

**Trabajo previo a la obtención de título de
Magister en**

**MAESTRÍA EN CIENCIA DE DATOS Y MÁQUINAS DE APRENDIZAJE CON
MENCIÓN EN INTELIGENCIA ARTIFICIAL**

AUTORES:

Buitrón Pambabay Samantha Carolina

Espinoza Bone Jose Manuel

Rosero Ordoñez Evelin Marisol

Salazar Grandes Mayra Cecilia

TUTORES:

Mora Cajas Karla Estefanía

Lopez Tito Juan Fernando

Modelo Ensemble basado en técnicas de aprendizaje
supervisado para la predicción de fallas electromecánicas
en sistemas de producción

Resumen

El presente trabajo desarrolla un modelo Ensemble basado en técnicas de aprendizaje supervisado para estimar la probabilidad de que una máquina presente al menos una falla electromecánica durante los siguientes 30 días. El estudio utilizó 10.527 registros históricos, los cuales fueron depurados, agregados por máquina y día, complementados con periodos sin fallas y enriquecidos mediante variables temporales de 7 y 15 días. Se evaluaron los algoritmos Regresión Logística, Random Forest y Gradient Boosting mediante validación temporal, optimización de hiperparámetros y selección de umbrales de clasificación. Posteriormente, se analizaron diferentes configuraciones Soft Voting, seleccionándose un Ensemble compuesto por Regresión Logística y Random Forest, con pesos de 1 y 4 y un umbral de 0,62. En el conjunto de prueba, el modelo alcanzó una precisión de 0,941, un recall de 0,919, una especificidad de 0,547, una balanced accuracy de 0,733 y un coeficiente MCC de 0,433. Los resultados evidenciaron un comportamiento más equilibrado que los modelos individuales y la línea base. Finalmente, el modelo fue integrado en un dashboard orientado a la priorización de máquinas y al apoyo de la planificación del mantenimiento.

Palabras clave: aprendizaje supervisado, modelo Ensemble, mantenimiento predictivo, fallas electromecánicas, Soft Voting.

Abstract

This study develops an Ensemble model based on supervised learning techniques to estimate the probability that a machine will experience at least one electromechanical failure within the following 30 days. The analysis used 10,527 historical failure records, which were cleaned, aggregated by machine and day, complemented with failure-free periods, and enriched through 7-day and 15-day temporal features. Logistic Regression, Random Forest, and Gradient Boosting models were evaluated using temporal validation, hyperparameter optimization, and classification-threshold selection. Different Soft Voting configurations were subsequently assessed, and the final Ensemble combined Logistic Regression and Random Forest with weights of 1 and 4, respectively, and a classification threshold of 0.62. On the test set, the model achieved a precision of 0.941, recall of 0.919, specificity of 0.547, balanced accuracy of 0.733, and an MCC of 0.433. These results showed a more balanced performance than the individual models and the baseline classifier. Finally, the model was integrated into a dashboard designed to prioritize machines according to their estimated risk and support maintenance planning.

Keywords: supervised learning, Ensemble model, predictive maintenance, electromechanical failures, Soft Voting.