International Journal of Electric Power and Energy Studies
ISSN: 3006-2004 (Print), ISSN: 3006-0826 (Online) | Volume 2, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijepes.v2n3.02
Journal homepage: https://wepub.org/index.php/IJEPES/index

Research on the Social Development of New Energy Vehicles based on XGBoost-SHAP and GA-biLSTM Models

Runqing Wang[1, *], Jincheng Zhang[1], Wenyan Shi[2], Xueying Li[1], Zhao Xue[1]

[1] College of Computer Science and Engineering, Dalian Minzu University, Dalian, China
[2] College of Mathematics, Liaoning Normal University, Dalian, China
[*]Corresponding author: Runqing Wang

ABSTRACT

China deeply promotes the implementation of new energy gas vehicle national strategy, new energy automobile industry continues to develop healthily, and gradually become an important force to promote the transformation and upgrading of the global automobile industry. In order to understand the impact of various domestic factors on the development of new energy vehicles, this paper firstly adopts the Delphi method to establish the index system of new energy vehicle development, and establishes the XGBoost-SHAP evaluation model to explore the development of China's new energy vehicles; secondly, it adopts the GA-biLSTM prediction model to forecast the time sequence of the comprehensive development indexes of new energy vehicle industry. Finally, through the social impact analysis of new energy vehicles, a simulation model is established to explore the impact of new energy vehicles on the ecological environment.

KEYWORDS

Delphi Method; XGBoost-SHAP; GA-biLSTM.

1. INTRODUCTION

In 2022, the development of China's new energy automobile industry continued the high-speed growth trend of the previous year, accounting for more than 60% of the global new energy vehicle sales. The development of new energy vehicle industry ushered in the golden period of marketisation [1]. From the export scale, China's new energy vehicles have gradually become an important support for China's auto exports, in 2022 the export li will reach 6.79 million units. From the perspective of technological development, China has become the first country to apply for new energy vehicle battery patents [2]. Now on the domestic factors and market environment on the development of China's new energy automobile industry is discussed.

First of all, we should establish a set of indicator systems for China's new energy vehicle development to determine the main factors affecting the development of China's new energy vehicles. After determining the indicators, we can test the correlation between the indicators and establish the XGBoost-SHAP model to analyze the main factors as characteristic variables. Secondly, we collect the relevant data of the past years and build a time series model to forecast the development indicators of new energy vehicles, and then comprehensively analyze the development trend of China's energy vehicles in the next ten years according to the forecast results. Finally, the impact of electric vehicles on the urban ecosystem is quantitatively analyzed by building a model.

2. EXPLORATION OF NEW ENERGY VEHICLE DEVELOPMENT BASED ON THE XGBOOST-SHAP MODEL

2.1. Correlation Analysis

Firstly, various factors that can cover the development of new energy vehicles are screened, and a set of indicator system for the development of new energy vehicles in China is established through the Delphi method [3].

The Pearson correlation coefficient algorithm is an algorithm used to calculate the strength of the linear relationship between two variables and measure the degree of linear correlation between two random variables.

The Pearson correlation coefficient is calculated as:

$$r=\frac{\sum_{i=1}^{9}(x_i-\frac{\sum_{i=1}^{9}x_i}{10})(y_i-\frac{\sum_{i=1}^{9}y_i}{10})}{9\times\sqrt{(\frac{\sum_{i=1}^{9}\left(x_i-\frac{\sum_{i=1}^{9}x_i}{9}\right)^2}{9})(\frac{\sum_{i=1}^{9}\left(y_i-\frac{\sum_{i=1}^{9}y_i}{9}\right)^2}{9})}} \quad (1)$$

Where $\left(x_i-\frac{\sum_{i=1}^{9}x_i}{9}\right)\left(y_i-\frac{\sum_{i=1}^{9}y_i}{9}\right)$ is the sum of the products of the deviations of X and Y ,is $\frac{\sum_{i=1}^{9}\left(x_i-\frac{\sum_{i=1}^{9}x_i}{9}\right)^2}{9}$ the variance of X, $\frac{\sum_{i=1}^{9}\left(y_i-\frac{\sum_{i=1}^{9}y_i}{9}\right)^2}{9}$ is the variance of Y.

