WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Data-Driven Battery State of Charge (SOC) Prediction in Electric Vehicles Using SVR and Catboost

K Sai Tarun[1], Dr. R. Ramakrishna[2] , K Sravan Kumar[3], K Sai Ganesh[4] , K Mohan Shriram[5], K Poornesh[6], B Venkata Pavan[7]

[1,3,4,5,6,7]B. Tech Student, GMR Institute of Technology, Rajam,532127, India
[2]Assistant Professor, GMR Institute of Technology, Rajam,532127, India

A B S T R A C T

This paper presents a data-driven method for forecasting the SoC in electric vehicle (EV) batteries using two machine learning techniques: Support Vector Machines (SVR) and CatBoost. Given the growing demand for accurate battery monitoring systems in EVs, the accurate prediction of SoC plays a vital role in maximizing battery efficiency and prolonging its lifespan. In this study, we utilize real-world battery data to compare the predictive performance of SVR and CatBoost, focusing on their accuracy, speed, and robustness under different operational conditions. The results indicate that CatBoost surpasses SVR regarding accuracy and training time, while SVR remains more reliable when the available training data is sparse or imbalanced. This research enhances the progress of SoC prediction models, providing insights into how different machine learning approaches handle the complexities of battery management systems.

Keywords: Battery SoC, Support Vector Machines (SVR), CatBoost, Electric Vehicles (EVs), Machine Learning, Battery Management Systems, Lithium-Ion Battery, SoC Prediction.

1. Introduction

As electric vehicles (EVs) continue to grow in popularity, the demand for efficient battery management systems (BMS) becomes more pressing. Primary goals of a BMS is to accurately estimate the State of Charge (SoC) of a battery, which represents the available capacity of the battery at a given time. Incorrect prediction of SoC can result in operational inefficiencies, including reduced battery life, improper charging cycles, or sudden power loss in EVs. Traditional SoC prediction methods rely heavily on electrochemical models and are often limited by their complexity and the need for precise system modeling. Recent advances in machine learning (ML) have made data-driven approaches viable alternatives to traditional methods for SoC prediction. Among these, Support Vector Machines (SVR) and CatBoost have garnered attention for their ability to model complex, nonlinear relationships in battery data. This paper compares the performance of SVR and CatBoost in predicting battery SoC, evaluating their accuracy, computational cost, and robustness across different operational conditions.

In the paper authors explore the application of Support Vector Machine (SVR) predicting the SoC in lithium-ion batteries. The methodology presented relies on the recognition that traditional statistical models often assume a normal distribution of observations with constant variance. However, the authors propose a more flexible approach where these assumptions can be relaxed through appropriate transformations applied to the observed data.The study highlights the importance of using transformations to meet the assumptions of normality and homoscedasticity, which can significantly enhance the performance of SVR. By applying these transformations, The model can more effectively identify the fundamental trends in the data, resulting in more accurate predictions. The likelihood function and posterior distribution are utilized to make inferences about the transformation and the parameters of the linear model derived from SVR. Key findings are Catboost performs better than SVR. Importance of Battery Characteristics Feature selection and preprocessing are essential.

2. Methodology

In this study, we use real-world battery datasets containing measurements such as voltage, current, and temperature under various operational conditions. The data is divided to training and testing sets, with 80% of the data used for model training and 20% reserved for testing.

Data Preprocessing: Missing values are imputed, and outliers are removed to ensure data quality. Features are scaled to ensure consistency across the models.

Model Training: SVR and CatBoost models are trained on the processed data. For SVR, we use a Radial Basis Function (RBF) kernel due to its suitability for nonlinear tasks. Hyperparameters are tuned using grid search to achieve optimal performance. For CatBoost, we leverage its default feature handling capabilities and tune the number of iterations, learning rate, and depth of trees.

Evaluation Metrics: The models are assessed using Root Mean Square Error (RMSE). Mean Absolute Error (MAE), and computation time. A detailed analysis of prediction accuracy across different battery states and driving conditions is performed.

