

Analysis of New Energy Electric Vehicle Development in China Based on Machine Learning

Danqing Wang^{#1,*}, Chengan Ouyang^{#,2}, Xinyang Peng^{#,3}

¹ Communication University of China, Beijing, China, 100024

² Southwest Forestry University, Kunming, China, 650224

³ Hangzhou Normal University, Hangzhou, China, 311121

* Corresponding Author Email: wdq992626988@outlook.com

[#]These authors contributed equally

Abstract. New energy vehicles are an inevitable trend in the future development of the automotive industry. Therefore, it is important to study the development prospects of new energy vehicles. This paper analyzes the main factors of new energy vehicles through the method of machine learning, and predicts the development of new energy vehicles in the next decade. The research reveals that international competition and collaboration, government policies, and technological advancements are the primary factors influencing new energy vehicles. At the same time, although the profit growth rate of new energy vehicles occasionally slows down, it still maintains a growth trend overall. It is expected that the research in this paper can contribute to the further development of the new energy vehicle industry.

Keywords: New energy electric vehicles, Machine learning, Prediction, Stacking.

1. Introduction

The automobile industry is an inseparable part of a country's economy, and as more and more countries pay more attention to environmental issues, the development of new energy vehicles has become an inevitable trend^[1,2]. The topic of new energy vehicles, as a significant trend for future development, has long captivated the attention of numerous scholars. The automobile industry is an inseparable part of a country's economy, and as more and more countries pay more attention to environmental issues, the development of new energy vehicles has become an inevitable trend^[3,4].

The topic of new energy vehicles, as a significant trend for future development, has long captivated the attention of numerous scholars. As early as the last century, McCarthy P S.^[5] conducted research on the relationship between car sales volume and factors such as income, quality, and market participants. Greenspan A. and Cohen D.^[6], on the other hand, employed statistical methods to forecast future car sales. In this century, Jingzheng Ren^[7] utilized a combined approach of Interpretive Structural Modeling (ISM) and fuzzy Decision-Making Trial and Evaluation Laboratory (DEMATEL) to investigate the key factors influencing the development of new energy vehicles. Ou, X.^[8] Based on an analysis of the development trends in vehicle technology, vehicle price, vehicle fuel efficiency, and fuel supply prices, the study employed life time cost model to predict the scale of development for New Energy Vehicles (NEVs), specifically passenger cars, across different scenarios.

However, few scholars have undertaken a comprehensive exploration of the factors influencing the development and future prospects of new energy vehicles using machine learning models. Therefore, this paper will study the main factors driving the development of new energy vehicles and the prospects for the development of new energy vehicles based on machine learning. It is hoped that this paper can provide a feasible research method and provide ideas for promoting the further development of new energy vehicles.

2. Model preparation

2.1. DBSCAN and UMAP model

Taking China as an example, firstly, this paper constructs the development loss parameters of new energy electric vehicles through the existing data of new energy electric vehicles. For special indicators that have an impact on the development loss data of new energy electric vehicles, such as the sales growth rate of new energy electric vehicles, the increase in losses, the current proportion of new energy electric vehicles, the proportion of vehicles in total, the total asset loss, the asset loss rate and other parameters.

Jump amplitude:

$$C = \frac{T_1 - T_0}{T_0} + \Delta T \quad (1)$$

Metric correlation:

$$N = 1 - \frac{6 \sum di^2}{n(n^2 - 1)} \quad (2)$$

Anomaly score:

$$r_{ij} = \max \left\{ \frac{y_{i+1} - y_1}{y_{n-j} - y_1}, \frac{y_n - y_{n-i}}{y_n - y_{j+1}} \right\} \quad (3)$$

Persistence score:

$$Q_{NC} = E^2 + \frac{Q_0 - Q_1}{\Delta Q} \quad (4)$$

where y, T are random variables, i, n, j are random real numbers, and Q is the output value, E is loss function.

In order to effectively analyze the composition of the development parameters of new energy electric vehicles, the loss parameters of 126,601 new energy electric vehicles are processed by us based on DBSCAN algorithm, the data are mainly from the National Bureau of Statistics. In the DBSCAN algorithm, all samples whose density can reach the core object are found to form "clusters". According to the given neighborhood parameters Eps and Minpts, all the core objects are determined, and an unprocessed core object was selected for each core object, and a clustering "cluster" was generated from the samples whose density can be reached. Here, only the new energy electric vehicle loss parameter data DBSCAN center was explained.