Through the above steps, the Pearson correlation coefficient between the two variables and can be calculated. The Pearson correlation heat map can quantitatively analyze the relationship strength between relevant development indicators of new energy electric vehicles. As shown in figure 1.

There is a high positive correlation between the sales volume and the number of charging piles, which can be used as an important indicator to evaluate the development of new energy vehicles. Secondly, there is a negative correlation between the sales volume of new energy vehicles and the number of their patent disclosures, reflecting some aspects of the development of China's new energy vehicle industry. In addition, it can be found that there is a kind of interdependence and interpenetration between these nine main factors, which are similar in some aspects, or influence and depend on each other in some aspects. They form a close relationship and jointly influence the development of new energy vehicles in China.

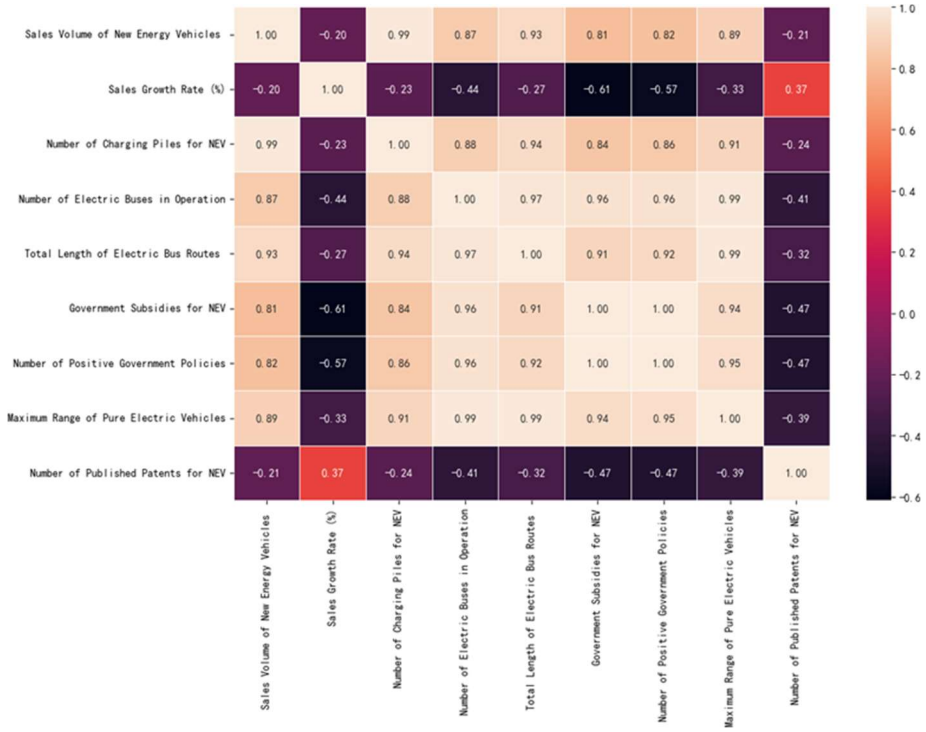

Figure 1. Analysis of correlation

2.2. XGBoost-SHAP Model

Build a prediction model with XGBoost:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F \tag{2}$$

Where, K denotes the number of trees, f denotes a function of the function F space representing an abstract structure like a tree. $\hat{y}_t$ denotes the final prediction result. We take the relevant indexes of China's new energy electric vehicle development as features and the indexes such as development level or sales volume as target variables for training.

We define the objective function as:

$$obj(\theta) = \sum_{i=1}^{10} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

Where l is our loss function and Ω is the penalty term.

In each iteration, XGBoost computes the first and second order derivatives of the loss function with respect to the current model prediction and then constructs a new decision tree based on these derivatives. The updated formula is as follows:

$$y^i(t) = y^i(t-1) + \eta * \sum [g' * h(x_i)] \tag{4}$$

SHAP (SHapley Additive explanations) values are a method for interpreting model predictions, and the SHAP values quantify for each influencing factor the degree to which it contributes to the model's predictions. After the XGBoost model has been trained, the SHAP method is utilized to explain the model's dependency and degree of influence on each influencing factor. As shown in figure 2.

