

Estimación de la vida útil restante de una bomba de ciclones Warman 650: comparación de cuatro modelos de aprendizaje automático

1.er Victor Isael Idrogo Clavo, 2.º Manuel Torre Muñoz

Facultad de Ingeniería Industrial y de Sistemas, Grupo 07

Universidad Nacional de Ingeniería

Lima, Perú

victor.idrogo.c@uni.pe, manuel.torre.m@uni.pe

Resumen. La bomba de ciclones Warman 650 es un componente crítico del circuito de molienda de una unidad minera a gran escala: más de la mitad de sus intervenciones son correctivas y cada hora de parada equivale a 9.7 t de producción perdida. Este trabajo evalúa si la vida útil restante (RUL, Remaining Useful Life) de la bomba PU-001 puede estimarse a partir de sus variables de proceso. Se emplearon 87 841 registros con resolución de un minuto, correspondientes a 62 días de operación continua, y ocho variables de condición. La variable objetivo se construyó como el tiempo restante hasta el último registro disponible y se entrenaron cuatro modelos bajo una partición temporal 80/20: Random Forest, XGBoost, regresión Ridge y una red neuronal multicapa. Los dos ensambles de árboles obtuvieron errores absolutos medios de 802.56 h y 725.42 h con coeficientes de determinación de -100.38 y -84.21. La red neuronal redujo ese error en 71.3 %, hasta 208.54 h, y fue el único modelo cuyas predicciones ingresaron parcialmente en el rango real del conjunto de prueba. La regresión Ridge, pese a poder extrapolar, resultó la peor de las cuatro con 922.68 h. El patrón obtenido confirma experimentalmente el diagnóstico propuesto: la limitación de los ensambles de árboles para extrapolar fuera del rango de entrenamiento es una causa efectiva del error, y la capacidad de extrapolar por sí sola no basta si la relación entre variables y objetivo no es lineal. No obstante, ningún modelo superó a un predictor trivial que repite el último RUL observado, con un error de 146.42 h. Se concluye que el rediseño del etiquetado y la incorporación de múltiples historiales de falla son condiciones necesarias para obtener un modelo desplegable.

Palabras clave. vida útil restante, mantenimiento predictivo, XGBoost, red neuronal, extrapolación, bombas de pulpa, minería.

I. INTRODUCCIÓN

El área de molienda concentra el riesgo operacional de una unidad minera a gran escala. En el circuito estudiado, dos molinos SAG reducen el tamaño del mineral y seis bombas Warman 650 transportan la pulpa resultante hacia siete hidrociclones; de este circuito dependen del orden de 140 000 t de mineral procesado por día. La detención no planificada de una de estas bombas interrumpe el flujo de mineral y compromete directamente la producción diaria.

El diagnóstico del problema arroja tres cifras que definen la magnitud económica: el 52 % de las intervenciones sobre la bomba no alcanzó la vida útil estimada de 1 000 horas, cada intervención implica hasta 10 horas de parada, y cada hora de parada representa 9.7 t de producción perdida. La baja anticipación de fallas genera, por tanto, un costo directo por mantenimiento correctivo y un costo de oportunidad por producción no realizada.

La estimación de la vida útil restante (RUL) es la herramienta habitual para convertir el mantenimiento

correctivo en mantenimiento basado en condición. El objetivo general de esta investigación es determinar la influencia de la técnica XGBoost del aprendizaje automático en la predicción del tiempo de vida útil restante de una bomba Warman 650. Los objetivos específicos son: (OE1) evaluar la precisión de XGBoost frente a otros métodos de predicción; (OE2) analizar el impacto de la calidad y cantidad de datos en la precisión; y (OE3) sentar las bases de un sistema de predicción en tiempo real que alerte a los operadores antes de la falla.

Este artículo reporta el modelado completo sobre la bomba PU-001 mediante cuatro algoritmos pertenecientes a tres familias distintas. Su contribución no es un modelo de alto desempeño sino un resultado negativo diagnosticado de forma experimental: se identifica la incapacidad de extrapolación de los ensambles de árboles como causa parcial del error, se verifica esa hipótesis entrenando modelos que sí extrapolan, se cuantifica la brecha frente a líneas base triviales y frente a la literatura, y se especifica el rediseño requerido. En prognosis industrial, donde los conjuntos con un solo historial de falla son la norma, este tipo de diagnóstico constituye un insumo metodológico reutilizable.

