Research on the Development Trend of New Energy Vehicles in China Based on Pearson Correlation Analysis and Polynomial Fitting

Jiahao Wang *, Gonghao Wei, Xun He

Aeronautics and Astronautics, Shenyang Aerospace University, Shenyang, China, 110136

* Corresponding Author Email: wjh2182790854@outlook.com

Abstract. The purpose of this paper is to analyse multiple factors that influence the development of new energy vehicles in China and to predict the future development trend. First, this paper selects 12 representative factors, including total vehicle sales, new energy vehicle penetration, pollutant emissions, economic indicators, etc., and uses Pearson correlation analysis to verify their relationship with new energy vehicle sales. The results show that nine of these indicators are significantly correlated with new energy vehicle sales, which provides a basis for subsequent modelling. Next, establishes a quadratic polynomial model using the polynomial function fitting method to predict the sales volume of new energy electric vehicles in China in the next 10 years. The model evaluation shows a good fit, with little difference between the predicted and actual values and a coefficient of determination close to 1. Finally, combining the model prediction results and the analysis of the influencing factors, this paper describe the trend of the future development of China's new-energy electric vehicle industry, including the influences of governmental support, technological advancement, environmental protection awareness, market competition, and infrastructure construction, and look ahead to the development of new-energy automobile market.

Keywords: New energy vehicle development; Pearson correlation analysis; polynomial fitting; sales forecast; industry development trend.

1. Introduction

This paper aims to analyze various factors influencing the development of new energy vehicles in China and provide predictions for their future development trends. The study begins by selecting 12 representative factors, such as total vehicle sales, new energy vehicle penetration, pollutant emissions, and economic indicators, and conducting Pearson correlation analysis to examine their relationships with new energy vehicle sales. The analysis reveals that nine of these indicators exhibit significant correlations with new energy vehicle sales [1], establishing a foundation for subsequent modeling.

To predict the sales volume of new energy electric vehicles in China over the next 10 years, a quadratic polynomial model is constructed using the polynomial function fitting method. The evaluation of the model demonstrates a strong fitting effect, as the predicted values closely align with the actual values, with a coefficient of determination close to 1. By combining the model's predictions with the analysis of influencing factors, this study outlines the future development trends of China's new-energy electric vehicle industry [2].

The analysis considers various aspects, including governmental support, technological advancements, environmental protection awareness, market competition, and infrastructure construction. These factors collectively shape the development prospects of the new-energy automobile market in China.

This study's innovative contributions lie in the comprehensive analysis of multiple factors influencing new energy vehicle sales, the establishment of a robust predictive model, and the exploration of the future development trends in China's new-energy electric vehicle industry. The data used in this research are sourced from reliable and authoritative databases, ensuring the validity and accuracy of the results. The findings hold significant implications for policymakers, industry stakeholders, and researchers, providing insights into the potential of the new-energy automobile market in China and guiding future decision-making processes.

2. China's New Energy Vehicle Development Trend Forecast

2.1. Selection of representative factors

In order to be able to measure the development of new energy vehicles, this paper regards the sales of new energy vehicles in previous years as data that can represent the development of new energy vehicles. As can be seen from Figure 1, the sales of new energy vehicles show a yearly increasing trend from 2012 to 2022.

