Prediction of the Development of China's New Energy Industry

Peitong Li *, Hu Wei, Jiaqi Fu

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou, China, 450001

* Corresponding author: 16692076357@163.com

Abstract. In the past decade, China's new energy vehicle industry has developed rapidly due to the influence of national policies. In recent years, China has started planning to cease policy support for the new energy vehicle industry. Therefore, predicting the future development of China's new energy industry is of great significance for policy-making in China. This paper selects 10 factors influencing the development of new energy electric vehicles in the past ten years and 4 indicators reflecting the development of new energy electric vehicles. We combine the Entropy Weight Method with TOPSIS to score the development status of new energy electric vehicles based on 4 indicators. Then, we use Principal Component Analysis to reduce the dimensionality of the ten influencing factors to three principal components. To quantify these effects, Lasso regression is performed using 3 principal components to prevent overfitting. Therefore, this article obtained a model for predicting the development status of China's new energy vehicle industry, with a goodness of fit of over 0.98, indicating a good fit of the model. In order to predict the development score for the next decade, this paper first predict the data of ten factors for the next decade. The factors that remain unchanged are not treated. Factors that change the trend significantly are predicted using GM (1,1) and the forecast is good. For stationary time series, this paper collects data from nearly 20 years and trains on Long Short-Term Memory Networks. R2 = 0.891 and RMSE < 0.1, Therefore, the model can be used to predict data for the next 10 years. After obtaining various data, this paper obtains the development score of new energy electric vehicles in the next 10 years according to the regression equation. Finally, this paper concludes that the development and progress of China's new energy industry will slow down in the next decade, with a score increase of about 20%.

Keywords: Lasso regression, Long Short-Term Memory Networks, Development of new energy industry.

1. Introduction

Under the background of global automobile industry facing transformation, the development of new energy automobile industry has a profound impact on the global energy structure and environmental sustainable development[1]. The Chinese government strongly advocates the green development, people pay more and more attention to low-carbon environmental protection, in the travel, began to choose pollution-free green transportation, in the new energy automobile industry, pure electric models occupy a dominant position, therefore, explore the main factors affecting the development of China's new energy and electric vehicles, predict the development of new energy and electric vehicle development on the ecological environment is imperative.

We found that there are many factors affecting the development of new energy electric vehicles, such as new energy car ownership, market share, market penetration rate, Car production and sales, and so on.

Traditional automobiles mainly use fossil fuels, which emit harmful gases when burned, bringing serious problems to the environment. Applying new energy to automobiles can reduce the burning of fossil fuels, reduce the impact of harmful gas emissions on the environment, and achieve the purpose of environmental protection. The use of new energy vehicles has a good impact on the ecological environment[2].

This paper has addressed the following issues.

(1) Analyze factors that affect the development of new energy electric vehi-cles in China and establish a mathematical model to show their impact on the development of new energy electric vehicles in China.

(2) Use the mathematical model to describe and predict the development of new energy electric vehicles in China in the following ten years.

2. Materials and Methods

2.1. Data acquisition and preprocessing

Through publicly available data from the National Bureau of Statistics of China, we have collected various indicators reflecting the development of new energy electric vehicles in China from 2013 to 2022, as well as various indicators affecting the development of new energy and electric vehicles in China from 2013 to 2022. The data are standardized by Z-Score method to ensure that different indicators have the same scale. The collected data is shown in Figures 1 and 2.

