Intelligent Analysis and Modeling of the Development Trend of New Energy Electric Vehicles in China Based on Machine Learning

*Haojiang Xi**

State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, China

ABSTRACT

Since 2011, China has actively promoted new energy electric vehicles with favorable policies, marking significant industry growth akin to the "China High Speed Rail." Using data-driven approaches like machine learning and probability statistics, we crafted models for three-level indicator weights using Analytic Hierarchy Process, correlation-based multiple regression, and ARIMA-based future predictions.

In response to Question 1, we gathered over 80 development-related datasets, cleaning and visualizing data for analysis. Complex correlation analysis was mitigated by categorizing data into three indicator levels, focusing on annual sales, policies, tech advancement, market economy, infrastructure, resources, and corporate capabilities. Weighted models revealed policy, infrastructure, and tech as top influencers, impacting new energy vehicle development.

Question 2 required extensive data collection for a 10-year forecast. Historical industry data including sales, production, policy shifts, and technological advancements was pivotal. Six indicators from Question 1 and sales volumes were integrated using TOPSIS, with predictive modeling employing correlation coefficients and machine learning models like BILSTM and SARIMA.

For Question 3, data collection and mathematical models assessed new energy vehicles' impact on traditional counterparts. Analysis indicated a negative impact, with some gasoline vehicle sales shifting but limited effect on factors like oil prices and emissions, signaling potential future trade-offs.

Question 4 involved analyzing foreign policies affecting China's new energy vehicle exports. Policy resistance variables assessed policy impacts on export volumes, with anomaly detection and ARIMA models predicting future impacts.

To gauge differences between new energy and traditional vehicles for a 1 million population city (Question 5), calculations indicated substantial resource and environmental savings from new energy vehicles.

Lastly (Question 6), leveraging these findings, an open letter advocating new energy vehicles' benefits and global contributions to the electric vehicle industry was crafted.

Keywords: Data-driven Modeling, Analytic Hierarchy Process, Multiple Regression, SARIMA, Impact Analysis, Policy Evaluation, Environmental Benefits

1 INTRODUCTION

The automobile industry, as one of the leading industries in the national economy, plays an important role in economic growth [1-2]. Due to the severe pollution caused by traditional automobile exhaust emissions, a large amount of environmental pollution has been caused; At the same time, traditional cars heavily rely on fossil fuels [3]. With the continuous increase in car ownership, the demand for fossil fuels is increasing, which exacerbates the contradiction between energy production and consumption in China [4-5]. At present, China's economy is transitioning from a stage of high-speed growth to a stage of high-quality development, and is in a critical period of optimizing economic structure and transforming development mode [6]. The high pollution and high energy consumption of traditional fuel vehicles are no longer suitable for the current economic development requirements [7]. As an energy saving and environmentally friendly product, new energy vehicles [8], with their dual advantages in energy conservation and alleviating environmental pressure, have become a major trend in the development of the global automotive industry and a priority choice for the development of the automotive industry around the world. However, as a new type of automotive product, new energy vehicles [9], There are significant differences in technology compared to traditional fuel vehicles, which affect consumers' acceptance of new energy vehicle products. China is currently in a transitional stage from policy driven to market driven in the new energy vehicle industry [10].

2 The Description of the Problem

2.1 How do we approximate the whole course of ?

Question 1 requires us to analyze the main factors that affect the development of new energy electric vehicles in China. As is well known, there are many factors that affect the development of new energy electric vehicles, and the interaction between them is complex. However, the required data is not provided in the title, so we need to collect it ourselves. Therefore, through literature review, we have established three levels of indicators to comprehensively evaluate the impact of many factors on the development of new energy electric vehicles in China. We use the Analytic Hierarchy Process to solve for the weights of the third level indicators. After obtaining the results, we use the weighted sum of the weights to reduce the influencing factors to six second level indicators. Then, we use a multiple regression model based on correlation coefficient, mutual information, and random forest model fusion to analyze the factors affecting the development of new energy vehicles.

Question 2 requires collecting relevant data on the development of China's new energy electric vehicle industry, and establishing a mathematical model to describe and predict the development trend of China's new energy electric vehicles in the next 10 years. This issue is similar to Question 1, but the focus of the problem may be more refined, requiring more data collection and more accurate modeling to predict the development situation in the next 10 years. We can collect historical development data of China's new energy electric vehicle industry, including but not limited to: sales volume, production volume, market share, policy changes, technological progress, charging facility construction, environmental protection policies, etc. The data may involve multiple sources, such as publicly available government data, industry reports, automotive manufacturer data, market research reports, etc. Data preprocessing includes handling missing values, outliers, data format conversion, and other operations to ensure data quality. Process time series data to ensure continuity and consistency in time intervals. Multiple mathematical models can be used to describe and predict the development of new energy electric vehicles in China in the next 10 years, such as time series models (such as ARIMA, SARIMA), machine learning models (such as LSTM, GRU, random forest, gradient boosting tree, etc.), regression models (linear regression, polynomial regression, etc.), etc.

Choosing a suitable model requires consideration of factors such as data characteristics, time series properties, and prediction objectives. Finally, conduct result analysis and interpretation, analyze the predicted results of the model, and evaluate the reliability and rationality of the model. Explain the prediction results of the model and explore the key factors that affect the future development of new energy electric vehicles, such as policy changes, technological progress, market demand, etc.

This question aims to explore the impact of new energy electric vehicles on the global traditional energy vehicle industry by collecting data and establishing mathematical models. To achieve this goal, the primary task is to collect multidimensional data on the new energy electric vehicle and traditional energy vehicle markets, including sales volume, market share, growth trends, and other information. At the same time, it is necessary to collect data on the traditional energy market, such as fuel prices, carbon emissions, crude oil production, etc. After data collection, establishing a mathematical model is one of the core steps, which can be explored through methods such as time series analysis, regression analysis, and causal analysis to explore the mutual influence and underlying mechanisms of the two types of cars in the market. The data analysis stage focuses on exploring the correlation, trends, and degree of influence between new energy electric vehicles and traditional energy vehicle markets. Finally, based on the results of the mathematical model and the findings of data analysis, targeted conclusions and recommendations are proposed, including recommendations for policy formulation and market strategies, to address the impact of new energy vehicles on the traditional energy vehicle industry. The above analysis steps will help to comprehensively understand and evaluate the impact of new energy electric vehicles on the traditional energy vehicle industry, providing important reference and decision-making support for future development.

This question aims to explore the impact of new energy electric vehicles on the global traditional energy vehicle industry by collecting data and establishing mathematical models. To achieve this goal, the primary task is to collect multidimensional data on the new energy electric vehicle and traditional energy vehicle markets, including sales volume, market share, growth trends, and other information. At the same time, it is necessary to collect data on the traditional energy market, such as fuel prices, carbon emissions, crude oil production, etc. After data collection, establishing a mathematical model is one of the core steps, which can be explored through methods such as time series analysis, regression analysis, and causal analysis to explore the mutual influence and underlying mechanisms of the two types of cars in the market. The data analysis stage focuses on exploring the correlation, trends, and degree of influence between new energy electric vehicles and traditional energy vehicle markets. Finally, based on the results of the mathematical model and the findings of data analysis, targeted conclusions and recommendations are proposed, including recommendations for policy formulation and market strategies, to address the impact of new energy vehicles on the traditional energy vehicle industry. The above analysis steps will help to comprehensively understand and evaluate the impact of new energy electric vehicles on the traditional energy vehicle industry, providing important reference and decision-making support for future development.

For Question 5, we are required to analyze the impact of electrification of urban new energy electric vehicles (including electric buses) on the ecological environment. We assume that the urban population is 1 million. Firstly, we calculate the ownership of new energy vehicles and traditional energy vehicles. Then, we use an energy conversion model to calculate the relative energy savings. We need to convert the energy consumption of electric and gasoline vehicles into Standard Coal Equivalent (SCE). The standard coal conversion coefficient represents the amount of standard coal equivalent to a unit mass of fuel or electricity. After knowing the number of new energy vehicles, comparing the power consumption of new energy vehicles with the fuel consumption of equivalent traditional vehicles can reveal the impact of new energy vehicles on resource conservation and reducing harmful gas emissions such as CO, HC, and NOx.

For Question 6, based on the calculation results of question 5, promote the benefits of new energy vehicles to citizens and the contributions of countries around the world to the electric vehicle industry.

3 MODELS

3.1 Question 1

3.1.1 Data Collection

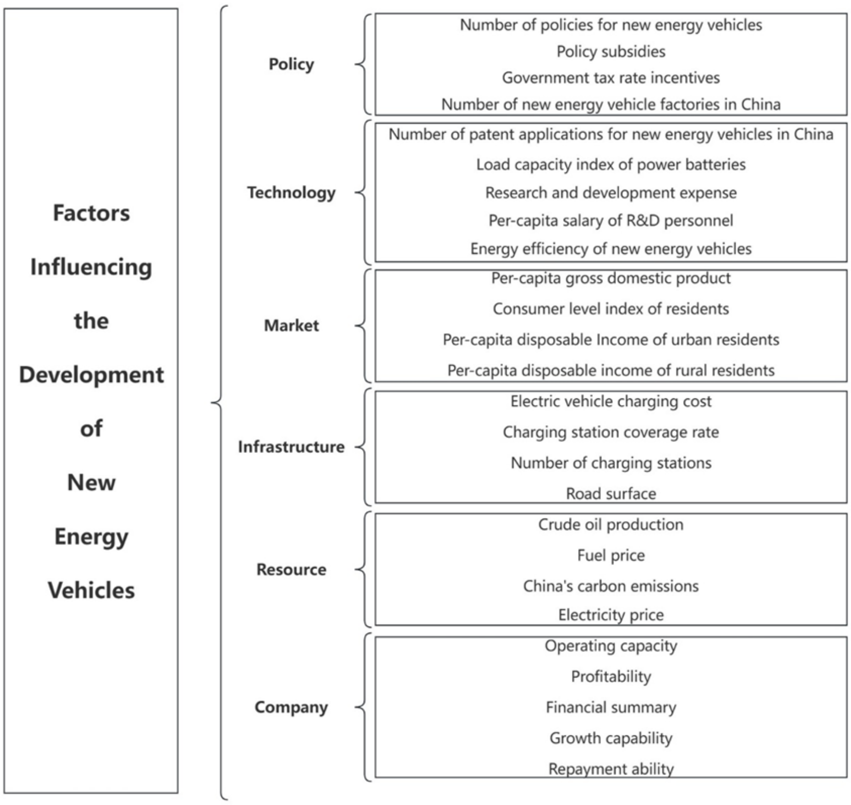


Figure 1: Schematic diagram of the three-level comprehensive factor evaluation index model

The development of new energy electric vehicles in China is influenced by various factors, including policies, technology, market, and society. To establish a comprehensive and accurate mathematical model and collect the main factors affecting the development of new energy electric vehicles in China as much as possible, we have established a three-level comprehensive factor evaluation index model. Firstly, we use the annual sales volume of new energy vehicles as a factor to measure the development of new energy vehicles. Six indicators, including policy, technological development, market economy development, infrastructure construction, resource environment, and the company's own capabilities, are used as secondary indicators. Each indicator has several tertiary indicators. It should be noted that the selection of these indicators was based on experience and relevant literature research.