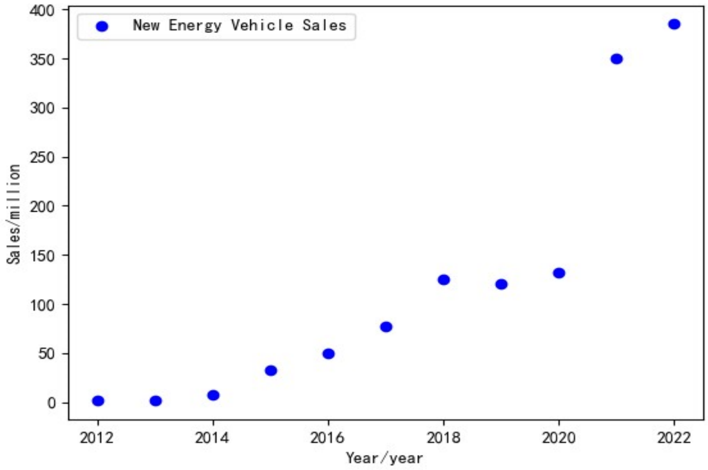

Figure 1 Change in China's new energy vehicle sales by year

By reviewing relevant information, this paper selects some representative factors affecting the development of new energy vehicles, including: total vehicle sales, penetration rate of new energy vehicles, China's annual sulfur dioxide emissions, nitrogen oxides emissions, soot emissions, particulate emissions, gross domestic product (GDP), GDP per capita, China's scientific research investment in electric power (new energy), the number of new energy policies, the number of public recharging pile holdings(10,000), China's new energy vehicle patent applications, and 12 other indicators. The changes in each indicator over the years are shown in Figure 2.

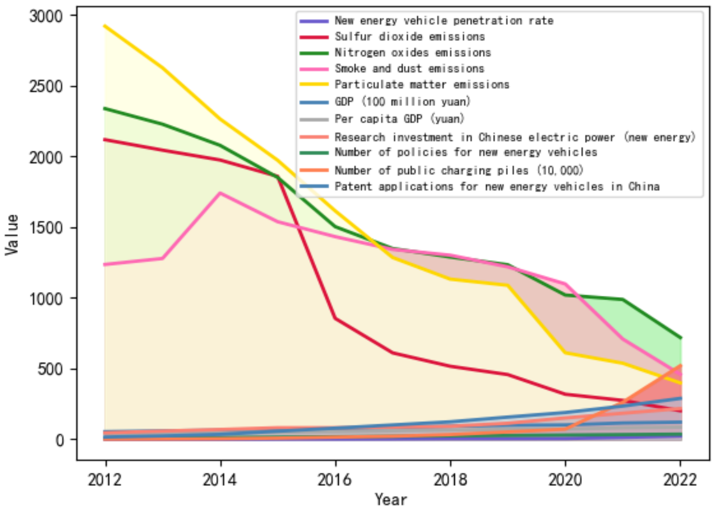

Figure 2 Changes in indicators by year

2.2. linear relationship test

In order to test whether these indicators are representative factors that can influence the development of new energy vehicles, a correlation analysis between these indicators is needed. Since the variable types of all these indicators are scalar (also known as fixed-distance) variables, this paper uses Pearson correlation analysis [3].

The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{1}$$

The above equation defines the overall correlation coefficient, which is often represented by a lowercase Greek letter ρ. Estimating the covariance and standard deviation of the sample yields the Pearson's correlation coefficient, which is often represented by the lowercase letter r.

$$r = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}} \tag{2}$$

r can also be estimated from the mean of the standardized scores of $(x_i + y_i)$ sample points to obtain an expression equivalent to the above equation:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma_X}\right)\left(\frac{Y_i-\overline{Y}}{\sigma_Y}\right) \tag{3}$$

where $\frac{X_i-\bar{X}}{\sigma_X}$, $\bar{X}$ and σ_X are the standardized score, sample mean and sample standard deviation of X_i sample, respectively.

Before using this method for analysis, first, it is necessary to verify whether there is a linear relationship between these variables. In this paper, the linear test shows that the indicators have a linear relationship; secondly, it is necessary to ensure that these variables do not have obvious outliers, because the Pearson correlation coefficient is susceptible to the influence of outliers. Before this paper, the collected data have been pre-processed, including the treatment of data outliers, so there is no need to consider the impact of data outliers on the results in this step.

2.3. Evaluation of Shapiro-Wilk test results

After completing the linear relationship test and outliers' treatment of the data, it is necessary to verify that these data conform to a normal distribution, which is an essential step in performing the Pearson correlation test. For this reason, this paper also uses SPSS to test whether these data show a deviation or conformity from normality, which is detected in this paper using the Shapiro-Wilk method [4]. The results of the test are shown in Table 1:

