper combines the specific pricing and sales of a supermarket with the massive data, explores the distribution pattern and correlation of sales volume of various vegetable categories, analyzes the relationship between the total sales volume and the cost-plus pricing, and through the construction of the optimization model, obtains the reasonable total daily replenishment volume and the optimal pricing strategy of each vegetable category in the coming week (July 1-7, 2023), in order to maximize the profit of the supermarket, and at the same time, to provide a broad range of fresh supermarkets with a certain degree of reference (data source: http://www.mcm.edu.cn/).

ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

2. Research on the distribution pattern and interrelationship of sales volume of different vegetable categories

2.1 Normal distribution test

The first step is to find the missing values in the dataset through the find function of MATLAB, and to determine some of the outliers through life experience, which improves the precision of the data in the subsequent runs and the accuracy of the arithmetic results. In addition, for the series that conform to the normal distribution, the outliers can be defined and screened using the 3σ principle, combined with manual determination to prevent the occurrence of misjudgment. For the series that do not pass the normal distribution, manual judgment can be made by directly listing the sequence in ascending order. By conducting in-depth analysis of the domain knowledge and business background of the data, further processing of misjudgments can be carried out. Therefore, to identify further ways to remove outliers, this paper performs a normal distribution test.

There are usually two methods of testing for normal distribution, one is the Shapiro-Wilk test, which is applicable to small sample data (sample size < 50), and the other is the Kolmogorov-Smirnov test, which applies to large sample data (sample size > 5000). Since the sample size N > 5000 for wholesale price (RMB/kg), this paper utilizes SPSSPRO to conduct the Kolmogorov-Smirnov test to determine the distribution mode.

Generally speaking, the smaller the P-value is, the more likely it is to reject the null hypothesis. If it presents significance (P<0.05), it means that the null hypothesis (the data conforms to normal distribution) is rejected, which means the data does not satisfy normal distribution. In addition, due to the difficulty in satisfying the test in real-world research, if the absolute value of sample kurtosis is less than 10 and the absolute value of skewness is less than 3, combined with a normal distribution histogram, the variable can be described as basically following a normal distribution [4]. This paper takes the wholesale price (RMB/kg) as an example, and the results of its descriptive statistics and normality test are shown in Table 1.

Table 1: Test results

Sample size	Median	Mean	Standard deviation	Skewness	Kurtosis	K-S test
55982	4.63	5.963	5.068	4.371	52.004	0.142

As can be seen from the above table, the K-S test yields a p-value of 0.142>0.05, then the original hypothesis is accepted, that is, the wholesale prices satisfy a normal distribution. According to the 3σ principle, when the given data set obeys a normal distribution, there is about 99.7% chance that a data point falls on the interval (μ-3σ, μ+3σ). Therefore, the probability of data points falling outside this range is about 0.3%, which is a small probability event and can be identified as an outlier. The formula for determining outliers is as follows.

$$P(|x-u|>3\sigma)\leq 0.003 \tag{1}$$

To further test normality and perform outlier detection, this paper plots the normality test histogram of the wholesale price (RMB/kg), as shown in Figure 1. If the obtained data histogram is extremely different from the bell shape, the normality distribution is rejected, which is an intuitive and practical method [5-6].

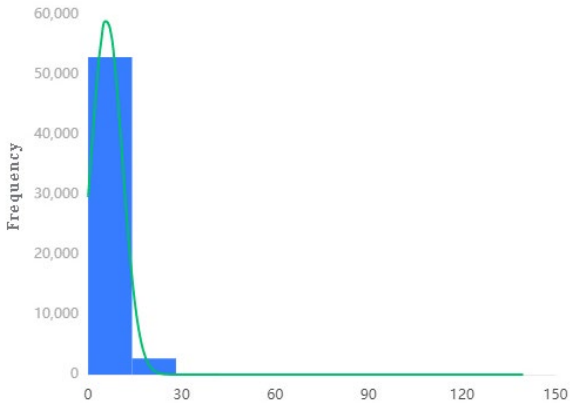

Figure 1: Normality test histogram of the wholesale price

ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

As can be seen from Figure 1, the distribution of wholesale prices shows a bell shape (high in the middle and low at both ends), indicating that the data, although not absolutely normal, can basically be accepted as a normal distribution. At the same time, according to the 3σ principle, 659 outliers were found in the wholesale price data and removed.

2.2 Visualization of sales volume of vegetable categories

The following figure shows the sales volume of each vegetable category of the superstore from 2020 to 2023.

