

Desarrollo y comparación de modelos de aprendizaje automático para la predicción del ciclo de vida útil restante (RUL) en motores turbofán: integración de Graph Neural Networks (GNN) y XG Boost.

Nombre del Equipo Turbinas ML

Autores

- Sara Illatopa Marín
- Jose Manuel López Arias
- Julio César Posada Torres
- David Álvarez Grisales

Fecha Noviembre 2025

Índice

Introducción.....	4
Marco Teórico y Antecedentes.....	5
Desarrollo Metodológico.....	6
Análisis Exploratorio de Datos (EDA) y Preparación de Datos.....	6
Selección de Modelos e Ingeniería de Características.....	11
Entrenamiento y Evaluación.....	15
Análisis de Resultados y Conclusiones.....	16
Conclusiones.....	17

Objetivo

Desarrollar y comparar rigurosamente el desempeño de dos enfoques de aprendizaje automático, XG Boost y Graph Neural Networks (GNN), para la predicción del Remaining Useful Life (RUL) en motores turbofán.

Objetivos específicos:

1. Implementar un modelo novedoso (GNN) para la predicción del RUL en motores turbofán, utilizando el conjunto de datos C-MAPSS de la NASA, como la contribución principal del proyecto.
2. Comparar sistemáticamente el desempeño, la generalización y la capacidad predictiva del modelo basado en GNN con un modelo clásico de alto rendimiento (XG Boost).
3. Realizar una ingeniería de características exhaustiva mediante el Análisis Exploratorio de Datos (EDA) y la cuantificación de la importancia de variables utilizando los métodos nativos de XG Boost y la Regresión Lasso (Lasso Regression).
4. Aplicar un Autoencoder para el filtrado de características, la reducción de ruido y la compresión de dimensionalidad, con el fin de optimizar las entradas para ambos modelos y mejorar la interpretabilidad de los resultados.
5. Evaluar y documentar los resultados del análisis comparativo utilizando métricas estándar de regresión, para determinar si el enfoque basado en GNN ofrece una mayor precisión predictiva sobre el modelo de gradient boosting.

Introducción

El mantenimiento predictivo representa un cambio de paradigma en la industria aeronáutica, superando las limitaciones estratégicas y económicas del mantenimiento preventivo tradicional, basado en horas de vuelo, y del mantenimiento correctivo, que resulta inviable debido a los inaceptables riesgos de seguridad. Este enfoque avanzado se fundamenta en el análisis continuo de datos operativos para anticipar fallas en componentes críticos antes de que ocurran, optimizando así la fiabilidad y la disponibilidad de los activos.

El problema central que aborda este proyecto es la estimación del **Ciclo de Vida Útil Restante** (Remaining Useful Life o RUL) en motores turbofán. El desafío técnico consiste en modelar un problema de regresión de alta dimensionalidad sobre series temporales multivariadas, donde el objetivo es predecir un horizonte temporal decreciente (RUL) a partir de las sutiles desviaciones en los datos de los sensores que anuncian una falla inminente.

Una solución precisa para la predicción del RUL genera un impacto transformador en múltiples dimensiones. En primer lugar, incrementa la seguridad operacional al permitir la identificación proactiva de patrones de degradación que podrían derivar en fallos catastróficos. A nivel económico, optimiza significativamente los costos al evitar reemplazos prematuros de componentes y minimizar las paradas no planificadas. Finalmente, mejora la eficiencia logística, facilitando una planificación precisa para la adquisición de repuestos y la programación de intervenciones de mantenimiento.

El objetivo principal de este proyecto es desarrollar y comparar rigurosamente el desempeño de dos enfoques de aprendizaje automático para la predicción del RUL. Se evaluará un modelo clásico de alta eficacia, **XG Boost**, frente a un modelo basado en **Redes Neuronales de Grafos (Graph Neural Networks - GNN)**. Para llevar a cabo este análisis comparativo, se utilizará el conjunto de datos estándar de la industria para este propósito: el dataset C-MAPSS, proporcionado por el Centro de Investigación Ames de la NASA.

Este informe detalla la metodología implementada, desde el análisis y preparación de los datos hasta el entrenamiento y evaluación de los modelos, para finalmente presentar un análisis de los resultados y las conclusiones derivadas del estudio.

Marco Teórico y Antecedentes

Un análisis de la literatura existente no sólo valida la selección de técnicas y herramientas, sino que también contextualiza la contribución del trabajo. En este caso, la investigación previa confirma la idoneidad del dataset NASA C-MAPSS como un estándar de la industria y la pertinencia de las técnicas de aprendizaje automático para abordar el problema de la estimación del RUL.

El mantenimiento predictivo se centra en el uso de modelos estadísticos y de aprendizaje automático para predecir el momento futuro de una falla. En el contexto de sistemas complejos como los motores turbofán, la estimación del RUL es un problema de regresión sobre series temporales. La literatura científica ha explorado diversas técnicas para este fin, incluyendo modelos tradicionales como la regresión multivariable, y enfoques más avanzados como *random forests*, *support vector machines* (SVM), y, más recientemente, arquitecturas de redes neuronales profundas como las Redes Neuronales Recurrentes (LSTM y GRU), que son especialmente efectivas para capturar dependencias temporales.

El dataset **NASA C-MAPSS** (Commercial Modular Aero-Propulsion System Simulation) es un conjunto de datos simulados que se ha consolidado como el principal referente para el *benchmarking* de algoritmos de pronóstico en este dominio. Simula las trayectorias de degradación de una flota de motores turbofán bajo distintas condiciones operativas y modos de falla, proporcionando un entorno controlado y realista para la validación de modelos.

Estudios previos clave que han utilizado este dataset y han informado nuestro enfoque incluyen:

- **Saxena & Goebel (2008):** Introdujeron el dataset C-MAPSS y establecieron su utilidad para la validación de algoritmos de pronóstico. Subrayan la importancia de contar con métricas estandarizadas para comparar de manera objetiva el rendimiento de los diferentes modelos de predicción de RUL.
- **Zheng et al. (2017):** Demostraron la superioridad de los modelos basados en *Long Short-Term Memory* (LSTM) sobre métodos más tradicionales, gracias a la capacidad de estas arquitecturas para capturar dependencias a largo plazo en los datos de los sensores.
- **Li et al. (2019):** Propusieron enfoques híbridos que combinan la ingeniería de características manual con redes neuronales profundas. Sus resultados mostraron mejoras significativas en la precisión de las predicciones del RUL en los diferentes subconjuntos de C-MAPSS.

Este marco de antecedentes confirma la viabilidad técnica del proyecto y establece la base sobre la cual se implementará una metodología rigurosa para comparar un modelo clásico con un enfoque innovador.

Desarrollo Metodológico

¿Cómo podemos estimar con la mayor precisión posible la vida útil restante de un motor turbofán, utilizando exclusivamente los datos generados por sus sensores durante la operación? Para responder a esta pregunta, se utilizó el dataset de simulación de motores a reacción NASA C-MAPSS.