Table. 1 DBSCAN center distribution

Number	Center 1	Center 2
H1	0.830949554	0.202297344
H2	0.536776037	0.13063839
H3	0.265030766	0.018999691
H4	1.537390804	0.564216366
H5	0.309729357	0.008865604
H6	0.090619698	0.001839619
H7	0.072795877	0.00333293
H8	0.400266647	0.008082933
H9	0.281236523	0.010664473
H10	0.34901391	0.202297344

The specific distribution of each center can be observed from Table 1, and the clustering of each set of data can be inferred. Then, the new energy electric vehicles loss parameter data was analyzed by UMAP algorithm, which is convenient for processing the relationship between the data parameters. In practice, since the fuzzy set membership gradually decays to almost zero, only the membership

degree of the closest neighbor of each point needs to be calculated. A fast and efficient way to calculate (approximate) nearest neighbors is required, even in high-dimensional data spaces. In practice, using the curve family of $\frac{1}{1+ay^{2b}}$, low-dimensional representations can be initialized using spectral embeddings since the Laplace-Beltrami operator for topological representations is an approximation of the Laplace-Beltrami operator of manifolds:

$$\rho_i = \min \{d(x_i, x_j) \mid 1 \leq j \leq k, d(x_i, x_j) > 0\} \quad (5)$$

$$\sum_{j=1}^k \exp \left(\frac{-\max(0, d(x_i, x_j) - \rho_i)}{\sigma_i} \right) = \log_2(k) \quad (6)$$

where each x_i , identify ρ_i and i, k, j are all arbitrary real numbers.

Calculation of distribution:

$$p_{i|j} = \exp \left(\frac{-\max(0, d(x_i, x_j) - \rho_i)}{\sigma_i} \right) \quad (7)$$

$$p_{ij} = p_{i|j} + p_{j|i} - p_{i|j} p_{j|i} \quad (8)$$

The weight values of each new energy electric vehicles loss parameter are obtained, and then the relationship between the new energy electric vehicles loss parameters was adjusted. The extent to which each component contributes can be seen by the total variance explained, as well as the percentage of variance, the required component was charged in excess of 82% cumulatively, after which the projection of the numerical data set was calculated and derived. According to the coefficient factor, and then according to the final situation of the accumulation of spatial distribution, and the projection map to define the distribution, the final definition is in the impact factors A, B, C, D, E, F, G, and the values of the factors are 0.25, 0.14, 0.61, 0.24, 1.23, -0.72, -1.03.

By analyzing and processing the distribution of development parameters of each new energy electric vehicle, the mixed exponent V was synthesized by our simulation. First, the characteristic values are decomposed:

$$J(a, b) = a^T S_{XY} b - \frac{\lambda}{2} (a^T S_{XX} a - 1) - \frac{\theta}{2} (b^T S_{YY} b - 1) \quad (9)$$

The associated target is then optimized by the Lagrangian coefficient:

$$S_{XX}^{-1} S_{XY} b = \lambda a \quad (10)$$

$$S_{YY}^{-1} S_{YX} a = \lambda b \quad (11)$$

Where S_{xx} is denoted as the sum of squares of deviations of x , S_{yy} is denoted as the sum of squares of deviations of y , and S_{xy} is denoted as the deviation product sums of x and y , a, b are arbitrary real numbers.

Through the analysis of the above weights and the three mixed cases of the new energy electric vehicle loss parameter index, the general fluctuation of the loss parameters related to new energy electric vehicles was finally discovered. Here, 5 time points, 10 time points, 20 time points, and 30 time points are used as examples. Firstly, a case of three wave parameters of mixed parameters was calculated.

Table 2 fluctuation parameters under different time divisions

Times	V ₁	V ₂	V ₃	V ₄
5X	-0.726	-0.198	0.292	0.316
10X	0.411	0.08	0.853	0.546
20X	0.075	0.88	0.201	-0.378
30X	-0.843	0.298	0.104	0.241

As illustrated in the Table 2, the positive and negative cases and magnitude of the V mixed index at different multiples can be inferred, which lays a foundation for the construction of the model in the future.