Table 1 Results of Shapiro-Wilk test

	Kolmogolov-Sminov [a]			Shapiro-Wilk		
	statistics	DOF	Distinctiveness	statistics	DOF	Distinctiveness
Penetration rate of new energy vehicle	0.314	11	0.003	0.003	11	0.001
total vehicle sales	0.154	11	0.200	0.200	11	0.406
CO2 emission	0.242	11	0.071	0.071	11	0.010
Nitrogen oxide emissions	0.162	11	0.200	0.200	11	0.530
Soot emission	0.233	11	0.097	0.970	11	0.290
Particulate matter emission	0.142	11	0.200	0.200	11	0.552
GDP(100million yuan)	0.127	11	0.200	0.200	11	0.692
GDP per capita(RMB)	0.128	11	0.200	0.200	11	0.695
China's investment in new energy research	0.248	11	0.057	0.057	11	0.106
Number of new energy vehicle policies	0.096	11	0.200	0.200	11	0.796
Number of public charging piles(10,000 units)	0.364	11	0	0	11	0
The number of patent applications for new energy vehicles in China	0.127	11	0.200	0.200	11	0.437

If the significance value of the indicator is greater than 0.05, then the H0 hypothesis is accepted, that is, the data of this variable is considered to obey the normal distribution. From the results of the above analysis, it can be judged that the nine indicators of the penetration rate of new energy vehicles, China's annual emissions of nitrogen oxides (NOx), soot and dust, particulate emissions, gross domestic product (GDP), GDP per capita, China's electric power (new energy) scientific research investment, the number of new energy policies, and China's new energy automobile patent applications conform to a normal distribution.

2.4. Pearson correlation test

After completing the above preparations, the correlation test can be started on these processed data. Again, SPSS was used for this process. Some of the data resulting from the processing is shown in the figure below and the rest of the data is placed in the annexure. The results of the correlation test are shown in Table 2.

Table 2 Pearson correlation test results

		Sales of new energy Pearson correlation vehicles	Penetration rate of new energy vehicle	Nitrogen oxide emissions	Soot emission	Particulate matter emission	GDP (100 million yuan)	GDP per capita (RMB)	China's investment in new energy research
Sales of new energy vehicles	Pearson correlation	1	0.964	0.839	0.906	0.831	0.924	0.927	0.958
	Sig.(Twin Tails)	0	0	0.001	0	0.002	0	0	0
	Number of cases	11	11	11	11	11	11	11	11

Through the results in the figure, it can be seen that the Pearson correlation coefficient of new energy vehicle sales and new energy vehicle penetration rate is 0.964, $P=0.000<0.05$. It is then concluded that there is a significant positive correlation between new energy vehicle sales and new energy vehicle penetration rate. Then the correlation of each index is obtained by drawing the correlation heat map, as shown in Figure 3.

