Economics and Public Policy
ISSN Print: 3104-4387
ISSN Online: 3104-4395
CD ISSN: 3104-4409
https://doi.org/10.63313/EPP.9001

Prediction of the Inventory of New Energy Vehicle Charging Piles in China Based on Support Vector Machine

Jun Li. Yumei Wang*
Anhui University of Finance and Economics, Bengbu 233030, China
Email: lijun200117@163.com

How to cite this paper: Li, J., & Wang, Y. M. (2025). Prediction of the inventory of new energy vehicle charging piles in China based on support vector machine. Economics and Public Policy, 1(1), 24-33. ISSN Print: 3104-4387; ISSN Online: 3104-4395.
https://doi.org/10.63313/EPP.9001
Published: 2025-06-24

Abstract

With the rapid development of China's new energy vehicle (NEV) industry, the charging pile, as a key supporting facility, plays a crucial role. Predicting the inventory of charging piles is of great significance for guiding the healthy de-velopment of the industry. This paper employs the Support Vector Machine (SVM) as the prediction model. Firstly, data on the inventory of NEV charging piles in China over the past six years were collected. Combined with factors such as the production and sales volume of NEVs, relevant policies, and public pa-tents, a prediction model was constructed. By selecting appropriate kernel func-tions and adjusting parameters, the SVM model demonstrated good predictive performance on the training data. Cross-validation and parameter optimization were conducted to ensure the stability and reliability of the prediction results. The prediction indicates that the inventory of NEV charging piles in China is expected to continue growing in the coming years.

Keywords

New Energy Vehicles; Charging Pile Inventory; Support Vector Machine; Predic-tion Model

1. Introduction

With the increasing global awareness of environmental protection and the advancement of renewable energy technologies, new energy vehicles (NEVs) have achieved rapid development and have become a major highlight of China's automotive industry in recent years. This development not only reflects the Chinese government's high emphasis on and strong support for the NEV industry but also indicates continuous technological progress and the gradual improvement of the industrial chain in China.

To encourage consumers to purchase NEVs, the government has implemented a series of measures, including subsidies and tax reductions, which have helped lower the prices of NEVs and thus increased consumer willingness to buy. Additionally, the government has intensified efforts to build infrastructure for NEVs, such as charging piles and battery-swap stations, providing strong support for the promotion and application of NEVs. As the number of NEVs increases, the demand for charging infrastructure also grows.

Domestic NEVs have made significant progress in terms of range and charging time, which has continuously improved consumer acceptance. The charging issue for NEVs is considered the “last mile” in promoting NEVs and is crucial for their widespread adoption.

Support Vector Machine (SVM) is a supervised learning model that excels in classification and regression problems. It distinguishes different categories or predicts continuous values by finding the optimal separating hyperplane between data points. SVM has good generalization ability for high-dimensional data and is suitable for handling complex nonlinear relationships. Accurate prediction of the inventory of NEV charging piles can help governments and related enterprises formulate more rational policies and business strategies to promote the healthy development of the NEV industry.

In summary, predicting the inventory of NEV charging piles in China based on the SVM model is not only theoretically significant but also has important guiding value for practical applications and market regulation.

2. Current Research Status

The Chinese new energy vehicle (NEV) market is currently experiencing rapid growth. With the government's supportive policies for NEVs and the increasing acceptance of environmentally friendly modes of transportation among consumers, the inventory of NEVs is expected to continue rising in the coming years. Most existing research has focused on NEVs themselves, with limited studies addressing the inventory of public charging piles for NEVs. As the number of NEVs increases, the demand for charging piles will also correspondingly rise. Researching and forecasting the demand for charging piles can help governments and enterprises plan the construction of charging infrastructure more rationally. With technological advancements, the charging speed, compatibility, and intelligence level of charging piles are also continuously improving. Studying these technological trends can help charging pile manufacturers and operators provide higher quality services.

3. Methodology

Support Vector Regression (SVR) is a regression method based on Support Vector Machine (SVM), used for predicting continuous numerical variables. SVR inherits the core concepts used by SVM for classification problems and adapts them to handle regression tasks. The goal of SVR is to find a function that minimizes the prediction error for the given data points. SVR attempts to identify an optimal fitting hyperplane that not only fits the data but also maximizes the margin with the data points.

