# Research on Precise Demand Prediction of New Retail Target Product Based on Dual Model

Jianzheng Xu[1], Zong Liu[1], Yinghui Xu[1], Jinpeng Hao[1], Jiaming Sun[1], Yunrui Lu[1], Bowen Pang[1]

[1a]193423020209@email.sau.edu.cn

[1b]m15065355661@163.com

[1c]2274224074@qq.com

[1d]1011518348@qq.com

[1e]2160623589@qq.com

[1f]1106760368@qq.com

[1g]1368578272@qq.com

[1]Shenyang Aerospace University, Shenyang, Liaoning, China

**Abstract:** In this study, the target data set is firstly obtained by data processing, and then the Grey Verhulst model and ARIMA model are respectively used for modeling and prediction research on the target data set, so as to calculate the predicted value. Then MAPE (average absolute percentage error) is calculated according to the predicted value. The typical characteristics and adaptability of Grey Verhulst model and ARIMA model are compared and analyzed. This study shows that the ARIMA model is more accurate than the Grey Verhulst model in the short term, and its prediction accuracy decreases sharply with the extension of time, which is suitable for the short term prediction. The accuracy of Grey Verhulst model is relatively stable, and the accuracy is improved with the extension of time, which is suitable for medium and long term prediction.

**Keywords:** Model prediction research, Grey Verhulst model, ARIMA model, Data analysis

## 1 Introduction

With the vigorous development of China's economy, the consumption mode of the consumer market has gradually changed from "material-oriented" to "customer-oriented" [5]. In the current new retail industry, people's demand is no longer the pursuit of practicality, but the consideration of individuation, fashion and beauty, and customers no longer regard cost performance as the only criterion for buying goods [6]. Driven by this special demand, the development of multiple varieties and small quantities has gradually become a new production mode for new retail enterprises [3]. This makes the jewelry and toys in the retail stores in the

ICIDC 2022, June 17-19, Qingdao, People's Republic of China
Copyright © 2022 EAI
DOI 10.4108/eai.17-6-2022.2322736

mall more dazzling, and also increases the great difficulty in the inventory management of the retail industry. Some studies have pointed out that one of the significance of accurate prediction of future demand is to effectively guide enterprise operation and supply and marketing plan [7].

Based on this research idea, this paper adopts grey Verhulst model and ARIMA model to predict the precise demand of new retail target products, and studies the prediction characteristics and applicability of the two models through comparative analysis. In addition, daily data simulation and prediction given in the official event annex are used to verify the feasibility of the model and the typicality and applicability of accurate demand prediction. On the one hand, it can solve the inventory management problem of the company's retail industry, and on the other hand, it can also adapt to the transformation of the company's production model. The research results not only solve the problem of "accurate demand prediction" for new retail enterprises, but also adapt to the development of the consumer market and the consumption pattern in the market.

The organizational structure of this paper is as follows: The second part is the empirical analysis of the model adopted in this paper and the introduction and discussion of the research results; The third part is the summary of the full text; The fourth part is an appendix.

## 2 Empirical analysis

### 2.1 Data Sources

In this paper, N products produced by a retail enterprise in East China are selected as the basic prediction data provided in annex D of the 10th MathorCup Mathematical Modeling Challenge for Universities in 2020.

Firstly, the target data set is obtained through data processing, and the Grey Verhulst model and ARIMA model are used to model and forecast the target data set respectively, so as to calculate the predicted value. Then MAPE (average absolute percentage error) is calculated according to the predicted value, and then the Grey Verhulst model and ARIMA model are compared and analyzed. The prediction characteristics and applicability of the model are summarized.

### 2.2 Prediction Of Grey Verhulst Model

The monthly sales volume of target subcategory has a certain random fluctuation, showing a non-stationary random process with a certain trend of change, and generally has saturation state. The Grey Verhulst model was established to forecast the sales volume in October, November and December 2019.

The original data of target subclass 27050401 are shown in Table 1.