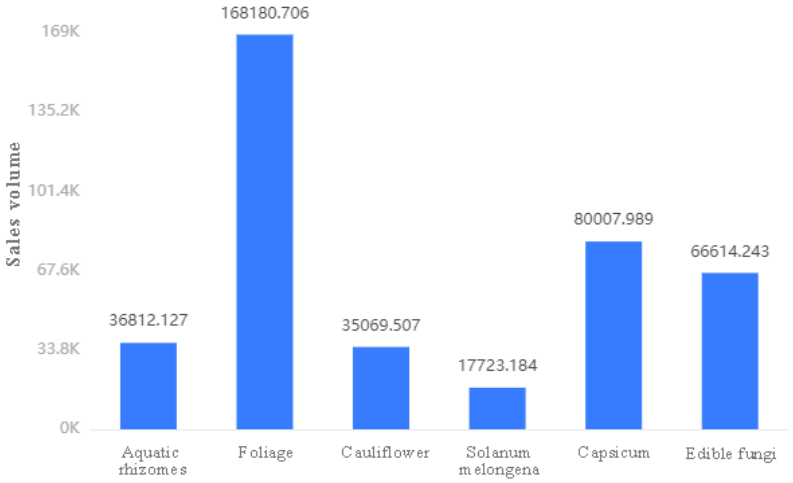

Figure 2: Total sales volume of each vegetable category in 2020-2023

Figure 2 reveals that the total sales volume of foliage vegetables is the highest among all categories, and the solanum melongena category is the lowest. In addition, the total sales of aquatic rhizome vegetables and cauliflower vegetables are relatively close. The preliminary inference to be drawn from this is that foliage vegetables were purchased more frequently for most of the time frame counted, while sales of solanaceous vegetables may have been concentrated in certain seasons or have been in a condition of being purchased less frequently for a long time. Moreover, the purchase volume of vegetables in the other four categories may also exhibit a certain periodicity.

In order to more intuitively reflect the change rule of sales between categories, this paper uses Excel to visualize the data and draw a line graph as shown below.

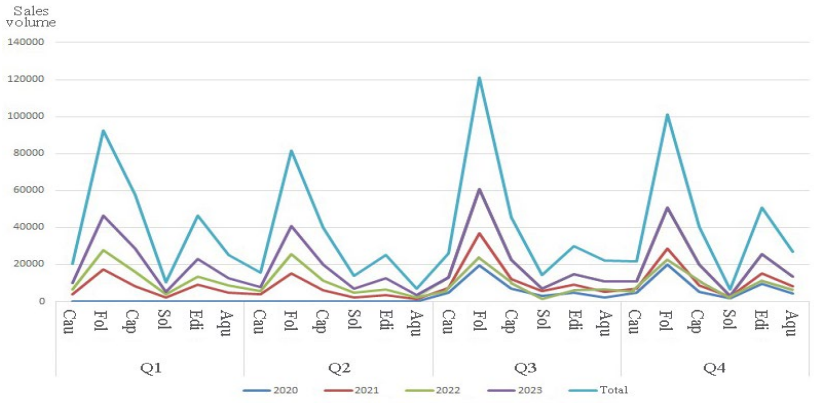

Figure 3: Quarterly sales volume of vegetable categories

The above figure shows the sales volume of six vegetable categories for each quarter from 2020 to 2023, in order from left to right: cauliflower, foliage, capsicum, solanum melongena, edible fungi, and aquatic rhizomes (shown in the first three letters of the name). According to Figure 3, the distribution pattern of sales volume of different vegetable categories can be obtained. As the year increases, the sales volume of each vegetable category generally increases, and there is no big difference in the proportion of sales volume among categories each year. For different quarters of the same year, the sales volume of each vegetable category changes differently, and the quarter in which the sales volume peaks is not the

ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

same. For example, sales of foliage vegetables peaked in the third quarter, and their sales in the first and second quarters, though lower than the other two quarters, were still significantly higher than those of other categories, indicating that the above conjecture is correct. Sales of edible fungi, on the other hand, would peak in the fourth quarter of the year, with significantly lower sales in the second and third quarters. In addition, the sales of cauliflower, solanum melongena, and aquatic rhizomes are almost flat in all quarters.

2.3 Correlation analysis

Spearman correlation coefficient can quantify the trend between two random variables and can be used to reflect the degree of correlation of data between regions [7]. After grasping the general relationship between vegetable categories, this paper imports the processed data into SPSS to carry out correlation test and judge their correlation with each other through the Spearman coefficient. The basic formula is as follows:

$$\rho=1-\frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

Notes: $d_i=x_i-y_i$.