3.1. Five-Fold Cross-Validation

Five-Fold Cross-Validation is a commonly used model evaluation method in machine learning and statistics, particularly suitable for small datasets, as it fully utilizes the data for both training and validation. The basic steps of Five-Fold Cross-Validation are as follows:

- Data Splitting: The entire dataset is evenly divided into five parts (or as equal as possible), known as "folds".
- Loop Validation: Five iterations of validation are conducted. In each iteration, one fold is selected as the validation set (for testing model performance), and the remaining four folds are combined as the training set (for training the model).
- Model Training and Evaluation: In each iteration, the model is trained using the training set and then evaluated on the validation set. Performance metrics such as accuracy, recall, F1 score, or other relevant indicators are typically used.
- Result Recording: The performance evaluation results of the model in each iteration are recorded.
- Performance Aggregation: After the five iterations are completed, the evaluation results are averaged to obtain the final performance estimate of the model.
- Variance Analysis: Optionally, the variance of each evaluation result can be calculated to assess the stability of the model's performance.

3.2. The advantages of Five-Fold Cross-Validation

- Full Utilization of Data: Since each data point is used for both training and validation, this method maximizes the use of limited data resources.
- Assessment of Model Stability: Through five independent training and validation processes, a better understanding of the model's performance on different subsets of data can be achieved, thereby evaluating the model's stability and reliability.

3.3. Model Parameters Introduction

- Margin: In SVR, the margin refers to the difference between the actual and predicted values. SVR aims to find a function with the largest margin, which helps improve the model's generalization ability.
- Loss Function: SVR typically uses the epsilon-insensitive loss function. This means that if the difference between the predicted and actual values is less than a threshold ε, the loss is zero. This allows the model to tolerate a certain level of error without penalizing the prediction.
- Regularization: To prevent overfitting, SVR includes a regularization term in the loss function, usually the L2 norm of the model weights. The coefficient C in front

of the regularization term is a hyperparameter that balances the trade-off between model complexity and fitting error.

3.4. Model Parameters Introduction

The dataset is split into training and testing sets. A default-parameter SVR model is then constructed and fitted on the training data. Predictions are made, and the Mean Squared Error (MSE) of the model is calculated.

4. Data Description and Preprocessing

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your journal for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

4.1. Authors and Affiliations

In order to conduct a predictive analysis of the trend and changes in the number of public charging piles for new energy vehicles in China, the following steps are adopted in this paper to collect and organize the data.The data collection period is from January 2018 to April 2024, totaling 64 months.

- Public Charging Pile Inventory Data: The data on the inventory of public charging piles for new energy vehicles (NEVs) were sourced from the China Charging Alliance, covering the monthly inventory of these charging piles.
- Policy Data: The data were sourced from the China Government Website and include the number of policies issued each year.
- NEV Sales Data: The data were sourced from the China Association of Automobile Manufacturers Statistics Information Network and include the monthly sales of NEVs in China over the 64-month period.

These datasets will provide the foundation for the subsequent analysis and forecasting.

4.2. Data Preprocessing

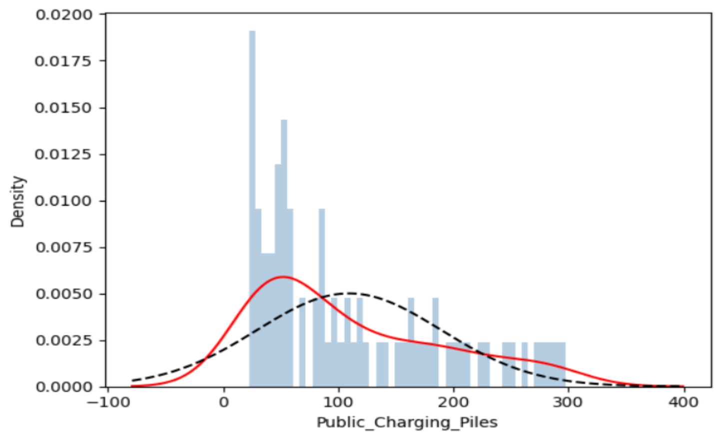

Figure 1. Histogram with Kernel Density Estimation (KDE) and Fitted Normal Distribution Curve

The dataset contains no anomalies or missing values and consists entirely of valid data.

The histogram, which includes Kernel Density Estimation (KDE) and a fitted normal distribution curve, reveals that the data distribution is severely right-skewed. Due to this skewness, the variable cannot be directly used for modeling. Typically, logarithmic transformation is applied to the data. Additionally, Z-score standardization is employed to transform the data into a standard normal distribution with a mean of 0 and a standard deviation of 1. Standardizing the data in this manner helps enhance the accuracy and robustness of the model.

