

Research of China's New Energy Vehicle Market Based on TOPSIS and LSTM Models

Huiyi Tu^{1,*}, Weiliang Li², Zijian Zheng³, Yongqiang Li⁴ and Xingle Zhou⁵

¹School of Economics and Management, Shanghai University of Applied Science and Technology, Shanghai, China

²School of law, Inner Mongolia University of Finance and Economics, Hohhot, China

³School of Data Science and Applications, Inner Mongolia University of Technology, Hohhot, China

⁴School of Finance and Taxation, Inner Mongolia University of Finance and Economics, Hohhot, China

⁵School of Tourism, Inner Mongolia University of Finance and Economics, Hohhot, China

* Corresponding Author Email: 1822762771@qq.com

Abstract. This research paper provides an in-depth analysis of the development trajectory of China's New Energy Vehicle (NEV) market, utilizing both the TOPSIS and LSTM models. The TOPSIS model is ingeniously applied to assess the development level of the NEV industry, identifying key factors such as government policies, charging infrastructure, public acceptance, technological progress, and economic indicators. Subsequently, the paper delves into the LSTM model, a specialized form of Recurrent Neural Network (RNN) designed for handling time-series data with long-term dependencies. The LSTM architecture, equipped with forget, input, and output gates, is trained to minimize Mean Square Error (MSE) and Root Mean Square Error (RMSE), thereby optimizing model performance and preventing overfitting. The predictive capabilities of the LSTM model are demonstrated through a forecast of NEV sales in China for the next decade, showcasing the model's effectiveness in capturing market dynamics and growth trends. The integration of TOPSIS for evaluative analysis and LSTM for predictive forecasting offers a robust framework for understanding and anticipating the complexities of the NEV market.

Keywords: New energy vehicles; LSTM models; Predictive models; TOPSIS methods; .Time series forecasting.

1. Introduction

China, as a pivotal player in the global NEV market, has experienced rapid growth in NEV adoption and manufacturing since the government introduced a series of preferential policies in 2011 [1]. The proliferation of NEVs in China is indicative of a broader trend towards greener transportation solutions, aligning with the country's commitment to sustainable development and emission reduction targets [2]. The Chinese NEV market has been bolstered by significant investments in research and development, the establishment of a robust charging infrastructure, and the growing consumer awareness of the environmental benefits associated with electric mobility [3].

This paper aims to explore the development trajectory of China's NEV market by employing a dual-model approach, integrating the TOPSIS and LSTM methodologies [4, 5]. The TOPSIS method offers a systematic framework for evaluating the development level of the NEV industry by quantifying the relative performance of various factors such as government policies, technological advancements, and market acceptance. On the other hand, the LSTM model, with its ability to capture long-term dependencies in data, provides an accurate forecasting tool for predicting the market's future trajectory based on historical sales and market share data. By combining the strengths of both models, this research endeavors to provide a comprehensive analysis of the NEV market in China, shedding light on the key drivers of growth and the potential challenges that may arise in the coming decade [6].

2. Data Preprocessing

We collect data related to China and the global new energy vehicle industry, including sales volume, market share, government policies, technological progress, etc. And ensure the reliability of the data source, prioritizing the collection of officially announced data sources. Check data integrity and eliminate records with too many missing values. Using python, the data is found to have no missing values.

The data is standardized or normalized so that different indicators are at the same order of magnitude for easy comparison and analysis. Identify and handle outliers using box plots. As shown in Fig. 1, it can be seen that there are no outliers in the data. The data set is divided into 70% training set and 30% test set for training and validation of machine learning models.