According to the derived correlation coefficient table, the following heat map of correlation coefficients is obtained:

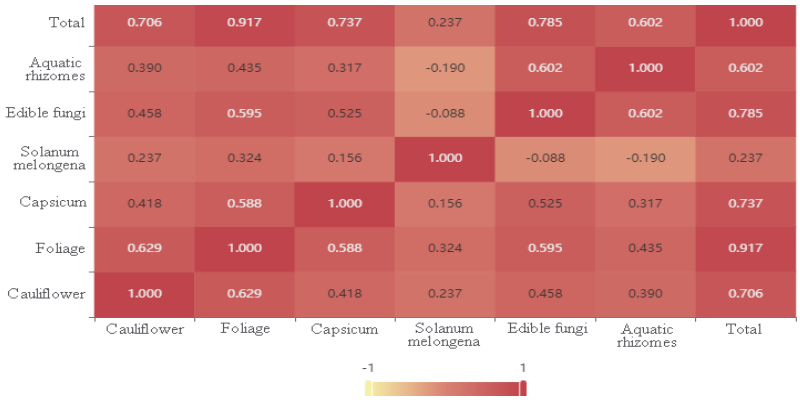

Figure 4: Heat map of correlation coefficient

The above Figure 4 indicates the magnitude of the correlation coefficient by the shade of the color. It can be seen that there is a particularly significant and positive correlation between foliage and cauliflower, as well as between aquatic rhizomes and edible fungi, which to a certain extent reflects the consumer's purchasing habits. Solanum melongena category has a low correlation with many categories, including capsicum, aquatic rhizomes, and edible fungi (negatively correlated with the latter two categories).

Then, this paper randomly selects a single product from the categories of foliage and cauliflower, respectively. Taking sweet cabbage (foliage category) and green stalks scattered flowers (cauliflower category) as an example for correlation analysis, the results obtained are as follows:

Table 2: Correlation analysis between sweet cabbage and green stem scattered flowers

		Green stem scattered flowers	Sweet cabbage
Green stem scattered flowers	Pearson correlation coefficient	1	.941**
	Sig.(two-tailed)		.000
	Number of cases	13	8
Sweet cabbage	Pearson correlation coefficient	.941**	1
	Sig.(two-tailed)	.000	
	Number of cases	8	8
**. At the 0.01 level (two-tailed), the correlation is significant.			

From Table 2, it can be seen that the correlation coefficient between sweet cabbage and green stem scattered flowers is 0.941, indicating a strong positive correlation between the two individual items,

ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

which is also in line with the conclusion drawn earlier on the correlation between these two categories.

3. The Relationship between Total Sales Volume of Vegetable Categories and Cost-Plus Pricing

Taking the example of vegetables from the cauliflower category, this paper uses the dataset obtained after data preprocessing and applies the MATLAB fitting toolbox to perform stepwise regression in order to establish a linear regression model between sales volume and cost-plus pricing. That is:

$$Y_i = \beta_0 + \beta_1 x_i \tag{3}$$

In this context, *1* is the coefficient for the independent variable, and *0* is the coefficient for the constant term. Y_i represents the pricing of cauliflower on the *i*th day, and x_i represents the total sales volume of cauliflower category vegetables on the *i*th day. We have imported the extracted data into MATLAB and utilized the regress function to perform linear regression, with the results shown in the following Table 3.

Table 3: Linear regression results (cauliflower category vegetables)

	Unstandardized coefficients		Standardized coefficients	t	P	R^2	F
	B	Standard error	Beta				
Constant	20.722	1.997	-	9.557	0.000***	0.236	F=20.34,P=0.000***
Sales(kg)	-15.342	3.43	-0.478	-4.521	0.000***		

Notes: *p<0.1, **p<0.05, ***p<0.01

As indicated by the table, the expression for the functional relationship is $Y_i = 20.722 - 15.342xi$. The population regression coefficient is not zero, indicating a significant correlation between variables at the significance level. The fit chart is shown below Figure 5. The blue line represents the actual values, while the green one represents the predicted values.