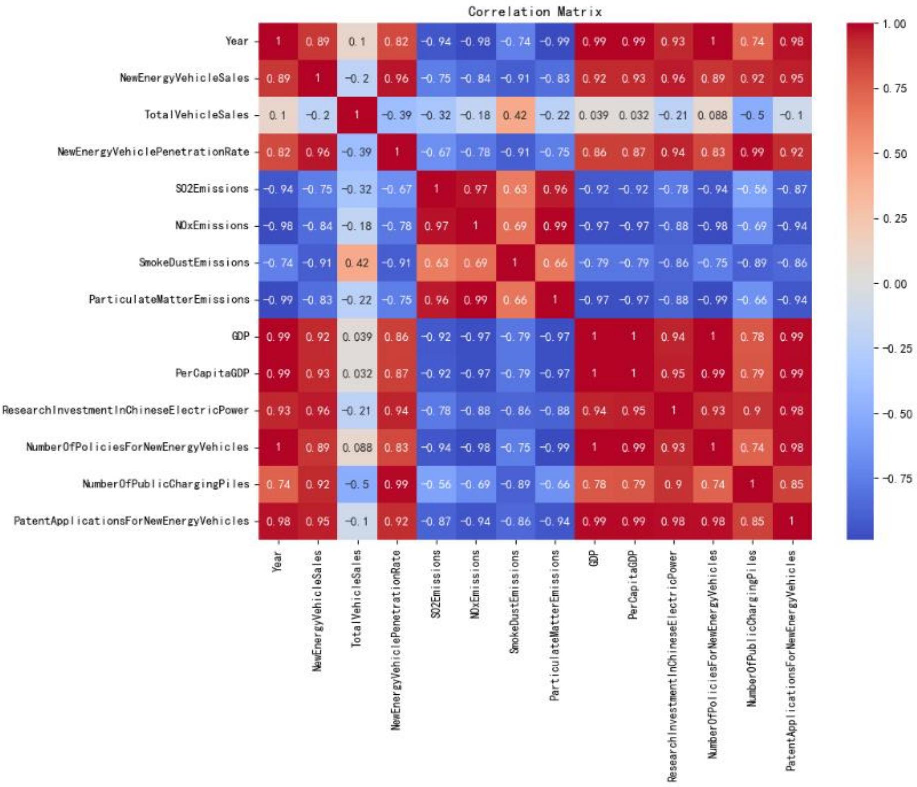

Figure 3 Indicator correlation heat map

3. Predictive modeling

3.1. Model Basics

Polynomial function fitting [5]: assuming that the given data are generated by M polynomial functions, choose the M polynomial function that is most likely to generate these data, i.e., choose a function among the M polynomial functions that has good predictive power for known as well as unknown data.

Least Squares (also known as Least Square Method) [6]: is a mathematical optimization technique. It finds the best function match for the data by minimizing the sum of squares of the errors. Using the least squares method makes it easy to find unknown data and minimize the sum of squares of the errors between these found data and the actual data.

First of all, this paper will be preprocessed data, which data include factors on the development of new energy vehicles and the year, for the prediction of the object and the year of the data, assuming that contains m points, called the sample point, for:

$$\{(x_1,y_1)(x_2,y_2)...(x_m,y_m)\} \tag{4}$$

That way a point in this point can be represented as:

$$(x_i,y_i), i=1,2,3,...,m \tag{5}$$

3.2. Polynomial order determination

According to the polynomial function:

$$\hat{y}=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+...+a_{n-1}x+a_n \tag{6}$$

To determine the most appropriate fitting order n, in this paper, the polynomial fitting results by comparing different orders are plotted as an image as shown in Figure 4:

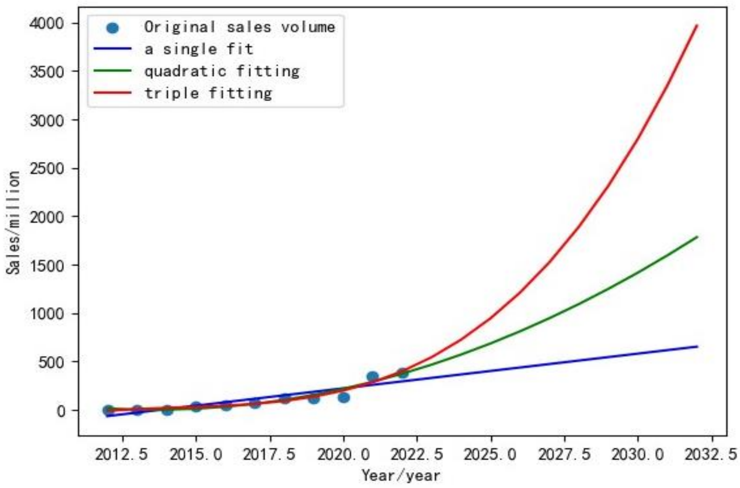

Figure 4 Comparison of model fits of different orders

As can be clearly seen from the above figure, the primary fit is underfitted, while the tertiary fit is clearly overfitted, so in this paper a quadratic polynomial model is chosen for the fit.

3.3. Solving for polynomial fit coefficients

According to the polynomial function:

$$\hat{y}=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+...+a_{n-1}x+a_n \tag{7}$$

It can be seen that the nth degree polynomial has $n+1$ unknown fit coefficients from a_0 to a_n. All that has to be done is to find these best $n+1$ fit coefficients.

Substituting the value of the horizontal coordinate of the sample point, x_i, into the assumed polynomial, $\hat{y}_i$, yields the nth degree polynomial in the vertical coordinate at the horizontal coordinate of the given sample point as:

$$\hat{y}_i = a_0 x_i^n + a_1 x_i^{n-1} + a_2 x_i^{n-2} + \ldots + a_{n-1} x_i + a_n \tag{8}$$

An indicator is needed to judge how much all the $\hat{y}_i$'s differ from the y_i's in the sample points, so the sum of the squares of the errors/residuals is used to characterize them:

$$\begin{aligned} \epsilon &= \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 \\ &= \sum_{i=1}^{m} [(a_0 x_i^n + a_1 x_i^{n-1} + a_2 x_i^{n-2} + \ldots + a_{n-1} x_i + a_n) - y_i]^2 \end{aligned} \tag{9}$$

Returning to the question of finding the best fit coefficients, if there is a set of fit coefficients ϵ that minimizes, then that set of fit coefficients can be considered the best.