Las características principales de este conjunto de datos son:

- **Fuente:** Propulsion Center of Excellence (PCoE), Ames Research Center, NASA.
- **Propósito:** Proveer datos simulados de la degradación progresiva de una flota de motores turbofán para el desarrollo y validación de algoritmos de pronóstico y diagnóstico.
- **Subconjuntos:** El dataset está dividido en cuatro subconjuntos (FD001, FD002, FD003 y FD004), que varían en el número de condiciones operativas (1 o 6) y modos de falla (1 o 2). Este proyecto se enfoca específicamente en el subconjunto **FD001**, que presenta una única condición operativa (nivel del mar) y un único modo de falla (degradación del compresor de alta presión).
- **Estructura:** Los datos se presentan en 26 columnas, que incluyen el identificador del motor (ID), el ciclo de tiempo (operación), 3 variables de configuración operacional y 21 lecturas de sensores que monitorean parámetros como temperatura, presión, RPM y flujo.
- **Archivos:** Para cada subconjunto, se proporcionan tres archivos clave:
 - **Train FD001.txt:** Contiene las trayectorias completas de 100 motores, desde un estado sano hasta el punto de falla. Se utiliza para el entrenamiento de los modelos.
 - **test FD001.txt:** Contiene trayectorias truncadas de otros 100 motores, finalizando en un punto aleatorio antes de la falla. Se utiliza para evaluar el desempeño predictivo del modelo.
 - **RUL_FD001.txt:** Proporciona los valores verdaderos del RUL para cada una de las trayectorias del conjunto de prueba, sirviendo como *ground truth* para la evaluación.

Análisis Exploratorio de Datos (EDA) y Preparación de Datos

El Análisis Exploratorio de Datos es un paso fundamental que precede a la modelización, ya que permite identificar patrones, anomalías, correlaciones y la relevancia preliminar de las variables.

En este análisis se usaron varios tipos de gráficas, se usó un histograma del RUL del conjunto de entrenamiento (train). El RUL indica cuántos ciclos de vida le quedan al motor antes de fallar. El histograma permite identificar si los datos están concentrados cerca del final de vida. Detectar si hay sesgos hacia motores más nuevos o más usados. Observar la

variabilidad del dataset, lo cual es clave para entrenar un modelo robusto. También observamos la distribución del RUL en el testeо.

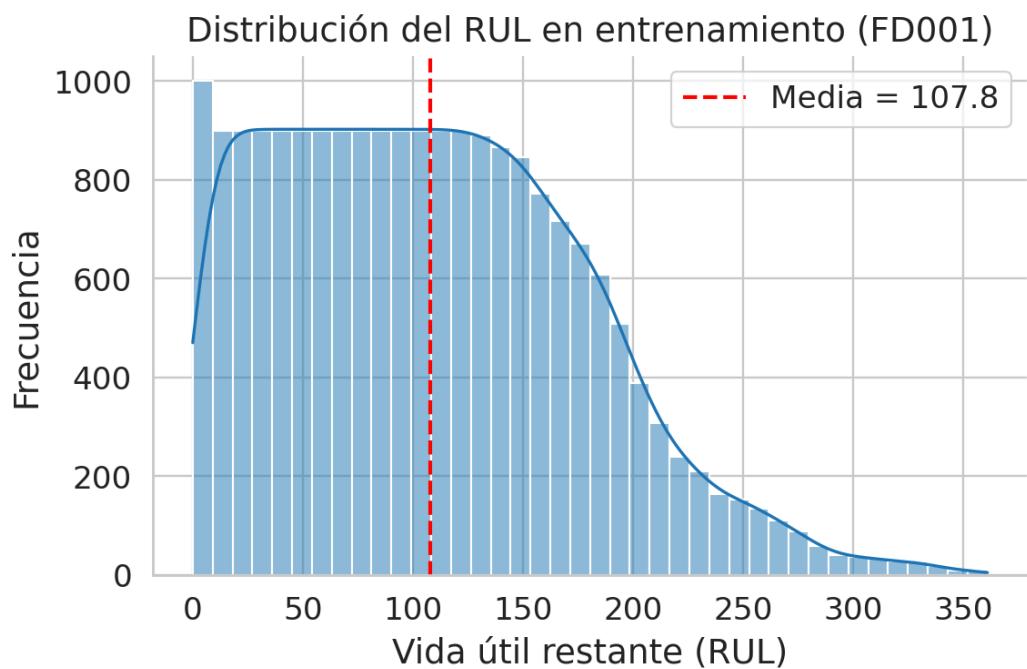


Figura 1

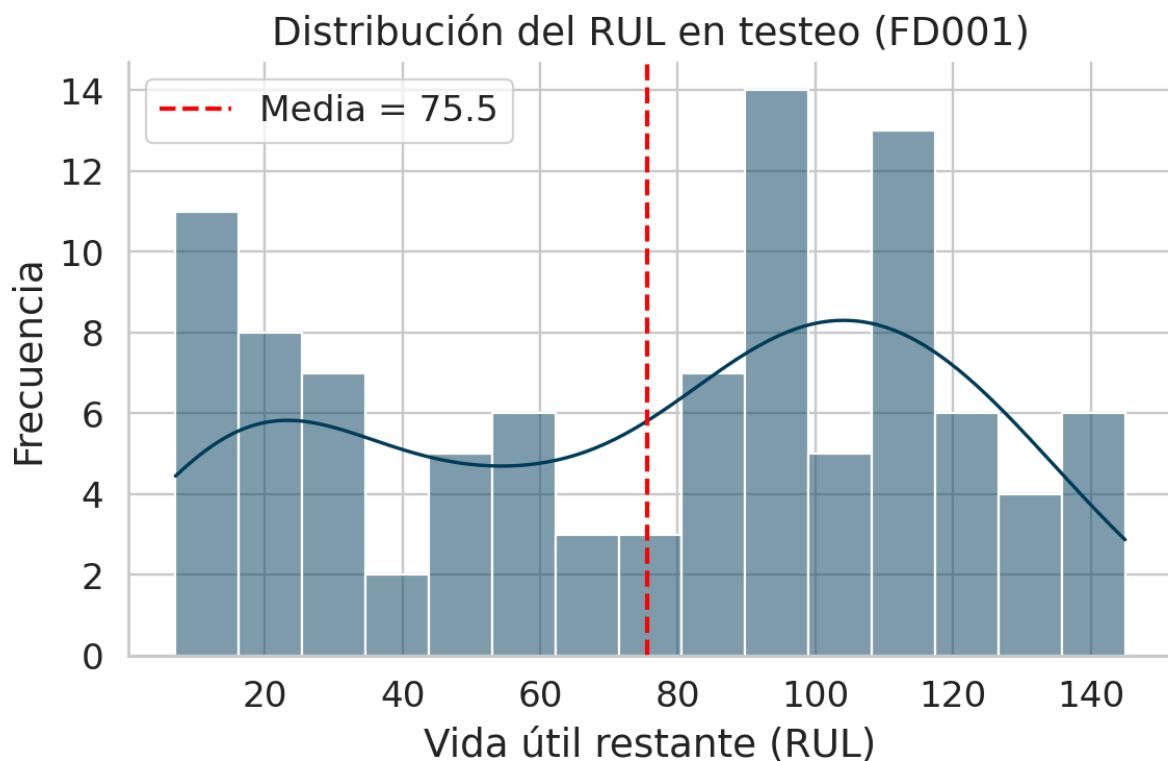


Figura 2

También usamos un mapa de calor para ver cómo se correlacionan las variables entre ellas, dándonos los siguientes resultados:

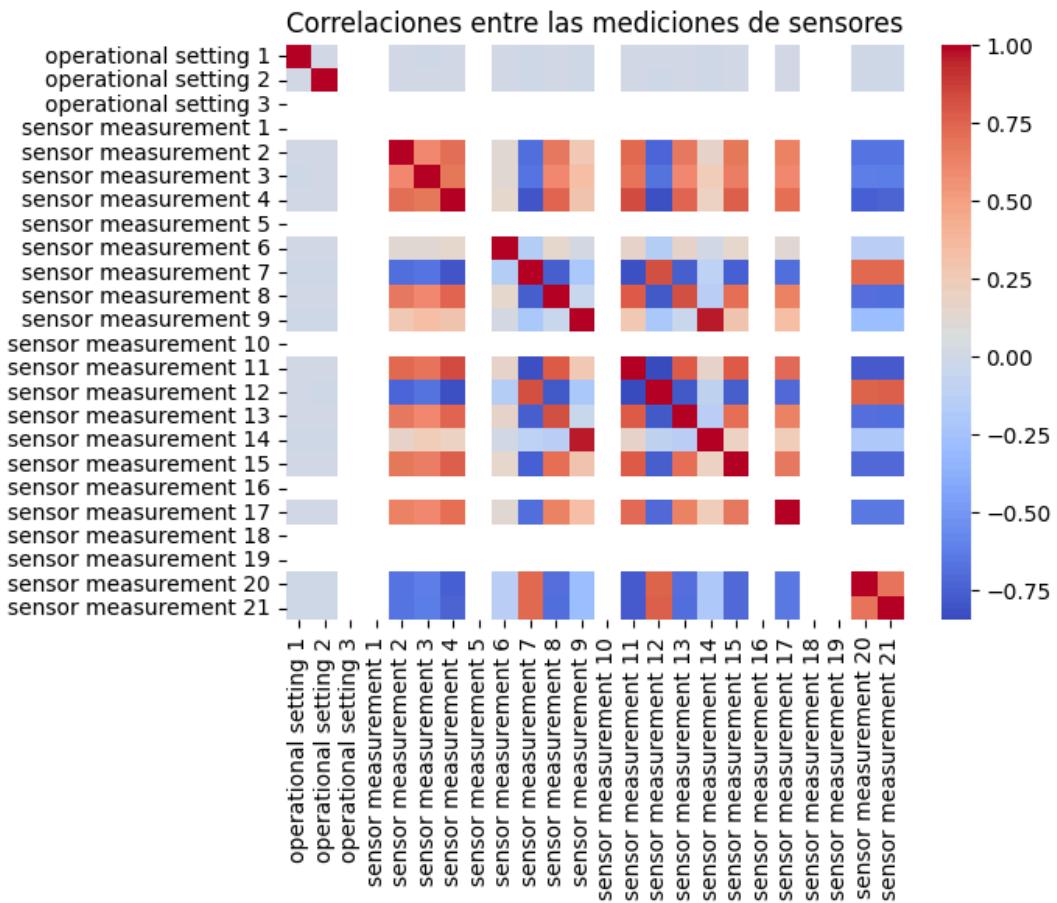


Figura 3

En este mapa de calor se observa la relación entre cada sensor.

El análisis inicial de los datos reveló que varios sensores presentaban varianza nula o muy baja, indicando que no aportan información significativa sobre el proceso de degradación y, por lo tanto, podían ser descartados para simplificar el modelo en ambos motores, motor 1 y 5.