Table 1: Raw data for target subclass 27050401.

| Month | 2018/1 | 2018/2 | 2018/3 | 2018/4 |
|---|---|---|---|---|
| Sales | 14629 | 18440 | 14223 | 13742 |
| Month | 2018/5 | 2018/6 | 2018/7 | 2018/8 |
| Sales | 22020 | 14989 | 16313 | 16711 |
| Month | 2018/9 | 2018/10 | 2018/11 | 2018/12 |
| Sales | 15684 | 18390 | 18285 | 24136 |
| Month | 2019/1 | 2019/2 | 2019/3 | 2019/4 |

| | | | | |
|---|---|---|---|---|
| Sales | 23254 | 19164 | 17378 | 12646 |
| Month | 2019/5 | 2019/6 | 2019/7 | 2019/8 |
| Sales | 15800 | 17201 | 15136 | 16976 |
| Month | 2019/9 | 2019/10 | 2019/11 | 2019/12 |
| Sales | 17899 | 19445 | 16360 | 17573 |

The monthly sales volume of target sub-category 27050401 in 2018-2019 is shown in Figure 1.

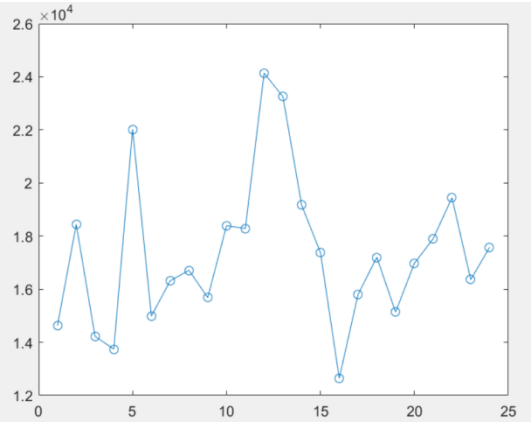


Figure 1: Monthly sales line chart of target sub-category 27050401 in 2018-2019.

(1) Set $x^{(1)}$ as the monthly original data series of target subcategory 27050401 in 2018-2019, i.e:

$$\begin{aligned} x^{(0)} &= (x_1^{(1)}, x_2^{(1)}, ..., x_{24}^{(1)}) \\ &= (14629, 18440, 14223, 13742, 22020, 14989, \\ &\quad 16313, 16711, 15684, 18390, 18285, 24136, 23 \\ &\quad 254, 19164, 17378, 12646, 15800, 17201, 1513 \\ &\quad 6, 16976, 17899, 19445, 16360, 17573) \end{aligned} \tag{1}$$

(2) Let it be a cumulative generation (1-IGO), given by

$$x_k^{(0)} = x_k^{(1)} - x_{k-1}^{(1)}, k = 2, 3, 4, ..., \tag{2}$$

Have to

$$
\begin{aligned}
x^{(0)} &= (x_1^{(0)}, x_2^{(0)}, ..., x_{24}^{(0)}) \\
&= (14629, 3811, -4217, -481, 8278, -7031, 13 \\
&24, 398, -1027, 2706, -105, 5851, -882, -409 \\
&0, -1786, -4732, 3154, 1401, -2065, 1840, 92 \\
&3, 1546, -3085, 1213)
\end{aligned} \tag{3}
$$

(3) For the adjacent mean generation, let

$$
z_k^{(1)} = 0.5(x_k^{(1)} + x_{k-1}^{(1)}), k = 2, 3, 4, ..., \tag{4}
$$

Have to

$$
\begin{aligned}
z^{(0)} &= (z_1^{(1)}, z_2^{(1)}, ..., z_{24}^{(1)}) \\
&= (0, 16535, 16332, 13983, 17881, 18504, 1565 \\
&1, 16512, 16198, 17037, 18338, 21210, 23695, \\
&21209, 18271, 15012, 14223, 16501, 16168, 16 \\
&056, 17437, 18672, 17902, 16967)
\end{aligned} \tag{5}
$$

(4) The parameter list

$$
\hat{a} = (a + b)^T \tag{6}
$$

Do the least square method, get

$$
\hat{\alpha} = (B^T B)^{-1} B^T Y = \begin{bmatrix} -0.0519 \\ 0 \end{bmatrix} \tag{7}
$$

(5) Verhulst model for

$$
\frac{dx^{(1)}}{dt} - 0.0519 x^{(1)} = 0 \tag{8}
$$

$\hat{x}^{(1)}$ (The predicted value of $x^{(1)}$) is

$$\begin{aligned}\hat{x}^{(1)} &= (\hat{x}_1^{(1)}, \hat{x}_2^{(1)}, ..., \hat{x}_{24}^{(1)}) \\ &= (14848.4,15062.8,15272.2,15476.5,15675 \\ &.5,15869.2,16057.7,16240.8,16418.5,16590 \\ &.8,16757.9,16919.6,17076.0,17227.2,17373 \\ &.3,17373.3,17650.2,17781.3,17907.5,18029. \\ &1,18146.0,18258.4,18366.4,18470.1)\end{aligned} \tag{9}$$