For the automotive company's own capabilities, including operational ability, profitability, growth ability, repayment ability, and financial situation, we have further subdivided them into the following table. Collect relevant data from all listed new energy vehicle companies, and merge the data of all companies using the average or sum method for a certain evaluation indicator.

Table 1: Performance evaluation indicators for new energy vehicle companies

Evaluating indicator	Specific indicators
Operational ability	Fixed asset turnover rate; Inventory turnover days; Current asset turnover rate; Total asset turnover rate; Business cycle; Accounts receivable turnover rate; Inventory turnover
Profitability	Inventory turnover days; Sales gross profit margin; Net profit margin from sales; Return on equity (diluted); Net profit margin on total assets; Return on investment capital; Return on total assets
Growth ability	Basic earnings per share increased year-on-year; The total operating revenue increased year-on-year; Year-on-year growth in operating revenue; Operating profit increased year-on-year; The total profit increased year-on-year; Net profit increased year-on-year; The net profit attributable to the parent company increased year-on-year; The net cash flow generated from operating activities increased year-on-year; Diluted earnings per share increased year-on-year; The net cash flow generated from operating activities per share increased year-on-year; Total assets increased relative to the beginning of the year; Net assets increased relative to the beginning of the year; The shareholder's equity attributable to the parent company has increased relative to the beginning of the year; Relative increase in net assets per share compared to the beginning of the year
Repay ability	Current ratio; Quick ratio; Conservative quick ratio; Property ownership ratio; Tangible net assets/total liabilities Tangible net assets/interest bearing liabilities; Tangible net assets/net debt; Earnings before interest, tax, depreciation and amortization/total liabilities, net cash flows from operating activities/current liabilities; Net cash flow from operating activities/interest bearing debt; Non current liabilities to working capital ratio; Cash flow interest coverage ratio
Financial situation	Total operating income; Total operating costs; Operating profit; Net profit; Total equity attributable to the owners of the parent company; Total assets; Total liabilities; Net cash flow generated from operating activities; Net cash flow generated from investment activities; Net cash flow generated from financing activities

3.1.2 Data preprocessing

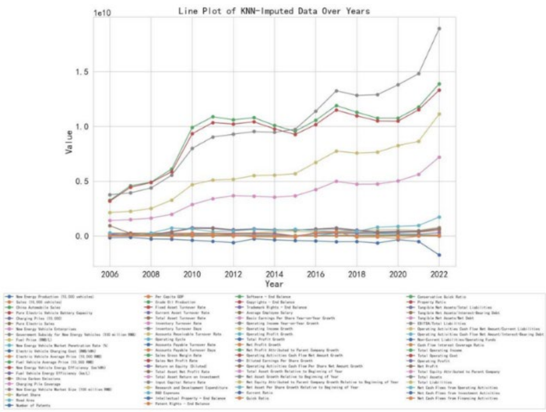


Figure 2: Overall trend chart

Firstly, we conducted an overall visual analysis of the collected data, and it can be seen that the majority of related collected data showed an overall increasing trend from 2006 to 2022.

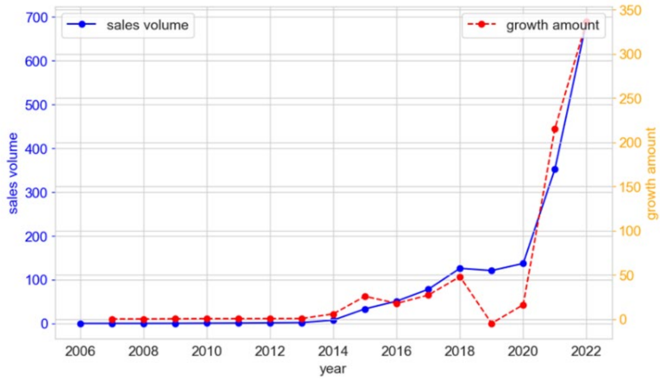
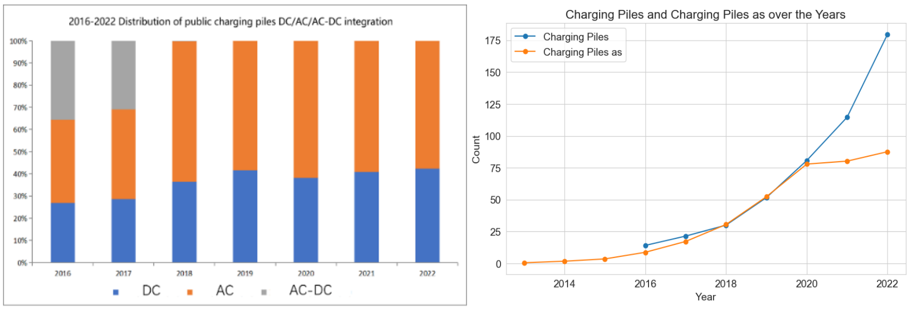


Figure 3: Line chart of sales volume and growth quantity of new energy vehicles

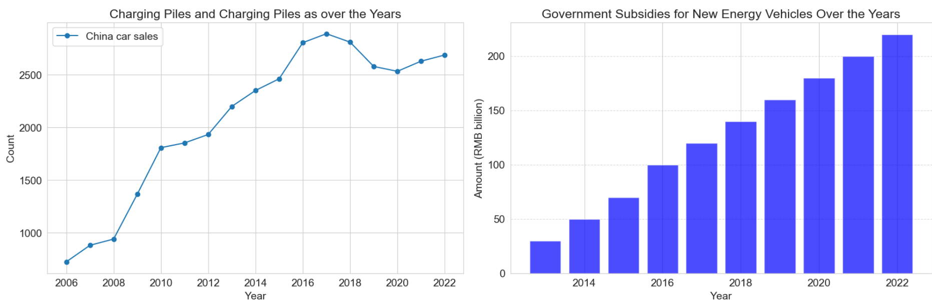
From the above figure, it can be seen that the number of new energy vehicles developed slowly before 2013, which was the initial stage of development; From 2013 to 2020, the sales of new energy vehicles steadily increased, marking a stage of stable development; After 2020, the sales of new energy vehicles have surged and developed rapidly, entering a stage of rapid development.



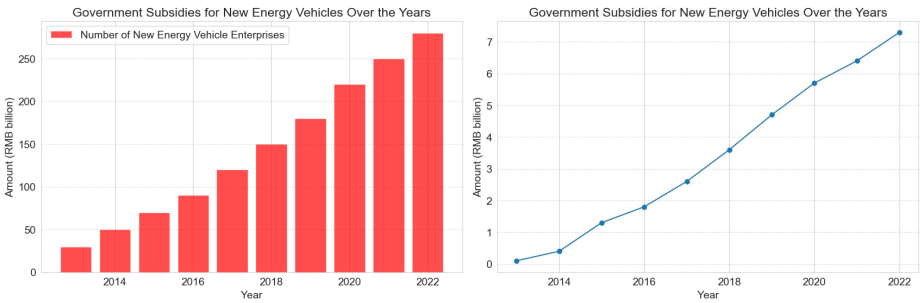
(a) **（b）**

Figure 4: a Distribution of DC and AC charging stations, and Figure b Quantity and coverage of charging stations

The above figure shows that charging stations are gradually being replaced by DC and AC types, while AC-DC types are gradually being phased out. The number of charging stations and the coverage rate of charging stations increased exponentially from 2013 to 2022, while the coverage rate of charging stations remained stable at around 90% in 2022.



(a) (b)



(c) (d)

Figure 5: a shows the sales figures of Chinese automobiles, while Figure b shows subsidies for new energy vehicles, Figure c represents the number of new energy vehicle companies, and Figure d represents the penetration rate of new energy vehicles

3.1.3 Establishment and solution of multiple regression models

After data integration, we found that there are many influencing factors, and conducting correlation analysis directly will result in a large number of complex and disordered results. After normalizing the secondary indicators, such as policy, technological development, market economy development, infrastructure construction, resource and environment, and the corresponding tertiary indicators of the automotive company's own capabilities, we use the Analytic Hierarchy Process to confirm the weights of the tertiary indicators. For each secondary indicator, multiplying and summing the weight values can obtain the value of each secondary indicator. Later, we will use the policy as an example to explain in detail.

3.1.4 Solving the weights of three-level indicators based on Analytic Hierarchy Process

After data integration, we found that there are many influencing factors, and conducting correlation analysis directly will result in a large number of complex and disordered results. After normalizing the secondary indicators, such as policy, technological development, market economy development, infrastructure construction, resource and environment, and the corresponding tertiary indicators of the automotive company's own capabilities, we use the Analytic Hierarchy Process to confirm the weights of the tertiary indicators. For each secondary indicator, multiplying and summing the weight values can obtain the value of each secondary indicator. Later, we will use the policy as an example to explain in detail.