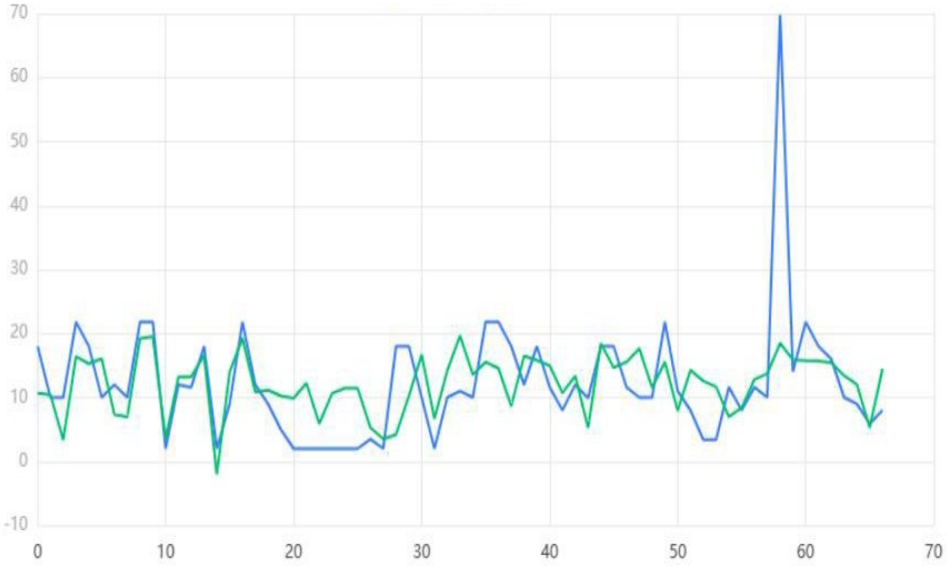

Figure 5: Fitted effect diagram

4. Research on Optimal Replenishment and Pricing Strategies for Specific Scenarios

Based on the dataset and the conclusions mentioned above, this paper solves for the example of the cauliflower category. It derives the daily replenishment total and pricing strategy for various vegetable categories that maximize the supermarket's profits for the upcoming week (July 1-7, 2023). For the rest of the vegetable categories, only the final calculated results are displayed.

First, set the decision variables as x_i (i=1 to 7), representing the total sales volume of the cauliflower category on day *i*. The supermarket adopts a cost-plus pricing method, the core idea of which is to combine the costs with the desired profit to determine the final sales price. Therefore, let c_i represent the

ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

wholesale price of the cauliflower category for that day, and w_i represent the daily cost markup percentage (profit margin).

The final pricing Y_i can be expressed as:

$$Y_i = (1 + w_i) * c_i \tag{4}$$

The total profit for cauliflower vegetables on day i is:

$$W_i = c_i * w_i * x_i \tag{5}$$

Therefore, the total profit for the supermarket in the coming week is:

$$W_{Total} = \sum_{i=1}^{7}(W_{i1} + W_{i2} + W_{i3} + W_{i4} + W_{i5} + W_{i6}) \tag{6}$$

The numbers 1-6 respectively represent categories of cauliflower, foliage, capsicum, solanum melongena, edible fungi, and aquatic rhizomes vegetables. When establishing the optimization model, the equality constraints are the relationships between sales volume and pricing, and the inequality constraints relate to the non-negativity of the associated variables. The model is as follows:

$$\begin{cases} Max\ W_{Total} = \sum_{i=1}^{7}\sum_{j=1}^{6} W_{ij} \\ y_{ij} = a_{j0} + a_{j1}x_{ij} \\ y_{ij} = (1 + w_{ij}) * c_{ij} \\ W_{ij} = w_{ij} * c_{ij} * x_{ij} \\ x_{ij}, y_{ij}, W_{ij} \geq 0 \\ w_{ij} \in R \\ i = 1,2,\ldots,7 \\ j = 1,2,\ldots,6 \end{cases} \tag{7}$$

Taking into account that the period required for forecasting is a complete week, including workdays and holidays, it is preliminarily judged to exhibit relative volatility. To ensure the maximization of supermarket profits, the primary goal should be to meet sales demands. This paper conducts an AHP (Analytic Hierarchy Process) test on sales data from the same period in previous years, dividing it into stationary and non-stationary series. The sales volume is then predicted separately using ARIMA models and seasonal ARIMA models, and the predictions are substituted into the fitted models for each category to derive the optimal pricing strategy.

The ARIMA model is a widely-used method for analyzing and modeling various time series data. This model is based on the notion that the time series to be predicted is generated by some stochastic process, and if the stochastic process generating the sequence does not change over time, then the structure of this stochastic process can be precisely characterized and described [8]. Using past observations of the series, one can deduce the future values of the series. The recognition of the seasonal order for the multiplicative seasonal model, that is, the cycle length S, can be obtained through autocorrelation and partial autocorrelation plots [9]. If substantial peaks in absolute value occur at points that are multiples of a changing cycle and exhibit oscillating changes, the time series can be described by a SARIMA model [10]. In the ARIMA model, the future value of the series is expressed as a linear function of the current and lagged values of lagged terms and random disturbance terms, and the general form of the model is as follows:

$$Y_t = c + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q} \tag{8}$$

Combining the aforementioned models and solution results, we can obtain the total daily replenishment volume for the next seven days and the optimal pricing strategy (presented in the form of cost markup ratio), which are shown in Tables 4 and 5, respectively.