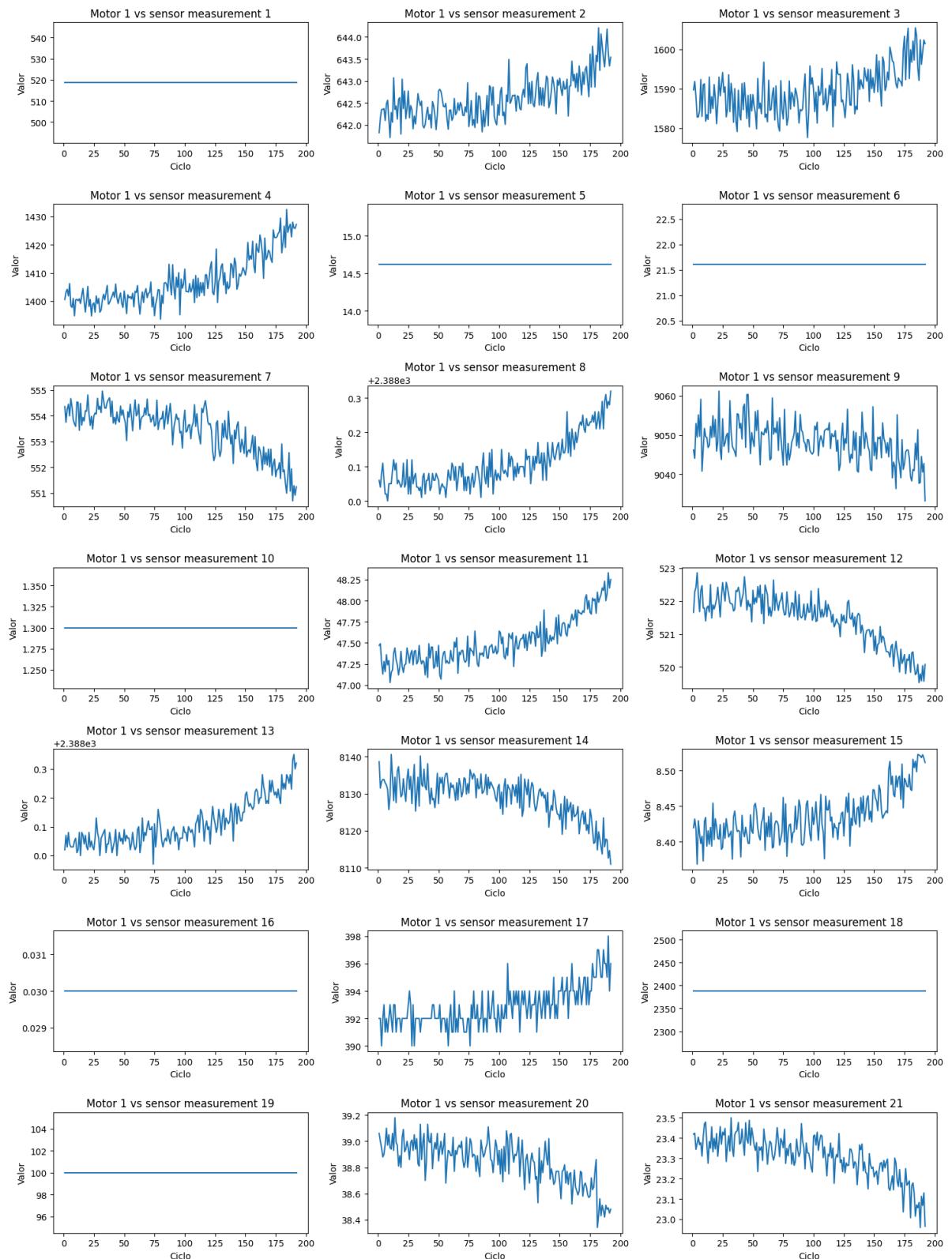


figura 4

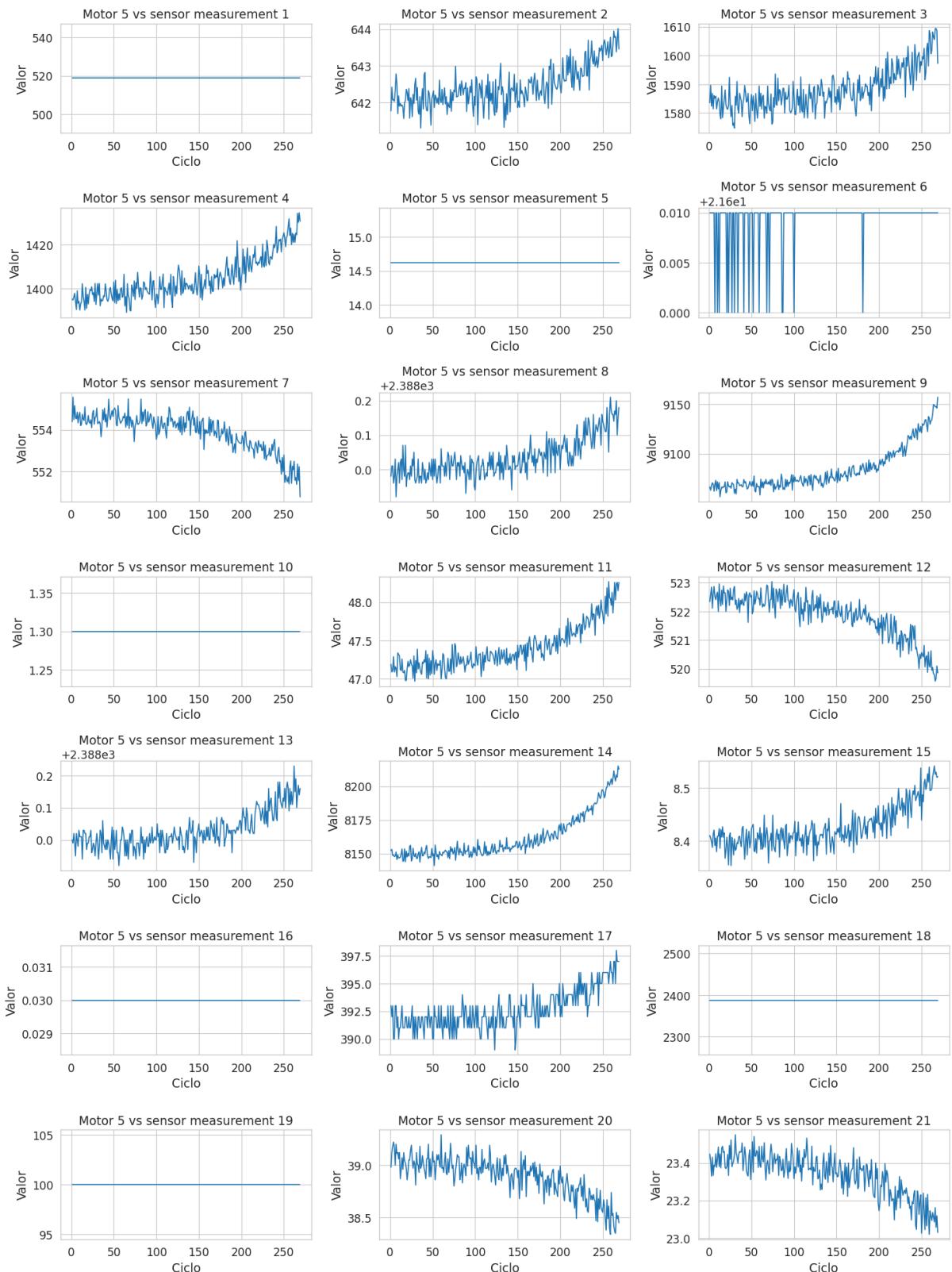


Figura 5

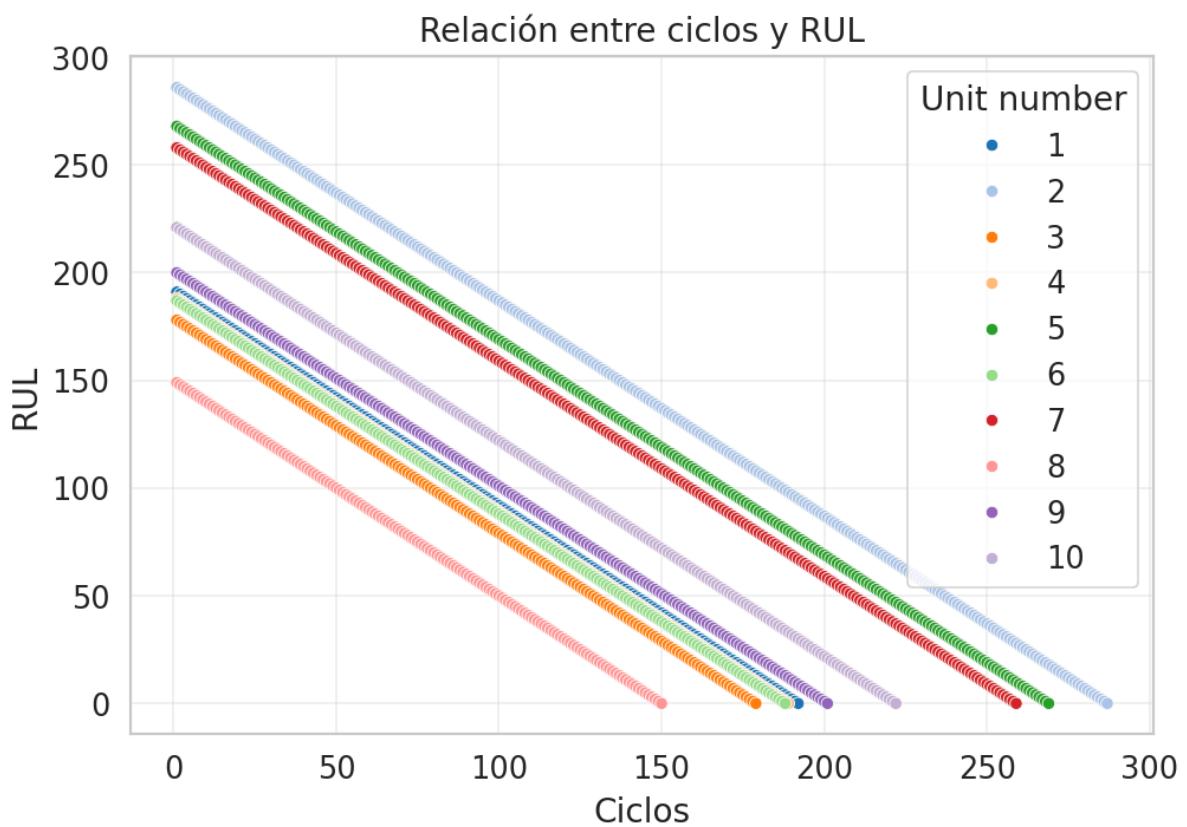


Figura 6

Selección de Modelos e Ingeniería de Características

La estrategia de modelado se centró en la comparación de un algoritmo clásico, conocido por su robustez, y un enfoque novedoso para este tipo de problema basado en grafos.