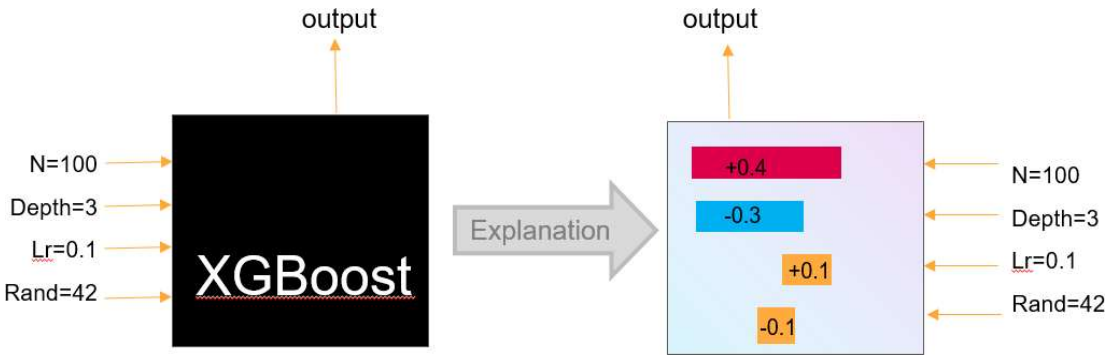

Figure 2. XGBoost-SHAP model

$$\Phi_i(f)=\sum\left[\frac{(|S|!*(n-|S|-1)!)}{n!}\right]*\left[f(S\bigcup i-f(S)\right] \tag{5}$$

The main steps to analyze the importance of indicators using the XGBoost-SHAP model are:

(1) We use the already trained XGBoost model as a parameter passed into the method to create an interpreter (explainer) object.

(2) Use the SHAP model to interpret the feature matrix X and calculate the SHAP value (i.e., feature importance), and analyze the importance of the individual indicators by the SHAP value.

(3) Explain and discuss the influence of individual indicators on the development of new energy vehicles based on the importance analysis.

2.3. XGBoost-SHAP Model Solving

We visualized the predictions of the XGBoost model compared to the actual values to demonstrate the model fit of the model (as shown in Figure 3).

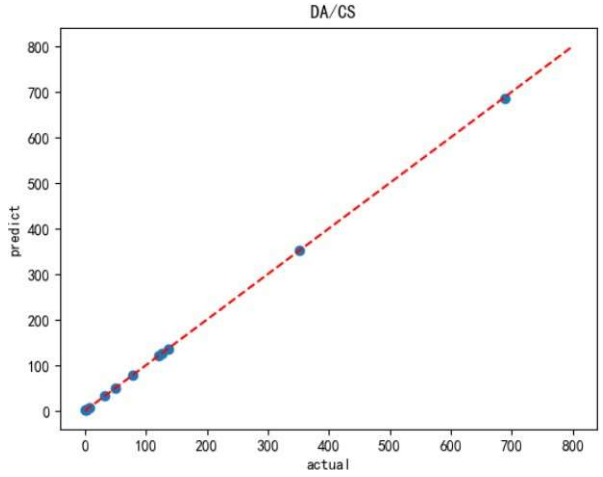

Figure 3. Model fitting renderings

The fitting effect of the XGBoost model, and intuitively shows that the fitting effect between the predicted results and the actual values is very good.

This shows that the XGBoost model is highly accurate in predicting sales volume, effectively capturing trends in the data and providing reliable predictions of actual values. This high fitting performance provides strong support for the credibility and usefulness of the models (show in Table 1).

Table 1. XGBoost fitting effects

	MAPE	R2
XGBoost	3.134686	0.999971

By taking the absolute value of the SHAP value of each sample and then taking the mean value, the SHAP value of each influencing factor was constituted, and a graph of all the importance of the influencing factors was plotted to show the importance of all the influencing factors (as shown in Figure 4).

