Journal of Engineering Innovations & Technology (2024)

Journal of Engineering Innovations & Technology

journal homepage: www.journal-eit.online

Paper Type: Original Article

Research on New Energy Vehicle Development Prediction based on Random Forest Model and gray Prediction

Fei Xu[1,*], Kexin Sun [1,*]

1.School of Mathematics and Statistics, Shandong University of Technology, Zibo, China

Abstract

This paper focuses on predicting the development of new energy vehicles (NEVs) using random forest model and gray Prediction models. New energy vehicles, including hybrid, pure electric, and fuel cell electric vehicles, have seen rapid growth due to their low pollution, low energy consumption, and strong peak load capacity. This research aims to analyze the factors influencing the NEV industry and forecast its future development. The gray Prediction method, known for addressing small sample sizes and incomplete data, is used to forecast the conservative quantity of NEVs in this paper. Meanwhile, the random forest model regression model evaluates the impact of various factors on the market share of traditional fuel vehicles. Key variables include production, fuel prices, charging stations, subsidies, and sales volumes. Results indicate that the NEV market in China will experience rapid growth over the next decade, with increasing market penetration and sales. Factors such as government subsidies and technological advancements significantly influence the traditional fuel vehicle market.

Keywords: New Energy Vehicle, Random Forest Model, Gray Prediction, Development Prediction

1 | Introduction

New energy vehicles refer to vehicles using advanced technical principles, new technologies and new structures, using unconventional vehicle fuels as power sources, and integrating advanced technologies in vehicle power control and drive. New energy vehicles include four categories: hybrid electric vehicles, pure electric vehicles, fuel cell electric vehicles and other new energy vehicles. As a kind of new energy vehicle, new energy electric vehicles have gained rapid development in recent years due to their characteristics of low pollution, low energy consumption and strong peak load capacity. New energy electric vehicles, including electric buses and household electric vehicles with less than 7 seats, have been welcomed by consumers and governments around the world. Since 2011, the government of a certain country has been actively promoting the development of new energy electric vehicles and has implemented a series of favorable policies. The new energy electric vehicle industry has made significant progress and has gradually become another symbol of the country, akin to its high-speed railway system. It has had a significant impact on the country's and the world's ecological environment and related industries. Therefore, it is of great significance to analyze the factors influencing the development of this country's new energy electric vehicle industry and predict its future development.

2 | Related Works

In recent years, with the continuous development and application of new energy technologies, the new energy vehicle (NEV) industry has garnered significant attention and research. Chen Menglong et al.[1] employed machine learning algorithms to predict stock prices in the NEV industry, demonstrating the broad prospects and potential value in this field. Shikang Wang[2] focused on the application of intelligent manufacturing technology in NEVs, highlighting its critical role in improving production efficiency and quality. In terms of valuation of listed NEV companies, Xu Huang[3] used neural network models to conduct research, providing a new valuation method. Jianxin Wei[4] evaluated the value of NEV enterprises using machine learning methods, offering insights for market positioning and strategic decisions. Jiale Chen[5] researched user satisfaction of small NEVs through sentiment analysis, gaining a deeper understanding of consumer demands and expectations for NEVs. Xuyang Li[6] conducted research on classifying the quality of NEV patents using machine learning models, aiming to enhance patent management and innovation capabilities. Yiqing Wang[7] predicted prices for pure electric second-hand cars using machine learning, providing a scientific basis for price formulation in the used car market. Hao Zhang[8] analyzed multi-channel retail demand and forecasted sales using machine learning, offering data support for retail enterprises' sales strategies. Additionally, Leqi Wu[9] studied the development trends of NEVs using internet search data and machine learning methods, revealing the importance of big data in industry research. Yingshuang Yin[10] forecasted sales of various types of cars based on integrated Baidu Index data and multi-source data fusion, showcasing the potential of multi-source data fusion technology in sales forecasting.

3 | Theory and Method

3.1 | Establishment of gray prediction model

The gray prediction method, characterized by its qualitative approach to time series prediction and nonlinear features, primarily addresses challenges such as small sample sizes, inadequate data information, and incomplete datasets. This method excels in calculating the correlation and trend of relevant factors, enabling the prediction of future development trends. Given that new energy vehicles exhibit these characteristics, the article employs the gray prediction method to forecast the conservative quantity of new energy vehicles. The outlined steps for this prediction are as follows:

Step 1: Identify the key factors that affect the development of new energy vehicles, and use the number of new energy vehicles in stock (10000 units) and sales of new energy vehicles (10000 units) from XX to YY as indicators to measure the development of new energy vehicles in China.

