Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 2, No. 3, 2024

New Energy Electric Vehicle Industry Development Trends Forecast

-- A Ten-Year Outlook for China Based on MLPRegressor

Yiying Liu*, Shichang Sun, Hongxi Liu, Liangyang Xu

College of Computer Science and Engineering, Dalian Minzu University, Dalian, China
* **Corresponding author**: Yiying Liu (Email: 2316542792@qq.com)

Abstract: This study focuses on developing the MLPRegressor model for predicting the future trends in China's new energy electric vehicle industry over the next decade. We employed a dataset comprising 16,500 records, encompassing 12 features, such as the number of public electric vehicles, the length of electric vehicle operational routes, and the number of electric vehicles owned by residents, for model training. We compared our model with traditional models like ARIMA, XGBRegressor, and Random Forest Regressor, using MAPE and MSLE as evaluation metrics. The final results revealed that our model exhibited superior performance in terms of loss, outperforming the other models significantly. This research not only provides businesses with more accurate strategies for developing the new energy electric vehicle industry but also serves as a crucial reference for the government in formulating environmentally friendly policies.

Keywords: MLPRegressor, new energy electric vehicle, MAPE, MSLE.

1. Introduction

New energy vehicles, in particular pure electric vehicles, plug-in hybrid electric vehicles and extended-range electric vehicles, which are powered by renewable energy sources[1], are gradually becoming the focus of the global automotive market due to their environmental and energy-saving characteristics, especially in China, the world's largest automotive market.This not only reflects the growing concern for environmental protection, sustainable development and energy efficiency, but also represents technological advancement and a shift in consumption patterns.

As electric vehicle technology continues to evolve and mature over time, coupled with strong government policy support, the challenges facing the market, such as battery range and charging speed issues, are gradually being addressed[2]. In 2020, the global sales of new energy vehicles reached 3.75 million units, with China occupying more than half of the market share，demonstrating its leading position in the automotive industry and highlighting the immense potential and growth space in the new energy vehicle market.

Against the escalating global climate crisis, predicting and analyzing the future development trajectory of China's new energy electric vehicles over the next decade is crucial for guiding industry development strategies and optimizing resource allocation. This study employed the MLPRegressor model to forecast China's ten-year development of new energy electric vehicles. Compared to traditional methods like ARIMA, XGBRegressor, and Random Forest Regressor, MLPRegressor offers higher flexibility and accuracy. ARIMA exhibits limitations in handling non-linear relationships and high-dimensional data. XGBRegressor requires manual parameter tuning and is unsuitable for high-dimensional data. Random Forest Regressor is prone to overfitting in certain situations. The deep neural network structure of MLPRegressor better captures complex non-linear relationships, providing higher flexibility and accuracy, which contributes to the more scientific formulation of electric vehicle development strategies, achieving sustainable development and environmental goals.

2. Dataset Processing

2.1. Dataset

We have collected a dataset comprising 16,500 records, with each entry encompassing 12 crucial features. These data cover vital factors influencing the usage of electric vehicles from 2012 to 2022. The features include global production of new automobiles, sales, and ownership of energy vehicles, public vehicles, the length of electric vehicle operational routes, the number of electric vehicles owned by residents, and public charging station numbers (categorized into public, DC, AC, and integrated DC-AC chargers). These data will be utilized in our study for an in-depth analysis of the development trends in the electric vehicle market.

2.2. Dataset Processing

Missing Value Handling:

In the research on the development of new energy electric vehicles, there are some deficiencies in the number of vehicles and charging piles in 2018. We used K-means clustering multiple interpolation method to fill in these missing values, and ensure the rationality of the data through correlation analysis. Ultimately, utilized a weighted averaging approach, we derived that in China, there were 0.06 million integrated DC-AC charging stations in 2021, 805,000 public charging stations in 2020, and the ownership of new energy vehicles was 2.613 million in 2018. This data processing method significantly enhanced the reliability and accuracy of the research results.