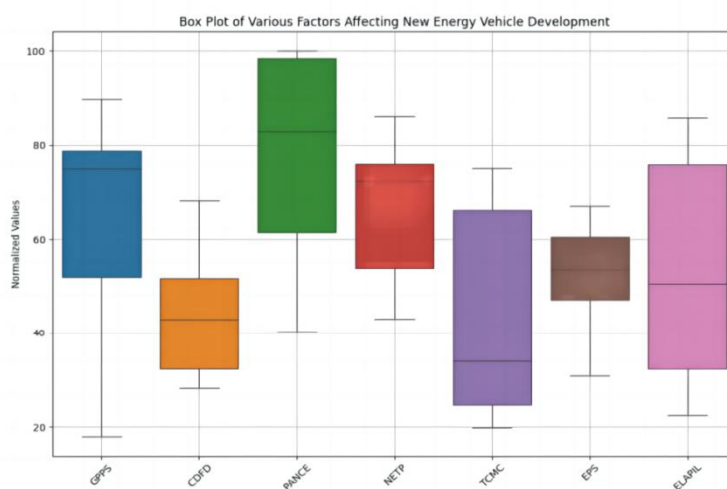


Fig. 1 Key Indicators Affecting the Development of China's New Energy Vehicle Industry

3. TOPSIS Method for Evaluating the Development Level of New Energy Vehicle Industry

3.1. Index Selection

After consulting a large number of relevant literature, it is found that the development of new energy vehicles may be affected by the following factors:

Government Policies and Subsidies(GPPS): Government policies and subsidies have a significant impact on the new energy vehicle market, including tax breaks, direct subsidies, research and development grants, etc.

Charging Infrastructure Development(CDFD): The availability and convenience of charging facilities directly affect consumers' willingness to buy new energy vehicles.

Public Acceptance of New Energy Vehicles(PANCE): Consumers' cognition, acceptance and purchase intention of new energy vehicles.

New Energy Vehicle Technological Progress(NETP): The level of technological advances in battery technology, drive systems, energy efficiency, etc.

Traditional Combustion Motorcar Competition(TCMC): The competitive relationship between traditional cars and new energy vehicles in terms of price, performance, brand, etc.

Environmental Protection Sense(EPS): The level of public awareness of environmental protection, which affects the likelihood of them choosing new energy vehicles.

Economic Level and Personal Income Level(ELAPIL): Economic development and personal income levels affect consumers' purchasing power and car purchase choices.

3.2. Establishment of TOPSIS method

3.2.1 Analysis of development indicators of new energy vehicles

We have collected the specific data of vehicle sales, ownership and market share of new energy electric vehicles from 2011 to 2022 as Table 1. Among them, CNEVS is the abbreviation of China's new energy vehicle sales, CTVS is the abbreviation of China's total automobile sales (tens of thousands of units), CNEVMS is the abbreviation of China's new energy vehicle market share (%).

Table 1 Car sales, ownership, and market share from 2011 to 2022

Year	CNEVS	CTVS(in10,000units)	Stock(in10,000units)	CNEVMS(%)
2011	0.8	1852	1.2	0.04
2012	1.2	1931.6	2.7	0.06
2013	1.7	2198.3	5.9	0.08
2014	7.5	2340.6	14.8	0.32
2015	33.1	2471.5	33.1	1.34
2016	50.7	2811.9	65.1	1.8
2017	77.7	2890.5	150.6	2.69
2018	125.6	2808.1	253.4	4.47
2019	120.6	2571.6	377.5	4.69
2020	136	2514.6	484.6	5.41
2021	351	2530	784.1	13.87
2022	688.7	2686.4	1310	25.64

3.2.2 Model building

The Topsis method is a decision support method used to identify the best solution from a set of alternative solutions. This method is based on the principle that the selected solution should be geometrically closest to the Ideal Best Solution, while being farthest from the Ideal Worst Solution. The steps are as follows:

(1) Data standardization: First, the data needs to be standardized to ensure that all indicators are equivalent when compared. This is usually done through Min-Max standardization, which converts all metrics into the same range from 0 to 1. The normalization formula is as follows: where x is an element in the original data matrix, $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values in column j , respectively.