Step 2: Perform data validation: Using $GM(1,1)$ model (solving a variable with a first-order differential equation) to test the data, first calculate the rank ratio of the sequence:

$$a(k) = \frac{x^0(k-1)}{x^0(k)} \tag{1}$$

Among them, $k = 2,3..n$,if the level ratios calculated by the above formula all fall within a tolerable coverage interval $X = \left(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}}\right)$, $GM(1,1)$ model can be established for gray prediction of the data.

Step 3: Building a gray Prediction Model: This article provides the following matrix formulas by defining gray derivatives, generating sequences, defining gray differential equations, and other processes using the matrix method:

$$B = \begin{bmatrix} (-z(2) & 1 \\ \vdots & \vdots \\ -z(n) & 1 \end{bmatrix} \tag{2}$$

$$Y = \begin{pmatrix} X^{(0)}(2) \\ \vdots \\ X^{(0)}(n) \end{pmatrix} \tag{3}$$

Among them, B is the data matrix and Y is the data vector. The least squares estimation parameter column of the gray differential equation satisfies the following equation:

$$u = [a, b]^T = \left(B^T B\right)^{-1} B^T Y \tag{4}$$

And the $GM(1,1)$ model can be represented as:$Y = Bu$. Among them, a is the development coefficient, which controls the development trend of the entire system, b is the gray action quantity, the size of the value reflects the relationship between data changes, and u is the influencing factor of external actions.

Step 4: Problem Solving:Perform an accumulation operation on the level ratio calculation formula in step 2 to weaken it to a disturbance of $X^{(0)}$:

$$x_k^{(1)} = \sum_{i=1}^{k} x_i^{(0)}, k = 1,2, \ldots, n \tag{5}$$

The sequence of generating equal weight critical values for $X^{(1)}$ is:

$$Z^{(1)} = \left\{z^{(1)}(2), z^{(1)}(3), \ldots, \ z^{(1)}(n)\right\} \tag{6}$$

$$Z^{(1)}(k) = \frac{1}{2}\left(x^{(1)}(k) + x^{(1)}(k-1)\right) \tag{7}$$

Where $z^{(1)}$ is the whitening background value of the model, according to the above formula, the corresponding white differential equation of the $GM(1,1)$ model can be obtained as:

$$x_0^{(k)} + az^{(1)}(k) = b \tag{8}$$

Next, establish a prediction model:

$$\hat{x}^{(1)}(k) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-a(k-1)} + \frac{b}{a}, k = 1,2, \ \ldots, \ n \tag{9}$$

3.2 | Indicator selection

With the increase of new energy vehicle technology and cost, consumers are more inclined to choose them, thus gradually reducing the market share of traditional cars. To address this issue, we selectively gather pertinent data related to new energy vehicles in China to scrutinize their impact on the traditional energy vehicle market. Data sources will include reputable outlets like the China Statistical Yearbook, New Energy Vehicle Association, and Ministry of Industry and Information Technology. The chosen independent variables will encompass the ratio of new energy vehicles to traditional fuel vehicles, new energy vehicle production, fuel prices, public charging station numbers, subsidies for new energy vehicles, count of new energy vehicle-related enterprises, average price of new energy vehicles, patent applications for new energy vehicles, market share, market penetration rate, and sales volume of new energy vehicles. The dependent variable will be the global market share of traditional energy vehicles. Our analysis will focus on elucidating the impact of these selected independent variables on the dependent variable, as shown in Table 1.