Analytic Hierarchy Process (AHP) is a multi-objective decision analysis method proposed by American operations researcher Thomas L. Saaty in 1970. AHP is widely used in complex decision-making problems, especially in engineering, management, economics, and other fields. This method helps decision-makers make decisions under multiple criteria and hierarchical structures, making problem handling more systematic and actionable. The basic principle of AHP is to decompose a complex decision-making problem into multiple levels, from the overall goal to specific decision-making factors, and then compare these factors pairwise through expert judgment to construct a judgment matrix. The judgment matrix reflects the relative importance between different factors. By calculating the eigenvectors and eigenvalues of the judgment matrix, the weights of each factor are ultimately obtained. These weights reflect the degree to which each factor contributes to the decision-making objectives.

1) Taking the second level policy indicators as an example, there are four third level indicators, namely the number of new energy vehicle policies, government subsidies, government tax incentives, and the number of new energy vehicle factories in China. A judgment matrix is established by comparing them pairwise:

$$\begin{bmatrix} 1 & 1.223 & 1.5233 & 2.3551 \\ 0.8176 & 1 & 1.2251 & 1.4562 \\ 0.6562 & 0.8163 & 1 & 1.0251 \\ 0.4246 & 0.6667 & 0.9756 & 1 \end{bmatrix}$$

2) The eigenvalues method calculates the weights of the four factors as follows:

[0.35004341 0.26658933 0.2079108 0.17545645]

3) Consistency check:

Consistency testing is to ensure that pairwise comparisons by experts are relatively consistent. The formula for calculating the Consistency Ratio (CR) is:

$$CR = \frac{\lambda_{max} - n}{n-1} \times \frac{1}{CI} \tag{1}$$

Among them, n is the order of the judgment matrix, and CI is the consistency indicator. Usually, CR is required to be less than a certain threshold, such as 0.1, indicating that the expert's comparison is relatively consistent. The CI value of the judgment matrix is: (0.005380197212223 7505+0j); The CR value of the judgment matrix is: (0.006045165406992978+0j), and passes the consistency test.

4) Comprehensive decision-making:

Similarly, we can obtain the final result by performing the same treatment on the other 5 secondary variables. We use weights to sum the data of the same secondary indicators for dimensionality reduction, and the final data obtained is the sales volume of new energy vehicles, policy, technological development, market economy development, infrastructure construction, resource environment, and the capabilities of the automotive company itself.

3.1.5 Analysis of Development Factors of New Energy Vehicles in China Based on Model Fusion

We use three methods to comprehensively analyze the impact of policies, technological development, market economy development, infrastructure construction, resource environment, and the capabilities of automobile companies on the development of new energy vehicles in China. The specific steps are as follows.

① We numbered the three methods of correlation coefficient, mutual information, and random forest as 1, 2, and 3. Assuming that the subsets of variables selected by methods 1-3 are A1, A2, and A3 respectively, and the number of variables in each subset is 6, they have been sorted according to the importance of correlation.

② The indicators for optimal variable selection are: (1) the number of times a variable appears (the more times it appears, the better) and (2) the comprehensive factors of variable ranking (the higher the variable ranking, the better). Due to the different dimensions evaluated by each method, a voting scoring method is used for sorting.

③ For a subset of variables, each variable is assigned a value of 6-1 in the order of occurrence, with each variable value decreasing by 1. That is, the variable ranked first is assigned a value of 6, and the variable ranked last is assigned a value of 1. Then assign values to all the variables that appear and sum them to obtain the set A4 of all variable subsets, and sort the variables in A4.

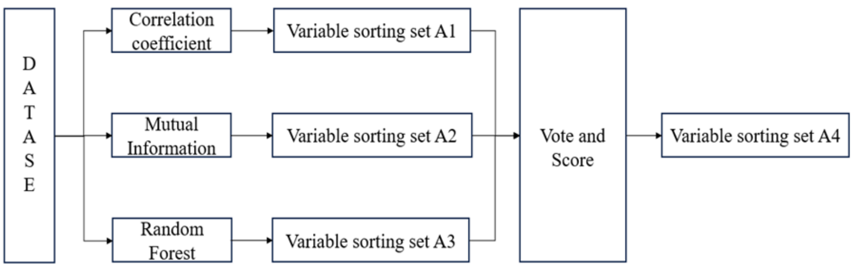


Figure 6: Flowchart for Exploring Factors Influencing the Development of New Energy Vehicles in China

Firstly, the method of correlation coefficient is used to explore the factors affecting the development of new energy vehicles in China. Pearson correlation coefficient is a statistical measure used to measure the strength and direction of the linear relationship between two variables. Its calculation is based on the covariance and standard deviation of two variables. The Pearson correlation coefficient is usually represented by the letter (r), and its calculation formula is as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

Among them, n is the sample size, x_i and y_i are the i th observation values of the two variables, $\bar{x}$ and $\bar{y}$ are the means of two variables.

The Pearson correlation coefficient ranges from -1 to 1: r = 1 indicates complete positive correlation (positive linear relationship), r = -1 indicates complete negative correlation (negative linear relationship), and r = 0 indicates no linear relationship.

From the figure below, we can clearly see that the top three factors affecting the development of new energy vehicles in China are technological development, policies, and infrastructure construction. These three factors far outweigh the impact of the company's own capabilities, market economy development, and resources and environment.

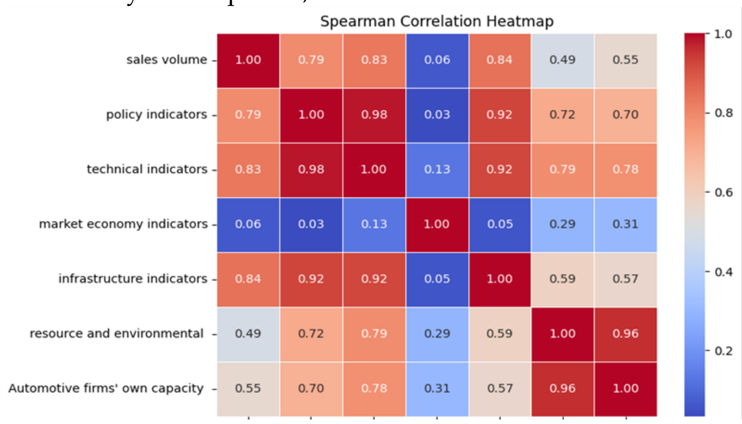


Figure 7: Correlation coefficient heatmap

Secondly, we use the method of mutual information, which is an information theory method used to measure the degree of correlation between two random variables. Mutual information measures the degree to which uncertainty in another random variable is reduced by knowing the value of one random variable. In information theory, mutual information can be used to measure the degree of interdependence between two random variables.

$$[I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right)] \tag{3}$$

$I(X;Y)$ is the mutual information between random variables X and Y. $p(x,y)$ is the joint probability of X and Y taking values of x and y. $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y, respectively.

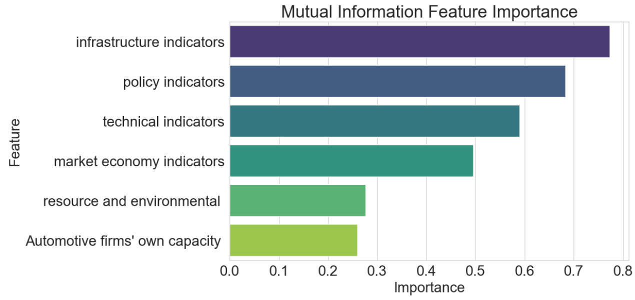


Figure 8: Random forest mutual influence result graph

Based on the results of the above three methods, we conducted a comprehensive evaluation of the factors that affect the development of new energy vehicles in China. We concluded that policy, infrastructure, and technological development are the three most important factors affecting the development of new energy vehicles in China.

Table 2: Quantitative table of factors that ultimately affect the development of new energy vehicles in China

Influence factor	Correlation coefficient	Mutual information method	Random forest	Total score	Weight
Policy indicators	4	6	5	15	0.238
Technical indicators	5	4	4	13	0.206
Market economy indicators	1	1	3	5	0.079
Infrastructure indicators	6	3	6	15	0.238
Resource and environmental	2	2	2	6	0.113
Automotive firms' own capacity	3	5	1	9	0.142

3.1.6 A Multiple Linear Regression Model Based on Weighted Correction of Independent Variable Coefficients

The formula for a multiple linear regression model can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon \tag{4}$$

Where Y is the dependent variable (response variable), $X_1, X_2, \ldots, X_p$ is the independent variable (characteristic variable), β_0 is the intercept (constant term), $\beta_1, \beta_2, \ldots, \beta_p$ is the coefficient of the independent variable, ϵ is an error term. The goal of the model is to find a set of coefficients $\beta_0, \beta_1, \ldots, \beta_p$ makes the model fit the training data best, i.e. minimizing the error term ϵ。

We substitute the quantitative score obtained in the previous text and fuse the coefficients of the independent variables with weights to obtain the final formula as follows:

$Y = 93.85 + -1084.47 * X_1 + 1945.70 * X_2 + -90.29 * X_3 + 130.04 * X_4 + -814.63 * X_5 + 368.72 * X_6$

Among them, X1 to X6 are respectively, policy indicators, technical indicators, market economy indicators, infrastructure indicators, resource and environmental, automotive firms' own capacity.

3.2 Question 2

3.2.1 Data preprocessing

Question 1 identified six important factors through analysis and literature search. Due to the presence of partial missing and outliers, the sample data with a high missing rate was first deleted, and then the Epanechnikov density estimation method was used to detect outliers in the data. Outliers below the threshold were removed, and some missing data was filled in using interpolation. Finally, standardize the data.

(1) Visual analysis

In order to explore the changes in sales and production of new energy vehicles over time and further optimize prediction strategies, a time series diagram was drawn as shown in the figure. It can be visually seen that sales and production increase over time, accompanied by sudden changes.