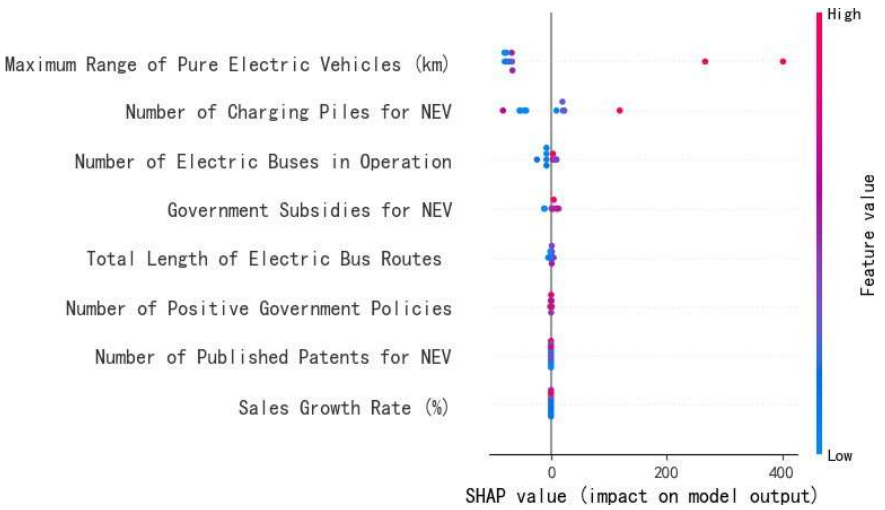

Figure 4. Influencing factors shap value

In summary, the SHAP value analysis shows that among the multiple factors driving the development of China's new energy vehicle industry, the construction of charging infrastructure, the number of new energy vehicles operated in public transportation, and the total length of operating routes are the most critical. Meanwhile, public acceptance and utilization also have some influence, while some other factors play a smaller role in the current model.

3. CONSTRUCTION OF GA-BILSTM BASED PREDICTION MODEL

To use regression models to predict time series data, you usually need to build lag features to use past data to predict future data.

The formula of the hysteresis characteristic is expressed as follows:

$$X_{t-n} = f(X, X_{t-1}, \cdots, X_{t-n+1}) \tag{6}$$

Among them, X_{t-n} represents the eigenvalue at time t, X_{t-n} represents the eigenvalue at time t and f represents the generation function of the lag feature.

In order to predict each indicator for the next 10 years, given the limited data, we decided to use a one-step lag feature construction method. This approach aims to use past data as features in order to build models.

3.1. Establishment of bi-LSTM Model

The bi-LSTM neural network structure model is divided into two independent LSTMs. The input sequences are input to the two LSTM neural networks in forward and reverse order respectively for feature extraction. The two output vectors (the extracted feature vectors) are spliced to form. The word vector serves as the final feature expression of the word. The model design concept of bi-LSTM is to make the feature data obtained at time t possess information between the past and the future [4].

The core structure of LSTM includes three main gate structures: forget gate, input gate, and output gate, as well as a cell state. These structures allow LSTMs to store, modify, and output information. The specific calculation process:

$$f(t) = \sigma(Wfh_{t-1} + Ufx_t + bf) \tag{7}$$

$$i(t) = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{8}$$

$$a(t) = \tanh(W_a h_{t-1} + U_a x_t + b_a) \tag{9}$$

$$o(t) = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{10}$$

f_t represents the forgetting gate, σ is the Sigmoid activation function, x_t represents the input of the current memory unit, i_t represents the input threshold, $\overline{C_t}$ represents the candidate value of the cell state at the moment, $\bar{C}_t$ represents the cell.

3.2. GA Algorithm Model Tuning

Genetic algorithms are a family of search algorithms inspired by the theory of natural evolution. By mimicking the processes of natural selection and reproduction, genetic algorithms can provide high-quality solutions to a variety of problems involving search, optimization, and learning. At the same time, they resemble natural evolution and thus can overcome some of the obstacles encountered by traditional search and optimization algorithms, especially for problems with large numbers of parameters and complex mathematical representations.

The specific operation process is shown in the follows figure 5:

Hyperparameter tuning of LSTM is performed based on GA, and the analysis results are analyzed using 5-layer cross-validation.