In the following, we first take the following n partial derivatives for a separately and make each partial derivative 0 (making the partial derivative 0 for the extremum of the multivariate function), i.e.:

$$\begin{cases} \dfrac{\partial \epsilon}{\partial a_0} = 0 \\ \dfrac{\partial \epsilon}{\partial a_1} = 0 \\ \ldots \\ \dfrac{\partial \epsilon}{\partial a_n} = 0 \end{cases} \tag{10}$$

where a term $\dfrac{\partial \epsilon}{\partial a_j} = 0, j = 0, 1, 2, \ldots, n+1$ in the above system of equations can be expanded and written as:

$$\frac{\partial \epsilon}{\partial a_j} = \sum_{i=1}^{m} 2 x_i^{n-j} [(a_0 x_i^n + a_1 x_i^{n-1} + a_2 x_i^{n-2} + \ldots + a_{n-1} x_i + a_n) - y_i] = 0 \tag{11}$$

Similarly, the other terms in the above system of equations can be expanded into equations containing the fitted coefficients a_0 to a_n. Then the $n+1$ equations in the system of equations can be solved for these $n+1$ fitted coefficients.

3.4. Model predictions

Based on the available data, this paper utilizes Python to build a second-order polynomial, which is fitted to obtain the fitting results in Figure 5:

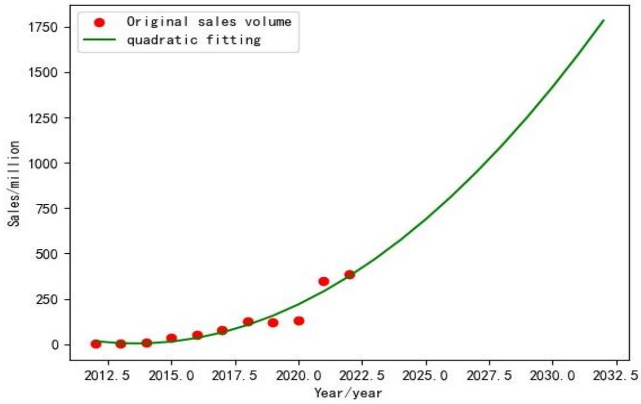

Figure 5 Model fitting results

The fitted polynomial equation is found to be:

$$y = 5.254x^2 - 2.116e+04x + 2.13e+07 \tag{12}$$

From the fitting result graph, it is initially judged that the fitting effect is better, so the prediction of China's new energy electric vehicle sales volume in the latter 10 years is carried out, and the results are shown in Table 3.

Table 3 Forecast results of sales volume of new energy electric vehicles

Year/year	Sales/million
2023	468.1763636
2024	572.2890909
2025	686.9102098
2026	812.0397203
2027	947.6776224
2028	1093.823916
2029	1250.478601
2030	1417.641678
2031	1595.313147
2032	1783.493007

4. Model Evaluation

4.1. Residual Analysis

First, in this paper, residual plots, which are the differences between predicted and observed values, are plotted in Python to check for patterns, heteroskedasticity, or outliers in the model [7].

The residual plot is shown in Figure 6.

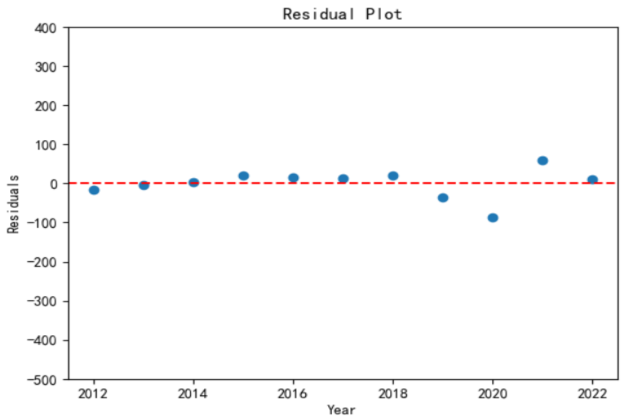

Figure 6 Plot of fitted residuals

From Figure 6, it can be seen that the model fits well and there is little variability in the data.