(7) Model accuracy check

①Residual qualifying model

Let the original sequence be

$$x^{(0)} = (x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}) \tag{10}$$

The corresponding Grey Verhulst model prediction (simulation) sequence is

$$\hat{x}^{(0)} = (\hat{x}_1^{(0)}, \hat{x}_2^{(0)}, ..., \hat{x}_n^{(0)}) \tag{11}$$

The residual error sequence is listed

$$\begin{aligned}\varepsilon^{(0)} &= (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) \\ &= (x_1^{(0)} - \hat{x}_1^{(0)}, x_2^{(0)} - \hat{x}_2^{(0)}, ..., x_n^{(0)} - \hat{x}_n^{(0)})\end{aligned} \tag{12}$$

The relative error sequence is

$$\Delta = (\Delta_1, \Delta_2, ..., \Delta_n) = \left(\left|\frac{\varepsilon_1}{x_1^{(0)}}\right|, \left|\frac{\varepsilon_2}{x_2^{(0)}}\right|, ..., \left|\frac{\varepsilon_n}{x_n^{(0)}}\right|\right) \tag{13}$$

For $k \le n$,

$$\Delta_k = \left|\frac{\varepsilon_k}{x_k^{(0)}}\right| \tag{14}$$

is called k point simulation relative error;

$$\overline{\Delta} = \frac{1}{n}\sum_{k=1}^{n}\Delta_k \tag{15}$$

is called the relative error; $1-\overline{\Delta}$ is called average relative accuracy; $1-\Delta_k$ is called k point simulation accuracy [4].

For a given $\alpha$, when $\overline{\Delta} < \alpha$ and $\Delta_k < \alpha$ are true, the model is called residual qualified model.

②Correlation degree qualified model

Let $x^{(0)}$ be the original sequence, $\hat{x}^{(0)}$ the corresponding grey model prediction (simulation) sequence, $\varepsilon^{(0)}$ the residual sequence, and $g$ the absolute correlation between $x^{(0)}$ and $\hat{x}^{(0)}$. For given $g_0 > 0$, there is $g > g_0$, then the model is called relational qualified model.

③Mean square deviation ratio qualified model

Let $x^{(0)}$ be the original sequence, $\hat{x}^{(0)}$ the corresponding grey model prediction (simulation) sequence, $\varepsilon^{(0)}$ the residual sequence, then, the mean and variance of $x^{(0)}$ are respectively

$$\begin{cases} \overline{x} = \dfrac{1}{n}\sum_{k=1}^{n} x_k^{(0)} \\ S_2^2 = \dfrac{1}{n}\sum_{k=1}^{n}(x_k^{(0)} - \overline{x})^2 \end{cases} \tag{16}$$

The mean and variance of $\varepsilon^{(0)}$ are respectively

$$\begin{cases} \overline{\varepsilon} = \dfrac{1}{n}\sum_{k=1}^{n} \varepsilon_k \\ S_2^2 = \dfrac{1}{n}\sum_{k=1}^{n}(\varepsilon_k - \overline{\varepsilon})^2 \end{cases} \tag{17}$$

The mean square error ratio is

$$C = \frac{S_2}{S_1} \tag{18}$$

④Small error probability qualified model

The probability of small error is zero

$$p = P\{|\varepsilon_k - \bar{\varepsilon}| < 0.6745S_1\} \tag{19}$$

For given $p_0 > 0$, at $p > p_0$, the model is called $p_0 > 0$ small error probability qualified model.

It can be seen from the above that given a set of values, a level of accuracy of the test model can be determined. The commonly used accuracy grades are shown in Table 2, which can be used as reference for testing models. In general, the most commonly used is the relative error test index.

Table 2: Grey Verhulst model accuracy test grade.

| Level | Indicators | | | |
|---|---|---|---|---|
| | Relative correlation degree | Absolute correlation | Mean square error ratio | Probability of small error |
| I | 0.01 | 0.90 | 0.35 | 0.95 |
| II | 0.05 | 0.80 | 0.50 | 0.80 |
| III | 0.10 | 0.70 | 0.65 | 0.70 |
| IV | 0.20 | 0.60 | 0.80 | 0.60 |

Mean relative error $\bar{\Delta} = 11.25\%$, then the model accuracy is level IV; the absolute correlation was 0.8657, the model precision is levelerII. The sales volume of target subcategory 27050401 from October to December was predicted, and the errors were as follows: 0.80%, 9.57%, 1.49%.