(2) Weighted normalization decision matrix: Then, weights need to be applied to the normalized data. In this study, we assume that all indicators have the same importance, so each indicator has the same weight. The formula for calculating the weighted standardized value is as (1)

$$v_{ij} = W_j \times x'_{ij}. \quad (1)$$

Among them, W_j is the weight of the j indicator:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}. \quad (2)$$

In Table 2, we assign corresponding weights to each indicator that affects the development of the new energy vehicle industry, which reflect the relative importance of each indicator in the overall evaluation. By multiplying the normalized data with the weights, we get a weighted normalized value.

Table 2. Weighted standardized data

Year	CNEVS	Stock	CNEVMS
2011	0	0	0
2012	0.0002	0.0004	0.0003
2013	0.0004	0.0012	0.0005
2014	0.0032	0.0035	0.0036
2015	0.0157	0.0081	0.0169
2016	0.0242	0.0163	0.0229
2017	0.0373	0.0381	0.0345
2018	0.0605	0.0642	0.0577
2019	0.0581	0.0958	0.0605
2020	0.0655	0.1231	0.0699
2021	0.1697	0.1994	0.1801
2022	0.3333	0.3333	0.3333

(3) Determine the ideal best and worst solutions: The ideal best solution is the maximum value on each index, and the ideal worst solution is the minimum value on each index.

$$v^+ = \{ \max(v_{ij}) | i = 1, 2, \dots, m; j = 1, 2, \dots, m \}. \quad (3)$$

$$v^- = \{ \min(v_{ij}) | i = 1, 2, \dots, m; j = 1, 2, \dots, m \}. \quad (4)$$

(4) Calculate the distance between each evaluation object and the ideal solution: For each evaluation object, the Euclidean distance from the ideal solution and the negative ideal solution is calculated. Let the distance between the i evaluation object and the ideal solution be D_i^+ and the distance between the evaluation object and the negative ideal solution be D_i^- . The calculation formulas are as follows.

$$D_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2}. \quad (5)$$

$$D_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}. \quad (6)$$

(5) Calculation of relative proximity: relative proximity is the approximation degree of the evaluation object relative to the ideal solution. The greater the relative proximity, the better the evaluation object. The calculation formula is as follows.

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}. \quad (7)$$

The calculation results are shown in the Table 3.

Table 3 TOPISSScore data table

Year	CNEVS	CTVS(in10,000units)	Stock(in10,000units)	CNEVMS(%)	TOPSISScore
2011	0.8	1852	1.2	0.04	0.000000
2012	1.2	1931.6	2.7	0.06	0.000868
2013	1.7	2198.3	5.9	0.08	0.002383
2014	7.5	2340.6	14.8	0.32	0.010368
2015	33.1	2471.5	33.1	1.34	0.042266
2016	50.7	2811.9	65.1	1.8	0.064161
2017	77.7	2890.5	150.6	2.69	0.109901
2018	125.6	2808.1	253.4	4.47	0.182526
2019	120.6	2571.6	377.5	4.69	0.218875
2020	136	2514.6	484.6	5.41	0.266008
2021	351	2530	784.1	13.87	0.548898
2022	688.7	2686.4	1310	25.64	1.000000

3.3. Analysis of influencing factors of new energy vehicle development

We used Pearson correlation coefficient to determine the correlation between indicators and development scores. The calculation formula is as follows:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (8)$$

Among them: r_{xy} is the correlation coefficient between two variables x and y ; x_i and y_i are observations of the variables x and y ; $\bar{x}$ and $\bar{y}$ are the mean values of variables x and y respectively; n is the number of observations.

When the absolute value of the correlation coefficient is closer to 1, the higher the correlation between the two, and the closer it is to 0, the lower the correlation. Positive value indicates positive correlation, and negative value indicates negative correlation. We visualized how each indicator correlates with all other indicators and development scores using a correlation heat map, a method of visualizing a matrix of correlation coefficients. It shows the size of the correlation coefficient between different variables through the change of color, which makes the correlation analysis more intuitive. The heat map of the relationship between indicators and development scores is shown in Fig. 2. To determine the relationship between the indicator and the development score, we only need to observe the rightmost column of this heat map.