5. Prediction

This paper employs the Support Vector Regression (SVR) model to predict the inventory of charging piles based on three indicators: the number of policies, monthly sales of new energy vehicles (NEVs), and the number of patent publications. The inventory of charging piles is treated as the dependent variable, while the other three indicators serve as independent variables. Regression analysis is conducted using these three indicators to fit the model, which is then used to make reasonable predictions of the inventory of charging piles.

5.1. Model Evaluation Metrics

There are various metrics for evaluating a model, with the most commonly used one being the Mean Squared Error (MSE):

$$MSE = \sum_{i=1}^{n} (y_i - f(x_i))^2 / \mathrm{N} \tag{1}$$

where y_i represents the true value of the data, and $f(x_i)$ is the predicted value generated by the model. In other words, the expected value of the squared differences between the estimated parameters and the true values is referred to as the Mean Squared Error. This is the most frequently used standard for assessing the precision of model fitting. When the MSE is larger, it indicates a greater deviation in the predictions, meaning that the predicted values are less accurate. In such cases, it is necessary to change the parameters or the model to find one that is more suitable for the given dataset.

5.2. Default Parameter Model

Initially, an SVR model with default parameters is constructed. The default parameter values are as follows: C=1.0 (the regularization parameter, which controls the complexity of the model. The default value of 1.0 indicates a balanced model, neither overly complex nor overly simple); ϵ=0.1 (specifies the width of the epsilon-insensitive zone. Loss is only calculated when the prediction error exceeds this value. The default value of 0.1 means the model is not overly sensitive to small prediction errors). λ =0.005 (when using the RBF kernel, the gamma parameter defines the influence range of a single training sample. The default setting of 'scale' means that the gamma value is related to the number of features in the training samples).

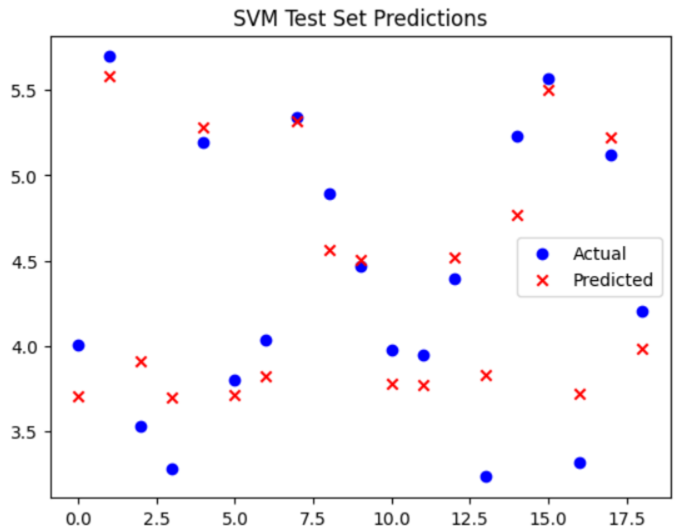

Figure 2. Comparison of Predicted Values and Actual Values with Default Parameter Model

The fitting error obtained from this model is 0.079, with 64 support vectors. In the figure, the red crosses represent the predicted values, while the blue dots indicate the true values. It is evident that the predicted values generally follow the same trend as the true values, although there are still significant errors in certain time periods.

5.3. Model Parameter Adjustment

Using five-fold cross-validation and grid search, the optimal values for the parameters C, ϵ, and γ in the Support Vector Regression (SVR) model were selected. After the five-fold cross-validation, the optimal values were determined as follows: the best penalty coefficient C=500, the best ϵ value of 0.1, and the best γ value of 0.005. The resulting Mean Squared Error (MSE) was 0.067. Compared to the model with default parameters, the MSE has decreased, indicating that parameter optimization is necessary when using SVR for prediction problems. The model was then refitted using these optimal parameters, resulting in the fitting graph shown below.