Table 1. Partially selected indicator data

Years	GDP	Ratio	Production	Oil price	Charging station	Subsidy
2012	39711	0.000120139	1.5	7.6	1.8	0
2013	43497	0.000332215	1.8	7.8	2.1	6.2
			……			
2021	80976	0.022758606	352.1	8.1	261.7	53.9
2022	85698	0.026139994	700.3	8.5	299.1	61.5
Years	**Enterprise**	**Market share**	**Permeability**	**Sales volume**	**Average car price**	**Patent**
2012	10	0.08	0.0007	1.3	38	3274
2013	300	0.9	0.001	1.8	35	4515
			……			
2021	2400	13.4	0.064	352.1	15	25642
2022	2600	25.6	0.073	688.7	13	27890

3.3 | Establishment of random forest model

Random forest model plays a pivotal role in the analysis of the impact of new energy electric vehicles on the traditional energy vehicle industry, as they can simultaneously consider multiple features. In our study, the selected data encompasses various features pertinent to both the new energy electric vehicle and traditional energy vehicle sectors, including fuel prices and subsidies for new energy vehicles. Leveraging these features as inputs, the random forest model effectively models and analyzes the relationship between new energy electric vehicle sales and other contributing factors. Moreover, when scrutinizing the impact of new energy electric vehicles on the traditional energy vehicle industry, non-linear factors and response relationships may come into play. The random forest model excels in non-linear modeling, thanks to its tree-like structure and random feature selection, enabling a more accurate fit to complex relationship patterns. Notably, this model possesses the capability to assess the importance of features. Analyzing the feature importance in the random forest model provides insights into which features exert the most significant influence on the market share of traditional fuel vehicles. This analysis is instrumental in identifying the key factors shaping the impact of new energy electric vehicles on the traditional energy vehicle industry. The algorithm steps for the random forest model used in this article are as follows:

Step 1: Use the TreeBagger function to build a regression random forest model with 100 decision trees, where each tree has a minimum leaf size of 5. The predicted output of the random forest model is the average of the predictions from all the decision trees, as shown in formula 10.

$$F(x) = \frac{1}{T}\sum_{i=1}^{T} f_i(x) \tag{10}$$

Where $F(x)$ is the predicted output of the random forest model, $f_i(x)$ is the predicted output of the i-th decision tree, and T is the total number of decision trees.

Step 2: Calculate the out-of-bag error rate (OBB) of the random forest model. $L(y_j, F(x_j))$ represents the loss function, which measures the difference between the true value y_j and the predicted value $F(x_j)$ of the j-th sample. N is the total number of samples, as shown in formula

11.

$$OOB = \frac{1}{N}\sum_{j=1}^{N} L(y_j, F(x_j)) \tag{11}$$

The importance of out of bag replacement of random forest model refers to the increase of out of bag error of random forest model after randomly replacing the value of a feature on the out of bag sample, that is:

$$VI_k = OOB - OOB_{\pi_k} \tag{12}$$

Where, VI_k is the importance of out of pocket substitution of the kth feature, OBB is the original out of pocket error of random forest model, OOB_{π_k} is the out of pocket error after randomly replacing the kth feature, π_k is the random permutation function of the kth characteristic.

Step3: Calculate the correlation degree R^2 between the predicted value and the real value.

$$R^2 = 1 - \frac{\sum_{j=1}^{N} (y_j - F(x_j))^2}{\sum_{j=1}^{N} (y_j - \bar{y})^2} \tag{13}$$

Where, y_j is the true value of the jth sample, $F(x_j)$ is the random forest model prediction value of the jth sample, and j is the mean value of the true value of the sample.

4 | Results and Discussions

4.1 | Results of gray prediction model

The predicted value can be obtained by subtracting the above equation, and the results are shown in Figure 1 and Table 1.

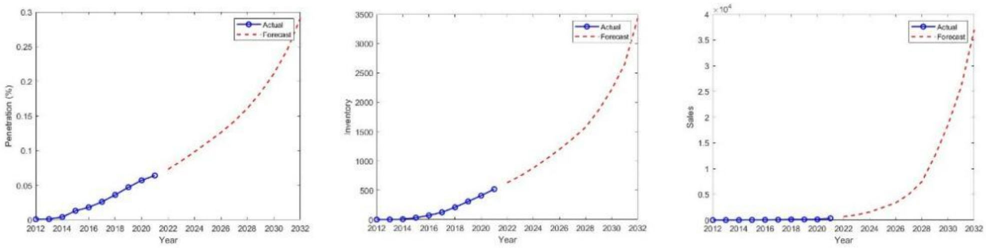

Figure 1. Results of gray prediction model

Table 2. Results of gray prediction model

Year	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
Penetration	8.5%	9.8%	11.2%	12.6%	14.2%	16.1%	18.4%	21.1%	24.5%	29.1%
Inventory	745	882	1033	1198	1379	1573	1867	2188	2511	2863
Sales	1015	1567	2450	3368	5014	7489	12524	18563	25909	36972