II. ESTADO DEL ARTE

A. Enfoques para la estimación del RUL

La literatura clasifica los métodos de prognosis en tres familias. Los enfoques basados en modelos físicos describen la degradación mediante ecuaciones de daño y requieren conocimiento detallado del mecanismo de falla. Los enfoques estadísticos y guiados por datos infieren la degradación a partir del historial de mediciones sin postular un modelo físico explícito; Si et al. [1] sistematizan esta familia y muestran que su aplicabilidad depende críticamente de la disponibilidad de trayectorias run-to-failure. Los enfoques híbridos combinan ambos. Lei et al. [2] ofrecen la revisión más citada del flujo completo, desde la adquisición de datos hasta la predicción del RUL, y destacan cuatro etapas: adquisición, construcción del indicador de salud, división de las etapas de degradación y predicción propiamente dicha.

Una observación transversal de estas revisiones es que la construcción de la variable objetivo es una decisión de modelado tan determinante como la elección del algoritmo. Heimes [3] introdujo el etiquetado lineal por tramos (piecewise-linear RUL), que satura el RUL en un valor máximo durante la fase de operación sana y solo lo hace decrecer linealmente cuando comienza la degradación observable. Este esquema es hoy práctica estándar porque

evita exigir al modelo que distinga estados sanos indistinguibles entre sí.

B. Métodos clásicos de aprendizaje automático

Sobre datos tabulares de condición, los métodos de referencia son las máquinas de vectores de soporte para regresión, los bosques aleatorios de Breiman [4] y los ensambles de potenciación del gradiente. XGBoost, propuesto por Chen y Guestrin [5], es la implementación más difundida de esta última familia: incorpora regularización explícita, submuestreo de filas y columnas, y un algoritmo de división consciente de la dispersión, lo que le otorga ventaja sistemática sobre datos tabulares. Su atractivo en prognosis industrial es doble: rendimiento competitivo con costo de cómputo moderado, e interpretabilidad directa vía importancia de características.

Estos métodos comparten, sin embargo, una limitación estructural relevante para este trabajo: al predecir mediante promedios de hojas, su salida está acotada al rango de la variable objetivo observado durante el entrenamiento. Un ensamble de árboles no extrapola; ante entradas cuyo objetivo real cae fuera de ese rango, devuelve necesariamente valores del rango aprendido. Los modelos lineales regularizados y las redes neuronales con salida lineal no presentan esa restricción, propiedad que este trabajo aprovecha para verificar experimentalmente el diagnóstico.

C. Aprendizaje profundo y arquitecturas temporales

La segunda corriente aplica redes profundas directamente sobre las series de sensores. Babu et al. [6] introdujeron redes convolucionales para regresión de RUL; Zheng et al. [7] mostraron que las redes de memoria de corto y largo plazo (LSTM) capturan dependencias temporales de mayor alcance; Ellefsen et al. [8] incorporaron preentrenamiento semisupervisado mediante autocodificadores para aprovechar datos sin etiquetar; y trabajos más recientes emplean codificadores de tipo Transformer con mecanismos de atención [9]. Zhao et al. [10] revisan esta línea y observan que la ventaja del aprendizaje profundo se manifiesta cuando existen múltiples unidades con historial completo de falla, condición que permite separar entrenamiento y prueba por unidad y no por tiempo. El perceptrón multicapa, aunque más simple que las arquitecturas anteriores, comparte con ellas la capacidad de aproximar funciones no lineales y de emitir predicciones fuera del rango de entrenamiento.

D. Conjuntos de referencia y protocolos de validación

El conjunto C-MAPSS de simulación de turbofán, publicado por Saxena et al. [11], es el banco de pruebas dominante: su subconjunto FD001 contiene 100 unidades de entrenamiento y 100 de prueba, cada una con su propia trayectoria run-to-failure. En rodamientos, la plataforma PRONOSTIA descrita por Nectoux et al. [12] cumple una función análoga. La característica compartida por ambos, y ausente en el presente caso de estudio, es la multiplicidad de trayectorias: la partición se realiza entre unidades, de modo que el conjunto de prueba cubre el mismo rango de RUL que el de entrenamiento y la tarea evaluada es interpolación, no extrapolación.