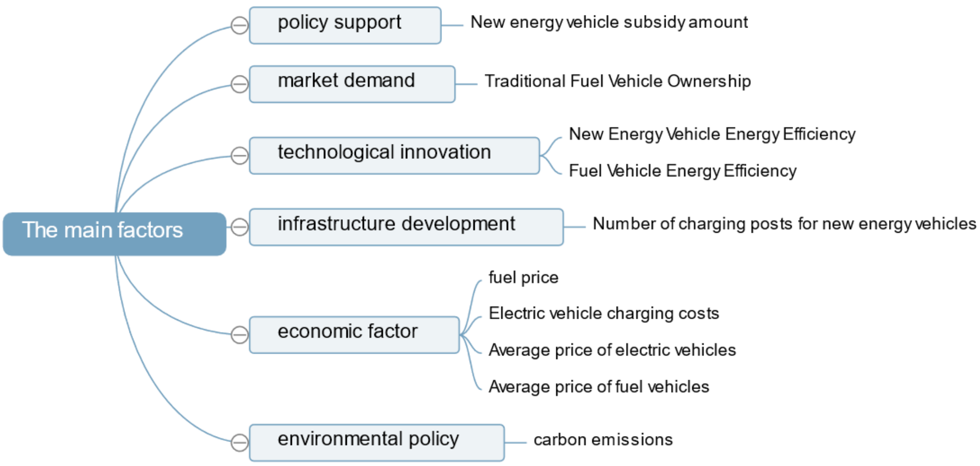

Figure 1. Influencing factors

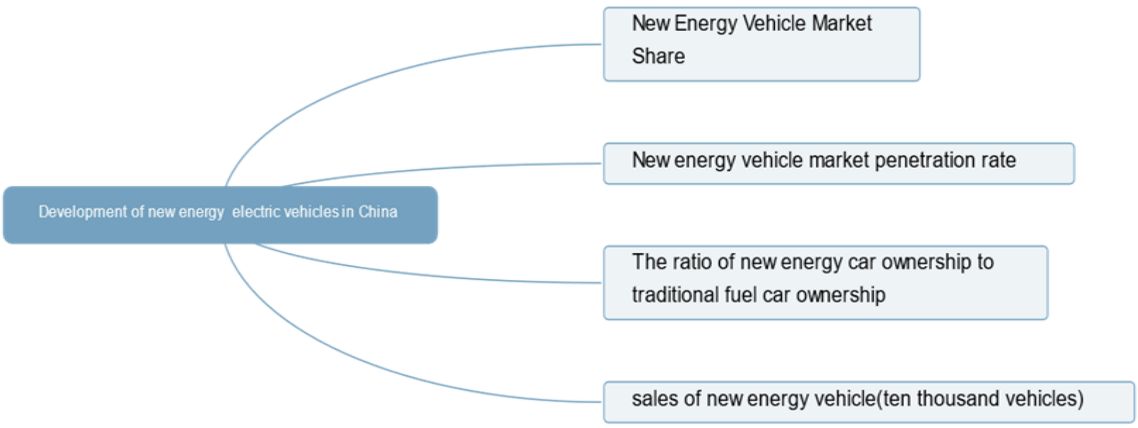

Figure 2. Development indicators

2.2. Method introduction

2.2.1. Analysis and Prediction Model Construction of New Energy Industry Development

The model construction process for this section is shown in Figure 3.