For LSTM, since the neural network contains a variety of hyperparameters, the settings of the hyperparameters will seriously affect the experimental results. Appropriate model parameter selection often requires a large amount of experimental verification. After GA algorithm tuning, the LSTM hyperparameter results are obtained as shown below (shown in Table 2).

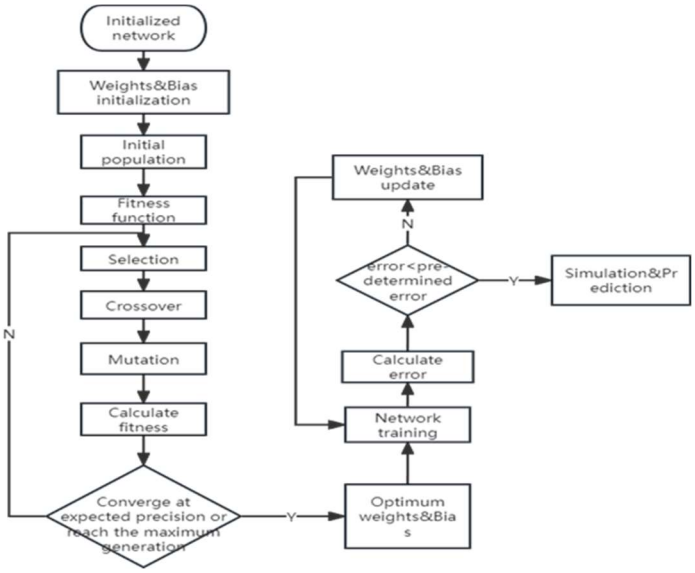

Figure 5. GA model map

Table 2. Parameter Settings

hyperparameters	set	hyperparameters	set
hidden_size	128	batch_size	64
learning_rate	0.001	num_epochs	100
Dropout ratio	20%	optimizer	Adam

The number of LSTM neurons is 128, and the network structure consists of 2 layers of LSTM layers. The output of each layer adopts 20% Dropout regularization. During the training process, each batch size is 64 and trained for 100 epochs. The optimizer uses the Adam algorithm with a learning rate of 10.001. 128 LSTM neurons can provide relatively strong model fitting capabilities. The 2-layer LSTM layer can handle some complex time series patterns. Dropout regularization can improve the generalization ability of the model. the batch size of 64 can ensure a certain training effect without excessive consumption of computing resources. The training time of 100 epochs can ensure that the model can fully learn the characteristics of the time series, while the Adam optimizer and the learning rate of 0.001 can help the model quickly (shown in Figure 6).

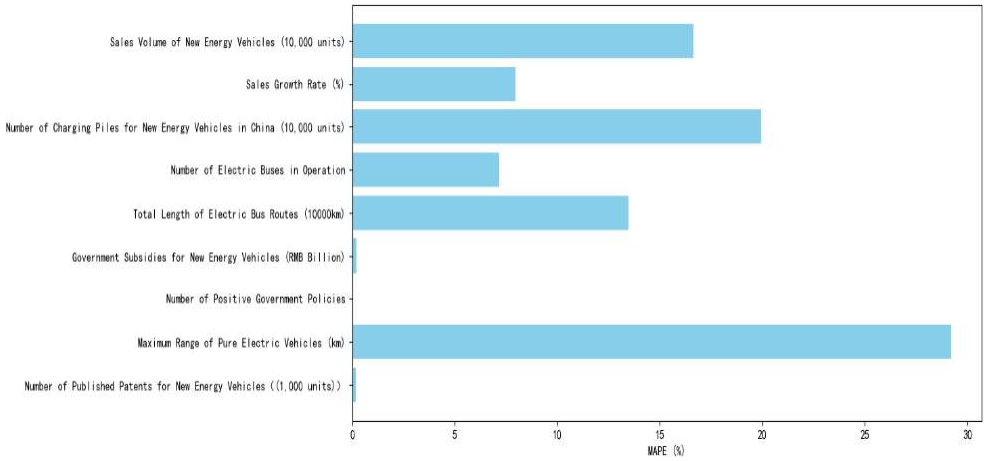

Figure 6. MAPE of different indicators for NEV industry

The figure shows the MAPE (Mean Absolute Percent Error) values, a measure of forecast accuracy, for different metrics. The lower the MAPE value, the more accurate the prediction. It can be observed from the chart:

The MAPE value of the maximum cruising range of electric vehicles is the highest, indicating that the prediction error of this indicator is relatively large. It may be that the cruising range is affected by many factors, such as technological development, policy changes, etc., making its prediction highly uncertain.