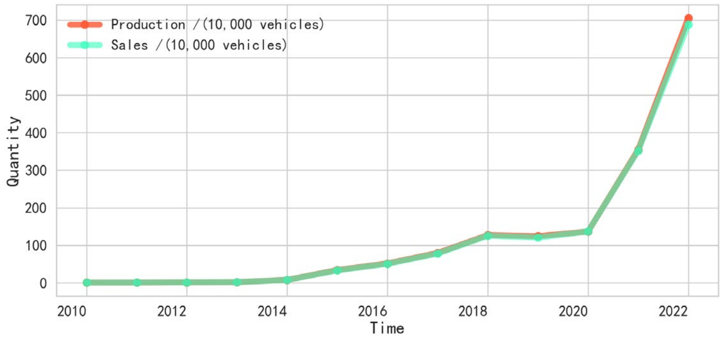


Figure 9: the sales and production of new energy vehicles over time

The following are the time series images obtained by normalizing the seven factors in the first question, all of which are correlated with time and grow over time. From the conclusion of Question 1, these factors reflect the development of new energy vehicles.

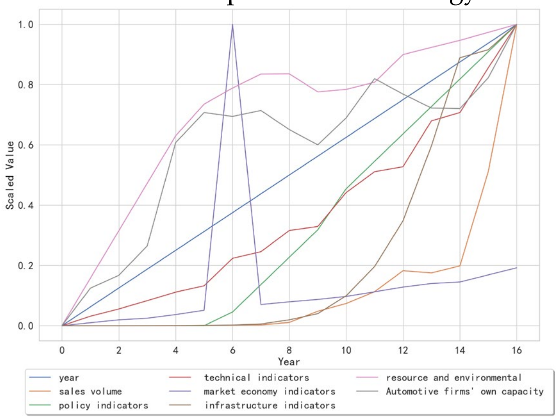


Figure 10: normalized time series images obtained from seven factors

(2) TOPSIS method

The relative closeness score reflects the degree of closeness of each sample (in this case, each year) to the ideal solution and negative ideal solution. In the TOPSIS method, a higher score indicates that the sample is closer to an ideal solution, while a lower score indicates that the sample is closer to a negative ideal solution.

These scores can be used to compare and evaluate the performance or characteristics of different years (or other evaluation objects). A higher relative proximity score may indicate that a year is closer to the ideal state in a given attribute or indicator, while a lower score may indicate that the corresponding year is further away from the ideal state in these indicators. Therefore, the relative proximity score can be used for sorting and selection in multi-attribute decision-making problems, helping to determine which years are more in line with expectations or superior. We will use TOPSIS to analyze the seven normalized parameters above, and the TOPSIS method chart below reflects the changes in the development of new energy vehicles in different years.

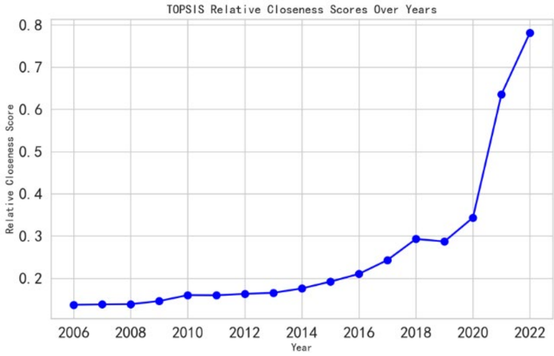


Figure 11: TOPSIS method diagram

3.2.2 Establishment and Solution of BILSTM, SARIMA, and Grey Prediction Models

(1) Construction of SSA Bilstm Prediction Model

SSA-BILSTM is a hybrid optimization algorithm that combines Salp Swarm Algorithm (SSA) and Bidirectional Long Short Term Memory Network Algorithm (BILSTM). This algorithm is mainly applied in the training and optimization process of neural networks.

The main component of LSTM (Long Short Term Memory) is memory cells, which are composed of input gates, forget gates, and output gates. These gates control the flow of information: the input gate determines the information transmitted to the memory cell at the current moment, the forget gate determines the information that should be retained or discarded in the memory cell at the previous moment, and the output gate controls the information output from the memory cell. These gates use sigmoid activation functions to regulate the flow of information, thereby controlling the input, storage, and output of information. LSTM also includes memory cells, two squeezing units, and two gating units. In addition to forward propagation calculation, it also uses backpropagation calculation to adjust network parameters to obtain the best solution, which helps to build a more optimized network structure. The structure diagram of the long short-term memory cell network is shown in Figure 4.7.

Bidirectional Long Short Term Memory Network (LSTM) is an extension of Long Short Term Memory Network (LSTM) that enhances the structure of LSTM, allowing for simultaneous learning of sequence information from the past to the future (forward) and from the future to the past (backward). In standard LSTM, the model can only process input sequence information along one direction (usually forward) in the time series. But in bidirectional LSTM, two independent LSTM layers process the input sequence simultaneously, with one layer processing the input in chronological order (forward) and the other layer processing the input in opposite chronological order (backward). This enables the network to obtain information about the context of the input sequence, rather than just partial information. Once the forward and backward hidden states are obtained through bidirectional LSTM, these states are connected and combined to provide a more comprehensive representation of sequence information. This combination makes the model more capable of understanding the context and relevant patterns in the input sequence, which helps improve the modeling and prediction capabilities of sequence data.

In traditional neural network training, backpropagation algorithm is a commonly used method that updates parameters by calculating the gradients of each parameter in the network, so as to make the output of the network as close as possible to the target value. However, the BILSTM algorithm may fall into local optima and converge slowly for complex non convex optimization problems.

In order to overcome the limitations of the BILSTM algorithm, the SSA-BILSTM algorithm combines the global search ability of the tunicate swarm algorithm with the local search and parameter update ability of the BILSTM algorithm. Specifically, the steps of the SSA-BILSTM algorithm are as follows:

1. Initialization: Randomly generate the initial weights and thresholds of the neural network, and initialize the position and state of the population of sea squirts.

2. Update of the population of tunicates: According to the principle of the SSA algorithm, individuals of tunicates adjust their positions based on the results of the evaluation function and converge to the vicinity of a more optimal solution through local search.

3. Backpropagation: Using the current position and state of the sea squirt population, execute a backpropagation algorithm to calculate the gradient of the neural network.

4. Parameter update: Based on gradient information, use the traditional BILSTM algorithm to update the weights and thresholds of the neural network.

5. Termination condition judgment: Determine whether the termination condition is met, such as reaching the maximum number of iterations or the error being less than the threshold. If the conditions are met, the algorithm ends; Otherwise, return to step 2 for the next iteration.

The SSA-BILSTM algorithm can search for the optimal solution on a global scale by combining the bottle sea squirt swarm algorithm and backpropagation algorithm, and accelerate the convergence process by utilizing the local search and parameter update capabilities of the BILSTM algorithm. This hybrid optimization algorithm can improve the training performance and convergence speed of neural networks, especially suitable for complex problems and large-scale neural network training tasks.

(2) Construction of ARIMA (Autoregressive Moving Average)

The ARIMA model (Autoregressive Moving Average) is a classic statistical model used for time series analysis and prediction. It combines two main time series features: autoregressive (AR) and moving average (MA).

Autoregressive (AR) section: The AR model uses the lagged values of observations (i.e. past observations) to predict the current value. It assumes that the current value is correlated with values from previous time points. The AR (p) model represents the relationship between the current value of a time series and its values at p past time points. This model is based on the autocorrelation of observed values and uses a linear relationship between the lagged value and the current value for prediction. Moving Average (MA) Part: The MA model uses the error term or residual of the observed values to predict the current value. It assumes that the current value is related to the random error at past time points.

The MA (q) model represents the relationship between the current value of a time series and the error at q past time points. This model is based on the randomness of observed values and uses a linear relationship between historical errors and current values for prediction.The ARIMA model combines AR and MA models: The ARIMA (p, q) model includes a combination of the AR (p) model and the MA (q) model. It can describe the autocorrelation and randomness of time series. The ARIMA model is used to describe stationary time series, which have both autocorrelation and randomness, meaning that the mean and variance of the series remain constant over time.

Parameters P and Q: P is the order of the autoregressive part, representing the number of lagged terms used in the model. Q is the order of the moving average part, representing the number of error terms used in the model.

The fitting and prediction of ARIMA models can be achieved through maximum likelihood estimation or other optimization methods. This model is widely used in the analysis and prediction of time series data, but usually requires the time series to be stationary. If the time series exhibits non stationarity, it may be necessary to first perform differential processing before applying the ARIMA model.

The AR (autoregressive) model (AR (p)) is represented as:

$$y_t = c + \phi_1 y_{\{t-1\}} + \phi_2 y_{\{t-2\}} + \cdots + \phi_p y_{\{t-p\}} + \varepsilon_t \tag{5}$$

Among them, y_t is the current value of the time series, c is the constant term, $\phi_1, \phi_2, \dots, \phi_p$ is the autoregressive coefficient, $y_{\{t-1\}}, y_{\{t-2\}}, \dots, y_{\{t-p\}})$ is the past value of a time series, ε_t is the error term.

The MA (Moving Average) model (MA (q)) is represented as:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{\{t-1\}} + \theta_2 \varepsilon_{\{t-2\}} + \cdots + \theta_q \varepsilon_{\{t-q\}} \tag{6}$$

Among them, y_t is the current value of the time series, ε_t is the error at the current time point, $\varepsilon_{\{t-1\}}, \varepsilon_{\{t-2\}} \dots$ is the past error value.

The ARIMA (Autoregressive Moving Average) model combines AR and MA models, represented as:

$$y_t = c + \phi_1 y_{\{t-1\}} + \phi_2 y_{\{t-2\}} + \cdots + \phi_p y_{\{t-p\}} + \varepsilon_t + \phi_p a_t \varepsilon_{\{t-1\}} + \theta_2 \varepsilon_{\{t-2\}} + \cdots + \theta_q \varepsilon_{\{t-q\}} \tag{7}$$

The ARIMA model includes two parts: autoregressive and moving average, combining past values and past error values of time series.

When selecting ARIMA parameters, autocorrelation graphs or information criteria (such as AIC/BIC) are usually used for judgment. However, when there are many models, the speed of traversing to find the optimal parameters is slower. To improve efficiency, it is possible to consider directly using the built-in automatic parameter selection function Auto Arima in Python, which greatly saves time.

3.2.3 Solution and Result

We used Bi LSTM network and ARIMS method to train the model on the data obtained from TOPSIES, and predicted the development of new energy electric vehicles in China in the next 10 years.