As shown in the figure above, it can be observed that the development trend of China's new energy vehicle market will show rapid growth in the next 10 years. This trend is manifested in market penetration, ownership and sales. As time goes by, the proportion of new energy vehicles in the Chinese market will continue to increase, and the market penetration rate will gradually increase. At the same time, the number of new energy vehicles will also increase rapidly, which means that more and more car owners choose to purchase and use new energy vehicles. This rapid growth trend will also be reflected in sales, and the sales of new energy vehicles will continue to rise. The accelerated development of this trend also indicates that the new energy market will usher in faster expansion and development in the future. As technology continues to advance and costs drop, the feasibility and competitiveness of new energy vehicles will continue to increase. The government's support policies for new energy vehicles and the increasing awareness of environmental protection will also further promote the prosperity of the new energy market. With the popularization of charging facilities and the increase in cruising range, consumers' acceptance of new energy vehicles will continue to increase, prompting the expansion of the market size.

4.2 | Results of random forest model

The specific parameter settings for the random forest model in this problem are shown in Table 3.

Table 3. random forest model Parameter Table

Parameter Name	Parameter value
Data splitting	0.7
Node split evaluation criterion	Mse
Maximum feature ratio considered during partitioning	None
Minimum number of samples for internal node split	2
Minimum number of samples for leaf nodes	1
Minimum weight of samples in leaf nodes	0
Maximum depth of the tree	10
Maximum number of leaf nodes	50
Number of decision trees	100
Sampling with replacement	True
Out-of-bag data testing	False

By solving the random forest model, the proportion diagram of the impact of 11 independent variables on the global market share of traditional fuel vehicles is drawn by using MATLAB software as shown in Figure 2.

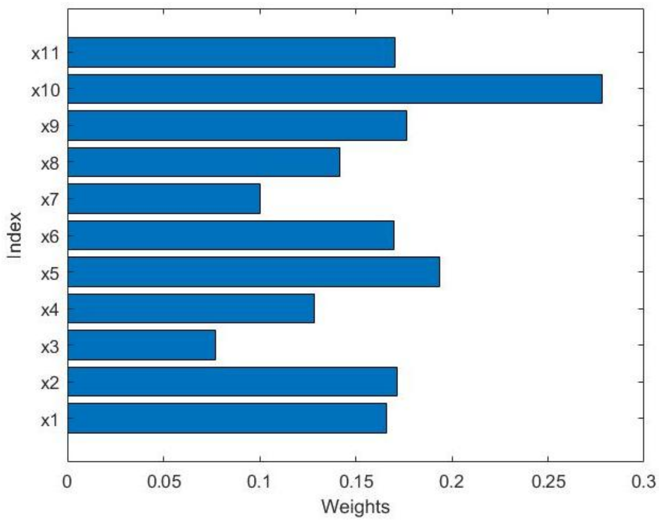

Figure 2. Influence specific gravity diagram

The detailed characteristic importance values of each variable are shown in Table 4.

Table 4. This caption has one line so it is centered

Index	Ratio	Production	Fuel price	Pcharging stations
Characteristic importance	0.166268	0.171466	0.0770041	0.12872
Index	**Subsidies**	**Enterprises**	**Market share**	**Penetration rate**
Characteristic importance	0.193739	0.169981	0.100504	0.141903
Index	**Sales volume**	**Average price**	**Patent applications**	
Characteristic importance	0.176783	0.278027	0.170233	

The generated fitting image is shown in the following Figure 3.

