

Predicción de la vida útil remanente de la batería de vehículos de movilidad sostenible

D. Acosta¹

¹Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia

D. Acosta (e-mail: dfernando.acosta@udea.edu.co).

ABSTRACT Regression models, supported by machine learning algorithms, allow for the prediction of numerical variables based on reported features for a given system. In the field of sustainable mobility, the analysis of battery lifespan has gained particular importance due to its potential impact on recycling and replacement strategies aimed at reducing environmental consequences. In this context, RUL represents the remaining lifespan of a battery. In this study, machine learning techniques, such as Support Vector Machines and Gradient Boosting, were employed to predict RUL of electric bicycle batteries in two environments. The predictions were based on an experimental dataset obtained from sensor evaluations conducted during different cycles. Support Vector Machines are known for their ability to find optimal decision boundaries, while Gradient Boosting is a technique that combines multiple weak predictors to create a strong ensemble model. Those methods achieved an RMSE of 181.5 and 183.9, respectively, through the exploration of different hyperparameters.

INDEX TERMS Remaining useful life, machine learning, support vector machine regression, gradient boosting regression.

I. INTRODUCCIÓN

Mientras el contexto mundial se enfoca en discusiones relativas al impacto ambiental de las actividades humanas, se hace fundamental buscar alternativas al uso de recursos energéticos no renovables y potenciar un uso sostenible de las energías renovables, involucrando, por ejemplo, el uso de vehículos eléctricos[1]. En dicho contexto, autoridades gubernamentales de países desarrollados impulsan estrategias relativas al uso de vehículos eléctricos (VE) que van desde reducción de impuestos hasta subsidios de compra e incluso infraestructura de carga[2]. Por otra parte, la producción de vehículos eléctricos y, en especial las baterías de estos pueden generar afectaciones ambientales que deben tratarse a través de estrategias que mejoren la posibilidad de reciclaje y uso en segunda instancia, además de desarrollos tecnológicos enfocados en el seguimiento de la vida útil de las baterías[3].

Las técnicas existentes de Machine Learning, proveen un entorno de amplias posibilidades enfocadas en la predicción de diferentes fenómenos. En este caso, se propone aplicar métodos de ML, modelos de regresión, en la estimación del estado de vida útil de la batería el RUL (Remaining useful life) [4] de un modelo de bicicleta eléctrica ajustando las relaciones entre factores relativos al desgaste de esta.

El presente informe contiene la estimación de la vida útil remanente de la batería de una bicicleta eléctrica a partir del análisis de dos baterías y diferentes variables reportadas en entornos experimentales. Para la predicción se usaron, ajustaron y compararon: Epsilon-Support Vector Regression, Gradient Boosting for regression, AdaBoost regressor y random forest regressor. Estos modelos se apoyaron en la librería SKLearn de Python y se implementaron a través de Notebooks.

En el contexto de la predicción de la vida útil remanente en baterías, se encontraron varios aportes relacionados: Pang [5] aplicó feature separable convolution (AFSC) y convolutional long short-term memory (ConvLSTM) además de Red AFSC y ConvLSTM, comparando MAPE, RMSE y MAE. Yuan [6] usó gradient boosting regression (GBR), support vector regression (SVR), Hist-GBR, AdaBoost, y linear regression (LR), comparando verage root mean square error (ARMSE) y relative error (RE). Zhang [7] utilizó SVR y Gaussian Process Regression (GPR) comparando MAPE y RMSE. Zhao [8] se apoyó en un modelo de conjunto basado en una estrategia de apilamiento con LR, Randon Forest Regressor (RFR), GBR y GPR. Incluso, otros autores como [9] se apoyan en estrategias de series de tiempo, incluyendo AutoRegressive Integrated Moving Average model with eXogenous predictors, ARIMAX y estrategias de regresión como

Linear Quantile Regression (LQR), Bootstrap Multiple Linear Regression (B-MLR) y Bayesian Bootstrap Multiple Linear Regression (BB-MLR).