Se emplearon métodos complementarios para evaluar la relevancia de cada variable y extraer la información más útil:

- Importancia de características de XG Boost: Cuantifica la contribución de cada sensor a la predicción del RUL según el modelo de gradient boosting.
- Coeficientes de Lasso CV: Utiliza la regresión penalizada para identificar las variables con mayor poder predictivo, llevando los coeficientes de las menos relevantes a cero.
- Autoencoder para reducción de dimensionalidad: Una red neuronal entrenada para comprimir la información de los sensores en un espacio latente de menor dimensión, filtrando el ruido y extrayendo las características más representativas.

Identificación de predictores clave:

El resultado más importante de esta etapa fue la identificación consistente, a través de XGBoost y Lasso, de que los sensores 11, 4, 9 y 12 son los predictores clave. La consistencia

de este hallazgo a través de dos métodos distintos otorga una alta confianza en que estos sensores capturan el proceso de degradación física central del compresor de alta presión. Este anclaje físico es crucial, ya que confirma que los modelos no están simplemente ajustándose al ruido, sino que están aprendiendo una representación significativa del modo de falla conocido para el subconjunto FD001.

Evaluación de la eficacia del Autoencoder.

El Autoencoder demostró ser una técnica de ingeniería de características muy eficaces. Al entrenar un modelo XG Boost utilizando las características comprimidas generadas por el autoencoder, se observó que el componente emb 1 concentró más del 67% de la importancia total, validando que fue capaz de encapsular la mayor parte de la señal útil en una única variable.

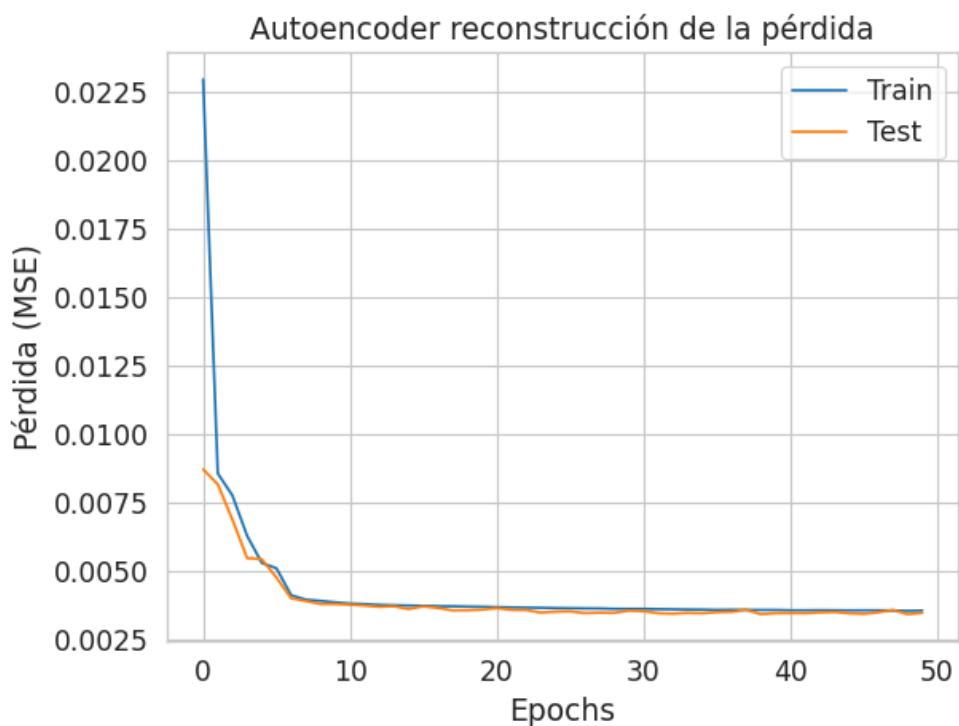


Figura 7

Configuración de Modelos para Comparación.

Con base en los análisis previos, se definió una lista definitiva de cinco modelos para ser entrenados y evaluados sistemáticamente:

1. XG Boost Baseline (sin embeddings): Este modelo sirve como punto de partida. Es entrenado directamente con los datos de los sensores preprocesados, sin aplicar una selección de características avanzada, para establecer una línea base de rendimiento.

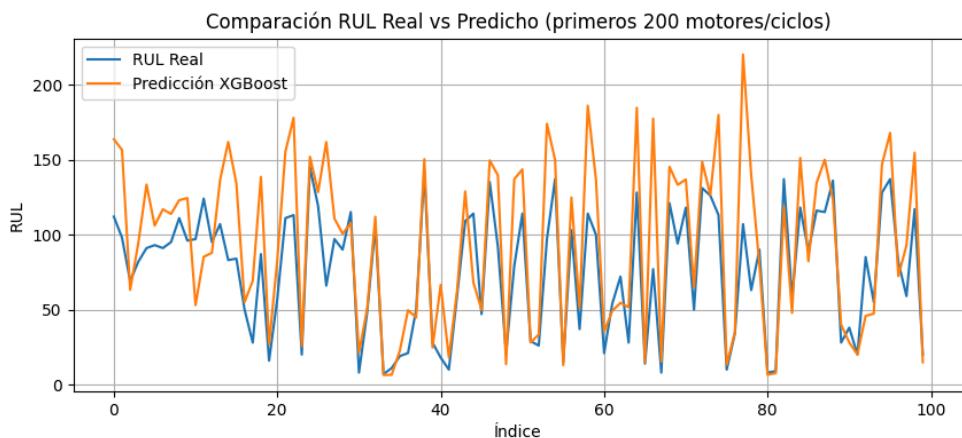


Figura 8

2.XG Boost con selección de características Lasso: Este modelo se entrena utilizando únicamente el subconjunto de sensores identificados como más relevantes por el modelo Lasso CV (sensores 11, 4, 9 y 12). Su objetivo es medir el impacto de una selección de características explícita en la precisión.