E. Prognosis en maquinaria rotativa y bombas de pulpa

En maquinaria rotativa, la vibración es el indicador de condición más utilizado; la norma ISO 20816-1 [13] establece las zonas de severidad que sustentan los umbrales de alerta empleados en la industria. En bombas centrífugas de pulpa, el

mecanismo dominante de degradación es el desgaste abrasivo del impulsor y del revestimiento, gobernado por la concentración de sólidos, la granulometría y la velocidad de operación, según describen Wilson et al. [14]. La consecuencia práctica es que la degradación de una bomba de ciclones depende tanto de la condición mecánica como del régimen de proceso, lo que hace que las variables de flujo, densidad y velocidad tengan valor predictivo pero también introduzcan confusión con el punto de operación.

F. Brecha identificada

La revisión permite delimitar la brecha que aborda este trabajo. Primero, la literatura de prognosis está fuertemente sesgada hacia conjuntos con múltiples trayectorias de falla, mientras que el dato industrial disponible en una operación minera típica corresponde a un único historial continuo por equipo. Segundo, son escasos los estudios que reportan y diagnostican explícitamente los resultados negativos derivados de esa asimetría. Tercero, casi no existen trabajos públicos de estimación de RUL sobre bombas Warman en circuitos de molienda. Este artículo aporta evidencia en los tres frentes.

III. MATERIALES Y MÉTODOS

A. Conjunto de datos

Se utilizó el registro histórico del sistema de control de la bomba de ciclones PU-001. El conjunto contiene 87 841 observaciones con resolución de un minuto, correspondientes a 62 días de operación continua entre el 1 de febrero y el 3 de abril de 2026. Tras descartar tres columnas vacías del archivo original, quedaron una marca de tiempo y ocho variables numéricas de proceso y condición: velocidad, flujo, presión, número de ciclones en servicio, densidad de pulpa, temperatura de motor, potencia y vibración en el porta rodamientos del lado bomba. No se detectaron valores nulos ni registros duplicados en las variables retenidas.

El conjunto conserva valores extremos atribuibles a errores de instrumentación en densidad y presión, que no fueron depurados en esta iteración; este punto se retoma en la discusión.

B. Etiquetado del estado de la bomba

Como capa descriptiva se clasificó cada registro en tres estados según la vibración del porta rodamientos: Normal por debajo de 0.065 mm/s, Alerta entre 0.065 y 0.087 mm/s, y Crítico a partir de 0.087 mm/s. La distribución resultante es de 83 566 registros normales (95.13 %), 3 395 en alerta (3.86 %) y 880 críticos (1.00 %).

C. Construcción de la variable objetivo

El conjunto no incluye una columna de RUL ni un registro de eventos de falla. La marca de tiempo original no era interpretable de forma uniforme, por lo que el eje temporal se reconstruyó a partir del primer instante válido asumiendo muestreo constante de un minuto. Se adoptó el supuesto de que el último registro disponible corresponde al fin de vida del componente, y el RUL de cada observación se definió como la diferencia en horas entre esa fecha y el instante de la observación. La variable resultante decrece linealmente desde 1 464 h hasta 0 h a lo largo del período (Fig. 1).

Este supuesto es la decisión de modelado más consecuente del estudio y se analiza críticamente en la Sección IV: al depender el RUL únicamente del índice temporal, la variable

objetivo no incorpora información sobre el estado real de degradación del equipo.

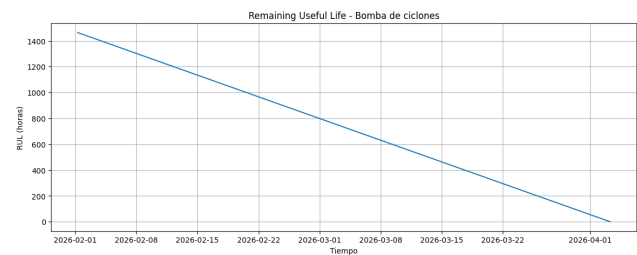


Fig. 1. Variable objetivo RUL construida como cuenta regresiva lineal hasta el último registro disponible.