Outlier Handling:

Given the diversity of data sources and to ensure the collected automotive data is at an average level, we implemented an outlier detection method. Specifically, we utilized the Pandas library to process the automotive data and conducted statistical analyses on global car production data. In the outlier testing, data points exceeding one standard deviation above or below the overall mean were considered outliers, as shown in the following formula.

$$A_v = \{x | x > (\mu + k \times \sigma) \text{ or } x < (\mu - k \times \sigma)\} \quad (1)$$

Where A_v represents outliers, k is the multiple of standard deviation used to determine the outlier threshold, and σ and μ denote the standard deviation and mean, respectively. Additionally, for ease of observation and visualization, Figure 1 visually presents the results of outlier detection.

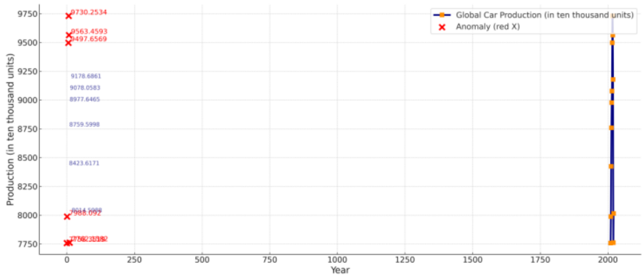

Figure 1. Conceptual Diagram of Outlier Analysis in Global Car Production.

We will further evaluate the existing outliers' authenticity and then employ a refined approach involving outlier correction and group processing to enhance the optimization process.

3. Construction of MLPRegressor

In the study of predicting the future use of new energy electric vehicles in China, MLPRegressor is a machine learning model based on feedforward neural networks. It is mainly used to solve the regression problem, that is, to predict the use of new energy electric vehicles in the future by learning historical data. MLPRegressor consists of an input layer, multiple hidden layers, and an output layer, each part with a specific function[3].

The input layer receives data on factors influencing the development of electric vehicles, such as vehicle ownership and the number of charging stations. The hidden layers, through multiple layers of neurons, perform non-linear transformations, learning and capturing high-dimensional features to enhance the model's expressive capability. The output layer provides the final usage predictions. MLPRegressor is trained using the backpropagation algorithm, optimizing model parameters to minimize the error between predicted and actual values.

3.1. Forward Propagation

Forward Propagation is a critical step in neural networks. Its structure is illustrated in Figure 2.

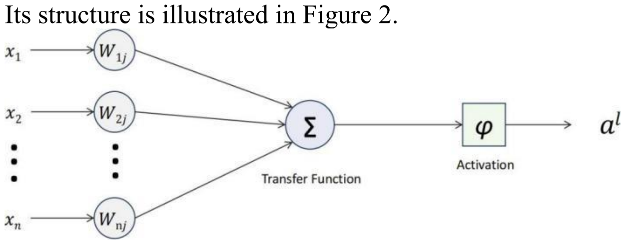

Figure 2. Structure of Forward Propagation in MLPRegressor

It takes features such as the number of charging stations in China, the ownership of electric vehicles by residents, and the operational routes of public electric vehicles as input data, which are then transmitted to the linear layer to generate the final results. The predicted values are calculated during training. The forward propagation formula for an L-layer network is as follows.

$$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]} \quad (2)$$

$$a^{[l]} = softmax^{[l]}(z^{[l]}) \quad (3)$$

The symbol $z^{[l]}$ represents the weighted sum for the Lth layer, $a^{[l-1]}$ is the normalized representation of features from the previous layer, and W and b denote the weight and bias, respectively. Additionally, we employed Leaky ReLU as the activation function, specifically setting the slope parameter, denoted as t, to 0.01. This configuration not only effectively avoids the gradient vanishing issue inherent in traditional ReLU activation functions but also enhances the model's ability to capture subtle variations in the data by preserving partial information from negative inputs.

$$f(x) = \begin{cases} x & if \ \ x \geq 0 \\ tx & if \ \ x < 0 \end{cases} \quad (4)$$

3.2. Model Training

MLPRegressor employs Mean Squared Error (MSE) as its loss function to quantify the difference between actual and predicted values. MSE imposes greater penalties for larger prediction errors, making the model more sensitive and accurate when predicting the future volume of new energy electric vehicles in China over the next decade. This approach not only aids in capturing and correcting biases in the model's predictions but also enhances the reliability of the forecasts.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (5)$$

Here, y_i and $\hat{y}_i$ represent the actual and predicted values, respectively. We conducted 120 iterations of model training and performed a visual analysis of the results, as depicted in Figure 3. Ultimately, we identified the optimal model at the 101st iteration.