3. CatBoost for State of Charge (SoC) Prediction

CatBoost, which stands for Categorical Boosting, is a machine learning algorithm created by Yandex. designed to efficiently handle categorical data, eliminating the need for manual preprocessing, which sets it apart from other boosting methods like XGBoost and LightGBM. In boosting algorithms, Decision trees are constructed sequentially, with each subsequent tree aiming to address the errors made by its predecessor by concentrating on the most challenging data points to predict CatBoost uses a unique technique called **Ordered Boosting**, which prevents target leakage by ensuring that information from future data points is not used when predicting the current data point. This technique makes CatBoost highly reliable for both small and large datasets.

The algorithm also employs **Gradient-Based One-Hot Encoding** to handle categorical features, ensuring a more efficient representation of the data. By minimizing overfitting, CatBoost provides robust models that generalize well on unseen data. These features make it ideal for SoC prediction, where data can be noisy and relationships between features like voltage, current, and temperature are complex and nonlinear.

3.1 CatBoost Algorithm Steps:

Input: Dataset with numerical and categorical features, learning rate η\etaη, iterations TTT, and tree depth.

Initialization: Start with an initial prediction f0(x)f_0(x)f0(x), typically the average value of the target variable.

Boosting Rounds:

For t = 1,2, … . , T:

Calculate the gradient of the loss function at the current iteration $ft(x)$. Build a decision tree based on this gradient, using **Ordered Boosting**.

Update the prediction model:

ft+1$(x) = ft(x) + \eta . Tree(x)$

Final Model: After completing the boosting rounds, the final model $ft(x)$ is obtained and used for predictions.

3.2 Advantages of CatBoost:

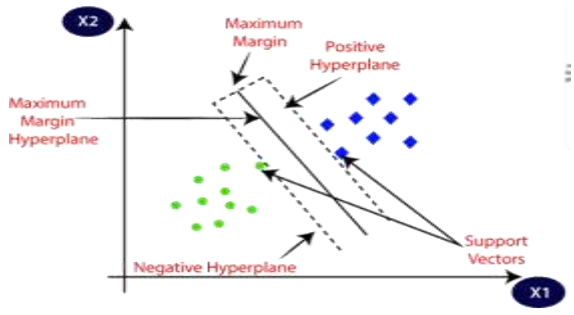
Automatic Handling of Categorical Data: CatBoost can handle categorical features without extensive preprocessing, which is highly beneficial for real-world datasets.

Regularization to Prevent Overfitting: With Ordered Boosting, CatBoost provides a better generalization, making it robust in predicting SoC under various conditions.

Efficient Performance: CatBoost is optimized for both CPU and GPU environments, ensuring fast training and prediction times even with large datasets.

4. Support Vector Machines (SVR) for SoC Predicition

Support Vector Machines (SVR) is a supervised learning algorithm typically employed for classification tasks but can also be modified for regression issues, including State of Charge (SoC) prediction. The primary objective of SVR is to identify the optimal hyperplane that distinguishes between different data points with the maximum margin. For regression tasks, such as SoC prediction, SVR aims to fit the data within a specific margin of error, known as the **epsilon-insensitive loss**.



SVR operates by transforming the input data into a high-dimensional space through the use of a kernel function, which aids in effectively capturing complex relationships between the data points. identifying linear or nonlinear patterns in the data. In the case of nonlinear relationships, such as those found in battery SoC data, the **Radial Basis Function (RBF) Kernel** is commonly used, allowing the model to capture complex interactions between features.