D. Partición de los datos

Se aplicó una partición cronológica 80/20, sin mezclado aleatorio, para respetar la naturaleza de serie de tiempo del problema: 70 272 observaciones para entrenamiento y 17 569 para prueba. El rango de RUL en entrenamiento va de 292.82 h a 1 464 h, mientras que en prueba va de 0 h a 292.80 h; los dos rangos son disjuntos.

E. Modelos entrenados

Se entrenaron cuatro modelos de regresión pertenecientes a tres familias, con el propósito de aislar el efecto de la capacidad de extrapolación sobre el error. Los dos primeros son ensambles de árboles y, por construcción, no extrapolan. Los dos últimos sí pueden emitir predicciones fuera del rango de entrenamiento, y se incorporan precisamente para contrastar esa propiedad. La Tabla I resume la configuración adoptada.

TABLA I. MODELOS ENTRENADOS Y CONFIGURACIÓN

Modelo	Familia	Configuración principal
Random Forest	Ensamble de árboles	200 árboles, profundidad 15, min. 10 muestras por división, 5 por hoja
XGBoost	Ensamble de árboles	500 estimadores, tasa 0.05, profundidad 6, peso mínimo por hijo 5, submuestreo 0.8
Ridge	Modelo lineal regularizado	Regularización L2 con alfa 1.0 sobre variables estandarizadas
Red neuronal (MLP)	Red neuronal	Capas ocultas 64 y 32, activación ReLU, optimizador Adam, tasa 0.001, detención temprana

Random Forest y XGBoost operan sobre las variables originales. Ridge y la red neuronal requieren estandarización, calculada únicamente con los parámetros del conjunto de entrenamiento. Semilla fija 42 en los cuatro casos.

El desempeño se evaluó con el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2). Para dimensionar los resultados se definieron además dos líneas base triviales, evaluadas sobre el mismo conjunto de prueba: la línea base A predice constantemente la media del RUL de entrenamiento (878.41 h); la línea base B predice constantemente el último valor de RUL observado en entrenamiento (292.82 h), es decir, supone que el equipo no envejece más allá del corte.

IV. RESULTADOS Y COMPARACIÓN CON OTROS AUTORES

A. Desempeño de los cuatro modelos

La Tabla II resume las métricas obtenidas sobre el conjunto de prueba, junto con el rango que abarcaron las predicciones de cada modelo. Este último dato resulta tan informativo como las métricas, ya que revela si el modelo pudo o no alcanzar la región de valores que debía predecir.

TABLA II. DESEMPEÑO SOBRE EL CONJUNTO DE PRUEBA (17 569 OBSERVACIONES)

Modelo	MAE (h)	RMSE (h)	R^2	Rango predicho (h)
Base A: media de entrenamiento	732.01	736.87	-74.99	878.4
Base B: último RUL observado	146.42	169.06	-3.00	292.8
Random Forest	802.56	851.09	-100.38	294 a 1 406
XGBoost	725.42	780.29	-84.21	95 a 1 486
Ridge	922.68	933.60	-120.99	331 a 2 225
Red neuronal (MLP)	208.54	233.20	-6.61	127 a 601

Rango real del RUL en el conjunto de prueba: 0 a 292.8 h. Las líneas base no utilizan ninguna variable de entrada.

Entre los dos ensambles de árboles, XGBoost superó a Random Forest en las tres métricas: redujo el MAE en 9.6 % y el RMSE en 8.3 %. Este resultado es consistente con la hipótesis específica HE1 y con la ventaja de la potenciación del gradiente sobre el promediado de árboles reportada por Chen y Guestrin [5]. Sin embargo, ambos presentan coeficientes de determinación fuertemente negativos, lo que indica que explican la varianza del objetivo peor que predecir simplemente la media del conjunto de prueba.