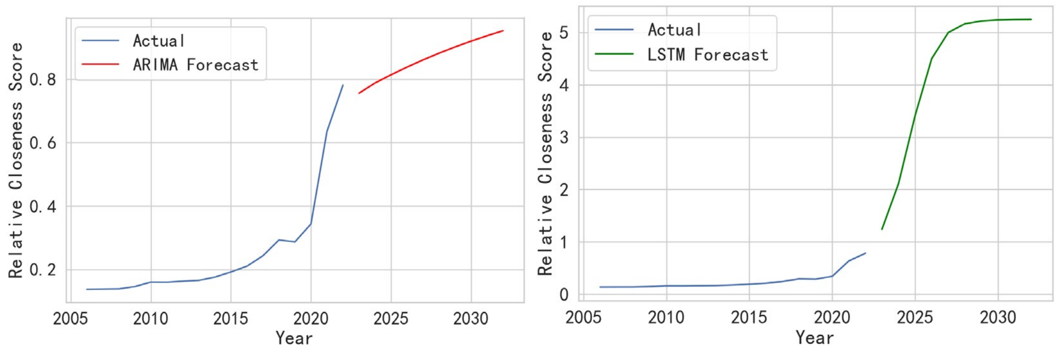


Figure 12: ARIMA prediction result chart; Bi LSTM prediction result graph

Compared with the results predicted by BILSTM and ARIMA respectively, BILSTM shows a trend of over learning and continues to grow at a high speed until 2027, followed by a slow growth, which is obviously contrary to the actual situation. This is because the new energy policies in most places have experienced a significant decline by 2025 and will not continue to surge. However, ARIMA's predicted results show a sustained slow growth, which has certain reference value and may slow down in the next decade, It was not reflected. Based on the prediction results of two models, we can conclude that the overall development of new energy vehicles will show a steady upward trend in the next decade, and it will slow down stagflation after 2027.

3.3 Question 3

The data in question three contains some missing values and outliers. Firstly, for sample data with high missing rates, delete them; Subsequently, apply 3 σ Principle based detection of outliers and removal of outliers; For partially missing data, KNN is used to fill in the missing values. Finally, normalize the data to compare various indicators on the same scale.

(1) Visual analysis

By handling outliers, the following data was obtained.

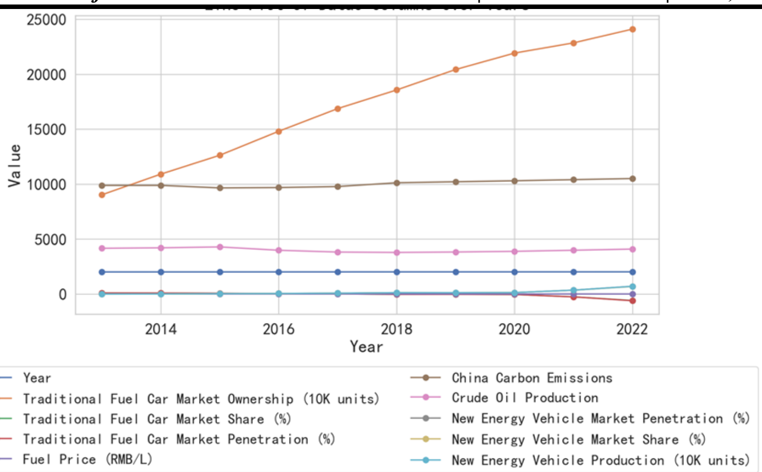


Figure 13: visualization processing of data for the global traditional energy vehicle industry using new energy electric vehicles.

In order to explore the difference between the sales of new energy vehicles and the ownership of traditional gasoline vehicles, we draw a dot matrix diagram to visually reflect the trend of changes in the two. By observing the trend of changes in the ownership of new energy vehicles and traditional gasoline vehicles, it is evident that the growth of gasoline vehicles has slowed down, while the proportion of new energy vehicles continues to rise.

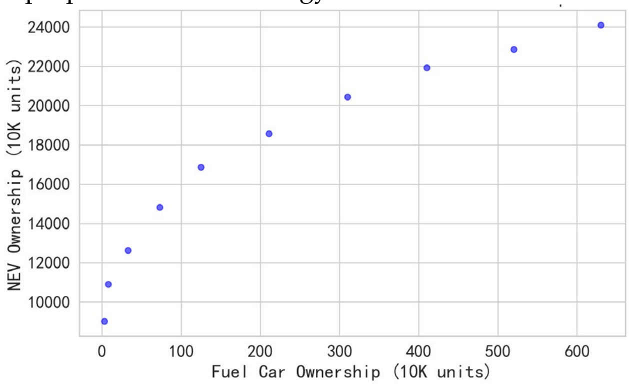


Figure 14: the difference between the sales volume of new energy vehicles and the ownership of traditional gasoline vehicles

3.3.2 Correlation analysis

(1) Pearson correlation analysis

Pearson correlation analysis mainly describes the unified change and movement trend of two sets of linear data by calculating the Pearson correlation coefficient. It can accurately reflect the strength of the linear correlation between two variables and can also be used to eliminate the existence of multicollinearity relationships. It is generally represented by r, and the formula is as follows.

$$r=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{8}$$

Among them, n represents the number of samples; Xi - the point i observation value corresponding to the variable X; Yi - the point i observation value corresponding to the variable Y; $\bar{X}$ - average of X samples; $\bar{Y}$ - average of Y samples.

Generally, we judge the correlation strength between variables by the range of correlation coefficients, as shown in the table:

Table 3: Pearson variable correlation classification table

relevant intensity	Range of variation of correlation coefficients
Strong Positive	0.8-1.0
Positive	0.6-0.8
Moderate	0.4-0.6
Weak	0.2-0.4
Very Weak or None	0.0-0.2

This question will calculate Pearson correlation for all data in the table and draw a heatmap.

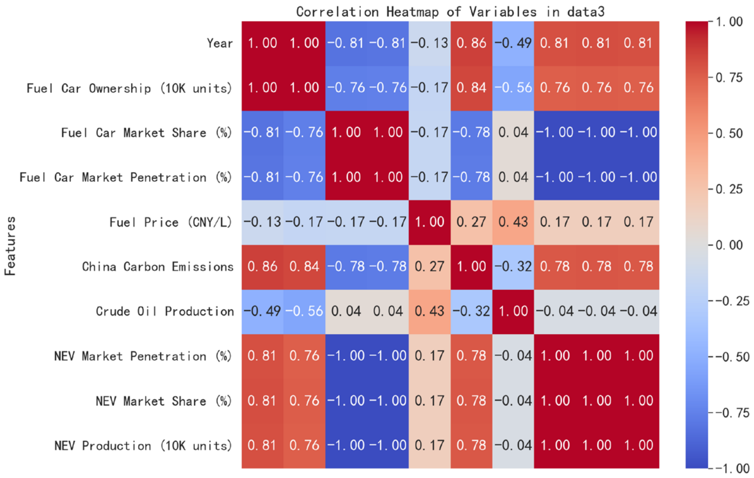


Figure 15: pearson correlation analysis obtained the correlation between the global traditional energy vehicle industry and new energy vehicles.

Pearson analysis obtained the correlation between various parameters. Due to the high correlation between market penetration rate, market share, and sales volume, we performed a deletion operation to remove these redundant features. Overall, it can be seen that the characteristics related to gasoline vehicles are negatively correlated with the sales volume of new energy vehicles.

(2) Spearman correlation analysis

Spearman rank correlation coefficient（ ρ ） The calculation formula for is as follows:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{9}$$

Among them, ρ is Spearman rank correlation coefficient, d is rank difference of each pair of data points, n is number of data points.

In our data, there are 5 parameters: market ownership of traditional fuel vehicles, fuel prices, China's carbon emissions, crude oil production, and new energy production. According to the above formula, calculate the Spearman rank correlation coefficient between these parameters. Arrange data points in ascending order and assign ranks to each data point. The calculation results of the Spearman correlation coefficient are shown in the following figure. Through Spearman correlation analysis, the correlation of the global traditional energy vehicle industry with new energy vehicles is obtained.

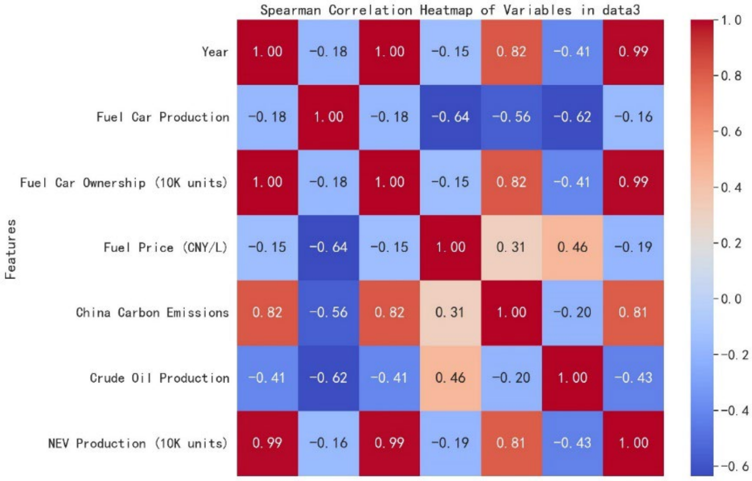


Figure 16 spearman correlation analysis obtained the correlation between the global traditional energy vehicle industry and new energy vehicles.

(3) Grey correlation analysis

Grey Relational Analysis (GRA) is a multivariate analysis method used to study the degree of correlation between variables. It evaluates the similarity or correlation between various variable data sequences by comparing their correlation.

The main idea is to convert data sequences of multiple factors into comprehensive measurement values, which can reflect the degree of similarity between each sequence. GRA can be used to handle situations with limited sample data and lack of correlation, helping to analyze and explore potential correlations between data.

The information reflected by a single correlation coefficient is scattered and requires centralized processing of the correlation information. The average method is used to quantitatively reflect the degree of correlation between the independent variable and the dependent variable.

$$\rho = \max_{i=1}^{n} \min_{k=1}^{m} \frac{|x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \varepsilon} \quad (10)$$

Among them, $x_0(k)$ is reference sequence, $x_i(k)$ is subsequence, m is the length of reference sequences and subsequences, n is number of subsequences, ε is the small positive number, ρ is the gray resolution coefficient, with a value of (0,1), is used to adjust the difference in output results.