This is where the time-domain coefficient T comes in and the principle of slight local denoising is used by us:

$$y = \alpha P_i + n_i, P_i \sim P_{(x)}, N_i \sim N_{(0, \sigma^2)} \quad (12)$$

After the local prominence is obtained, the loss of the local rate of change is studied:

$$L_1(X, Y; \theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|y_i - \hat{y}(x_i)\|_2^2 + \beta KL(\hat{\rho} \| \rho) + \frac{\lambda}{2} (\|W\|_F^2 + \|W'\|_F^2) \quad (13)$$

Finally, the T distribution at different times is established:

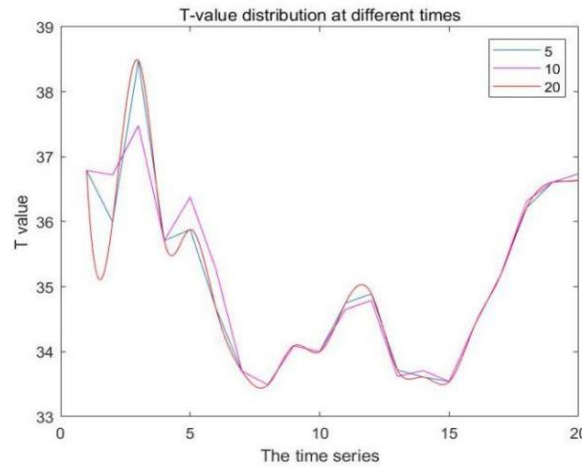


Fig. 1 Distribution of T values over time

As shown in the Fig.1, although at different multiples, the T -values are all relatively close. Therefore, in the process of fitting, the model will still have a good fit when dealing with the values of different multiples. The construction of the DBSCAN model has been completed, and the model is relatively stable and has good fit.

2.2. Fusion prediction model

On the basis of the Meta-model, a future trend algorithm for regional decision-making of new energy electric vehicles was established for the analysis of future factors. By analyzing the adaptability of machine learning models to data, AUC was analyzed and compared, and the results are shown in Table 3.

Table 3 Comparison of AUC conditions of each algorithm

Algorithm	Before the reference is adjusted	After tuning parameters
CATboost	0.8131	0.8402
XGBoost	0.8216	0.8414
Lightgbm	0.8017	0.8206
LR	0.7931	0.8310

From the Table 3, it can be inferred that although the performance of each individual model has been improved to a certain extent after tuning, the overall fitting effect is not particularly ideal, so it was decided to choose the idea of fusing machine learning models for prediction, so an ensemble classification model based on stacking ensemble learning was chosen. Finally, this paper adopts the ensemble classification model of Stacking ensemble learning, selects the Xgboost, Catboost, LR and Lightgbm algorithms with higher AUC on the basic classifier in the first layer, and selects the Lightgbm algorithm in the second layer, so that the processing can better realize the establishment of a risk anomaly early warning model, and the AUC of the fused machine algorithm will be more ideal.

The final output AUC was shown below, and the stacking ensemble model scores higher on the test data than all the basic classifiers. Moreover, the AUC of the ensemble model is 86.37, which also indicates that the model is relatively stable and there is no overfitting state.