In conclusion, the Grey Verhulst model has high accuracy. Ten target subclasses were predicted and summed up, and MAPE values were calculated as shown in Table 3.

Table 3: The target subclass predicted values of MAPE in October, November and December 2019

| Month | Original value | Predictive value | MAPE |
|---|---|---|---|
| October | 51778 | 49786.8 | 0.2209 |
| November | 42637 | 50606.4 | 0.2523 |
| December | 46289 | 51006.1 | 0.1822 |

### 2.3 Prediction of ARIMA Model

The ARIMA model is applied to the sales time series data of a target subclass with mixed trend. According to the changing trend of product sales reflected in the time series, the data is processed smoothly, and the sales in a future period of time is analyzed and predicted with the help of ARIMA model.

Select the continuous sales data of target subcategory 27050401 from 2018 to 1019 in monthly statistical unit. The data is an isometric sequence sorted from 2018/1 to 2019/9, and there is no correlation between sales volume. In order to observe the sales data of target subclass 27050401 more intuitively, time series diagrams were made respectively with time and sales volume as horizontal and vertical coordinates, as shown in the figure below. According to the chart, there were significant fluctuations between May and June 2018 and December 2018 to March 2019, mainly reflected in partial seasonal trends.

(1) Data checking and preprocessing

In solving the sales forecasting problem, because the time series analysis model has many restrictions on the actual random series, and attaches importance to the requirements of stationarity, so before analyzing the model, the stationarity of time series should be determined first, and the difference processing method should be used to make the non-stationary time series stable. From the time series diagram, we can intuitively judge that the sales data of this small class has a non-stationary sequence.

(2) The differential

First-order differential calculation of sales data is shown in Figure 2.

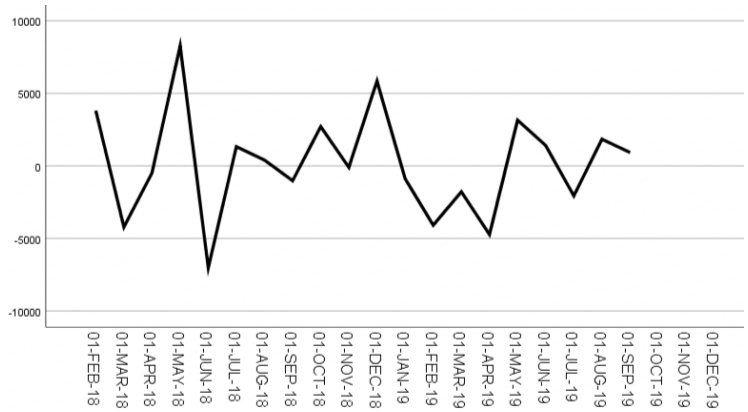


Figure 2: First order difference diagram of sales data.

The sequence of sales data of the target subclass 27050401 after difference fluctuates around the value 0, which is in line with the characteristics of stationary series. Intuitively, the sequence after difference processing can be preliminarily judged as stationary series.

(3) ARIMA model recognition and grading

The stationary time series data processed by difference has met the data stationarity requirements of ARIMA prediction model. According to the correlation characteristics of ARIMA model, the values of P and Q are preliminarily identified by reading the autocorrelation graph (ACF) and partial autocorrelation graph (PACF) of the stationary time series processed by difference, as shown in Figure. 3 and Figure. 4.

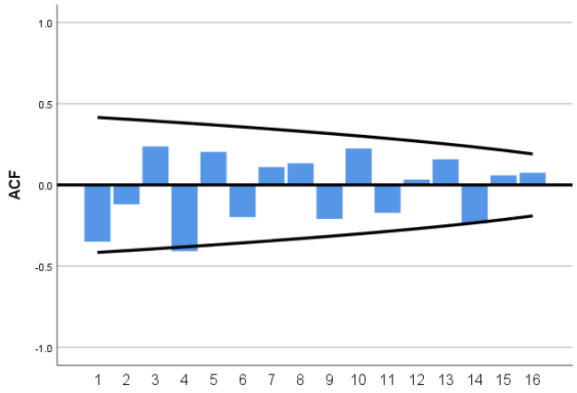


Figure 3: Autocorrelation graphs of stationary time series.