This method includes the following steps:

Construct a correlation coefficient matrix: Calculate the correlation coefficients between the seven factors obtained after processing based on the pre-set correlation coefficient calculation formula. This step mainly involves quantifying the correlation between various factors and forming a correlation coefficient matrix.

Determining correlation coefficients: Usually, methods such as grey correlation are used to calculate the correlation coefficients between various factors, which represent the degree of similarity between the factors.

Correlation coefficient ranking: Sort or evaluate the correlation coefficients to determine which factors are more relevant or similar to other factors.

Result analysis: Based on the analysis of correlation coefficients, we can identify important influencing factors or factors with high correlation. Overall, the correlation is not significantly different from the previous two methods, except that the coefficients of this method are positive. The overall correlation between the sales of gasoline vehicles and new energy vehicles is weak, and the correlation with oil prices is strong.

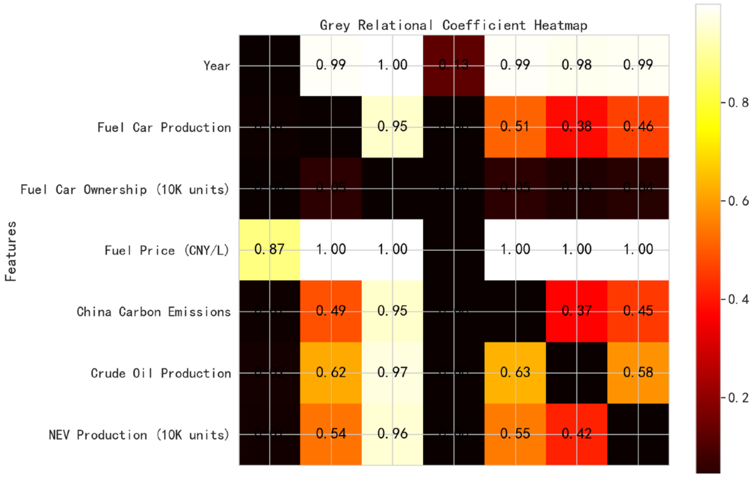


Figure 17: grey correlation analysis obtained the correlation between the global traditional energy vehicle industry and new energy vehicles.

Through a comprehensive evaluation of three correlation analysis methods, it is found that new energy vehicles have a restraining effect on the sales of gasoline vehicles. The promotion of new energy vehicles has a certain impact on the global traditional energy vehicle industry. However, at the same time, traditional energy related factors such as oil prices and carbon emissions have not been effectively suppressed and are still in an upward stage, which may continue to cause pollution to the environment.

3.3.3 Significance test

T-test, also known as student t-test, is a statistical method used to test whether the difference in mean between two samples is significant. It is suitable for determining whether there is a significant difference in the mean values between two groups of samples when the sample size is small (less than 30) and follows a normal distribution.

The basic principle is to determine the significance between the means by calculating the ratio between the mean difference of two sets of samples and their standard error. The formula for t-test is as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{11}$$

Among them, $\bar{X}_1$ and $\bar{X}_2$ are the mean values of two samples, s_1 and s_2 are the standard deviation of two samples, n_1 and n_2 are the size of two samples.

If the t-value is greater than the critical value, the null hypothesis can be rejected, that is, there is a significant difference in the mean between the two samples. Otherwise, the null hypothesis cannot be rejected, which assumes that there is no significant difference in the mean between the two samples.

We separately calculated all the data in data3 that were traversed through a for loop for each column, and performed t-tests on each column with 'NEV Production (10K units)'. There is used the scipy.stats.ttest ind() function to perform an independent sample t-test to calculate the t-statistic and p-value between each feature and the 'NEV Production (10K units)' column. The results are shown below, and there are significant differences between each feature and NEV Production (10K units), indicating a significant trend difference.

Table 4: the independent-sample t test

Feature	t-statistic	p-value	Significance
Year	27.1267	4.73E-16	Significant difference

Fuel Car Production	21.45107	2.87E-14	Significant difference
Fuel Car Ownership (10K units)	10.27919	5.84E-09	Significant difference
Fuel Price (CNY/L)	-2.26793	0.035878	Significant difference
China Carbon Emissions	82.95854	1.04E-24	Significant difference
Crude Oil Production	43.49009	1.10E-19	Significant difference
NEV Production (10K units)	0	1	No significant difference

3.3.4 Linear regression fitting

Linear regression is a statistical learning method used to establish a linear relationship model between the independent variable (feature) and the dependent variable (target). It is one of the simplest and most commonly used methods in regression analysis, suitable for predicting continuous target variables.

The linear regression model is based on the following assumptions:

Assuming a linear relationship between the dependent variable (target variable) and the independent variable.

Assuming that the residual of the model (the difference between predicted and true values) follows a normal distribution.

The linear regression model can be represented as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{12}$$

Among them, y is dependent variable (target variable); x_1, x_2,... , x_n are independent variable (characteristic); β_0 is the intercept (the value of the model at $x = 0$); β_0, β_1, β_2 are the coefficient of the independent variable, indicating the degree of influence of the independent variable on the dependent variable; ϵ are the error term that represents the parts that the model cannot explain.

The goal of a linear regression model is to find appropriate coefficients that minimize the error between the predicted results of the model and the actual observed values. Here, the least squares method is used to estimate these coefficients, which determines the optimal coefficient value by minimizing the sum of squared residuals. We take the market share of new energy vehicles (%) as the independent variable and the market share of gasoline vehicles (%) as the dependent variable. As shown in the figure, with the increase of new energy vehicles, the market space of gasoline vehicles is further compressed.

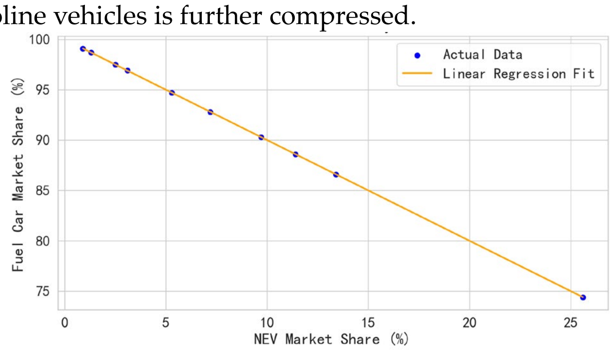


Figure 18: NVE market share

The fitting line obtained from the linear regression model shows a significant positive correlation between the decline in market share of traditional fuel vehicles and the growth in market share of new energy vehicles. In other words, the results obtained from the model show that as the market share of traditional fuel vehicles decreases, the market share of new energy vehicles shows a significant growth trend.

This trend may imply an increase in consumer interest in new energy vehicles, possibly due to their environmental, economic, or other advantages. The decline in market share of traditional fuel vehicles may be driven by factors such as environmental awareness, government policy encouragement, and the development of new technologies, which may indirectly lead to an increase in market share of new energy vehicles.

3.3.5 Question 3 Conclusion

Through Pearson analysis, Spearman analysis, grey correlation analysis, significance test comprehensive evaluation, and linear regression fitting, we can conclude that new energy vehicles have a restraining effect on the sales of gasoline vehicles. The promotion of new energy vehicles has a certain impact on the global traditional energy vehicle industry. However, at the same time, traditional energy related factors such as oil prices and carbon emissions have not been effectively suppressed and are still in the upward stage, May cause continuous pollution to the environment, etc. The rapid development of new energy electric vehicles is a blow to the global traditional energy vehicle industry, and more and more people are choosing to purchase new energy vehicles. The traditional vehicle market is under pressure, and the trend of this ebb and flow may continue in the future.

3.4 Question 4

3.4.1 Data preprocessing

(1) Policy factor analysis

Read the export volume of new energy vehicles for each month from October 2021 to September 2023 (28 months), use time points and new energy vehicle export data, and organize them into a time series format to ensure the temporal order and accuracy of the data. Regarding the policy evaluation effect of question 4, traditional policy effect evaluation cannot be conducted due to the inability to find a control group without implementing resistance policies. The following are domestic policies for new energy vehicles in various countries, which have greatly promoted the development process of new energy.

Table 5: new energy vehicle policies in some countries

Country	Policy
the Netherlands	Electric vehicle registration is free of charge, road administration does not collect taxes, and specific cities provide subsidies
Norway	The government stipulates that purchasing pure electric vehicles enjoys multiple benefits, such as exemption from all taxes and fees (including 25% value-added tax), exemption from urban tolls and parking fees in public parking lots, and driving on dedicated bus roads. Imported electric vehicles are also exempt from import tariffs.
Germany	Consumers who purchase hybrid cars priced at 60000 euros can receive a subsidy of 3000 euros, with a maximum subsidy of 400000 vehicles. In order to reduce insurance costs, the government allowed consumers who purchased electric vehicles to share license plates with another car in their homes between 2016 and 2020.
Britain	Purchasing eligible electric and hybrid vehicles can enjoy a subsidy of 4500 euros or 8000 euros. Qualified plug-in hybrid vehicles can receive a subsidy of 2500 euros. The conditions include carbon dioxide emissions below 50g/km, range exceeding 70 miles, or carbon dioxide emissions between 50g/km and 75g/km with a selling price of no more than 60000 euros.
France	Purchasing electric vehicles or hybrid electric vehicles can receive government subsidies, which vary depending on emissions and vehicle models, with a maximum subsidy amount of 6300 euros.
Spain	Purchasing electric vehicles can receive different amounts of government subsidies, depending on the vehicle model and type. For example, purchasing an electric passenger car can receive a subsidy of up to 5500 euros, while purchasing an electric truck and bus can receive subsidies of 8000 euros and 20000 euros respectively.