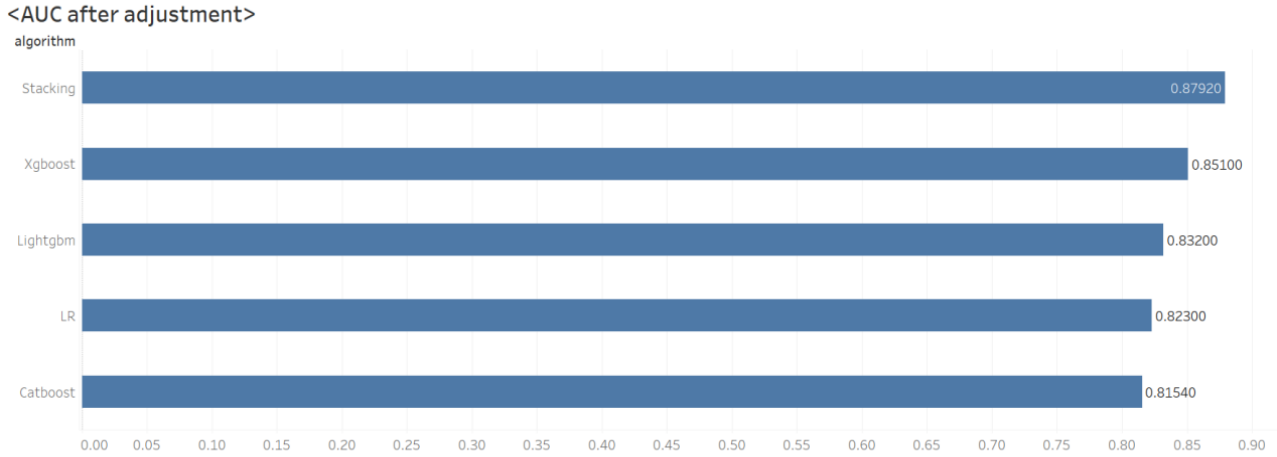


Fig. 2 Comparison of the adjusted AUC conditions of each algorithm

It can be inferred from Fig. 2 that the stacking model integrates the performance advantages of the four models in the Table 3 and fully exploits their respective advantages, so the stacking model constructed has better effect and strong generalization ability. After scrutinizing the model's sensitivity and specificity, the algorithm was parameterized to evaluate and adjust the stability of the model. The AUC of the algorithm after parameter adjustment is 86.31%, which is also close to 0.88, and the authenticity of the algorithm is relatively high, which indicates that the model has a good fitting effect.

After that, in order to perform a detailed correlation analysis on the time period, the neural network algorithm was built to establish the parameter Q for research. In order not to pay too much attention to particularity, set here that the hidden layer has a total of L nodes, the input layer has M nodes, the number of samples is W , and the output layer has P nodes, then the input vector is $[x_1, x_2, x_3, \dots, x_M]$, W_{ij} is the node at the input layer, and the node connection weight value at the hidden layer is respectively, then the input and output of the hidden layer node are:

$$net_j = \sum_{i=1}^N w_{ji} x_i, O_j = f_i(net_j) \quad (14)$$

The input layer nodes after that are:

$$y_k = f_2 \left(\sum_{j=1}^L w_{jk} O_j \right) \quad (15)$$

After that, the mean squared error analysis is carried out, and its formula is:

$$MSE = \frac{1}{N} \sum_{t=1}^N (O_t - Y_t)^2 \quad (16)$$

where O_t is the predicted output value of the neural network, Y_t is the real value, and N is the number of samples.

This paper defines the decision accuracy P_{bp} of BP networks:

$$P_{bp} = 1 - \frac{\sum_t |O_t - Y_t|}{\sum_t |Y_t|} \quad (17)$$

Self-built point accuracy P_Z :

$$P_{bp} = 1 - \frac{\sum_t |O_t - Y_t|}{\sum_t |Y_t|} \quad (18)$$

After that, BP neural network was used to train the data, and finally the correlation model between the decision coefficient and the power generation margin of new energy electric vehicles in two time periods was established.

3. Results and discussion

Based on the above model establishment rules, The impact model of new energy electric vehicles was established, and the impact of different factors on new energy electric vehicles was analyzed according to this model, and the preliminary rules were obtained, and the classification was given. The characteristics of different new energy electric vehicles are described, q is different from the prediction parameter Q above, where q is the evaluation score of each index.

Table. 4 characteristics affecting the development categories of new energy electric vehicles.

Series Index q	Grade	Specific characteristics
$q < 87.5$	Excellent	Government policy support
$87.5 < q < 147$	Good	Technological innovation and R&D investment
$147 < q < 300$	Mild	Market demand
$q < 87.5$	Moderate	Industrial chain construction
$87.5 < q < 147$	Severe	International cooperation and competition
$147 < q < 300$	Serious	Urban management and policy guidance