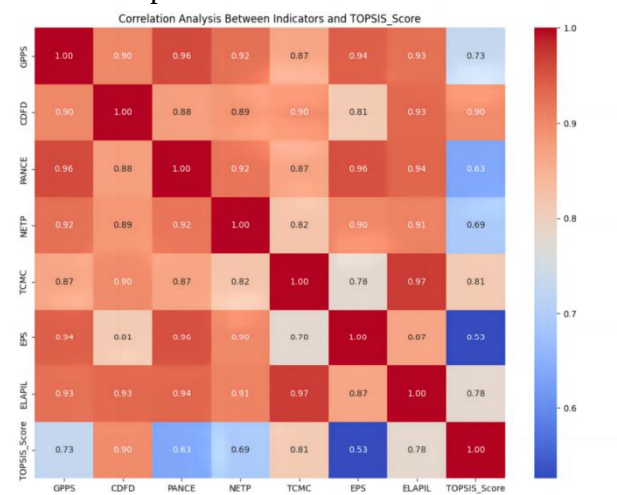


Fig. 2 Heat map of the relationship between indicators and development scores

The analysis results show that EPS and development score TOPSIS_ The correlation of score is low, but because its value is not very low, it is uncertain whether it is not used as an independent variable of multiple regression, so we respectively studied the fitting effect with or without EPS as an independent variable.

Multivariate polynomial regression is a statistical method that models the relationship between different combinations of multiple independent variables (including their higher order terms and interaction terms) and the dependent variable. The general form of the model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \cdots + \beta_{nn} X_n^2 + \beta_{12} X_1 X_2 + \cdots + \epsilon. \quad (9)$$

The fitting effect is shown in Fig. 3.

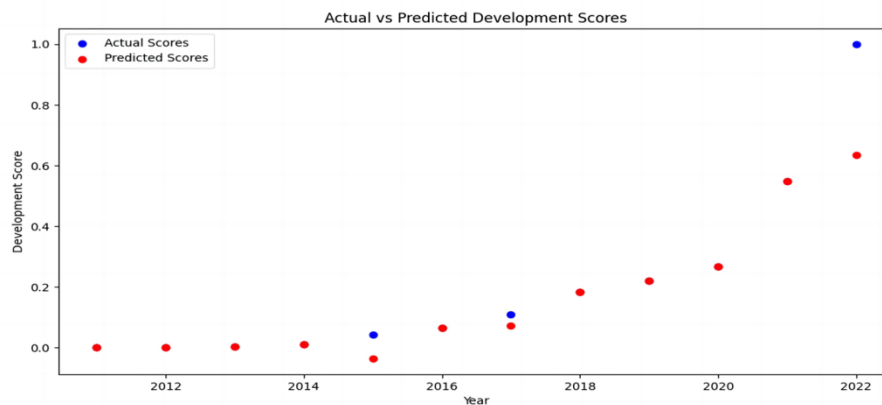


Fig. 3 Fit effect diagram

The analysis shows that the development of new energy vehicles is more affected by the objective environment such as technological progress and infrastructure development, and is related to the economic level.

4. LSTM model for New Energy Vehicle Market Trajectory

4.1. LSTM network

LSTM network is a special kind of recurrent neural network (RNN), which is suitable for processing and predicting long-term dependent data of interval and delay in time series. LSTM optimizes the long-term dependence problem of traditional RNN by introducing three gating structures: forget gate, input gate and output gate to control the retention and forgetting of information. The LSTM network architecture is shown in Fig. 4