4.2. Goodness-of-fit test

In order to better evaluate the model fitting effect, this paper again through the calculation of the decision factor, which refers to an important indicator to judge the goodness of fit, the value is between 0 and 1, the larger indicates that the fitting effect is better [8].

The calculation formula is:

$$R^2 = \sum(\hat{y}_i - \bar{y})^2 / \sum(y_i - \bar{y})^2 = 1 - \sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y})^2 \tag{13}$$

where $\hat{y}_i$ denotes the model fit prediction, y_i denotes the true value of the data, and $\bar{y}$ denotes the average of the true values.

The data is brought in and calculated to obtain:

Decision factor: R^2=0.9223598409383409

Since R^2 is close to 1, it can be seen that the model fits better and has good explanatory power.

5. Development analysis and result

5.1. Development analysis

From the existing data using Python to draw various factors on the development of China's new energy electric vehicle industry changes in the area of the map, due to space reasons this paper only shows Figure 7.

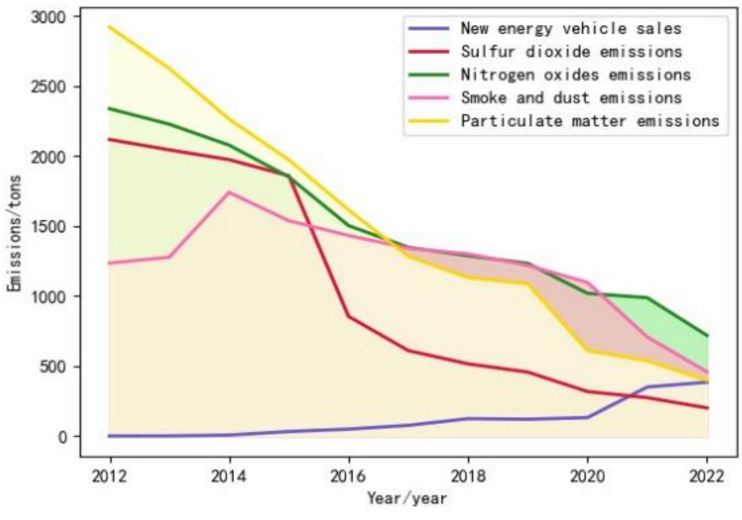

Figure 7 Change in sales of new energy vehicles for environmental pollutants

It is clear from Figure 7 that as the sales of new energy electric vehicles in China increase, the emissions of pollutants are decreasing.

From each figure, it is found that the penetration rate of new energy vehicles, China's annual NOx emissions, soot emissions, particulate emissions, GDP, GDP per capita, China's scientific research investment in electric power (new energy), the number of new energy policies, and China's new energy automobile patent applications all have different beneficial effects on the future development of China's new energy and electric vehicle industry [9].

5.2. Analyzed result

The development of new energy electric vehicles in China has been on a high growth trend over the past few years and is expected to continue to grow rapidly and mature over the next 10 years [10].

This paper combines the results of the forecast of China's new energy electric vehicle sales in the next ten years can summarize its development into several points. As shown in Figure 8.

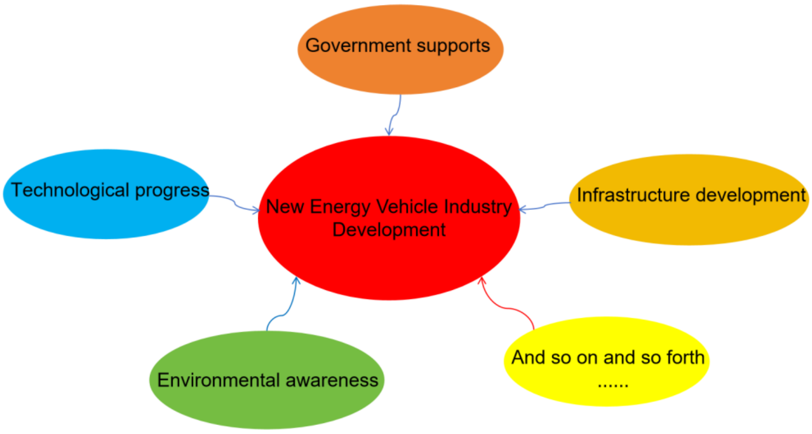

Figure 8 Description of the future development of China's new energy electric vehicle industry

(1) Government support: The Chinese government has been committed to promoting the development of new energy vehicles and encouraging consumers to purchase new energy vehicles through various incentives, such as subsidies for car purchases, free parking, free highway access, etc., which will continue to promote the development of the new energy vehicle market.