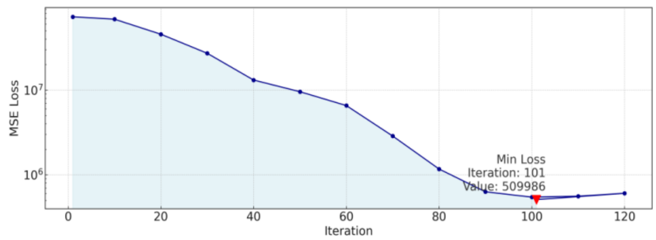

Figure 3. Variation of MSE Loss with Different Iteration Counts

3.3. Backpropagation

In the study predicting the future production of electric vehicles in China, the weights and biases in the MLPRegressor model are optimized through the backpropagation algorithm. Utilizing the chain rule, the error from the output layer is efficiently propagated to the internal layers of the network, ensuring a continual reduction in error throughout the entire training process. This sustained self-optimization process contributes to more accurate predictions when dealing with new and unseen data. Specifically, when following the steps outlined below, we can compute the error at the output layer and further propagate the error to obtain the error for each layer. The formulas are as follows:

$$\frac{\partial \mathrm{Loss}}{\partial \mathrm{z}^{[l]}} = \left(\frac{\partial \mathrm{Loss}}{\partial \mathrm{a}^{[l]}}\right) \cdot \left(\frac{\partial \mathrm{a}^{[l]}}{\partial \mathrm{z}^{[l]}}\right) \quad (6)$$

We utilize error signals to minimize the loss function to update the weights and bias parameters for each layer, which is a critical learning process for neural networks and is accomplished through the following two formulas:

$$W^{[l]} = W^{[l]} - \alpha \frac{\partial Loss}{\partial W^{[l]}} \quad (7)$$

$$b^{[l]} = b^{[l]} - \alpha \frac{\partial Loss}{\partial b^{[l]}} \quad (8)$$

4. Solution and Performance Analysis of the MLPRegressor Model

Our study considered several vital factors to forecast the development of China's new energy electric vehicles over the next decade. These factors encompass the number of public buses and electric vehicles, the nationwide ownership of new energy vehicle charging stations, the year-end average ownership of electric-assisted bikes per hundred households, the ownership of operating public buses, the passenger volume of buses and trams, the total mileage of operating bus and tram routes, as well as sales figures and sales growth rates.

For a more intuitive prediction of the future trends in China's new energy vehicle industry over the next decade, we selected six representative influencing factors for visual analysis. The detailed records of these analyses are presented in Figures 4, Figures 5, and Figures 6.