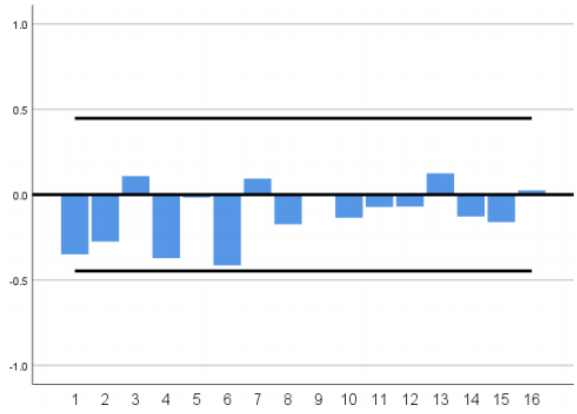


Figure 4: Partial autocorrelation of stationary time series.

The fourth order in the ACF diagram is truncated, and the first order in PACF is truncated. It can be preliminarily judged that the sequence after difference is suitable for ARIMA (4,1,1) model.

(4) ARIMA model

$$\Delta X_t = 0.194\Delta X_{t-1} + \varepsilon_t - 0.98\varepsilon_{t-1} \tag{20}$$

(5) ARIMA model adaptability test

In order to obtain the adaptability of the model, white noise test was performed on the residual sequence of the model through the values of autocorrelation function (ACF) and partial correlation function (PACF), as shown in Figure 5. The test results show that the residue does not have obvious autocorrelation independently. Therefore, residual error can be regarded as white noise. And it is suitable for the ARIMA model of sales data.

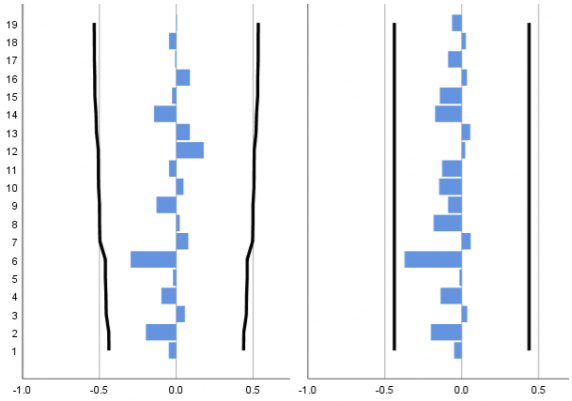

Figure 5: ARIMA model adaptability test diagram.

(6) ARIMA model prediction

Since the ARIMA model built for this sales data is adaptable, the ARIMA model can be used to forecast the sales data for the next three months based on the historical data for the first 21 months of 2018-2019. [1] The ARIMA model fitting and prediction lines are shown in Figure 6, and the data are obtained in Table 4.

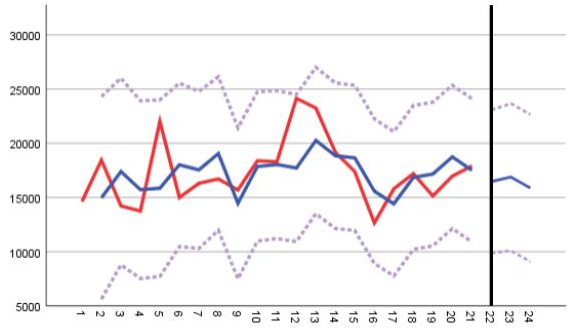

Figure 6: Image of projections based on data for the first 21 months of 2018-2019 for target subcategory 27050401.

Table 4: The target subclass predicted values of MAPE in October, November and December 2019

| Month | Original value | Predictive value | MAPE |
|---|---|---|---|
| October | 19445 | 16473 | 0.01266 |
| November | 16360 | 16883 | 0.01046 |
| December | 17573 | 15876 | 0.00709 |

Output graphs of the top ten target subcategories by sales volume to visually see if the seasonality and correlation of the data is significant. As can be seen from Figure 7, only the top three target subcategories of sales have obvious seasonality, while the remaining target

subcategories of sales are relatively stable, which means that the prediction results of these data may be more accurate.