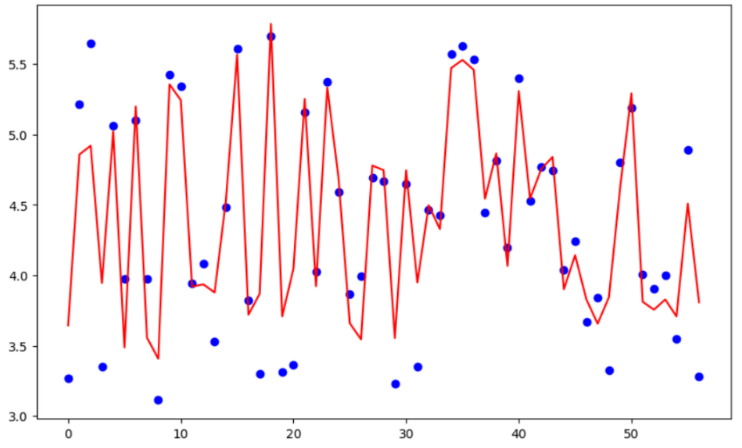

Figure 3. Fitting Plot after Parameter Adjustment

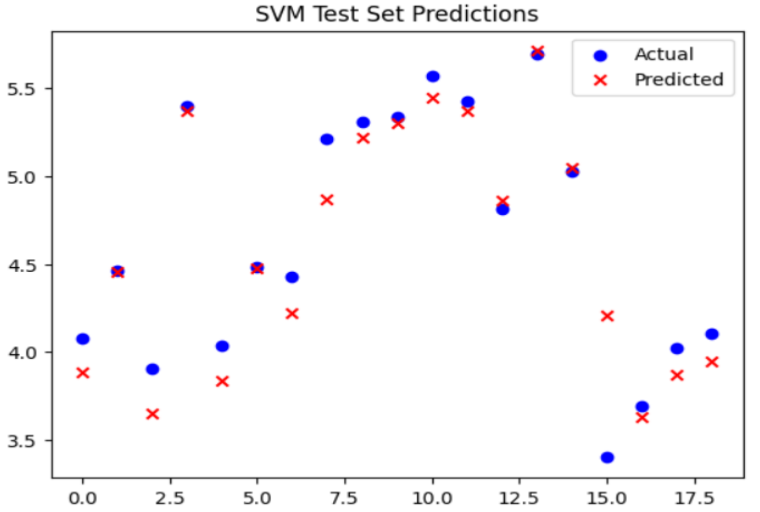

Figure 4. Comparison of Predicted and Actual Values after Parameter Adjustment

The prediction graph above was obtained using the optimized regression model. It is evident that after adjusting the parameters to their optimal values, the predicted values from this model are much closer to the true values than those from the default parameter model, with smaller errors.

Through the fitting and prediction analysis of the inventory of new energy vehicle (NEV) charging piles in China, it can be seen that the Support Vector Machine (SVM) model provides good predictive performance for this time series.

Limitations of Prediction:

Support Vector Regression (SVR) is a powerful regression prediction method that performs well in many applications, but it also has some drawbacks. The performance of SVR largely depends on the choice of kernel function and parameter settings (such as the penalty parameter C and kernel function parameters), which may require extensive experimentation and expertise. For linear data, SVR may not perform as well as traditional linear regression methods. The choice of kernel function is crucial for the performance of the SVR model, but it is often not intuitive and needs to be determined through methods such as cross-validation. Additionally, SVR models have poor interpretability compared to simpler regression models, which can be a disadvantage in applications where model interpretability is required.

6. Prediction

In this paper, the Support Vector Regression (SVR) model was employed to predict the inventory of new energy vehicles (NEVs) in China. By comparing the actual data with the predictions made by the SVR model, we found that SVR demonstrates high accuracy and stability in predicting the inventory of NEVs in China. This indicates that SVR can effectively capture the trends and patterns of the NEV market growth. During the process of establishing the SVR model, we discovered that selecting an appropriate kernel function and tuning parameters (such as C and ϵ) are crucial for enhancing the model's predictive performance. Through methods like cross-validation, we determined the optimal model parameters.

The prediction results of the SVR model reveal a rapid growth trend in China's NEV market. With the support of government policies and increasing consumer acceptance of environmentally friendly transportation, the inventory of NEVs is expected to continue rising in the coming years. The SVR model's predictions also highlight the significant role of policy support in driving the development of NEVs. The future trends predicted by the SVR model suggest that as battery technology, charging technology, and intelligence levels continue to improve, the performance and convenience of NEVs will be further enhanced, attracting more consumers.

Although SVR exhibited good predictive performance in this study, considering its potential limitations in certain situations, future research could explore comparisons and combinations with other predictive models to improve the robustness and accuracy of predictions. In summary, the SVR model provides an effective tool for predicting the inventory of NEVs in China, and its results offer valuable references for government policy-making, corporate strategic planning, and market analysis.