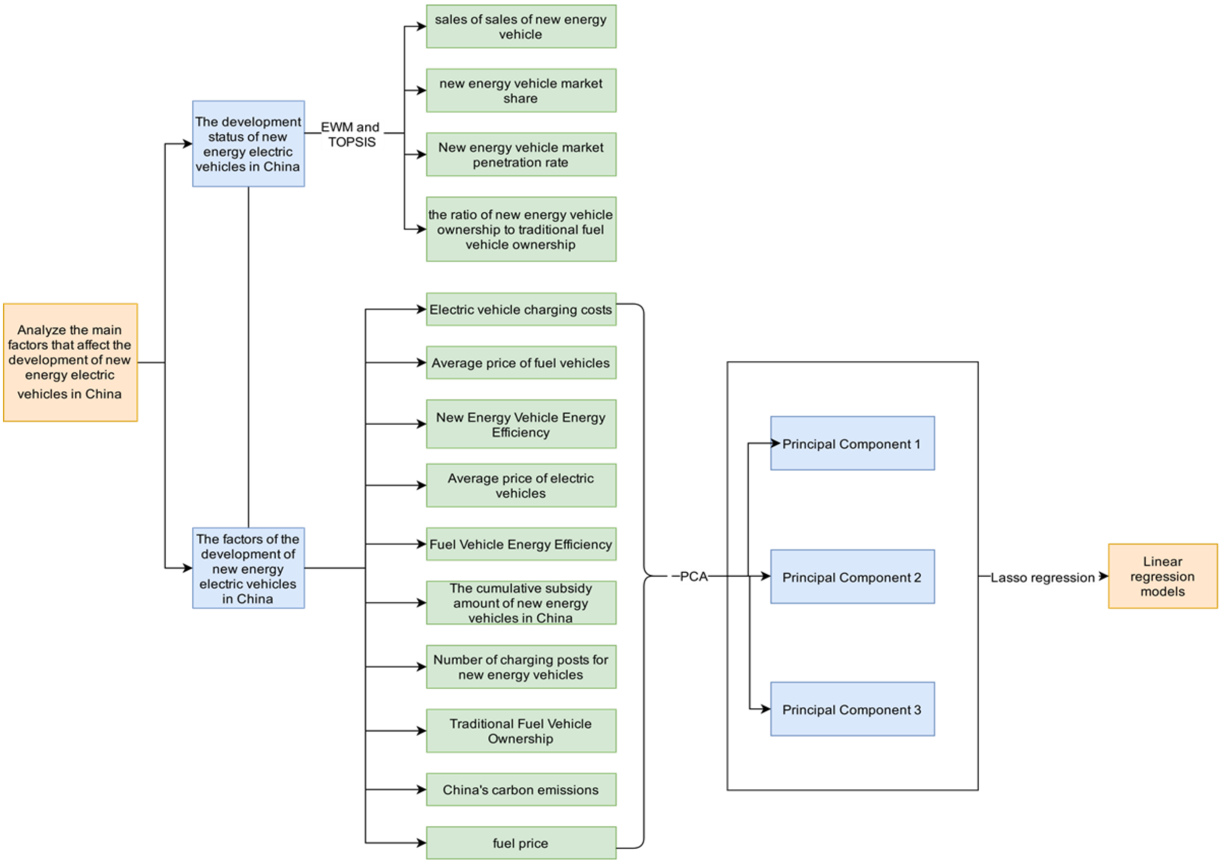

Figure 3. Influencing factors

2.2.2. Prediction of the Future Development of China's New Energy Industry

The model construction process for this section is shown in Figure 4.

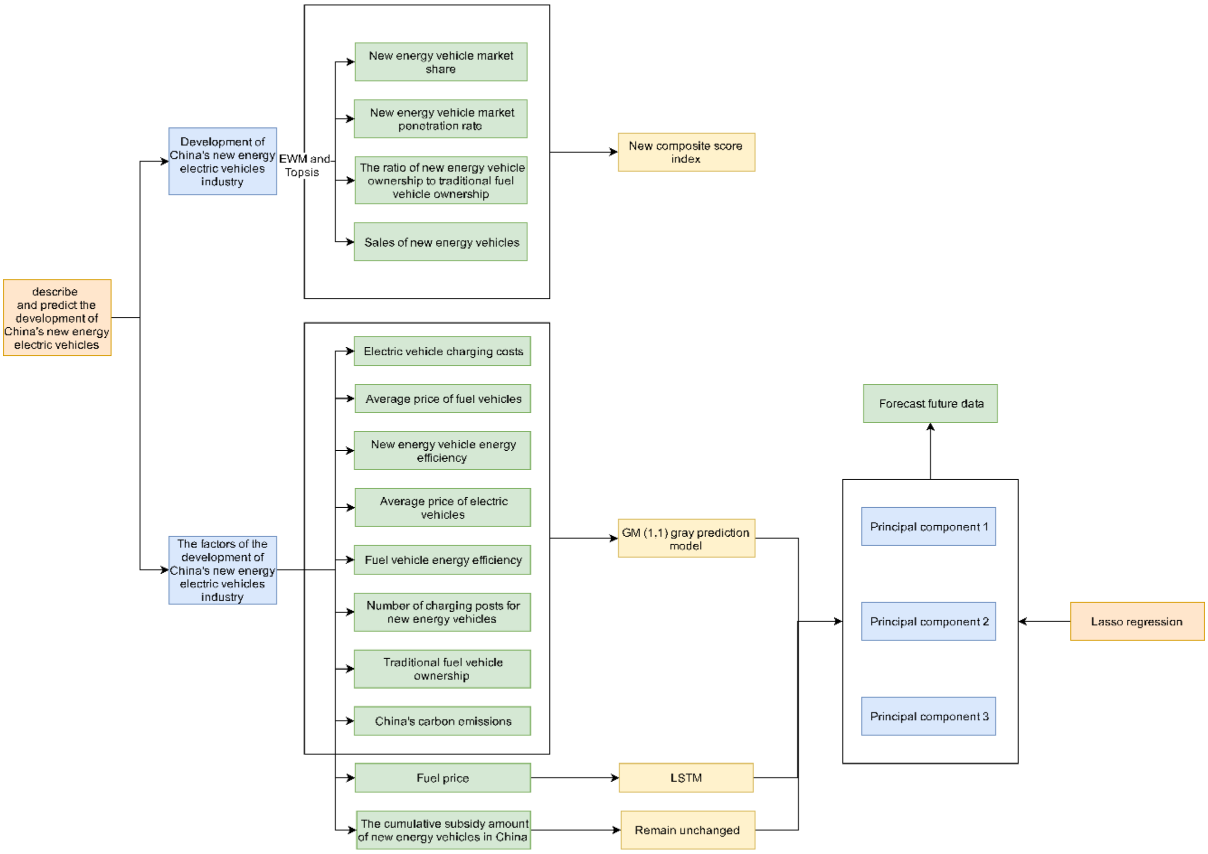

Figure 4. Influencing factors

2.3. Model evaluation indicators

We choose R^2 as the model evaluation metric.