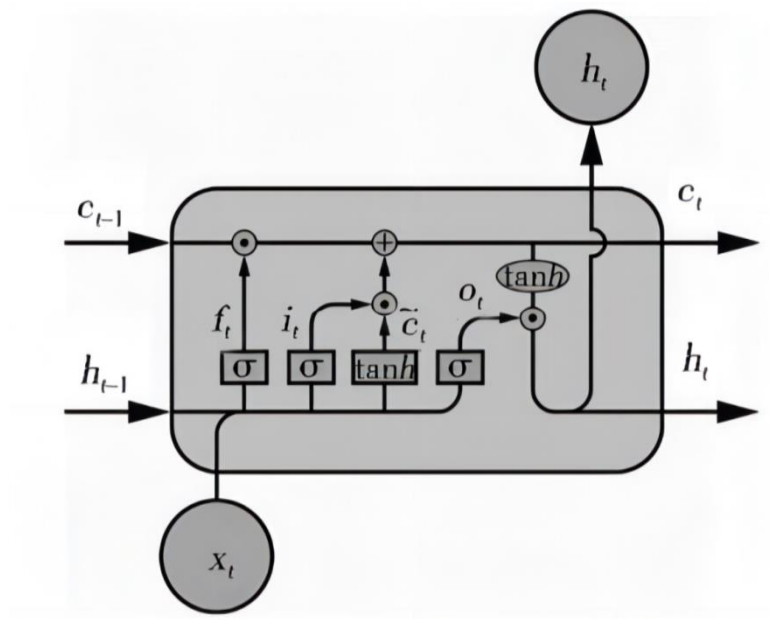


Fig. 4 Illustration of LSTM Network Architecture

4.2. Model training

In the process of model training, mean square error (MSE) and root mean square error (RMSE) are used as loss functions to evaluate and optimize the model performance. As shown in Figure 5, when the training rounds reached 23, the validation set loss did not decrease. After 23 rounds of training, the model stopped early to prevent over-fitting.

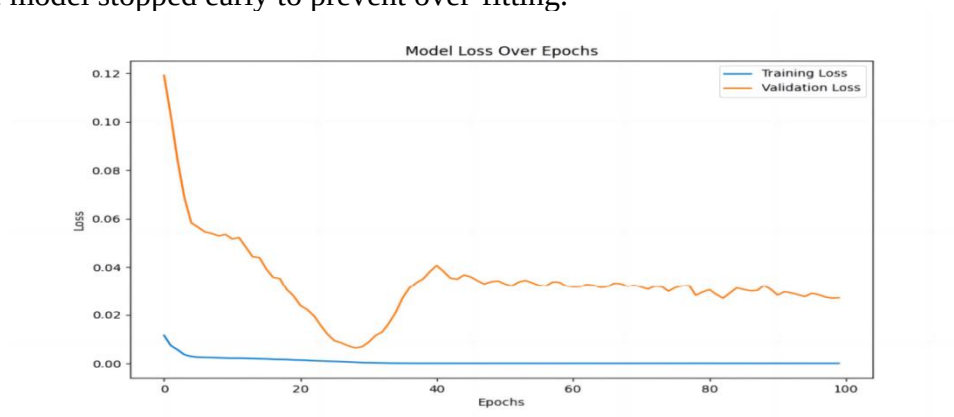


Fig. 5 Training analysis

4.3. Loss analysis

The change of loss value in the training process of the model is shown in the Fig. 6. It can be seen that with the increase of training times, both training loss and validation loss decrease, indicating that the performance of the model gradually improves during the process of learning the data set.

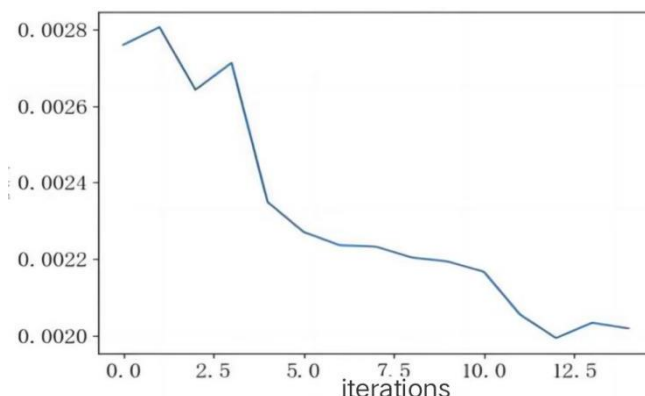


Fig. 6 Training analysis

4.4. Result

According to the forecast made by the LSTM model, the sales volume of new energy electric vehicles in China will show a continuous growth trend in the next decade. The comparison between historical sales and forecast sales shows that the model can effectively capture the dynamics of market development.