Boycott policy timeline:

Table 6: boycott policy

Time	Country/Region	Event
June 2022	Germany, Japan	Oppose new energy
September 2022	America	Passing the Inflation and Reduction Act
March 2023	China	Multiple spontaneous combustion of NEV
June 2023	America	Leading the way in boycotting Chinese NEV after competition failure

(2) Visual analysis

We assume that the export volume of new energy vehicles in China is a natural growth trend over time, but the fluctuations that may occur may be influenced by various policies. Therefore, we set a variable for policy resistance, which is recorded as 1 if there is policy resistance, and 0 otherwise. We have implemented corresponding policy resistance markers for a series of time points, assuming that these changes have an impact on export volume. Based on this, we established a model with year and policy resistance variables as independent variables, and export volume as dependent variable, to evaluate the impact of policy resistance on export volume, and modeled it in the manner of question 3.

Visualize the collected data on the export volume of new energy vehicles for each month from October 2021 to September 2023 (28 months), with the red dots indicating where policy events occurred. It can be visually observed that every time policy boycott occurs, the sales of new energy vehicles during that period will experience a decline.

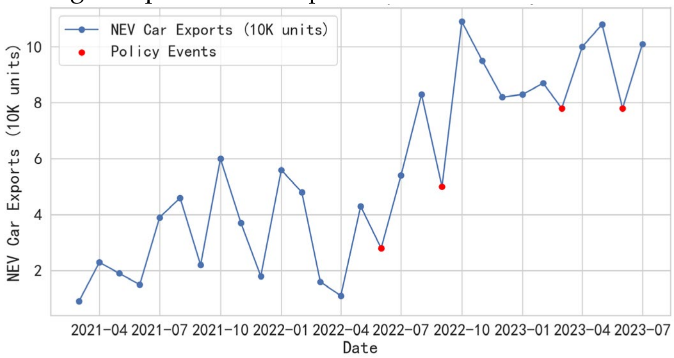


Figure 19 time series plot of NEV car exports with policy events

(3) Anomaly detection

Anomaly detection is the process of identifying observation values in a dataset that are different from the majority of the data. It detects the presence of abnormal export volume near key event time points to confirm whether there are policy events causing abnormal changes in export volume.

The results of Anomaly detection are represented as colors in the scatter plot. In this code, the Anomaly detector uses the Isolation Forest model, which marks outliers as -1 and normal points as 1. In the scatter plot, colors are based on these marker values.

The blue dots (warm tones in cmap='coolwarm ') represent the points marked as Anomaly (-1). The red dot (the cool tone in cmap='coolwarm ') represents the point marked as Normal (1). This color coding enables visualization of the distribution of outliers and normal points in time series data. It can be seen that outliers are usually the low and high points of sales, and cannot effectively reflect policy turning points.

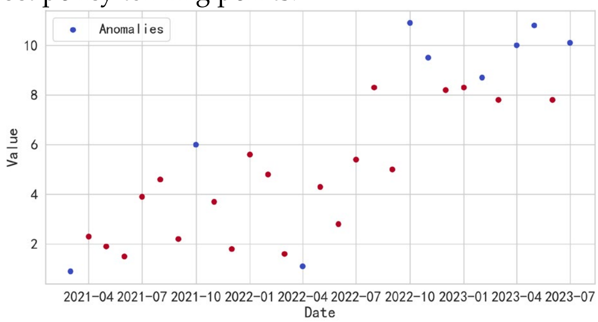


Figure 20: anomaly detection in time series data

(4) T-test

Through T-test, the result table is provided below, where const corresponds to the intercept term, where x1 and x2 are independent variables 1 and 2, respectively. From the P-value, it can be preliminarily seen that x1 (policy resistance variable) is statistically significant and not zero (P=0.019), while x2 (year) is very significant (P=0.000). This means that in this model, the impact of year on the export volume of new energy vehicles is very significant. In this case, the resistance policy variable has statistical significance, indicating that it may have a certain impact on the export volume of new energy vehicles.

Table 7: t-test result table

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6254	0.653	0.958	0.347	-0.716	1.967
x1	-2.432	0.973	-2.499	0.019	-4.433	-0.431
x2	0.348	0.04	8.675	0	0.266	0.43

3.4.2 Time series prediction analysis

Time series models can be used to predict the future export volume of new energy vehicles, in order to estimate the potential impact of policy events. Here, we use ARIMA to predict the next five months, taking the export volume of new energy vehicles and policy impact as model inputs, and obtain the following prediction result graph. It can be seen that the changes in the following months are relatively stable, indicating that under normal circumstances, policies will have a sustained impact on the future.

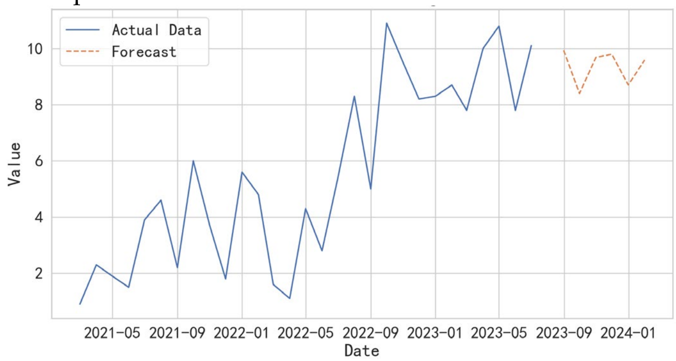


Figure 21: ARIMA Forecasting

3.4.3 Question 4 Conclusion

Through data visualization analysis, anomaly detection, t-square detection, and ARIMA time series prediction, we can draw the following conclusions:

Time series prediction: The ARIMA model can effectively predict the future export volume of new energy vehicles. This model can estimate the potential impact of policy events on future export volumes.

Anomaly detection: Through anomaly detection, it was found that there was a significant decline in the export volume of new energy vehicles during policy events. This indicates that every time a policy boycott occurs, the export volume is affected to a certain extent.

T-test: After T-test, it was found that the resistance policy variable has statistical significance, which means that policy resistance may have a certain impact on the export volume of new energy vehicles.

Based on these analysis results, it can be concluded that foreign policies have a significant impact on the export volume of new energy vehicles, and this impact is sustained over time. Therefore, in the future, relevant countries need to fully consider their potential impact on the export of new energy vehicles when formulating relevant policies, in order to better promote the development of the new energy vehicle industry. China also needs to make corresponding plans and guidance for the production, import and export of new energy vehicles in accordance with the new energy policies of relevant countries.

3.5 Question 5

In recent years, the number of cars in China has surged, followed by excessive exhaust emissions during driving due to the excessive number of cars. For new energy vehicles, due to the change of energy structure, they have the advantages of less waste emissions and environmental protection during driving, but their energy consumption and pollution in the fuel and power production phase are different from those of traditional fuel vehicles.

The data of the city with a population of 1 million assumed in this article is similar to the city of Yichang in mainland China, and this city is assumed to be named "City-A". Data collection was conducted on the ownership of various types of new energy electric vehicles in A City in 2022. Based on the "General Principles for Comprehensive Energy Consumption Calculation" and "Limits and Measurement Methods for Pollutant Emissions from Light Vehicles (China III and IV stages)", the energy-saving and emission reduction effects of new energy electric vehicles in City-A were calculated.

3.5.1 Calculation of ownership of new energy vehicles and traditional energy vehicles

To analyze the impact of electrification of urban new energy electric vehicles (including electric buses) on the ecological environment, we must have an accurate calculation of the number of new energy vehicles and traditional energy vehicles in cities with a population of 100, and based on this, we can explore the differences between the two. By exploring the changes in the ownership of new energy vehicles and traditional energy vehicles, we can compare their different impacts on the environment.

Assuming that City-A has a population of 1 million, compared to the population of 3.896 million in Yichang City in 2020, an effective estimate of the number of different vehicles in City-A can be made, specifically:

(1) The estimation method for the number of logistics vehicles in 2015 is based on the modern logistics development plan of Yichang City, which includes 10 comprehensive logistics parks and 9 professional logistics centers. On average, there are 500 logistics vehicles in each central park, totaling 19500 vehicles. Other express delivery companies have 500 logistics vehicles, totaling 20000 vehicles. Therefore, City-A is converted to 5133 vehicles based on the equal proportion of population.

(2) Estimated number of tourist vehicles in 2014: According to the statistical bulletin of Yichang Tourism Bureau, Yichang received a total of 408.5 million domestic and foreign tourists in 2014, a year-on-year increase of 23.03%. According to 300 days per vehicle × 8 hours of operation, with an average capacity of 20 people per trip, requiring 4085.01 types of tourist vehicles × 10000/300/8/20, with a 10% increase. In 2015, Yichang City had 936 tourist buses, so the number of tourist buses in City-A was 240.

(3) By analogy, the number of private cars, buses, cleaning vehicles, etc. in Yichang City will be calculated one by one. The conversion results for City-A based on Yichang City are shown in the table below.

Table 8: number of various types of cars in City-A

Motorcycle type	Cars in 2017	NEVs in 2017	Cars in 2018	NEVs in 2018	Cars in 2020	NEVs in 2020
Bus	400	141	441	205	523	334
Taxi	797	159	805	402	821	657
Logistics vehicle	5660	1132	5943	2971	6551	5241
Tourist vehicle	265	53	278	139	306	245
Sanitation truck	84	17	95	47	122	98

Private car	203267	6098	215745	10787	240650	24065
Total	210472	7600	223306	14552	248973	3064

Calculation method for electric vehicles: Taxi, logistics vehicles, tourist vehicles, and sanitation vehicles are calculated based on a conversion rate of 20% in the first two years, 50% in the third year, and 80% by the end of the 13th Five Year Plan period; Private cars are calculated at 3% in the first two years, 5% in the third year, and 10% at the end of the 13th Five Year Plan period; The conversion rate of buses is determined according to the relevant regulations of the Ministry of Transport, the Ministry of Finance, and the Ministry of Industry and Information Technology on the proportion of new energy vehicles in buses

3.5.2 The increase in new energy electric vehicles has an impact on the environment

According to relevant information, we have learned that the driving range and total energy of the power battery pack of various types of new energy electric vehicles fluctuate in different ranges, so their power consumption per hundred miles also varies. We perform a weighted average on this to determine a value.

Table 9: energy consumption and annual average operation of vehicles over a hundred miles.