$$R^2 = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2} = 1 - \frac{\Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2} \tag{1}$$

3. Model establishment and solution

3.1. Analysis and Prediction Model Construction of New Energy Industry Development

3.1.1. The entropy weight method (EWM)

Calculate the information entropy of each factor.

$$\begin{cases} E_j = -\sum_{i=1}^{n} p_{ij} \cdot \log(p_{ij}) \\ w_j = \dfrac{1 - E_j}{k - \sum_{j=1}^{k} E_j} \\ S_i^E = \sum_{j=1}^{k} w_j \cdot x_{ij} \end{cases} \tag{2}$$

Among them, $p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$ is the proportion of factor j in year i, E_j is the proportion of factor j in each year. We can calculate the weight and the weighted score. The weight data is shown in the table1.

Table 1. The results obtained by method EWM

index	information entropy (e)	information utility (d)	Weight (%)
Min-max standardization of SN	0.85	0.15	22.224
Min-max standardization of NMS	0.79	0.21	31.012
Min-max standardization of NMPR	0.858	0.142	20.946
Min-max standardization of the ratio of ROO	0.826	0.174	25.817

According to the results, the weight of the standardization of the market share of new energy vehicles is 31.012%. It shows that since 2013, the market share of new energy vehicles have changed greatly, and it is reflected in the trend of increasing year by year. In addition, the weight of the ratio of new energy vehicle ownership to traditional fuel vehicle ownership is 25.817 %, the weight of the standardization of new energy vehicle sales is 22.224% and the weight of the penetration rate of new energy vehicle market is 20.946%.

3.1.2. TOPSIS

(1) Determine the positive and negative ideal solutions:

For each factor, determine its maximum and minimum values, among them the positive ideal solution (A_{*j}) is the maximum value, negative ideal solution (A_{-j}) is the minimum value.

(2) Calculate similarity and the comprehensive score

$$\begin{cases} D_i^* = \sqrt{\sum_{j=1}^{k} (x_{ij} - A_{*j})^2} \\ D_i^- = \sqrt{\sum_{j=1}^{k} (x_{ij} - A_{-j})^2} \\ S_i^T = \dfrac{D_i^-}{D_i^- + D_i^*} \end{cases} \tag{3}$$

3.1.3. Calculate the final comprehensive score

$$S_i = \alpha \cdot S_i^T + (1-\alpha) \cdot S_i^E \tag{4}$$

Among them, S_i^T is the comprehensive score of every year calculated by TOPSIS method, S_i^E is the comprehensive score of every year calculated by the entropy weight method and α is the weight.

This method can comprehensively consider the calculation of weight and the measurement of similarity, and evaluate different indicators more comprehensively.

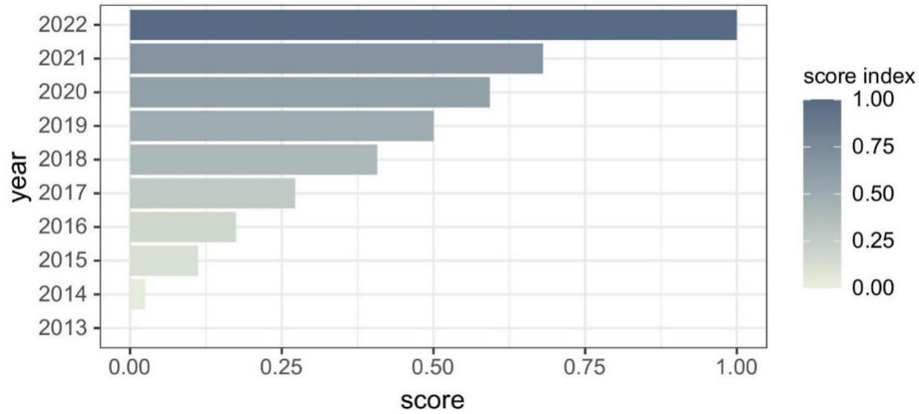

Figure 5. The composite score of each year

Figure 5 shows the comprehensive scores of each year, from which it can be seen that the highest score is in 2022.

3.1.4. Principal component analysis

First of all, due to many factors, there is a greater possibility of collinearity between different factors. In order to test whether there is collinearity among them, the correlation heat map shown in Figure 6.