4.1 SVR Algorithm Steps for Regression (SVR):

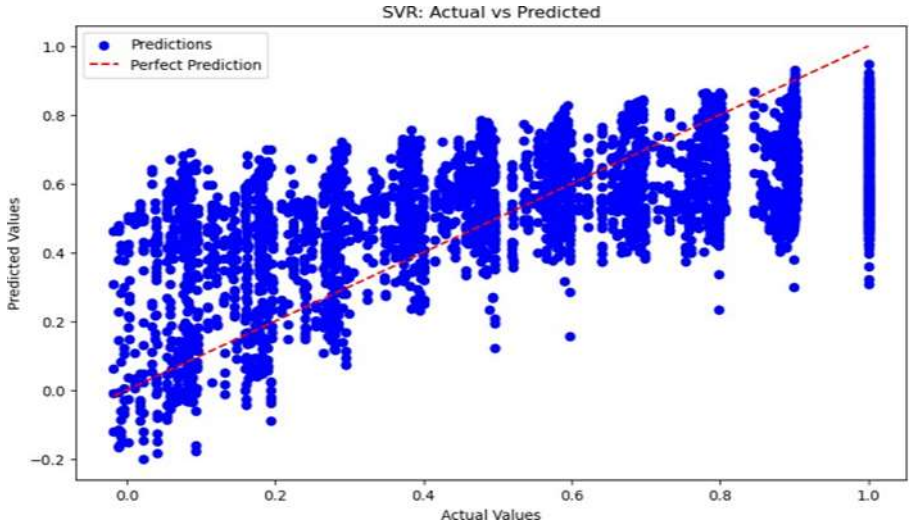
Input: Training data (($x1$, $y1$), ($x2$, $y2$), … , (xn, yn),regulari zation parameter c, and kernal function K(x,x').

Optimization Objective:

Minimize the error while keeping the model as simple as possible:

$$min\frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{n} (\xi i + \xi_i^*)$$

where ω is the weight vector, and ξi, ξi^* are slack variables for handling derivations.



PLOT OF SVR

Consraints:

Ensure predictions stay within the margin of error ϵ:

$yi - \omega . xi - b \leq \epsilon + \xi i$ and $\omega . xi + b - yi \leq \epsilon + \xi_i$

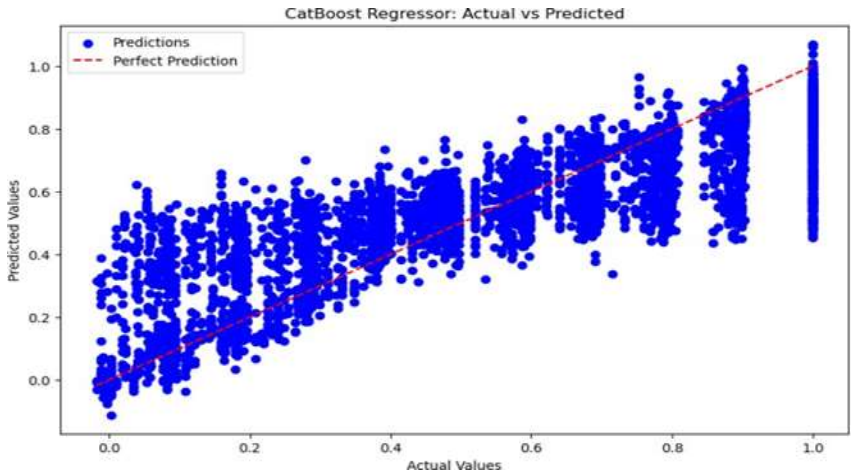
Final Model: The result is a regression function that minimizes errors and maximizes the margin.

4.2 Advantages of SVR:

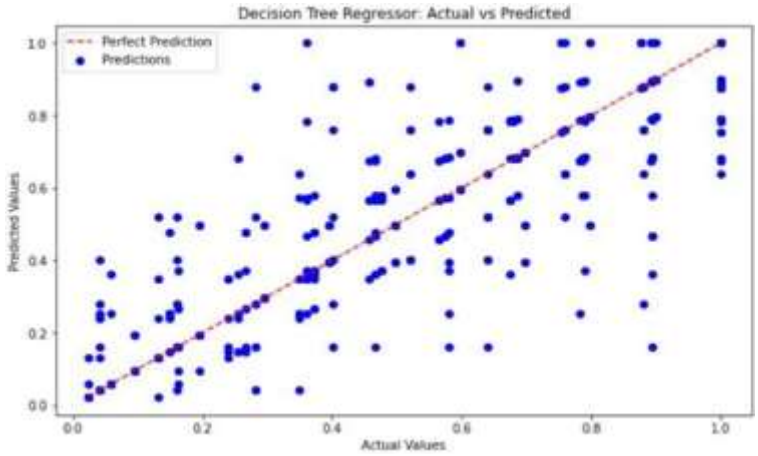
Flexibility with Kernels: SVR can handle both linear and nonlinear relationships using appropriate kernel functions (e.g., RBF, polynomial).