El resultado más relevante es el de la red neuronal, que redujo el MAE de XGBoost en 71.3 %, de 725.42 h a 208.54 h, y el RMSE en 70.1 %, de 780.29 h a 233.20 h. Su coeficiente de determinación, aunque sigue siendo negativo, mejora de -84.21 a -6.61, y su RMSE representa el 15.9 % del horizonte de vida frente al 53.3 % de XGBoost. Fue además el único modelo cuyas predicciones descendieron de forma sostenida por debajo de las 300 h, ingresando parcialmente en el rango real del conjunto de prueba (Fig. 2).

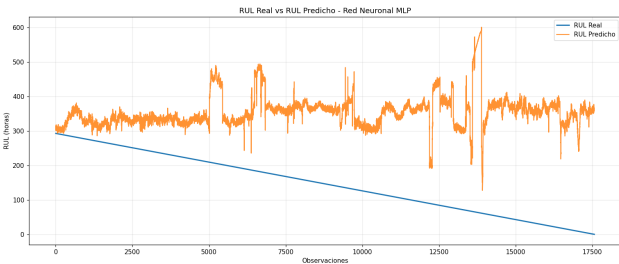


Fig. 2. RUL real frente a RUL predicho por la red neuronal. Las predicciones ingresan parcialmente al rango real, pero no reproducen el descenso hacia cero.

La regresión Ridge, en cambio, resultó la peor de las cuatro configuraciones, con un MAE de 922.68 h y un R^2 de -120.99. Sus predicciones alcanzaron 2 225 h, valor que excede en más de siete veces el máximo real del conjunto de prueba. Este resultado es informativo: Ridge sí puede extrapolar, de modo que su fallo demuestra que la capacidad de extrapolación no basta por sí sola cuando la relación entre las variables de proceso y el objetivo no es lineal.

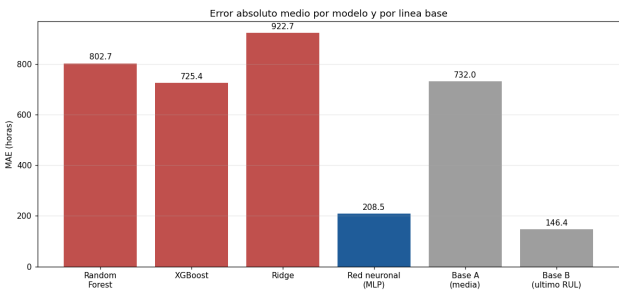


Fig. 3. Error absoluto medio de los cuatro modelos y de las dos líneas base.

B. Comparación con líneas base triviales

La comparación con predictores triviales sigue siendo el criterio decisivo. La línea base A, que ignora las ocho variables de entrada y predice siempre la media del RUL de entrenamiento, obtiene un MAE de 732.01 h, mejor que Random Forest y Ridge y prácticamente equivalente a XGBoost. La línea base B, que predice siempre el último valor de RUL observado en entrenamiento, obtiene un MAE de 146.42 h y un RMSE de 169.06 h.

Ningún modelo entrenado superó a la línea base B. La red neuronal se aproxima considerablemente, con un MAE 42.4 % superior al de esa regla, mientras que los restantes modelos quedan por encima en un factor de cinco a seis. Se sigue que, incluso en su mejor configuración, el conjunto de variables disponibles no aporta información suficiente sobre la degradación bajo el esquema de etiquetado adoptado.

C. Diagnóstico del error

El patrón conjunto de los cuatro resultados permite descomponer el error en dos causas separables. La primera es la incapacidad de extrapolación de los ensambles de árboles. Las Figs. 3 y 4 muestran que Random Forest y XGBoost mantienen sus predicciones aproximadamente entre 300 h y 1 450 h a lo largo de todo el conjunto de prueba, mientras que el RUL real desciende de forma monótona hasta cero. Al sustituir esta familia por una red neuronal, que no está sujeta a dicha restricción, el error se redujo en más del 70 %. La hipótesis queda verificada de forma experimental y no solo argumentada.