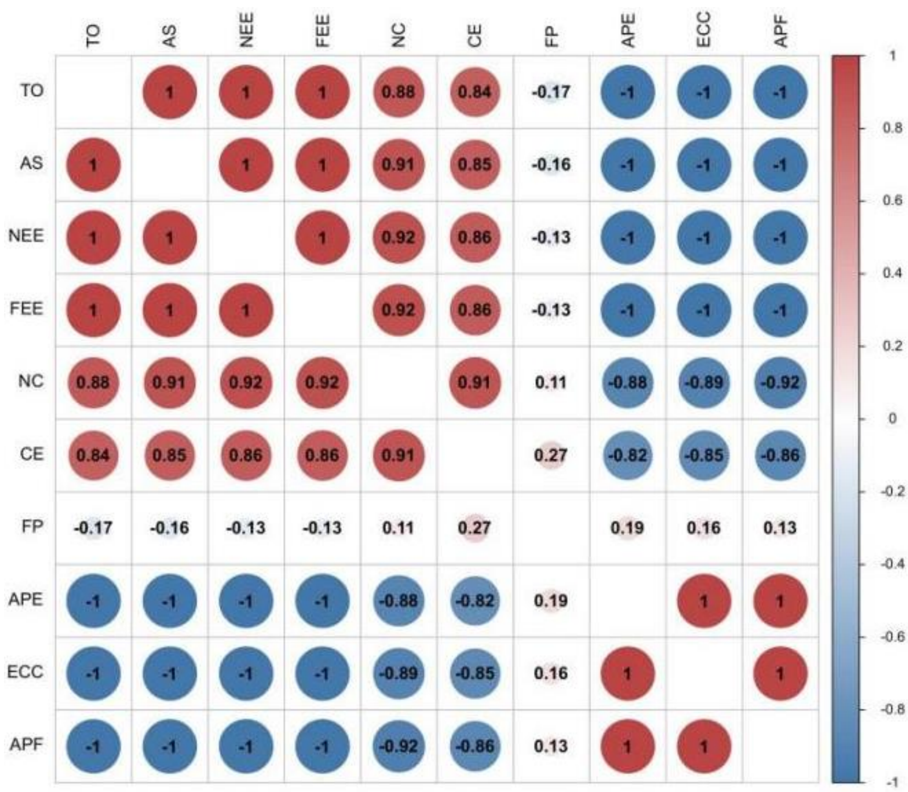

Figure 6. The correlation heat map

As can be seen from Figure 6, many factors are closely related to each other and the absolute value of the correlation coefficient is close to 1. Therefore, this paper concludes that there is a strong linear correlation within the data.

Before the principal component analysis, data were examined using the KMO test and Bartlett's test. The results are as follows:

Table 2. KMO test and Bartlett test

The value of KMO		0.74
Bartlett's test	Approximate chi-square	793.404
	df	45
	P	0.000***
Annotation:***, **, * represent 1 %, 5 %, 10 % significance level respectively.		

From table 2, $\mathrm{KMO} = 0.74 > 0.77$and $\mathrm{P} < 0.01$, data is suitable for using PCA algorithm to reduce dimension.

Principal component analysis (PCA) is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. It accomplishes this reduction by identifying directions, called principal components, along which the variation in the data is maximal[3]. This method can effectively reduce information redundancy, highlight the most important structure in the data and provide a more concise perspective for data analysis.

The results of principal component analysis are as table 3.

Table 3. Explanation of variance table

principal component	characteristic root	Explanation of variance (%)	Cumulative (%)
1	8.605	86.051	86.051
2	1.207	12.065	98.116
3	0.103	1.032	99.148
4	0.081	0.806	99.954

Considering the effect of dimension reduction, the critical condition is the variance interpretation rate of 1 %. Finally, this paper take the first three components to reduce the dimension of the data.

Table 4. Component Matrix Table

variable name	Component1	Component2	Component3
FEE	0.116	-0.031	-0.199
AS	0.116	-0.058	-0.284
APF	-0.116	0.031	0.199
NEE	0.116	-0.031	-0.199
ECC	-0.115	0.064	0.638
CE	0.103	0.319	0.952
NC	0.109	0.186	2.339
TO	0.115	-0.074	-0.828
APE	-0.115	0.092	0.817
FP	-0.011	0.817	-1.165

Table 4 shows the three principal component data. Next, we make a regression of the development score based on these three data.

3.1.5. Lasso regression

Lasso regression(Least Absolute Shrinkage and Selection Operator) is an improved method of linear regression by adding L1 norm as regularization term to the loss function. Its main characteristic is to compress some coefficients to zero in the process of fitting, so as to achieve the purpose of variable selection, simplify the model and avoid over-fitting. Lasso regression is very suitable for processing data sets with multicollinearity or the number of features is greater than the number of samples. By adjusting the size of the regularization parameter, Lasso regression can flexibly balance the deviation and variance, and effectively improve the generalization ability of the model.