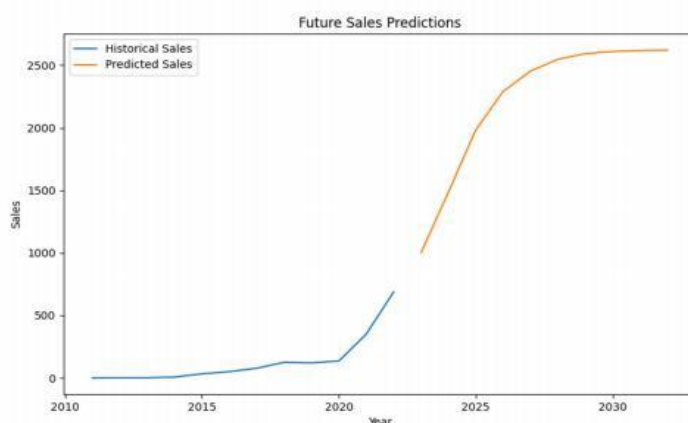


Fig. 7 Sales forecast for new energy vehicles in China in the next decade

Detailed projections for the next decade are shown in the Table 4.

Table 4 Forecast data for sales of new energy vehicles in China in the next decade

Year	CNEVS	Stock	CNEVMS
2023	1048.9	2271.3	36.4
2024	1601.1	3355.4	54.6
2025	2172.5	4434.3	72.1
2026	2510.9	5063.6	81.5
2027	2681.4	5380.2	85.8
2028	2771.9	5548.3	88
2029	2812.5	5624.6	88.9
2030	2827.4	5653.4	89.3
2031	2833.3	5665.1	89.4
2032	2836	5670.4	89.5

Through the prediction of LSTM model, it can be seen that China's new energy vehicle market will continue to grow in the next decade. Forecast data show that sales will rise year by year at a significant rate, reflecting the increasing market demand for new energy vehicles and the driving role of technological progress. This growth trend may be influenced by a variety of factors, including

government policy support, increased consumer acceptance of environmental protection and new technologies, and improved cost-effectiveness of new energy vehicles.

5. Conclusion

In conclusion, this research paper has presented a comprehensive analysis of the development of China's New Energy Vehicle (NEV) market using the TOPSIS and LSTM models. The TOPSIS method has effectively identified and evaluated the key factors influencing the NEV industry's growth, providing a clear understanding of the relative importance of government policies, charging infrastructure, public acceptance, technological progress, and economic indicators. The LSTM model, with its capability to handle long-term dependencies in time-series data, has successfully forecasted the sales trajectory of NEVs in China for the next decade, demonstrating a continuous growth trend.

References

- [1] Yeung, Godfrey. "'Made in China 2025': the development of a new energy vehicle industry in China." *Area Development and Policy* 4.1 (2019): 39-59.
- [2] Ma, Ye, et al. "Comprehensive policy evaluation of NEV development in China, Japan, the United States, and Germany based on the AHP-EW model." *Journal of cleaner production* 214 (2019): 389-402.
- [3] Liu, Chang, et al. "The capital market responses to new energy vehicle (NEV) subsidies: An event study on China." *Energy Economics* 105 (2022): 105677.
- [4] Çelikkilek, Yakup, and Fatih Tüysüz. "An in-depth review of theory of the TOPSIS method: An experimental analysis." *Journal of Management Analytics* 7.2 (2020): 281-300.
- [5] Yu, Yong, et al. "A review of recurrent neural networks: LSTM cells and network architectures." *Neural computation* 31.7 (2019): 1235-1270.
- [6] Liu, Yingqi, and Ari Kokko. "NEV technology in China." *Chinese Management Studies* 6.1 (2012): 78-91.