Motorcycle type	Electric vehicle (kW·h /100 km)	Fuel vehicle (L/100 km)	Annual average operating volume (10000 km)
Bus	110	30	10
Taxi	20	10	15
Logistics vehicle	50	20	10
Tourist vehicle	110	30	8
Sanitation truck	40	20	5
Private car	20	10	2

Based on the number of various types of electric vehicles obtained in the previous text, the energy consumption of new energy vehicles can be compared with the fuel consumption of equivalent traditional vehicles to obtain the energy-saving situation of new energy vehicles in City-A for three years. To calculate the relative energy savings, it is first necessary to convert the energy consumption of electric and gasoline vehicles into Standard Coal Equivalent (SCE). The standard coal conversion coefficient represents the amount of standard coal equivalent to a unit mass of fuel or electricity.

It should be noted that the density of finished oil for vehicles is 0.725 (kg/L), the conversion coefficient of fuel to standard coal is 1.4714 (kgce/kg), the conversion coefficient of electrical energy to standard coal is 0.1229 (kgce/kWh), the power consumption of electric vehicles is 14300 (kWh/100km), and the fuel consumption of fuel vehicles is 39000 (L/100km).

Convert the power consumption of electric vehicles and fuel consumption of gasoline vehicles to standard coal consumption. Here is a summary of these two conversion processes:

Conversion of standard coal consumption for electric vehicles:

Standard coal consumption for electric vehicles (kWh)=Electricity consumption for electric vehicles (kWh) × The conversion coefficient of electrical energy to standard coal, specific calculation:

Electric vehicle standard coal consumption = 14300 Wh/100 km × 0.1229 kgce/kWh

Conversion of standard coal fuel consumption for fuel vehicles:

Standard coal fuel consumption for fuel vehicles (thousand liters)=Fuel consumption for fuel vehicles (liters/100km) × Density of finished oil for vehicles (kg/L) × The conversion coefficient of fuel to standard coal, specific calculation:

Standard coal fuel consumption for fuel vehicles = 39000 L/100 km × 0.725 kg/L × 1.4714 kgce/kg

These two calculation processes involve energy conversion and conversion of standard coal for comparison between different forms of energy. Here, the standard coal consumption of electric vehicles is expressed in kilowatt hours, while the standard coal fuel consumption of gasoline vehicles is expressed in kiloliters. This standardization process helps to compare energy efficiency between different types of vehicles.

Calculate the relative energy savings:

$$\text{Relative energy saving/t kgce} = \frac{\text{Standard coal fuel consumption for fuel powered vehicles}}{\text{Standard coal consumption of electric vehicles}}$$

Table 10: energy saving situation table for new energy vehicles

Motorcycle type	Vehicle ownership	Electric vehicle power consumption (10000 kWh)	Corresponding fuel consumption of fuel models (kL)	Relative energy saving (t kgce)
Bus	205	2255	6150	4003.6705
Taxi	402	1198	6030	4966.8128
Logistics vehicle	2971	14855	59420	45808.3635
Tourist vehicle	139	1223	22872	22253.0071
Sanitation truck	47	94	470	386.4904
Private car	10787	4314	21574	17741.554
Total	14552	23939	116516	95159.8983

It can be seen that in 2018, new energy vehicles can save 95159 tons of standard coal, which has a very positive significance for environmental protection and resource development. In addition, we can also explore the impact of air pollution. The classification criteria are that taxis and private cars meet the design passenger capacity of no more than 6 people (including drivers) and the maximum weight does not exceed 2500kg, using the first category of vehicle standards; The remaining models adopt the second type of vehicle standard, and specific parameters are based on gasoline engine parameters.

Table 11: Vehicle Emission Limit Table

Reference mass (RM) limiting value (g /km)	limiting value (g /km)		
	CO	HC	NOx
First class vehicles	1	0.1	0.08
Second class vehicles (1760<RM)	2.27	0.16	0.11

After calculation, A can reduce CO emissions by 1168.02 tons, HC emissions by 92.84 tons, and NO_x emissions by 92.84 tons, resulting in a total reduction of approximately 1328.7 tons in harmful gas emissions.

Table 12: emission reduction for new energy vehicles

Motorcycle type	Vehicle ownership	Emission reduction of electric vehicles/t		
		CO	HC	NO_x
Bus	205	46.55	3.28	2.26
Taxi	402	60.52	6.05	4.84
Logistics vehicle	2971	731.09	51.53	35.43
Tourist vehicle	139	27.30	1.92	1.32
Sanitation truck	47	6.80	0.48	0.33
Private car	10787	295.76	29.58	23.66
Total	14552	1168.02	92.84	67.84

The driving phase of traditional automotive fuels causes a large amount of CO emissions, which is the main cause of global warming; Because a large amount of NO_x is generated during the fuel and electricity production stages, which is the main cause of acid rain; Unburned HC has a negative impact on the human respiratory system and the environment, indicating that promoting new energy electric vehicles can effectively improve urban energy consumption and environmental issues.

3.6 Question 6

Dear citizens

Hello everyone!

With the increasing severity of global climate change, the advantages of new energy electric vehicles in environmental protection, energy efficiency, and reducing exhaust emissions have become increasingly apparent. Therefore, I hope everyone can have a better understanding and acceptance of new energy electric vehicles, and contribute to protecting the environment and promoting sustainable development.

Compared to traditional cars, new energy electric vehicles have many advantages. Firstly, they do not emit harmful gases and cause minimal environmental pollution. Secondly, electric vehicles are driven by electricity, which has higher energy efficiency and can effectively reduce energy consumption. With the continuous development of electric vehicle technology, its range and performance have also been greatly improved, which can fully meet daily travel needs. According to relevant research, the carbon emissions per kilometer driven by electric vehicles are only one tenth or even lower than those of traditional fuel vehicles. We expect to reduce CO emissions by 1168.02 tons, HC emissions by 92.84 tons, and NOx emissions by 92.84 tons in one year after promoting new energy electric vehicles in cities. This measure will reduce harmful gas emissions by approximately 1328.7 tons in total, which is equivalent to contributing a generous gift to the fresh air in our city.

The energy-saving characteristics of electric vehicles have also been widely recognized, and some models have a range of over 300 kilometers. The use of electric vehicles can also effectively reduce vehicle maintenance costs, as electric vehicles have fewer mechanical components, and the difficulty and cost of repairs are relatively low. The electric vehicle industry has made positive contributions to the economy, environment, society, and other aspects, and has also driven the development of battery production, charging facility construction, and other related industrial chains, creating many new employment opportunities. The popularization of electric vehicles helps alleviate urban air pollution problems, improve the quality of life of residents, promote technological innovation and industrial upgrading, and inject new vitality into the global economy.

Various countries have achieved some successful cases in promoting electric vehicles. For example, countries such as Norway and the Netherlands have achieved large-scale popularization of electric vehicles and continuously promoted their development through policy guidance and infrastructure construction. At the same time, research institutions and enterprises in various countries are constantly innovating and developing technology, promoting the progress of electric vehicle technology and reducing production costs.

Finally, I hope everyone can actively support and participate in promoting the development of new energy electric vehicles. Although the popularization of electric vehicles still faces some challenges, such as charging facility construction, battery technology, and cost issues, with the continuous progress of technology and increased policy support, these problems will be effectively solved. At the same time, we can also start from ourselves and contribute to protecting the environment and promoting sustainable development by purchasing and using electric vehicles.

Let's work together to build a better and more sustainable future! Thank you all!

4 Model Promotion and Evaluation

4.1 Strength and Weakness

(1) Multiple models ARIMA and LSTM were used for comparative analysis, resulting in high reliability of the model, and model optimization and combination were carried out.

(2) Three different correlation analysis methods were used and compared comprehensively with the testing methods, resulting in high reliability of the correlation results.

(3) Using the TOPSIS method to analyze the normalized factors and evaluate the performance of different years on m0ultiple indicators, in order to help determine which years are more in line with expectations or superior.

(5) The running mathematical software is relatively simple and straightforward, and the solution of the established model can be solved using only Python software.

(6) The model considers multiple influencing factors as features and performs feature engineering to solve the problem of data irregularity. It comprehensively considers various information and improves the comprehensiveness of the model.

4.1.2 Weakness

(1) Hypotheses have been made for complex problems, which cannot effectively cover all situations of the problem.

(2) Insufficient time may result in slightly insufficient optimization of the model.

(3) Model limitations: The model is based on current features and data, and may not take into account other potential influencing factors.

REFERENCES

[1] He, R., Zheng, C.-Y., & Zhao, H.-H. (2021). A Novel Incentive Mechanism with Government Subsidy for the Key Technologies R&D of New Energy Automobile Industry. Mathematical Problems in Engineering.

[2] Karimi, H. R., Xi, J., & Yan, F. (2016). Special Issue on "Recent Developments on Modeling and Control of Hybrid Electric Vehicles". Asian Journal of Control.

[3] Mao, Y., Li, P., & Li, Y. (2023). Exploring the promotion of green technology innovation in the new energy vehicle industry: An evolutionary game analysis. Environmental Science and Pollution Research.

[4] Ren, X., et al. (2019). Location of Electric Vehicle Charging Stations: A Perspective Using the Grey Decision-making Model. Energy.

[5] Ye, L., Qin, L., & Jia Hui, L. (2021). The impact of government subsidies on the green innovation capability of new energy automobile companies. IOP Conference Series: Earth and Environmental Science.

[6] Hua, Y., & Dong, F. (2022). How can new energy vehicles become qualified relays from the perspective of carbon neutralization? Literature review and research prospect based on the CiteSpace knowledge map. Environmental Science and Pollution Research.

[7] Xiaoguang, W., et al. (2023). Fuel vehicles or new energy vehicles? A study on the differentiation of vehicle consumer demand based on online reviews. Marketing Intelligence & Planning.

[8] Xu, F., & Zhang, S. (2019). Automobile newborn - new energy vehicle. Journal of Physics: Conference Series.

[9] Yangyang, W., et al. (2023). Price competition and joint energy-consumption reduction technology investment of new energy and fuel vehicles under the double-points policy. Managerial and Decision Economics.

[10] Li, W., Long, R., & Chen, H. (2016). Consumers' evaluation of national new energy vehicle policy in China: An analysis based on a four paradigm model. Energy Policy, 99, 33-41.