In order to avoid overfitting, Lasso regression is performed on the data, as shown in Figure 7.

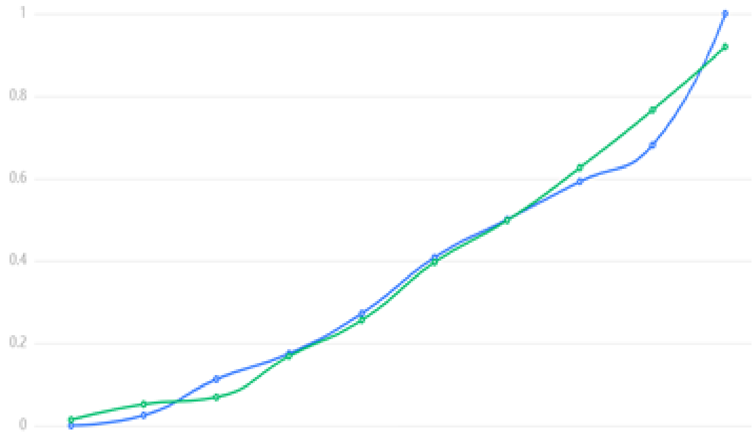

Figure 7. Comparison Chart

The regression results are shown in Table 5.

Table 5. Lasso regression results

Variable name	Normalization factor	Non-normalized coefficients	R^2
intercept	0.376	0.376	0.98
component 1	0.315	0.315	
component 2	0.032	0.032	
component 3	0.032	0.032	

After calculation, $R^2 = 0.98$ which reflects that the model has a high goodness of fit, avoids over-fitting and has a strong model interpretation ability. This paper analyzes the influencing factors affecting the development of new energy electric vehicles in China, and constructs a mathematical model reflecting the quantitative relationship.

3.2. Prediction of the Future Development of China's New Energy Industry

3.2.1. GM (1,1) gray prediction model

The principal component is composed of 10 factors. According to China 's policy[4-6], China has canceled subsidies for new energy electric vehicles in 2022. Therefore, it is believed that the cumulative subsidy amount in the next 10 years will remain unchanged. Through the Figure 8, we found that in addition to fuel prices, other indicators have a clear increasing or decreasing trend. The GM (1,1) grey prediction model is used for fitting. The fitting effect diagram is shown in Figure 8.

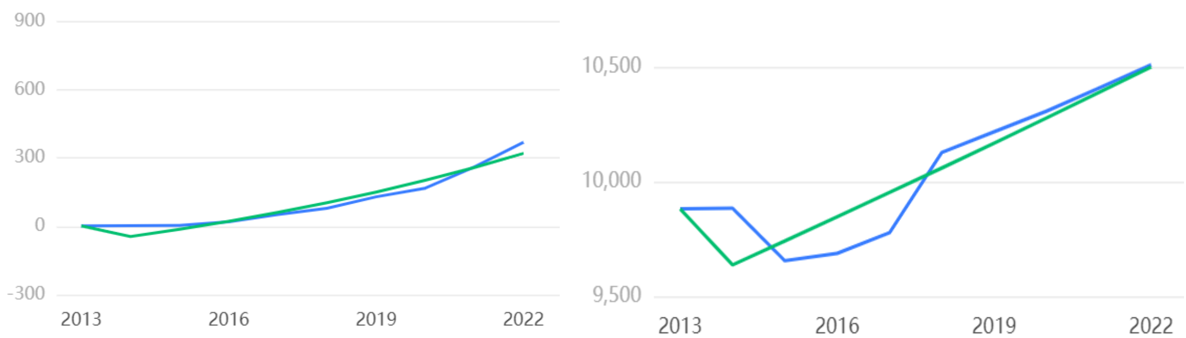

Figure 8. NC fitting diagram and CE fitting diagram

Taking the number of charging piles for new energy vehicles as an example, class ratio checking is carried out first. The test results are shown in Table 6.

Table 6. Results of class ratio checking

Year	original value	Ratio Value	Sequence value after translation conversion	The ratio value after translation conversion
2013	2.2	-	372.2	-
2014	3	0.733	373	0.998
2015	4.9	0.612	374.9	0.995
2016	20	0.245	390	0.961
2017	52	0.385	422	0.924
2018	80	0.65	450	0.938
2019	130	0.615	500	0.9
2020	168	0.774	538	0.929
2021	261.7	0.642	631.7	0.852
2022	370	0.707	740	0.854

If the the ratio value after translation conversion are all in the interval$\left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}}\right)$, we can know that this data is suitable for using grey prediction model to predict. n refers to the length of the data sequence. In this topic, $\mathrm{n} = 10$, so the interval is (0.834,1.199). Obviously, this data is suitable for using GM grey prediction model.