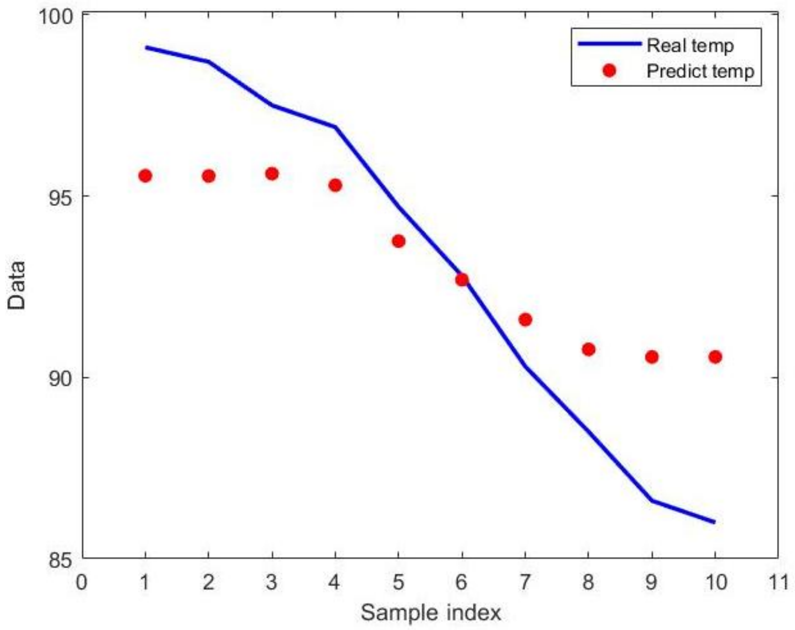

Figure 3. random forest model fitting image

Through analysis, among the factors related to new energy vehicles, the price of energy vehicles has the greatest impact on the global market share of traditional fuel vehicles; Secondly, the Chinese government's financial subsidies for the purchase of new energy vehicles have a great impact on the market of traditional fuel vehicles.

5 | Conclusion

Through the research in this paper, we applied random forest model and gray prediction model to make in-depth analysis and prediction of the development of new energy vehicles in China. The results show that the new energy vehicle market will maintain rapid growth in the next decade, and the market penetration rate and sales volume will be significantly increased. Research shows that government subsidies and technological progress have a significant impact on the traditional fuel vehicle market, and these factors also play an important role in the promotion and popularization of new energy vehicles. The research in this paper not only provides a valuable reference for policy makers, but also provides data support and analysis basis for industry stakeholders when making strategic decisions. With the continuous progress of technology and the gradual maturity of the market, China's new energy vehicle industry is expected to achieve sustainable development in the future. In order to further promote this process, it is suggested that more measures should be taken in policy support, infrastructure construction and public awareness promotion to ensure that the dual advantages of new energy vehicles in environmental protection and economic benefits can be fully played.

References

[1] Chen M., Fan C., Wu Z. Stock Price Prediction in the New Energy Vehicle Industry Based on Machine Learning Algorithms. Journal of Jilin Institute of Business, 93-100. DOI: 10.19520/j.cnki.issn1674-3288.2024.01.007 (2024).

[2] Wang S. Analysis of Intelligent Manufacturing Technology Application in New Energy Vehicles. Automobile Testing Report, 46-48 (2023).

[3] Huang X. Valuation Study of Listed Companies in New Energy Vehicles Based on Neural Network Models. Master's Thesis, Suzhou University. DOI: 10.27351/d.cnki.gszhu.2023.002297 (2023).

[4] Wei J. Value Assessment of New Energy Vehicle Enterprises Based on Machine Learning Methods. Master's Thesis, Jiangxi University of Finance and Economics. DOI: 10.27175/d.cnki.gjxcu.2023.001347 (2023).

[5] Chen J. User Satisfaction Study of Small New Energy Vehicles Based on Sentiment Analysis. Master's Thesis, East China Jiaotong University. DOI: 10.27147/d.cnki.ghdju.2023.000666 (2023).

[6] Li X. Patent Quality Classification Prediction of New Energy Vehicles Based on Machine Learning Models. Master's Thesis, Tianjin University of Technology. DOI: 10.27360/d.cnki.gtlgy.2022.000950 (2022).

[7] Wang Y. Price Prediction of Pure Electric Second-hand Cars Based on Machine Learning. Master's Thesis, Shandong University. DOI: 10.27272/d.cnki.gshdu.2022.005242 (2022).

[8] Zhang H. Analysis and Sales Forecasting of Multi-channel Retail Demand Based on Machine Learning. Master's Thesis, Guizhou University. DOI: 10.27047/d.cnki.ggudu.2022.003030 (2022).

[9] Wu L. Research on the Development of New Energy Vehicles Based on Internet Search Data and Machine Learning Methods. Master's Thesis, China University of Petroleum (Beijing). DOI: 10.27643/d.cnki.gsybu.2022.001860 (2022).

[10] Yin Y. Research on the Sales Forecast of Several Types of Cars Based on Baidu Index and Multi-source Data Fusion. Master's Thesis, Xiangtan University. DOI: 10.27426/d.cnki.gxtdu.2022.001444 (2022).