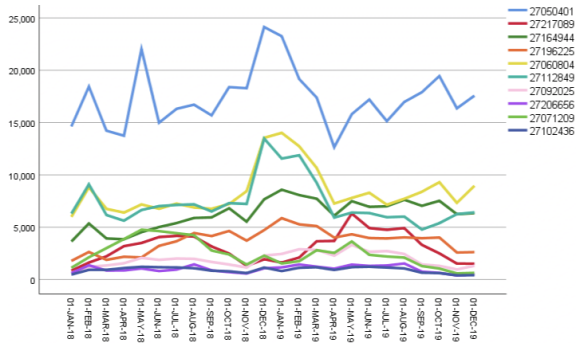


Figure 7: Sales chart of top ten target categories.

The same method is applied to the remaining 9 target subcategories according to the processing method of sales data of target subcategory 2705401, and the prediction is made for the last three months of 2019, and the MAPE value is calculated. [2] Finally, the original value, predicted value and MAPE value of the ten target subclasses were summed up monthly, as shown in Table 5.

Table 5: The sum of the top ten target categories in October, November and December 2019 predicted value MAPE.

| Month | Original value | Predictive value | MAPE |
|---|---|---|---|
| October | 51778 | 46459 | 0.1889 |
| November | 42637 | 41668 | 0.4385 |
| December | 46289 | 36016 | 0.6234 |

## 2.4 Comparative Analysis of The Two Models

MAPE predicted by the two models in Table 6, the accuracy of ARIMA model is higher than that of gray model in the short term, and its prediction accuracy will decrease sharply with the extension of time. The accuracy of Grey model is relatively stable, and the accuracy is improved with the extension of time. Therefore, ARIMA model is suitable for short-term forecasting and gray model is suitable for medium and long term forecasting. In this case, sales data are more saturated than seasonal.

Table 6: Comparative analysis table of the two models.

| Month | MAPE of Grey Verhulst model | MAPE of ARIMA model |
|---|---|---|
| October | 0.2209 | 0.1889 |
| November | 0.2523 | 0.4385 |
| December | 0.1822 | 0.6234 |

## 3 Conclusions

This paper aims to solve the problem of accurate demand prediction for new retail target products. In this paper, grey Verhulst model and ARIMA model are used to model and forecast the target data set respectively, so as to calculate the predicted value. Then MAPE (average absolute percentage error) is calculated according to the predicted value, and then the grey Verhulst model and ARIMA model are compared and analyzed. It is found that ARIMA model has higher accuracy than grey Verhulst model in the short term, and its prediction accuracy will decrease sharply with the extension of time, which is suitable for short-term prediction. The accuracy of grey Verhulst model is relatively stable, in the empirical analysis, the average relative error is 11.25%, the model accuracy of level 4, the absolute correlation of 0.8657, the model accuracy of level 2, and with the extension of time, the accuracy is improved, suitable for medium and long term prediction. In this study, sales data are more saturated than seasonal.

## References


[1] Hao Li. (2021). Prediction of Coal Price based on ARIMA Model. Frontiers in Economics and Management, 2(8), pp. 155-163.

[2] Ma Minglu and Wang Qiang. (2022). Analysis of the impact of epidemic situation on total electricity consumption of the United States under the prediction scenario. IOP Conference Series: Earth and Environmental Science, 983(1),

[3] Qiaoyu Shen, Junfeng Wei & Aifen Feng. (2020). GM(1,1) Model for Grey Prediction of New Retail Sales Based on Principal Component Analysis. Statistics and Applications (06). doi:10.12677/SA.2020.96109.

[4] Shoujun Li et al. (2020). A novel varistructure grey forecasting model with speed adaptation and its application. Mathematics and Computers in Simulation, 172(C), pp. 45-70.

[5] Shujun Teng, Huiwen Zheng & Baisen Liu. (2018). Research on the influencing factors of supermarket merchandise sales based on multiple regression analysis. National circulation economy (14), 6-9. doi:10.16834/j.cnki.issn1009-5292.2018.14.003.

[6] Xiaomei Huang. (2013). Application of unary linear regression analysis in supermarket product sales. Science and technology information(11), 77-78. doi:CNKI:SUN:KJXX.0.2013-11-054.

[7] Zhang Bin et al. (2021). Research on Long Short-Term Decision-Making System for Excavator Market Demand Forecasting Based on Improved Support Vector Machine. Applied Sciences, 11(14), pp. 6367-6367.