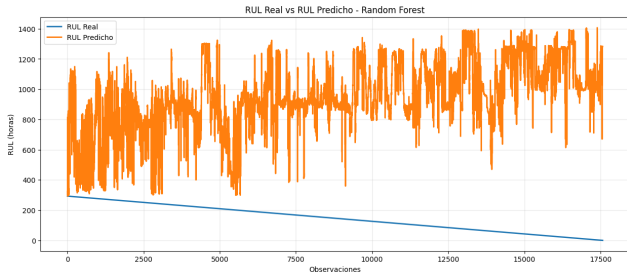


Fig. 4. RUL real frente a RUL predicho por Random Forest. La predicción permanece confinada al rango visto en entrenamiento.

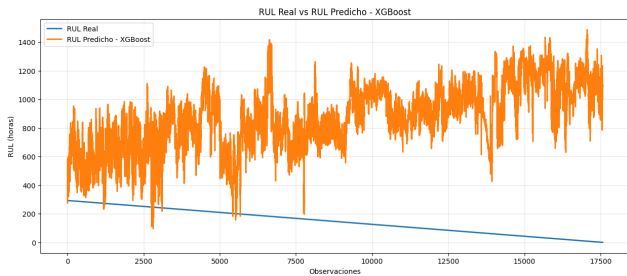


Fig. 5. RUL real frente a RUL predicho por XGBoost. El patrón de fallo es el mismo que en Random Forest, con menor dispersión.

La segunda causa es más profunda y ninguna familia de modelos puede corregirla. Al haberse definido el RUL como una cuenta regresiva del índice temporal sobre un único historial, la variable objetivo es una función determinista del tiempo y no del estado del equipo. Aprenderla equivale a aprender la correspondencia entre una condición de proceso y una fecha del calendario, correspondencia que carece de validez fuera del periodo observado. Esto explica que la red

neuronal, pese a su mejora sustancial, tampoco reproduzca el descenso hacia cero y permanezca por encima de la línea base B. Es exactamente el problema que el etiquetado lineal por tramos de Heimes [3] fue concebido para mitigar y que la partición por unidades de C-MAPSS [11] evita por diseño.

D. Importancia de las variables

La importancia de características ofrece información diagnóstica sobre qué señales dominan la separación del espacio de entrada (Tabla III). Los dos ensambles coinciden en situar el flujo en primer lugar y la presión en último. La vibración del porta rodamientos, que la literatura de maquinaria rotativa señala como indicador de condición por excelencia [2], [13], ocupa una posición intermedia. El modelo Ridge confirma este patrón desde otra perspectiva: sus coeficientes de mayor magnitud absoluta corresponden al flujo, la velocidad y la potencia, mientras que el coeficiente de la vibración es dos órdenes de magnitud menor.

Esta coincidencia entre familias distintas es coherente con el diagnóstico anterior: al no ser el RUL una función del estado mecánico, los modelos se apoyan mayoritariamente en variables de régimen de proceso que actúan como identificadores indirectos del momento del calendario, no como indicadores de degradación.

TABLA III. IMPORTANCIA DE VARIABLES POR MODELO

Variable	Random Forest	XGBoost	Ridge (coef.)
Flujo (m³/h)	0.259	0.209	-594.4
Temperatura de motor (°C)	0.218	0.132	181.5
Velocidad (%)	0.184	0.129	416.1
Vibración porta rodamientos	0.122	0.109	9.0
Densidad (%)	0.096	0.086	0.7
Potencia (kW)	0.065	0.089	199.3
Ciclones (#)	0.031	0.178	-4.4
Presión (kPa)	0.026	0.069	-89.3

Las columnas de Random Forest y XGBoost muestran importancia normalizada, con suma igual a 1. La columna de Ridge muestra el coeficiente sobre variables estandarizadas, cuyo signo indica la dirección del efecto.

E. Comparación con otros autores

La comparación con la literatura debe formularse con cautela, ya que los trabajos de referencia se evalúan sobre conjuntos distintos y no existe un banco de pruebas público para bombas Warman en molienda. Lo que sí admite comparación es el error relativo al horizonte de vida y, sobre todo, el protocolo de validación empleado. La Tabla IV sintetiza esa comparación.