II. METODOLOGÍA

El conjunto de datos: Estudiantes asociados al Grupo ALIADO de la Universidad de Antioquia recolectaron información de 62 ciclos de carga y descarga (1 característica: ciclo) de 2 baterías (1 característica: batería) y dos motores (1 característica: motor) de bicicleta eléctrica para un total de 91.811 registros, a través del seguimiento de pruebas sobre condiciones de tráfico real en dos vías con diferentes particularidades y, por motivos climáticos, algunas pruebas se llevaron a cabo en laboratorio (1 característica: lugar). Estos 62 ciclos contienen información de corriente, voltaje, potencia y temperatura en la batería (4 características) y mediciones de estas variables al motor en sus 3 polos (12 características). Se recolectaron además la humedad ambiental y la temperatura ambiente (2 características). Por otra parte, se tiene información de la aceleración en 3 ejes (3 características) y la duración, distancia, velocidad máxima y velocidad media (4 características). También se conoce el peso del conductor (1 característica). Esta información se recolectó a lo largo de un año.

El preprocesamiento: Se realizó el preprocesamiento en Python y se eliminaron ciclos completos de acuerdo con la bitácora de seguimiento además de registros que presentaron fallas de recolección. Algunos registros se imputaron mediante interpolación, se cambió algunos nulos por primeros registros. Finalmente se seleccionaron solamente los resultados experimentales de una de las vías analizadas y se tomaron únicamente las variables relativas al motor y la batería, excluyendo las variables ambientales y adicionales, incluyendo 49537 registros y 16 variables. Se calculó la variable objetivo RUL:

$$RUL_{i,j} = t_{n,j} - t_{i,j} + 1 \quad \forall j \quad (1)$$

donde i es el registro de tiempo en cada ciclo j y n es el máximo tiempo en el ciclo j .

Se realizó una partición aleatoria 80% de entrenamiento y 20% de prueba. Con este 80% se aplicaron diferentes pruebas según el caso.

Experimentos: Se implementaron modelos SVR, GBR, ABR, RFR. Al SVR se le realizó un GridSearch para encontrar parámetros adecuados, modificando kernel (linear, poly, rbf y sigmoid), C , gamma y grado; posteriormente se aplicó un testeo para revisar la capacidad de predicción mediante RMSE. Al resto de modelos, se les realizó una búsqueda gráfica de parámetros óptimos a partir de los niveles de error (RMSE y R^2) según cambios en diferentes hiperparámetros como en Fig. 1, pero no solo se buscó el valor que garantice el

mejor error, sino que se encontró aquel que no implicara cierto nivel de sobreajuste. Luego, se midió su eficiencia final mediante CrossValidation.

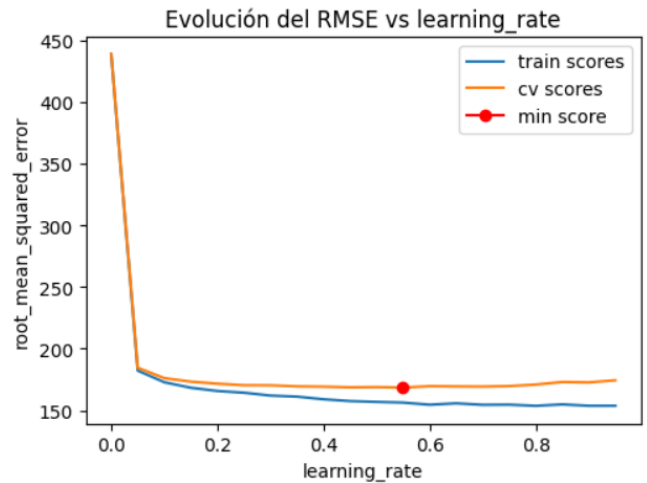


Figura 1. Verificación del error mediante RMSE modificando el hiperparámetro de learning rate en Gradient Boosting Regressor.

Se usó RMSE como medida de error decisiva para todos los modelos con un 20% de datos de prueba:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

III. RESULTADOS Y DISCUSIÓN

TABLA 1 Resultados de los modelos implementados		
Modelo	Mejores hiperparámetros	RMSE
SVR	C=10 gamma='scale' kernel='rbf'	181.5
GBR	learning_rate=0.2 max_depth=3 n_estimators=100	183.9
ABR	learning_rate=0.2 n_estimators=50 loss='exponential'	245.3
RFR	Overfit, se descartó	

Como se observa en la Tabla 1. el modelo con mejor desempeño corresponde a un SVR con $C=10$, gamma='scale' y kernel='rbf'. La búsqueda de los mejores hiperparámetros tuvo un alto costo computacional, lo que podría implicar la decisión de tomar un método más eficiente en principio, sin embargo, el entrenamiento directo del modelo no tiene gran diferencia con respecto al GBR. El RFR presentó overfit por lo que fue descartado, con la posibilidad posterior de reevaluar

los hiperparámetros iniciales. El R2 de los modelos fue superior a 0,8.