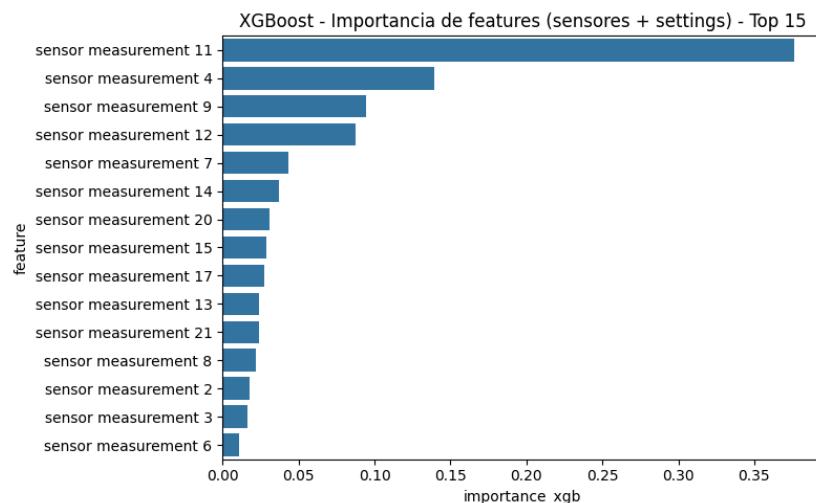


Figura 9

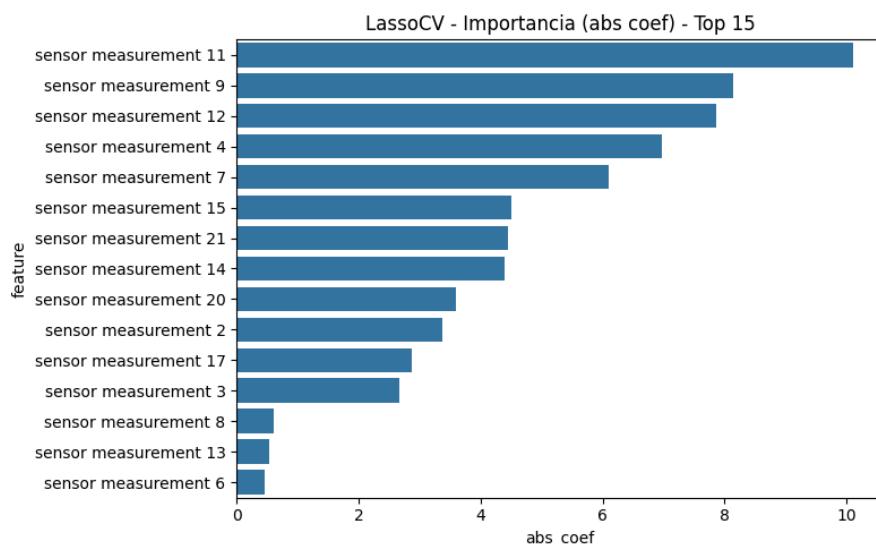


Figura 10

3.XG Boost con Embeddings: Utiliza como entrada las características comprimidas y filtradas generadas por el Autoencoder (el espacio latente) en lugar de los datos de los sensores crudos. Busca evaluar la eficacia de la reducción de dimensionalidad como método de ingeniería de características.

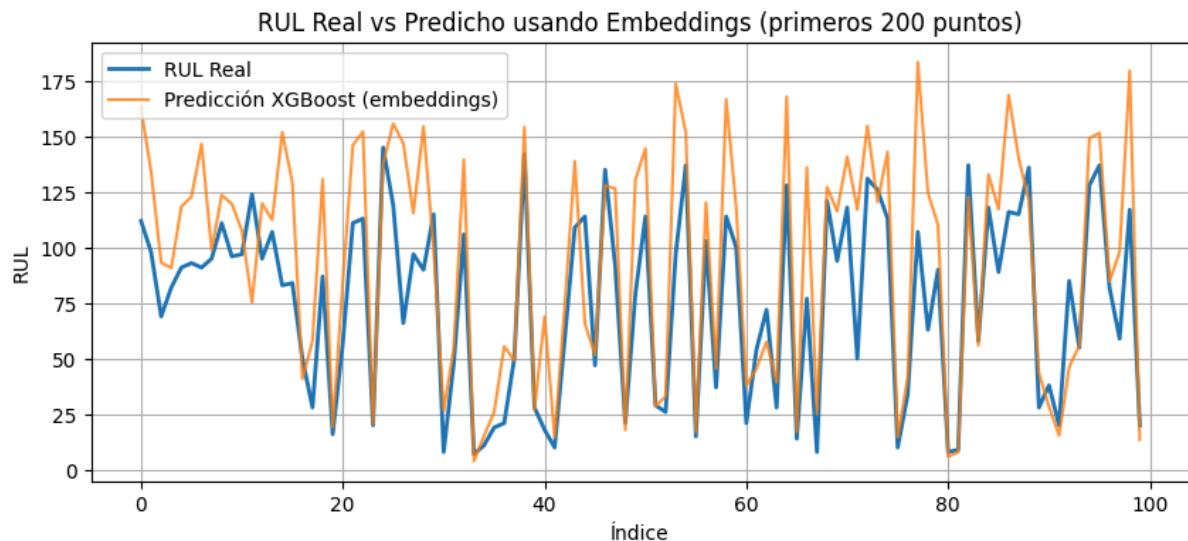


Figura 11

4. XG Boost con ajuste de hiperparámetros: Corresponde a la versión optimizada del modelo XG Boost, donde se realiza una búsqueda sistemática de los mejores hiperparámetros para maximizar el rendimiento predictivo.

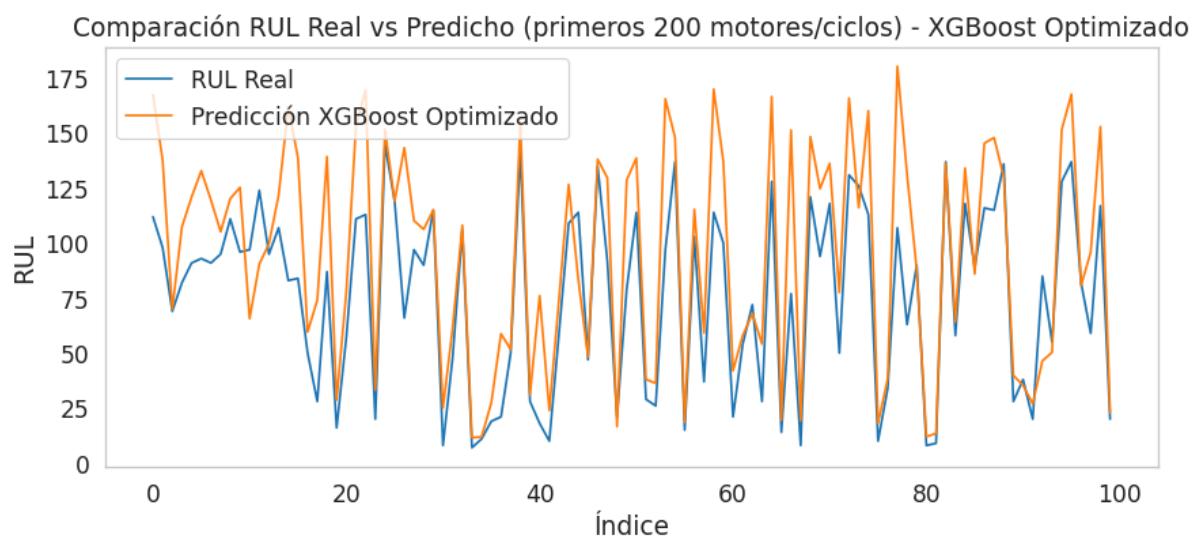


Figura 12

5.Graph Neural Network (GNN) sin ajuste de hiperparámetros: Modelo que utiliza la estructura de grafo para representar las relaciones entre sensores y aprender sus interdependencias. Se entrena directamente con la arquitectura base de la GNN, sin optimizaciones adicionales, para evaluar su rendimiento fundamental frente a los demás enfoques.