TABLA IV. POSICIÓN DEL PRESENTE TRABAJO FRENTE A LA LITERATURA

Trabajo	Conjunto / dominio	Método	Error rel.
Babu et al. [6]	C-MAPSS FD001	CNN profunda	≈ 12–13 %
Zheng et al. [7]	C-MAPSS FD001	LSTM	≈ 11 %
Ellefsen et al. [8]	C-MAPSS FD001	Autocod. + LSTM	≈ 9 %
Nectoux et al. [12]	PRONOSTIA	Referencia	—
Este trabajo	PU-001 (1 traj.)	XGBoost	53.3 %
Este trabajo	PU-001 (1 traj.)	Red neuronal	15.9 %

Los valores de la literatura corresponden al RMSE reportado sobre el horizonte útil del conjunto respectivo y se incluyen como orden de magnitud; deben verificarse contra la fuente original antes de la versión final. El error de este trabajo es RMSE sobre el horizonte total de 1 464 h.

La incorporación de la red neuronal acorta la distancia con la literatura de forma apreciable, de un error relativo del 53.3 % a uno del 15.9 %. La comparación, sin embargo, no es simétrica y no autoriza a declarar equivalencia. Los trabajos

citados disponen de cien trayectorias run-to-failure independientes, particionan entre unidades y aplican etiquetado lineal por tramos; en ese escenario la tarea evaluada es de interpolación. El presente trabajo dispone de una sola trayectoria y particiona en el tiempo, lo que convierte la tarea en extrapolación. A ello se suma un criterio que los trabajos citados sí cumplen y este no: superar a los predictores triviales sobre el propio conjunto de prueba.

V. CONCLUSIONES

A. ¿El modelo resuelve el problema o tiene impacto en el negocio?

No. En su estado actual ninguno de los cuatro modelos debe desplegarse. El requisito operativo derivado del contexto es anticipar la falla con margen suficiente para programar una intervención de hasta 10 horas dentro de la ventana de planificación de mantenimiento, lo que exige un error del orden de decenas de horas. El mejor modelo obtenido, la red neuronal, presenta un MAE de 208.54 h, equivalente a 8.7 días, y sigue siendo superior al de una regla que no consulta ningún sensor. Un sistema de alertas construido sobre estas predicciones generaría avisos desalineados con la condición real del equipo, con el riesgo de erosionar la confianza de los operadores en el mantenimiento predictivo, un costo organizacional difícil de revertir.

El trabajo sí produce, en cambio, dos entregables con valor inmediato. El primero es la capa de clasificación por umbrales de vibración, que identifica el 1.00 % de registros en estado crítico y el 3.86 % en alerta: sin ser un modelo de RUL, constituye una regla de monitoreo trazable y auditable, alineada con la práctica normativa [13], desplegable de inmediato. El segundo es la identificación experimental de la familia de modelos adecuada para este tipo de problema, que reduce el espacio de búsqueda para la siguiente iteración.

B. ¿Es el modelo "mejor" que el de otros autores?

No. Sostener lo contrario no sería defendible con la evidencia disponible. Frente a los autores revisados, el mejor modelo presenta un error relativo al horizonte de vida entre 1.2 y 1.8 veces mayor, y, a diferencia de ellos, no supera a predictores triviales sobre su propio conjunto de prueba, que es el criterio mínimo de validez de cualquier modelo de regresión. Adicionalmente, los trabajos citados validan sobre decenas o centenas de trayectorias independientes, mientras que aquí se dispone de una sola, por lo que las métricas obtenidas no tienen la representatividad estadística necesaria para sustentar una comparación formal.

Las comparaciones internas que el estudio sí sustenta son dos. Bajo condiciones idénticas de datos, partición y métricas, XGBoost fue superior a Random Forest en las tres métricas, lo que respalda de forma parcial la hipótesis específica HE1. Y la red neuronal fue superior a las tres alternativas restantes por un margen amplio, lo que constituye evidencia directa de que la familia de modelos es un factor determinante en este problema.