The MAPE value of the number of active government policies is the lowest, indicating that the prediction accuracy of this indicator is very high. This may be because the issuance of policies is less affected by market fluctuations and external factors.

Forecasting the next 10 years, the results are as follows (shown in Figure 7):

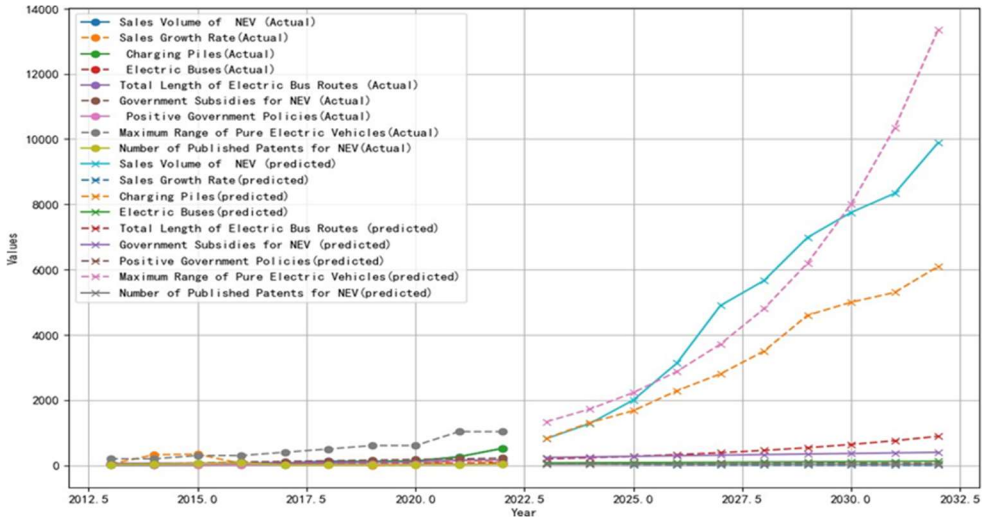

Figure 7. Electric vehicle data

According to the figure above, we can observe the future development of various indicators in China's new energy vehicle (NEV) industry:

The rapid development of charging infrastructure has provided strong support for the widespread use of new energy vehicles, and the sales growth rate is expected to gradually decline, possibly because the market is gradually becoming saturated. The government may gradually increase the amount of government subsidies for new energy vehicles based on market demand and industrial development to better support the development of the industry. Over time, battery technology continues to make breakthroughs, including increased energy density, improved charge and discharge efficiency, and extended battery life. These technological advances allow automakers to produce higher-performance battery packs that increase the range of pure electric vehicles. In summary, China's new energy vehicle industry will continue to maintain strong growth momentum in the next decade, with infrastructure construction and market acceptance being the key factors driving this trend.

4. ANALYZING THE IMPACT OF NEW ENERGY VEHICLES ON THE ECOLOGICAL ENVIRONMENT THROUGH SIMULATION MODELING

Through the research and related literature, it is concluded that the model focuses on three key environmental impact factors: carbon dioxide emissions, air quality (PM2.5 and PM10 concentrations), and noise pollution [5].

Reduction of carbon dioxide: Knowing that the essence of new energy vehicles is the electrification of vehicles, we define the electrification rate of vehicles as the ratio of the number of electric vehicles to the total number of vehicles, which is:

$$Electri\ \ fication\ \ rate\ \ of\ \ vehicle = \frac{Total\ \ number\ \ of\ \ electric\ \ vehicles}{Total\ \ number\ \ of\ \ cars} \times 100\% \quad (11)$$