The grey model construction index is shown in table 7.

Table 7. grey model construction index

Development coefficient (a)	Grey action quantity (b)	A posterior difference ratio (c)
-0.094	275.294	0.051

According to the development coefficient and grey action quantity, the grey prediction model can be constructed. $\mathrm{c} = 0.051 < 0.35$, the model has high precision.

The average relative error of the fitting model is 0.853 %, meaning that the model fitting effect is good. The average relative error of other factors fitting model is less than 10 % which shows that the fitting effect is good and the model is effective.

3.2.2. LSTM

The Long Short-Term Memory (LSTM) algorithm is a type of recurrent neural network that is designed to handle sequential data[7-10]. It uses a combination of memory cells and gates to store and update information over time, allowing it to learn long-term dependencies in the data. Its structural diagram is shown in Figure 9.

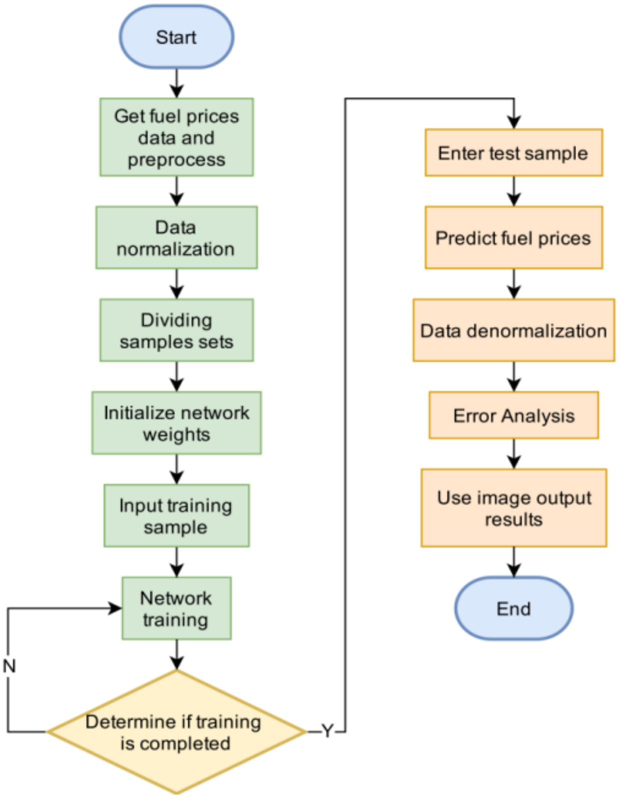

Figure 9. LSTM Algorithm Flowchart

It is found that the fuel price has good time series characteristics and certain stability. The average fuel data of each year in the past 20 years were collected and predicted using LSTM. The delay step is set to 2, the number of LSTM layers is 5, and the output is across a time point.

The predicted results of this test set are shown in Figure 10.

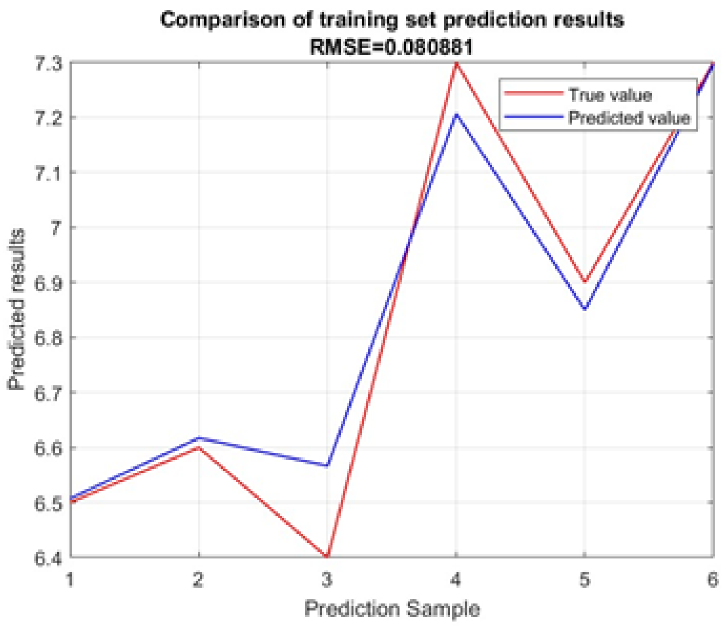

Figure 10. Comparison

Finally, in the training set, $R^2 = 0.89$, the mean square error of the test set is less than 0.1. We believe that the model is effective and can be used to predict fuel prices in the next decade.