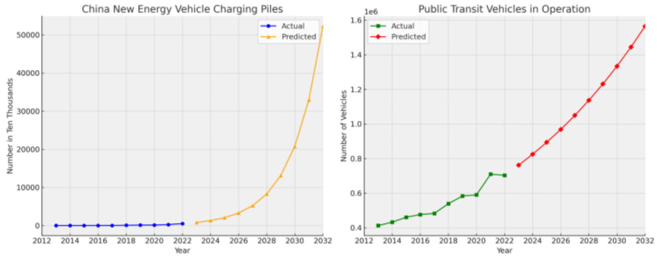

Figure 4. Forecast Analysis of Charging Stations and Public Bus Operation

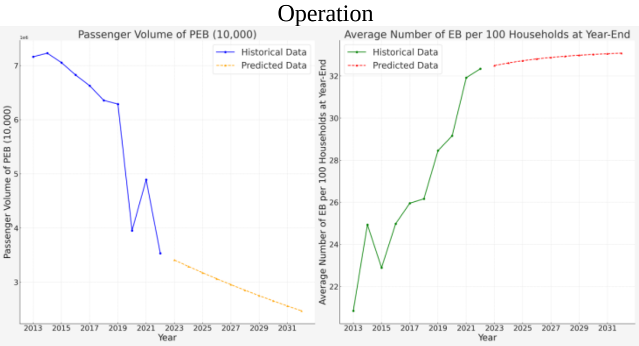

Figure 5. Forecast Analysis of Public and Residential Trams

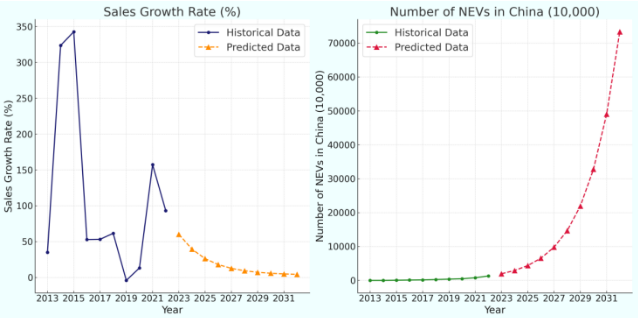

Figure 6. Forecast Analysis of Sales Growth Rate and Inventory Relevance

Our predictions indicate that the Chinese new energy vehicle market will experience significant growth over the next decade. This growth encompasses the continual expansion of charging infrastructure, an increase in electric public transportation, and the widespread adoption of personal electric mobility.

5. Comparative Experiments

To effectively predict the future trends in China's electric vehicle industry over the next decade and validate the predictive efficacy and uniqueness of our model based on features such as the number of charging stations, household electric vehicle ownership, and the number of electric vehicles in public transportation, we conducted a series of comparative experiments. Our model was meticulously compared in these experiments with ARIMA, XGBRegressor, and Random Forest Regressor. We utilized MAPE (Mean Absolute Percentage Error) and MSLE (Mean Squared Logarithmic Error) as performance evaluation metrics to comprehensively analyze and showcase the performance of each model in terms of prediction accuracy and reliability. All models in the comparative experiments were trained based on the same data features. We selected the optimal performance achieved by each model as the evaluation metric, and the results are presented in Table 2:

Table 2. Comparison of Model Performance Metrics

Model	ARIMA	MLPRegressor	XGBRegressor	RForest Regressor
MAPE(%)	*11.89*	*9.44*	*11.89*	*12.25*
MSLE	*0.0190*	*0.0098*	*0.0179*	*0.0982*

From the above diverse model prediction results, MLPRegressor demonstrates robust performance in forecasting the development of China's new energy electric vehicles over the next decade. Specifically, in terms of the MAPE evaluation metric, MLPRegressor exhibits an error rate of 9.44%, which is lower compared to ARIMA at 11.89%, XGBRegressor at 11.89%, and Random Forest Regressor at 12.25%. Regarding MSLE, MLPRegressor outperforms other models with an impressively low value of 0.0098, which indicates that considering various factors such as ownership, the number of charging stations, public vehicles, the length of electric vehicle operation routes, and the number of electric vehicles owned by residents, MLPRegressor has a significant advantage in prediction accuracy and error control.

6. Conclusion

In this study, we aimed to accurately predict the future trends of China's new energy electric vehicles over the next decade, involving multidimensional features such as the number of charging stations, the number of electric vehicles in public transportation, the length of electric vehicle operation routes, and the private ownership of electric vehicles by residents. We utilized the MLPRegressor model for predictions and validated the superior performance of our model through comparisons with classical models like ARIMA, XGBRegressor, and Random Forest Regressor, based on the MAPE and MSLE metrics. This outcome showcases the efficient accuracy of the MLPRegressor model in the predictive analysis of the new energy electric vehicle sector, holding practical significance for guiding the industry's future development directions and resource allocations.

References

[1] Li C ,Feng D ,Chengfeng Z , et al.Spatial distribution patterns and influencing factors of China's new energy vehicle industry[J].Journal of Cleaner Production,2022,379.

[2] Yang G .The Effect of Government Subsidies on Innovation Capability of the New Energy Vehicle Industry[J].Nature Environment and Pollution Technology,2023,22(4):1857-1866.

[3] Jürgen Schmidhuber.Deep learning in neural networks: An overview[J].Neural Networks,2015.