ISSN 2616-5902 Vol. 6, Issue 4: 269-275, DOI: 10.25236/AJBM.2024.060439

Table 4: Daily replenishment total when the profit of the supermarket is maximized

Time	Cauliflower	Foliage	Capsicum	Solanum melongena	Edible fungi	Aquatic rhizomes
1	53.354	211.456	75.144	9.791	65.201	21.827
2	20.122	260.457	108.957	8.732	66.912	11.628
3	41.145	261.424	109.551	5.622	31.853	15.661
4	39.265	164.145	43.092	10.835	49.309	9.709
5	50.924	262.481	110.027	3.295	46.763	18.394
6	31.45	260.473	108.927	6.624	58.331	10.224
7	29.175	203.124	70.087	6.153	61.965	14.355

Table 5: Pricing strategy when the profit of the supermarket is maximized

Time	Cauliflower	Foliage	Capsicum	Solanum melongena	Edible fungi	Aquatic rhizomes
1	50.06%	33.25%	75.14%	96.00%	33.06%	84.08%
2	25.25%	27.43%	108.95%	41.80%	100.03%	44.21%
3	50.15%	42.88%	109.55%	64.84%	107.01%	88.24%
4	87.36%	20.38%	43.09%	105.51%	86.00%	32.74%
5	41.12%	91.23%	110.02%	95.34%	92.50%	52.91%
6	39.27%	38.74%	108.92%	109.11%	91.30%	33.92%
7	52.17%	55.48	70.08%	84.10%	62.41%	38.11%

5. Conclusion

According to historical sales data, this paper visualizes the sales volume of each vegetable category in a specific year, analyzes its distribution law, calculates the Spearman coefficient to judge its inherent relationship, and verifies it through the analysis of the relationship between individual products. It was found that with the increase of years, the sales volume of each category generally increased, but the relative ratio of sales volume among categories was not much different in each year. For different quarters in the same year, the sales volume changes of each vegetable category were different. Overall, the sales volume of foliage vegetables was the largest, and that of solanum melongena was the smallest. There was a strong positive correlation between foliage vegetables and cauliflower, edible fungi and aquatic rhizomes. In order to better achieve the goal of maximizing profits, MATLAB and SPSSPRO software were used to obtain the functional relationship between sales volume and cost-plus pricing. By constructing optimization models and solving them, the optimal daily replenishment total and pricing strategy in the given specific situation were obtained. Similar analytical processes can also be extended to find optimal strategies applicable to other periods, thereby providing a reference for daily operation of fresh supermarkets to offer more convenient services and fresher vegetables for community residents.

References

[1] Liang Shuting. Study on the Inventory Strategy of Community Fresh Food Retail Outlets Affected by Both Time and Stock Quantity on Spoilage Rate [D]. Beijing Jiaotong University, 2020.

[2] Mao Lisha. Study on Pricing Strategies and Production-Marketing Models of Vegetable Wholesale Markets from a Supply Chain Perspective [D]. Central South University of Forestry and Technology, 2023.

[3] Lu Yajie. Research on Dynamic Pricing Issues of High-Quality Fresh Vegetables in Our Country's Supermarkets [D]. Beijing Jiaotong University, 2010.

[4] Rakesh S, Suvin P S, Chakravortty D, et al. Closing the Loop and Eco-Friendly Cutting Oils [C]//International Tribology Conference.2015.

[5] Chen Jun. Normality Test of Data and Practical Application of Excel/SPSS/Stata Software [J]. Journal of Sichuan Vocational and Technical College, 2019, 29(03):157-161.

[6] Chi Shengchao, Zhang Zhenyuan. Statistical and predictive analysis of drainage water quality characteristics in drainage tunnels based on normal test [J]. Northern Transportation, 2021, (02): 88-91+94

[7] Wu Junying, Lu Xin, Liu Hong, et al. Ultra-Short-Term Multi-Regional Electric Power Load Forecasting Based on the Spearman-GCN-GRU Model [J]. China Electric Power, 2024.

[8] Jindong C, Wenda L, Niancheng Z. Research on Pricing Model and Strategy of Electric Vehicle Charging and Discharging Based on Multi View[J].Proceedings of the CSEE, 2018.

[9] Wang Yan. Application of Time Series Analysis. [M]. Beijing: Renmin University of China Press, 2005.

[10] Zhou Xiangyu, Li Si. Short-term Forecasting of Courier Service Volume in Jiangsu Province Based on TOPSIS Criteria and SARIMA Model [J]. Science, Technology, and Industry, 2023, 23(17):136-142.