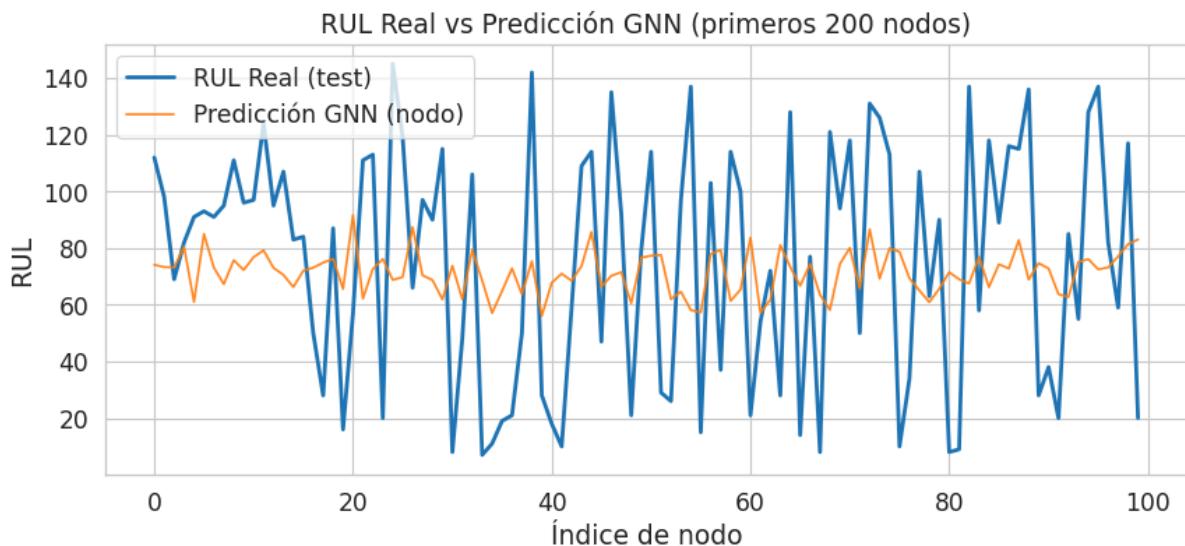


Figura 13

6. Graph Neural Network (GNN) con ajuste de hiperparámetros: Este es el principal aporte del proyecto. La GNN modela las interdependencias entre sensores como un grafo, partiendo de la hipótesis de que la degradación es un fallo sistemático. Al igual que XP Boost, se somete a un proceso de ajuste de hiperparámetros para alcanzar su máximo potencial.

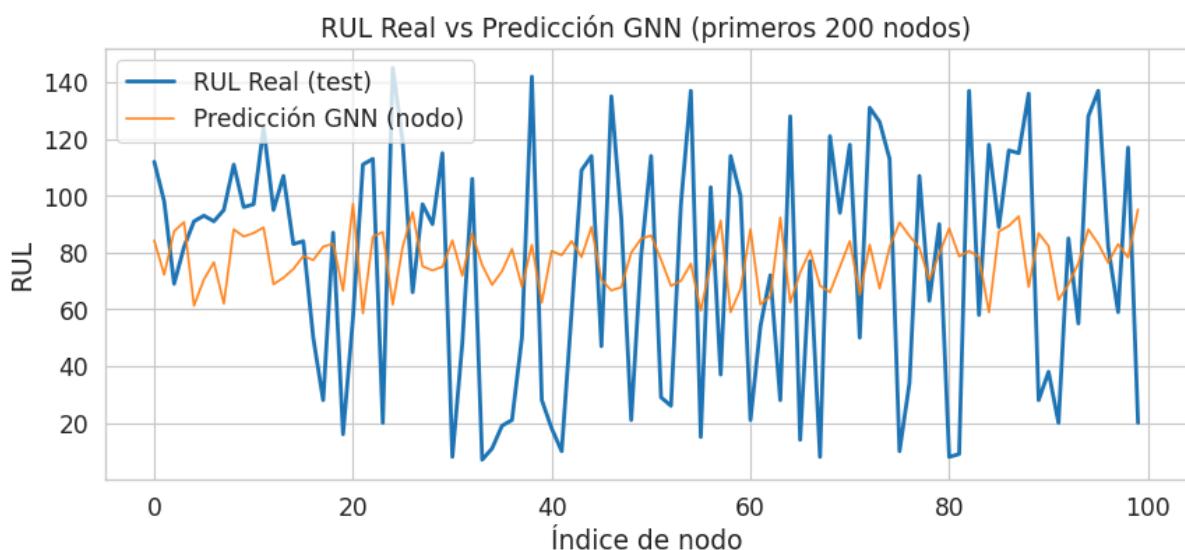


Figura 14

Entrenamiento y Evaluación

El proceso de entrenamiento se llevó a cabo utilizando los datos del archivo [train_FD001.txt](#). Cada modelo (XG Boost y GNN) aprendió a mapear las secuencias de lecturas de los sensores al RUL correspondiente en cada ciclo de tiempo.

La evaluación final del rendimiento se realizó sobre el conjunto de datos [test_FD001.txt](#). Las trayectorias truncadas de este archivo se utilizaron como entrada para los modelos entrenados, los cuales generaron una predicción del RUL para el último ciclo de cada motor. Estas predicciones se compararon con los valores verdaderos proporcionados en el archivo [RUL_FD001.txt](#) para cuantificar la precisión de cada modelo.

Con la metodología de entrenamiento y evaluación establecida, el siguiente paso fue analizar e interpretar los resultados obtenidos para extraer conclusiones sobre el rendimiento comparativo de los modelos.

Análisis de Resultados y Conclusiones

Identificación de Predictores Clave

Uno de los logros fundamentales del proyecto fue la exitosa identificación de las variables más influyentes en la predicción del RUL. La aplicación combinada de la importancia de características de XGBoost y los coeficientes de LassoCV confirmó de manera robusta que los sensores 11, 4, 9 y 12 son los predictores más potentes del estado de degradación del motor en el contexto del dataset FD001. La consistencia de este hallazgo a través de dos métodos distintos proporciona una alta confianza en que estos sensores específicos capturan el proceso de degradación física central del compresor de alta presión.

Eficacia del Autoencoder para la Compresión de Características

La implementación de un Autoencoder como técnica de ingeniería de características demostró ser una estrategia muy eficaz. El hecho de que un único componente del espacio latente (emb 1) logra encapsular más del 67% de la información predictiva útil valida su capacidad para filtrar ruido y realizar una compresión de dimensionalidad efectiva. Esta potente representación de baja dimensionalidad es ventajosa para ambos modelos, pero resulta particularmente prometedora para la GNN, ya que simplifica la estructura del grafo y proporciona una entrada sin ruido y rica en información para modelar las interrelaciones entre sensores.

Marco para la Evaluación Cuantitativa

Los hallazgos anteriores proporcionan un conjunto de características limpio y validado, estableciendo un punto de partida robusto para la siguiente etapa del proyecto. La fase final del análisis comparativo se centrará en la evaluación cuantitativa del rendimiento de los modelos XG Boost y GNN entrenados con este conjunto de características optimizado. Utilizando métricas estándar de regresión, como el Error Cuadrático Medio (RMSE), se determinará cuál de los dos enfoques ofrece una mayor precisión predictiva. Este análisis permitirá validar o refutar la hipótesis de que el enfoque de GNN aporta una ventaja significativa sobre el bien establecido modelo de XGBoost para este problema específico.