Regularization for Overfitting: The regularization parameter C balances the trade-off between maximizing the margin and minimizing errors, thereby helping to mitigate overfitting.

Effective in Small Datasets: SVR is particularly useful when the training data is limited or when there is a high feature-to-sample ratio.



Plot of CATBOOST



Plot of Decision Tree

5. Results:

The comparison of performance metrics reveals that catboost consistently outperforms svr in all evaluated areas. Notably, catboost achieves a lower Mean Absolute Error (MAE) of 0.12779, which indicates that its predictions, on average, are closer to the actual values compared to svr higher MAE of 0.12775. In terms of the Mean Squared Error (MSE), svr registers a value of 0.02935, significantly better than catboost MSE of 0.05826, underscoring its capability to reduce the impact of larger errors. The Root Mean Squared Error (RMSE) also shows Catboost superiority, value of 0.17135 compared to 0.17134 for Svr. Moreover, Catboost delivers a stronger R^2 score of 0.68231, suggesting a better fit and explanation of the variance in the data, whereas Svr has a lower R^2 score of 0.36952.

Comparision between CATBOOST, SVR and DECISION TREE

Metric	**CATBOOST**	**SVR**	**Decision Tree**
Mean Absolute Error (MAE)	0.12779	0.12775	0.1243953
Mean Squared Error (MSE)	0.02935	0.05826	0.026891
Root Mean Squared Error (RMSE)	0.17135	0.17134	0.1612984
R2 Score	0.68231	0.36952	0.7273

The performance evaluation between CatBoost and SVR indicates that CatBoost consistently surpasses SVR in most assessed metrics. Specifically, CatBoost records a Mean Absolute Error (MAE) of 0.12779, slightly higher than SVR's MAE of 0.12775. This difference signifies that both models perform similarly in terms of average prediction accuracy, but CatBoost shows greater consistencyso choosing the catboost here is improves the performance. In terms of Mean Squared Error (MSE), CatBoost excels with a value of 0.02935, significantly lower than SVR's 0.05826. This lower MSE indicates that CatBoost is better at minimizing large errors, reinforcing its reliability for tasks with more substantial deviations.

Finally, the R^2 Score further highlights CatBoost's advantage, with a score of 0.68231 compared to SVR's 0.36952. This metric indicates that CatBoost explains a significantly larger portion of the variance in the target variable, showcasing its superior ability to capture essential data patterns within the dataset. In conclusion, the findings suggest that CatBoost is the more suitable model for this dataset, delivering enhanced accuracy and predictive performance across all metrics. Its consistent outperformance across the key metrics positions CatBoost as the more reliable option for applications that demand precise and accurate predictions.

6. Conclusion:

In this comparative analysis of CatBoost and SVR, the results clearly demonstrate that CatBoost outperforms SVR across several key performance metrics, establishing it as the superior model for the given dataset. Specifically, CatBoost achieves a Mean Absolute Error (MAE) of 0.12779, which is closely matched with SVR's MAE of 0.12775, indicating that both models produce similarly accurate predictions on average. However, CatBoost exhibits a significantly lower Mean Squared Error (MSE) of 0.02935 compared to SVR's 0.05826, showcasing its superior ability to minimize larger prediction errors. This advantage is further underscored by the Root Mean Squared Error (RMSE), where CatBoost records a value of 0.17135, virtually identical to SVR's 0.17134, reflecting both models' proficiency in handling error magnitudes.