(2) Technological progress: with the continuous progress of science and technology, the range of new energy electric vehicles and charging technology will be further improved, which will enhance the convenience and popularity of new energy vehicles.

(3) Environmental awareness: As the global awareness of environmental protection increases, more and more consumers will choose environmentally friendly vehicles, and the new energy vehicle market will continue to expand under this trend.

(4) Market competition: With the entry of major domestic and foreign automobile manufacturers, the competition in the new energy vehicle market will be further intensified, and the competition among automobile brands will drive the emergence of new products.

(5) Infrastructure construction: In the next 10 years, it is expected that China will further increase the construction of charging infrastructure for new energy vehicles to promote the popularization of new energy vehicles.

6. Conclusions

The results of this paper show that twelve representative factors are closely related to the sales of new energy vehicles, which provides a basis for the subsequent model establishment.

The quadratic polynomial model established based on polynomial function fitting has a better effect on predicting the sales of new energy electric vehicles in China in the next ten years, which verifies the effectiveness of the model.

Combined with the model prediction results and the analysis of influencing factors, this paper describes the trend of the future development of China's new energy electric vehicle industry, pointing out the continuous influence of government support, technological progress, environmental awareness, market competition and infrastructure construction. The results of this research provide an important reference basis for the policy formulation and future development planning of China's new energy vehicle industry, and also provide useful insights for the development of related industry enterprises.

In the future, China's new energy electric vehicle market will continue to grow, bringing huge economic and social benefits. Driven by various factors such as policy, technology and market demand, the development of new energy vehicles will usher in more opportunities and challenges and is expected to become one of the important engines for the development of China's automotive industry. China’s new energy electric vehicle market will continue to grow in the next 10 years and driven by a variety of factors such as policy, technology and market demand, the development of new energy vehicles will usher in more opportunities and challenges and is expected to become one of the important engines for the development of China's automotive industry.

References

[1] WANG Jinzhao, YIN Jianyuan. Status quo, problems and trends in the development of new energy vehicles in China[J]. Southern Agricultural Machinery,2019,50(20):1.

[2] Zhang Changling. Analysis of China's new energy vehicle market mechanism and development trend outlook[J]. China Economic Review, 2021, (Z1):104-111.

[3] DONG Libeng, NIE Qinghao, SUN Xiaokun, et al. Analysis of the influence of shield tunneling parameters on surface settlement based on Pearson's correlation coefficient method[J]. Construction Technology (in Chinese and English),2024,53(01):116-123.

[4] Ouyang Q, Zheng B, Wu Shengpan, et al. Function and potential mechanism of dicarbonyl/L-xylulose reductase in the development of renal cell carcinoma[J]. Journal of Clinical Urology,2024,39(04):314-320.

[5] Li Jianwen,Wang Eibu. Polynomial function fitting for speech synthesis of Chinese tones[J]. Journal of Xi'an University of Science and Technology,2021,41(03):506-515.

[6] HUANG Bo,HE Jian. Fuzzy soft-switching control technology of hybrid vehicle braking mechanism based on predictive feedback[J]. Journal of Langfang Normal College (Natural Science Edition),2024,24(01):47-53.

[7] Jiang Zhongqing, Zhang Yue, Zhou Yi, et al. Abnormal runoff determination method based on signal processing technology and residual analysis [J]. Jilin Water Resources, 2024, (04): 37-41.

[8] Song Jiawei, Xun Baoqian, Qin Tao, et al.Rock blasting lumpiness prediction based on IGWO-CatBoost model[J].Blasting Equipment,2024,53(02):56-64.

[9] Wei Guo, Jiang Hongmei, Wei Keding, et al. Factors Influencing the Development of China's New Energy Vehicle Industry and Suggestions for Countermeasures [J]. Equipment Manufacturing Technology, 2023, (12): 136-139.

[10] Gong Mengze. China Electric Vehicle Hundred People Conference held multiple ministries to speak out on expanding the development advantages of new energy vehicles[N]. Securities Daily, March 18, 2024 (B02)