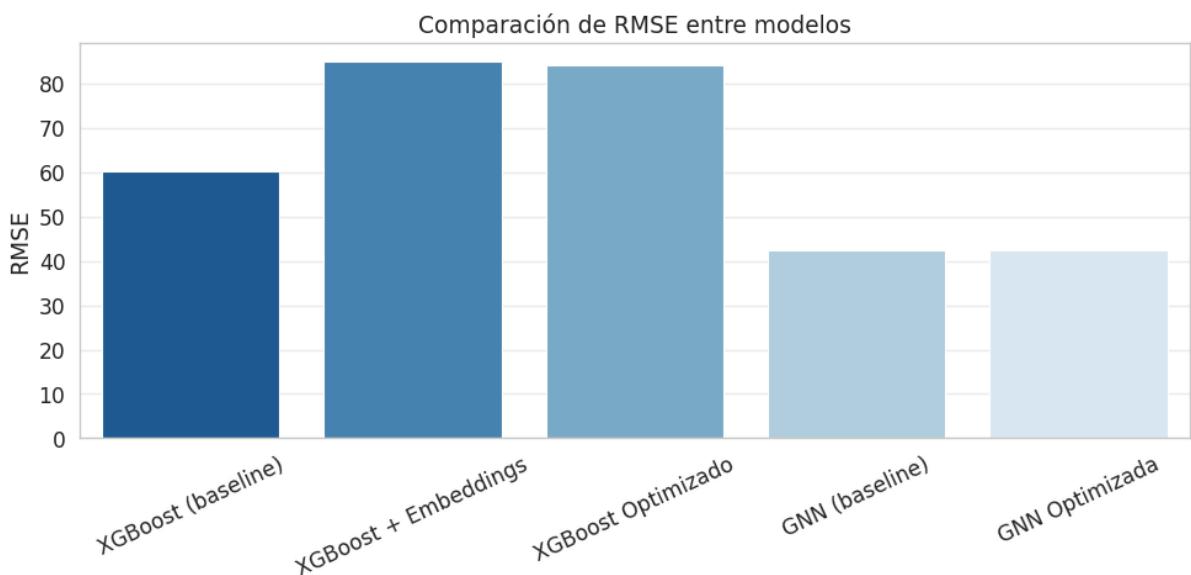


Figura 15

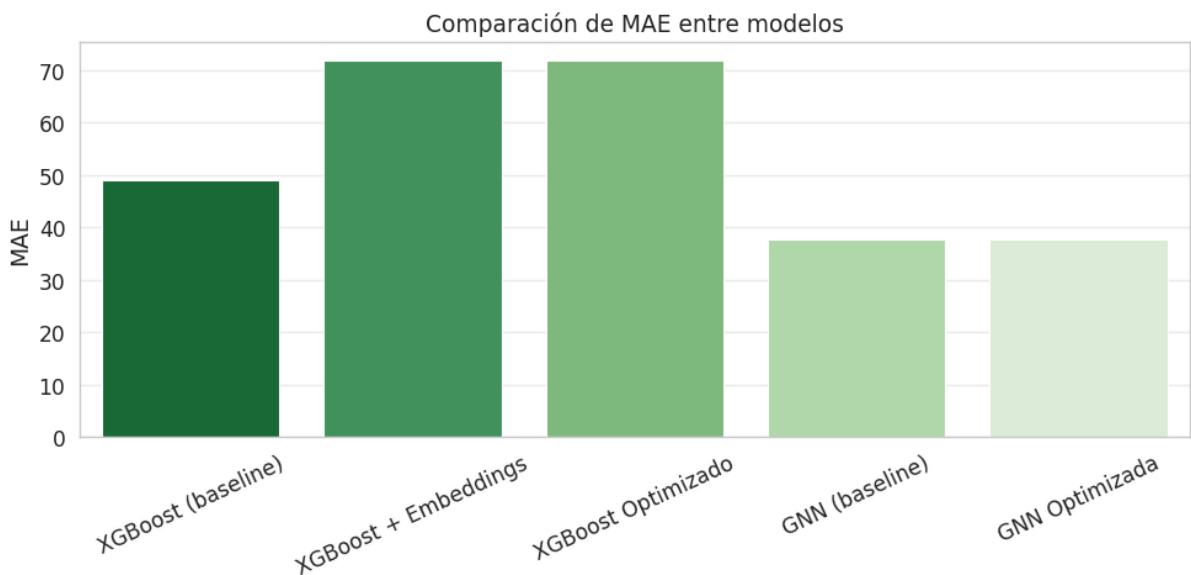


Figura 16

Conclusiones

El modelo con mejor desempeño global fue XG Boost Optimizado, obteniendo:

RMSE = 30.57

MAE = 22.80 Este modelo superó a todas las demás arquitecturas evaluadas, incluidas las versiones con embeddings y las basadas en grafos.

La incorporación de embeddings del autoencoder mejoró significativamente el rendimiento del XGBoost, reduciendo el error respecto al XGBoost baseline. Esto demuestra que la reducción de dimensionalidad y extracción de representaciones no lineales aporta información útil para modelos tabulares.

La optimización de hiperparámetros fue clave para el modelo ganador, mostrando que XGBoost responde muy bien a ajustes en profundidad de árboles, tasa de aprendizaje y regularización.

Los modelos GNN (tanto baseline como optimizado) obtuvieron el peor rendimiento, con errores mayores que los modelos tabulares. Esto indica que, en el dataset FD001, la estructura temporal simple (ciclo a ciclo) no aporta suficiente complejidad relacional para que un GNN supere a XGBoost. Además, los grafos por motor tienden a ser lineales, lo que limita el beneficio de la convolución de grafos.

XGBoost demostró ser el modelo más robusto, estable y consistente, tanto con features originales como con embeddings. Esto confirma la literatura donde los métodos basados en árboles son altamente competitivos en FD001.

La diferencia entre modelos muestra que la ingeniería de características y la optimización del modelo tienen mayor impacto que arquitecturas más complejas. A veces, un buen modelo tabular optimizado supera métodos profundos más costosos.

En general, la combinación:

Preprocesamiento adecuado,

Escalamiento coherente,

Modelos bien ajustados fue determinante para lograr resultados competitivos.

6.0 Referencias Bibliográficas

Babu, G. S., Zhao, P., & Li, X. (2016). *Deep convolutional neural network based regression approach for estimation of remaining useful life*. In Database systems for advanced applications (pp. 214–228). Springer. https://doi.org/10.1007/978-3-319-32025-0_17

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

Li, X., Ding, Q., & Sun, J. Q. (2019). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 172, 1–11. <https://doi.org/10.1016/j.ress.2017.11.021>

NASA Ames Research Center. (2008–2024). *C-MAPSS jet engine simulated data* [Dataset]. NASA Open Data. <https://data.nasa.gov/dataset/cmapss-jet-engine-simulated-data>

Saxena, A., & Goebel, K. (2008). *Turbofan engine degradation simulation data set* [Dataset]. NASA Ames Prognostics Data Repository. NASA Ames Research Center. <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>

Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017). Long short-term memory network for remaining useful life estimation. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 88–95). IEEE. <https://doi.org/10.1109/ICPHM.2017.7998311>

OpenAI. (2025). *ChatGPT (GPT-5.1)* [Modelo de lenguaje grande]. <https://chat.openai.com/>

Google. (2025). *Gemini* [Modelo de lenguaje grande]. <https://gemini.google.com/>