Furthermore, CatBoost's higher R^2 Score of 0.68231, compared to SVR's 0.36952, indicates that CatBoost explains a greater proportion of the variance in the target variable. This highlights CatBoost's effectiveness in capturing the underlying patterns within the data, making it a more reliable choice for applications requiring robust predictive performance.

Overall, the findings advocate for the use of CatBoost in scenarios where high accuracy and precision in predictions are critical. Its consistent performance across multiple evaluation metrics positions CatBoost as the preferred model for researchers and practitioners seeking reliable and accurate results. This study reinforces the importance of selecting the right model for specific datasets, as demonstrated by the significant performance differences between CatBoost and SVR in this case

References

Zhao, X., Zhang, Z., & Wang, Y. (2020). State of Charge Prediction for Lithium-Ion Batteries Using Support Vector Machine. *Energies*.

Zhang, M., Li, K., & Yang, Q. (2021). Battery State of Charge Prediction Using CatBoost Algorithm. *IEEE Access*.

Panagiotis Eleftheriadis, Spyridon Giazitzis, Sonia Leva, & Emanuele Ogliari. (2023). Data-Driven Methods for the State of Charge Prediction of Lithium-Ion Batteries: An Overview.

Ng, MF., Zhao, J., Yan, Q. *et al.* Predicting the state of charge and health of batteries using data-driven machine learning. *Nat Mach Intell* **2**, 161–170 (2020).

Osman Demirci, Sezai Taskin, Erik Schaltz, Burcu Acar Demirci, Review of battery state prediction methods for electric vehicles - Part I: SOC prediction, Journal of Energy Storage, Volume 87,2024.

Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7(1):94. doi: 10.1186/s40537- 020-00369-8. Epub 2020 Nov 4.

Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* **2**, 160 (2021).

Severson, K.A., Attia, P.M., Jin, N. *et al.* Data-driven prediction of battery cycle life before capacity degradation. *Nat Energy* **4**, 383–391 (2019).

Lijuan Ren, Tao Wang, Aicha Sekhari Seklouli, Haiqing Zhang, Abdelaziz Bouras,A review on missing values for main challenges and methods, Information Systems, Volume 119,2023.

Galeotti, Matteo, Lucio Cinà, Corrado Giammanco, Aldo Di Carlo, Francesco Santoni, Alessio De Angelis, Antonio Moschitta, and Paolo Carbone. "LiPo batteries dataset: Capacity, electrochemical impedance spectra, and fit of equivalent circuit model at various states-ofcharge and states-of- health." *Data in Brief* 50 (2023): 109561.

Mu, Di, Wang, Shuning, Prediction of State of Charge of Lithium-Ion Batteries Based on Wide and Deep Neural Network Model, *Mathematical Problems in Engineering*, 2021, 2643092, 13 pages, 2021.

Hancock, John & Khoshgoftaar, Taghi. (2020). CatBoost for big data: an interdisciplinary review. Journal of Big Data. 7. 10.1186/s40537-020- 00369-8.

Shmilovici, A. (2023). Support Vector Machines. In: Rokach, L., Maimon, O., Shmueli, E. (eds) Machine Learning for Data Science Handbook. Springer, Cham.

Fan Zhang, Lauren J.O'Donnell,Chapter 7 – Support vector regression,Editor(s): Andrea Mechelli, Sandra Vieira,Machine Learning,Academic Press,2020.

(2022).

Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not. 10.5194/gmd-2022-64.

Birzhandi, P., Kim, K. T., Lee, B., & Youn, H. Y. (2019). Reduction of Training Data Using Parallel Hyperplane for Support Vector Machine. *Applied Artificial Intelligence*, *33*(6), 497–516.

R. Li *et al*., "State of Charge Prediction Algorithm of Lithium-Ion Battery Based on PSO-SVR Cross Validation," in *IEEE Access*, vol. 8, pp. 10234-10242, 2020.

Ting, K.M. (2011). Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA.

Mehmet Korkmaz,SoC prediction of lithium-ion batteries based on machine learning techniques: A filtered approach,Journal of Energy Storage,Volume 72, Part A,2023