The forecast results are shown in table 8.

Table 8. Forecast results of fuel prices

Year	fuel price	Year	fuel price
2024	8.2825860	2029	8.6531566
2025	8.4997539	2030	8.8231564
2026	8.9524547	2031	8.6465959
2027	8.3565165	2032	8.8649556
2028	7.9845638	2033	9.1265598

We use Z-score to standardize the data of each factor in the next ten years. At this time, we get the data of the influencing factors in the next ten years. Combined with the regression model established by Lasso regression, we can get new Composite score index of China 's new energy electric vehicle industry in the next ten years.

It can be found that the new composite score index is greater than 1, while the composite score index obtained in question 1 is less than 1. This is because the composite score index is the result of min-max standardization. Merge the scores from these twenty years and standardize to obtain standard data. The score data is shown in the table 9.

Table 9. New composite score index

Year	New composite score index	Standardized score
2024	1.051936293	0.873509939
2025	1.054954644	0.866919777
2026	1.066737908	0.885800947
2027	1.069999473	0.880205466
2028	1.071612238	0.882375062
2029	1.092277272	0.907008399
2030	1.115018623	0.925892426
2031	1.134873009	0.942379172
2032	1.167058376	0.969105352
2033	1.204263679	1

From table 9, it can be seen that the growth rate of China's new energy electric vehicle industry index is gradually slowing down. In the next decade, the comprehensive score of China's new energy industry will increase by about 20%.

4. Conclusion

The growth rate of China's new energy electric vehicle industry index is gradually slowing down. The reason may be that the country stops policy support for new energy vehicles in the later period. In the next decade, the comprehensive score of China's new energy industry will increase by about 20%. It can be seen that even without policy support, China's new energy industry can still develop well. This article collects a large amount of influencing factor data and uses reasonable models to measure and predict the development status of new energy vehicles in China. However, in the development indicators, the indicators collected in this article are still relatively few and not comprehensive enough, and the measurement of development level may have a certain degree of one-sidedness. Further work in the future can consider collecting more development indicators to better measure the development status of the new energy vehicle industry.

References

[1] Feng Meng. Research on the Development Trend and Strategy of Energy Saving and New Energy Vehicles in China [J]. Automotive Test Reports, 2023 (2): 60 - 62

[2] Huifang Tian. Development Trends, Challenges and Implications of the Global New Energy Vehicle Industry [J]. China Development Monitor, 2023 (6): 12 - 17.

[3] Ringnér M. What is principal component analysis? [J]. Nature biotechnology, 2008, 26 (3): 303 - 304.

[4] Ministry of Finance, Ministry of Industry and Information Technology, Ministry of Science and Technology, Development and Reform Commission. Notice on the financial subsidy policy for the promotion and application of new energy vehicles in 2022: Financial Construction [2021] NO.466 [EB/OL]. (2021-12-31) [2023-11-24].

[5] Yu Shi, Hua Zhang, Zhihan Yu. Energy conservation and emission reduction benefit analysis and environmental impact assessment of the whole life cycle of electric vehicles. [J]. Resources and Industries, 2021, 23 (2): 100 – 109.

[6] Huijie Z, Jie D, Xihui C, et al. Understanding innovation of new energy industry: Observing development trend and evolution of hydrogen fuel cell based on patent mining [J]. International Journal of Hydrogen Energy, 2024, 52 (PA): 548 - 560.

[7] Zhang J. Research on the Development of Supply Chain Finance Driving the Development of New Energy Industry Chain—Taking BYD Supply Chain Finance as an Example [J]. Financial Engineering and Risk Management, 2023, 6 (10):

[8] Deng P. Analysis on the Integrated Development of New Energy Industry and Vocational Education [J]. Advances in Vocational and Technical Education, 2023, 5 (8):

[9] Tong Y, Ziwei Y, Chen X. Research on China's fiscal and taxation policy of new energy vehicle industry technological innovation [J]. Economic research - Ekonomska istraživanja, 2023, 36 (2):

[10] Zehui G, Shujie S, Yishan W, et al. Impact of New Energy Vehicle Development on China's Crude Oil Imports: An Empirical Analysis [J]. World Electric Vehicle Journal, 2023, 14 (2): 46 - 46.