IV. CONCLUSIÓN

Desde la perspectiva de modelamiento, el conjunto de datos permite una aproximación interesante a la predicción si se evalúa R2, puesto que los modelos rondan un valor superior a 0,8.

Se requiere soportar las conclusiones en conocimiento específico en el área de electrónica para analizar las conclusiones puesto que, por el momento, no se tiene un artículo como comparación directa y la eficiencia basada en el RMSE podría depender de las exigencias del fabricante o los topes establecidos por los académicos.

El modelo SVR puede implicar alto costo computacional al momento de evaluar cambios en los hiperparámetros pero su desempeño es superior. Esta relación de costo beneficio debe ser tomada en cuenta al momento de elegir un modelo que requiera cierta capacidad dinámica de reentrenamiento. Este modelo fue el mejor pero su diferencia no es radical en comparación con el GBR, el cual se muestra como una alternativa adecuada.

Las diferentes estrategias de búsqueda de hiperparámetros enriquecen de manera significativa la capacidad predictiva de los modelos y su aplicación provee una herramienta más confiable.

Se espera posteriormente implementar otros modelos más robustos y validar las implicaciones de las predicciones obtenidas, llegando a conclusiones basadas en opinión de expertos.

La variable objetivo, calculada desde cada registro fue la más adecuada, puesto que, si se calcula un RUL por ciclo, se hace necesaria una transformación de las variables en medidas estadísticas que dependen del comportamiento intrínseco y, no solo abordan complejidades estadísticas, sino que implican una reducción muy significativa del tamaño de los datos, haciendo poco factible la aplicación de modelos de aprendizaje automático.

REFERENCIAS

- [1] N. Ding, K. Prasad, y T. T. Lie, “The electric vehicle: a review”, 2017.
- [2] J. A. Sanguesa, V. Torres-Sanz, P. Garrido, F. J. Martínez, y J. M. Marquez-Barja, “A review on electric vehicles: Technologies and challenges”, *Smart Cities*, vol. 4, núm. 1. MDPI, pp. 372–404, el 1 de marzo de 2021. doi: 10.3390/smartcities4010022.
- [3] G. Harper *et al.*, “Recycling lithium-ion batteries from electric vehicles”, *Nature*, vol. 575, núm. 7781. Nature Publishing Group, pp. 75–86, el 7 de noviembre de 2019. doi: 10.1038/s41586-019-1682-5.
- [4] J. Hong, Z. Wang, W. Chen, L. Wang, P. Lin, y C. Qu, “Online accurate state of health estimation for battery systems on real-world electric vehicles with variable driving conditions considered”, *J Clean Prod*, vol. 294, abr. 2021, doi: 10.1016/j.jclepro.2021.125814.
- [5] X. Pang *et al.*, “A novel hybrid model for lithium-ion batteries lifespan prediction with high accuracy and interpretability”, *J Energy Storage*, vol. 61, may 2023, doi: 10.1016/j.est.2023.106728.
- [6] J. Yuan, Z. Qin, H. Huang, X. Gan, S. Li, y B. Li, “State of Health Estimation and Remaining Useful Life Prediction for a Lithium-Ion Battery with a Two-Layer Stacking Regressor”, *Energies (Basel)*, vol. 16, núm. 5, mar. 2023, doi: 10.3390/en16052313.
- [7] Y. Zhang y M. Zhao, “Cloud-based in-situ battery life prediction and classification using machine learning”, *Energy Storage Mater*, vol. 57, pp. 346–359, mar. 2023, doi: 10.1016/j.ensm.2023.02.035.
- [8] J. Zhao, H. Ling, J. Liu, J. Wang, A. F. Burke, y Y. Lian, “Machine learning for predicting battery capacity for electric vehicles”, *eTransportation*, vol. 15, ene. 2023, doi: 10.1016/j.etrans.2022.100214.
- [9] A. Bracale, P. De Falco, L. P. Di Noia, y R. Rizzo, “Probabilistic State of Health and Remaining Useful Life Prediction for Li-Ion Batteries”, *IEEE Trans Ind Appl*, vol. 59, núm. 1, pp. 578–590, ene. 2023, doi: 10.1109/TIA.2022.3210081.