C. Conclusiones metodológicas

Primero, la elección de la familia de modelos importa más que el ajuste de hiperparámetros dentro de una misma familia. Pasar de Random Forest a XGBoost redujo el MAE en 9.6 %; pasar de XGBoost a una red neuronal lo redujo en 71.3 %. Segundo, la capacidad de extrapolar es necesaria pero no suficiente: Ridge extrapola y aun así obtuvo el peor resultado

del estudio, porque la relación entre las variables de proceso y el objetivo no es lineal. Tercero, la definición de la variable objetivo debe reflejar el estado del equipo y no el paso del calendario; un RUL construido como cuenta regresiva del índice temporal impone un techo de desempeño que ninguna familia de modelos supera. Cuarto, toda evaluación debe incluir líneas base triviales: sin ellas, la mejora del 71.3 % obtenida por la red neuronal podría haberse reportado como un éxito, cuando en realidad sigue por debajo de una regla constante. Estas conclusiones responden directamente a los objetivos específicos OE1 y OE2.

D. Trabajo futuro

La ruta de corrección se especifica en cinco acciones. (1) Ampliar el conjunto a múltiples historiales de falla, varias bombas del circuito y varios ciclos de intervención por bomba, y particionar entre trayectorias en lugar de en el tiempo. (2) Redefinir la variable objetivo mediante etiquetado lineal por tramos [3], saturando el RUL durante la fase sana. (3) Incorporar variables de ventana móvil sobre la vibración, como valor eficaz, curtosis, pendiente y dispersión en ventanas de 1 a 24 horas, que capturan la tendencia de degradación en lugar del nivel instantáneo. (4) Depurar los valores físicamente imposibles detectados en densidad y presión, y etiquetar explícitamente los períodos de parada. (5) Escalar la familia identificada como más adecuada, avanzando del perceptrón multicapa hacia arquitecturas recurrentes o convolucionales sobre ventanas de señal [6], [7], con validación cruzada temporal por bloques e intervalos de predicción. Una alternativa complementaria, de menor exigencia de datos, es reformular el problema como clasificación binaria de "falla en las próximas N horas", que no requiere extrapolación y se ajusta mejor al uso operativo previsto en el objetivo específico OE3.

AGRADECIMIENTO

Los autores agradecen a la empresa minera que facilitó el registro histórico de operación de la bomba PU-001.

REFERENCIAS

- [1] X.-S. Si, W. Wang, C.-H. Hu y D.-H. Zhou, "Remaining useful life estimation: A review on the statistical data driven approaches", *European Journal of Operational Research*, vol. 213, n.º 1, pp. 1–14, 2011.
- [2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan y J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction", *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
- [3] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation", en *International Conference on Prognostics and Health Management (PHM)*, 2008.
- [4] L. Breiman, "Random forests", *Machine Learning*, vol. 45, n.º 1, pp. 5–32, 2001.
- [5] T. Chen y C. Guestrin, "XGBoost: A scalable tree boosting system", en *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
- [6] G. S. Babu, P. Zhao y X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life", en *Database Systems for Advanced Applications (DASFAA)*, pp. 214–228, 2016.
- [7] S. Zheng, K. Ristovski, A. Farahat y C. Gupta, "Long short-term memory network for remaining useful life estimation", en *IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 88–95, 2017.
- [8] A. L. Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov y H. Zhang, "Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture", *Reliability Engineering & System Safety*, vol. 183, pp. 240–251, 2019.

- [9] Y. Mo, Q. Wu, X. Li y B. Huang, "Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit", *Journal of Intelligent Manufacturing*, vol. 32, pp. 1997–2006, 2021.
- [10] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang y R. X. Gao, "Deep learning and its applications to machine health monitoring", *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.
- [11] A. Saxena, K. Goebel, D. Simon y N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation", en *International Conference on Prognostics and Health Management (PHM)*, 2008.
- [12] P. Nectoux et al., "PRONOSTIA: An experimental platform for bearings accelerated degradation tests", en *IEEE International Conference on Prognostics and Health Management (PHM)*, 2012.
- [13] ISO 20816-1:2016, *Mechanical vibration: Measurement and evaluation of machine vibration, Part 1: General guidelines*, International Organization for Standardization, 2016.
- [14] K. C. Wilson, G. R. Addie, A. Sellgren y R. Clift, *Slurry Transport Using Centrifugal Pumps*, 3.^a ed. Nueva York: Springer, 2006.