We give the average annual CO_2 emissions of a fuel vehicle as N ton, and the average annual emissions of an electric vehicle as M. Let the total number of electric vehicles be Z, then we can estimate the total value CO_2 of reduction from electrification of all vehicles, that is:

$$Reduction\ \ of\ \ CO_2 = Z \cdot (M - N) \quad (12)$$

Optimization of air quality: It can be found that air quality is closely related to the emission of pollutants, the popularity of electric vehicles, the lower the emission of pollutants, so based on these two variables to establish an air quality optimization model:

$$\begin{aligned} &Air\ \ quality\ \ optimization \\ &= f(Total\ \ number\ \ of\ \ electric\ \ vehicles, Pollutant\ \ discharge) \end{aligned} \quad (13)$$

Where f denotes the impact function of the increase in the total number of electric vehicles on air quality optimization.

Reduction of noise pollution: electric vehicles produce lower noise levels, and when the total number of electric vehicles increases, the noise level decreases accordingly, then a model of the degree of noise pollution reduction is established based on these two variables:

$$\begin{aligned} &Noise\ \ pollution\ \ reduction \\ &= g(Total\ \ number\ \ of\ \ electric\ \ vehicles, Noise\ \ level\ \ di\ \ f\ \ ference) \end{aligned} \quad (14)$$

Where g denotes the impact function of the increase in the total number of electric vehicles on the reduction of noise pollution.

In summary, we evaluate the overall impact of new energy-electric vehicles on the urban environment of millions of people based on the above three influencing factors and establish a mathematical model:

$$\begin{aligned} &Total\ \ environmental\ \ impact \\ &= h(CO_2 reduction, Air\ \ quality\ \ optimization, Noise\ \ pollution\ \ reduction) \end{aligned} \quad (15)$$

Where h is a function of an integrated assessment of environmental impacts.

After solving by assumptions, it is understood that electric vehicles have a significant role in reducing greenhouse gas emissions. The widespread use of electric vehicles can improve urban air quality, thereby enhancing the overall quality of the living environment. However, to achieve these goals, the actual effect depends on a variety of factors, including the utilization rate of electric vehicles, the cleanliness of electricity and urban traffic management. Through the above measures, it is expected that the role of electric vehicles in reducing greenhouse gas emissions will be further realized and the quality of the urban living environment will be enhanced.

5. CONCLUSION

The study first Delphi method to establish a set of China's new energy vehicle development indicator system, the establishment of the XGBoost-SHAP model research found that the new energy tram-related technology innovation, infrastructure improvement, and policy support plays a key role in its development. Afterward, the GA-bLSTM model was used to forecast the time series of the indicators for the comprehensive development of China's new energy vehicles, and it was found that China's new energy vehicle industry will continue to maintain strong growth momentum in the next ten years. A simulation model is established to explore the impact of new energy-electric vehicles on the ecological environment. The quantitative results show that the popularity of electric vehicles will effectively reduce the level of urban environmental pollution.

REFERENCES

[1] Zhang Yejia. Advantages and Worries of China's New Energy Vehicles: Current Situation and Suggestions for the Development of China's New Energy Vehicle Industry[J]. Intelligent Networked Vehicles, 2023,(05):64-67.

[2] Bai Wanrong; Wei Feng; Zheng Guangyuan; Wang Baohui. Research on TCN-BiLSTM based intrusion detection algorithm[J]. Computer Science,2023,50(S2):941-948.

[3] Xu Dan; Guo Dandan; Liu Jing; Zhang Yanwen; Meng Yuan; Zhang Boxiong; Deng Liting. Construction of evaluation index system for water conservancy scientific and technological achievements in Hebei Province based on Delphi method[J]. Hebei Water Conservancy,2023,(10):36-38+42.

[4] Xiao, H. P.; Wang, S. H.; Chen, L. L.; Fan, Y. C.; Wan, J. H.. An optimized network model for slope deformation prediction fusing GA and LSTM and its application[J/OL]. Geodesy and Geodynamics,1-8[2023-11-26]https://doi.org/10.14075/j.jgg.2023.08.152.

[5] Song Xiaoming; Qin Jiarui; Xing Yingchun; Zhou Xiaoyan. Research on the development of China's new energy vehicle industry chain under the background of "double carbon"[J]. Value Engineering,2023,42(22):166-168.