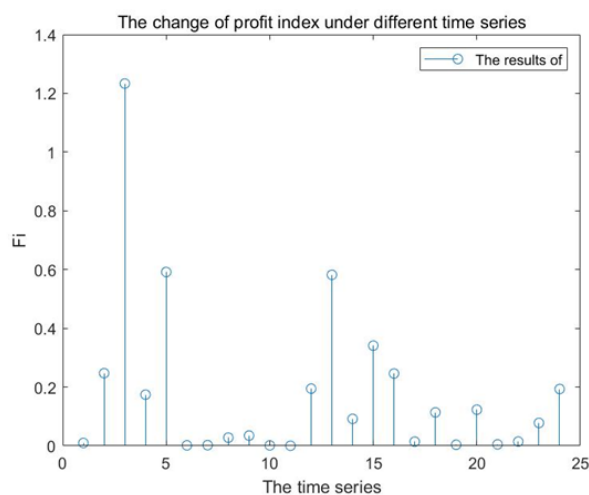


Fig. 3 The change of profit index under different time series

As can be seen from Table 4, international cooperation and competition, urban management and policy guidance have a significant impact on the development of new energy vehicles, which are the main factors affecting the new energy vehicle industry. In addition, technological progress, macroeconomic policies, industrial chain is an important factor in the development of new energy vehicles, and market factors have a relatively slight impact on the development of the new energy vehicle industry. Therefore, countries should strengthen cooperation in this field, jointly promote the construction of new energy vehicles, healthy competition is also very important for the further development of new energy vehicles, competition is conducive to the elimination of backward production capacity, promote technological reform and innovation. Governments should also adopt appropriate policies to further promote the development of new energy vehicles, the city governments

should adopt reasonable management methods, appropriate release of incentive policies, the central government should also set a good macro tone^[10], for the micro policy of new energy vehicles to point out a good direction, while strengthening the infrastructure construction of new energy vehicles, improve the corresponding supporting industrial chain.^[3] Although market factors have little impact on the overall development of the industry, we cannot ignore the incentive role of the market for enterprises, and the market is also the main source of funds for enterprises^[9], so it should also be paid attention to a certain extent and improve the market system.

Combined with the analysis of the influencing factors in the early stage, as illustrated in Fig.3, although it may be affected by factors such as slowdown in technological progress, policy support, and market saturation, new energy vehicles may have a deceleration period, but the overall profit index shows a fluctuating trend. However, overall profitability continued to show growth as the rate of change remained positive. The overall profit growth rate fluctuates around 0.2, and there is also the best period for the development of new energy, and the profit growth rate can even reach more than 1. It can be seen from the above that new energy vehicles still have great potential.

4. Conclusions

To sum up, this paper mainly adopts the research method of machine learning, uses DBSCAN algorithm and UMAP algorithm to perform cluster analysis and dimensionality reduction visualization, and then uses stacked model and BP neural network to explore the development law of new energy vehicles. The results show that international cooperation and competition, urban management and policy guidance are the main factors affecting the new energy vehicle industry. At the same time, the profit index of new energy vehicles fluctuates, but the overall trend is still up.

References

- [1] Liu Z, Hao H, Cheng X, et al. Critical issues of energy efficient and new energy vehicles development in China[J]. *Energy Policy*, 2018, 115: 92-97.
- [2] Chen Z, Zhang K, Jia S. Influence factors and risk analysis of new energy vehicles from the perspective of system[J]. *Environmental Science and Pollution Research*, 2021, 28: 62891-62906.
- [3] Gong H, Wang M Q, Wang H. New energy vehicles in China: policies, demonstration, and progress[J]. *Mitigation and Adaptation Strategies for Global Change*, 2013, 18: 207-228.
- [4] Yu F, Wang L, Li X. The effects of government subsidies on new energy vehicle enterprises: The moderating role of intelligent transformation[J]. *Energy Policy*, 2020, 141: 111463.
- [5] McCarthy P S. Market Price and Income Elasticities of New Vehicle Demands[J]. *The Review of Economics and Statistics*, 1996, 78 (03): 543-547.
- [6] Greenspan A, Cohen D. Motor Vehicle Stocks, Scrappage, and Sales[J]. *The Review of Economics and Statistics*, 1999, 81 (03): 369-383.
- [7] Jingzheng Ren. New energy vehicle in China for sustainable development: Analysis of success factors and strategic implications [J]. *Transportation Research Part D: Transport and Environment*, March 2018, Pages 268-288
- [8] Ou, X., Zhang, Q., Zhang, X. and Zhang, X. (2013) China's New Energy Passenger Vehicle Development Scenario Analysis Based on Lifetime Cost Modelling. *Low Carbon Economy*, 4, 71-79. Doi: 10.4236/lce.2013.42008.
- [9] Raymond Byrne M, Polonsky M J. Impediments to consumer adoption of sustainable transportation: alternative fuel vehicles[J]. *International Journal of Operations & Production Management*, 2001, 21(12): 1521-1538.
- [10] Zhang X, Bai X. Incentive policies from 2006 to 2016 and new energy vehicle adoption in 2010-2020 in China[J]. *Renewable and Sustainable Energy Reviews*, 2017, 70: 24-43.