Due to the limitations of my knowledge, there are many areas in this paper that need improvement and refinement. In the predictive analysis, using a single model may not capture all the characteristics of the data. Constructing models such as ARIMA (Autoregressive Integrated Moving Average) and traditional regression models, and comparing them with SVR, can provide a more comprehensive perspective. By comparing the predictive performance of different models using statistical indicators such as Mean Squared Error (MSE) and the coefficient of determination (R2), we can assess which model is more suitable for the current dataset and predictive task. Combining the predictions from different models can enhance the accuracy and robustness of the forecasts. This can be achieved through simple averaging or more complex weighted methods, where each model's weight is based on its predictive performance.

Another important aspect for improving the model is feature engineering. Identifying and constructing features that are useful for the predictive task can significantly enhance model performance. Fine-tuning the model parameters using techniques such as Grid Search or Random Search to find the optimal combination of parameters is also crucial. Enhancing the interpretability of the model is vital for stakeholders to understand the predictive logic and results of the model, which is crucial for its acceptance and application. Finally, feature selection is a key step in building predictive models, directly affecting the model's performance and the accuracy of the predictive results. Using statistical tests, model-based feature selection methods, or ensemble methods like Random Forest to evaluate the relevance of features to the predictive target is essential. Analyzing the correlations between different features to avoid selecting highly correlated features, thereby reducing multicollinearity issues, is also important. Combining domain expert knowledge in feature selection is crucial, as some seemingly insignificant features may hold significant importance in specific contexts.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Wang, for her invaluable guidance and unwavering support throughout my research. Professor Wang is not only a knowledgeable and dedicated mentor but also a kind and approachable person who has provided tremendous help in both my academic and personal life. Her profound expertise, insightful suggestions, and patient encouragement have been instrumental in shaping this work. Beyond academics, her warmth and care have made my research journey enjoyable and fulfilling.

I am also sincerely thankful to my family for their unconditional love and support, which have been my greatest motivation. My heartfelt appreciation also goes to my friends and colleagues for their companionship and stimulating discussions.

Finally, I acknowledge all the researchers whose work has contributed to this study. Without their foundational efforts, this research would not have been possible.

References

[1] Bae, S. and Kwasinski, A. (2012) Electric Vehicle Charging Infrastructure Planning: A Review of Modeling Approaches. Renewable and Sustainable Energy Reviews, 16, 5143-5152.
https://doi.org/10.1016/j.rser.2012.05.022

[2] Gnann, M. and Plötz, P. (2019) Forecasting the Demand of Electric Vehicles for Private and Shared Mobility. Transportation Research Part D: Transport and Environment, 71,

156-172.
https://doi.org/10.1016/j.trd.2018.12.011

[3] Chen, H., Luh, P.B. and Zhao, C. (2004) Support Vector Regression for Electric Load Forecasting. IEEE Transactions on Power Systems, 19, 1825-1834.
https://doi.org/10.1109/TPWRS.2004.835637

[4] Al-Sumaiti, A.K., Ahmed, M.H. and Salama, M. (2020) A Review of Machine Learning Applications in Energy Systems. IEEE Access, 8, 212518-212541.
https://doi.org/10.1109/ACCESS.2020.3040394

[5] Liu, J., Chen, H. and Zhang, X. (2021) Charging Infrastructure Demand Modeling for Plug-in Electric Vehicles: A Review of Approaches and Methods. Energy, 223, 120044.
https://doi.org/10.1016/j.energy.2021.120044

[6] Li, Y., Zhang, L. and Li, W. (2020) Optimal Planning of Electric Vehicle Charging Stations Considering Traffic Flow and Power Grid Constraints. Applied Energy, 278, 115654.
https://doi.org/10.1016/j.apenergy.2020.115654

[7] Wang, Y., Infield, D. and Wang, Y. (2021) Short-Term Electric Vehicle Charging Demand Prediction Using Deep Learning Methods. Energy, 214, 118874.
https://doi.org/10.1016/j.energy.2020.118874

[8] Zhang, L., Brown, T. and Samuelsen, G.S. (2018) Evaluation of Charging Infrastructure Requirements for Plug-in Electric Vehicles under Different Market Penetration Scenarios. Transportation Research Part D: Transport and Environment, 65, 591-611.
https://doi.org/10.1016/j.trd.2017.10.006

[9] Smets, K., Verdonk, B. and Jordaan, E.M. (2019) Support Vector Machines for Regression: A Comparison of Optimization Techniques. Neural Computing and Applications, 31, 2603-2615.
https://doi.org/10.1007/s00521-017-3214-2

[10] Dong, M., Liu, C. and Lin, Z. (2022) Policy Incentives and the Adoption of Electric Vehicles: A Global Review. Renewable and Sustainable Energy Reviews, 156, 111939.
https://doi.org/10.1016/j.rser.2021.111939