



**Universidad Tecnológica de Panamá**  
Maestría en Analítica de Datos



**Curso:**  
Modelos Predictivos

**Avances del Proyecto Final:**

Aplicación de modelos predictivos para evaluar la vida útil restante de motores  
de turbofán.

**Facilitador:**  
Juan Marcos Castillo, PhD

**De Hermoso, Ashley 8-853-1297**

**Año**

2025

# **Aplicación de modelos predictivos para evaluar la vida útil restante de motores de turbofán.**

## **1. Introducción**

En la industria de la aviación, la confiabilidad de las aeronaves representa un factor crítico, ya que cualquier falla o defecto en sus componentes puede resultar en la detención de los aviones, afectando tanto las operaciones de transporte. En este contexto, la capacidad de predecir con anticipación el momento en que un componente podría fallar se convierte en una herramienta valiosa para las aerolíneas, al permitir una mejor planificación del mantenimiento y una reducción en los costos operativos asociados a tiempos fuera de servicio.

A continuación, analizare un conjunto de datos simulados correspondientes a la degradación de cien motores de aeronaves y estimare la vida útil restante de los motores mediante la aplicación de distintos modelos predictivos estudiados durante el curso

## **2. Justificación**

El mantenimiento predictivo en la aviación se ha convertido en una herramienta fundamental para mejorar la disponibilidad de las aeronaves, reducir costos y prevenir fallas imprevistas, asegurando la confiabilidad de una aerolínea.

El presente proyecto tiene como objetivo analizar un conjunto de datos simulados relacionados con la degradación de cien motores de aeronaves. Estos datos comprenden múltiples ciclos de operación medidos a través de sensores que registran temperatura, presión y velocidad, entre otras. Cada motor tendrá registrados hasta que ciclo fallo. Este estudio servirá como base para evaluar el desempeño de los modelos y su aplicabilidad en escenarios reales de mantenimiento predictivo.

Este estudio puede representar un beneficio directo a las aerolíneas, ya que la implementación de modelos predictivos eficaces optimiza los recursos del departamento de mantenimiento, mejora la seguridad operacional y contribuye a una gestión más eficiente de la flota.

### **3. Antecedentes**

Como he mencionado el mantenimiento predictivo ha ido ganando auge en la industria de la aviación por la capacidad de anticipar fallas y optimizar recursos. En el estudio “**Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation**”, desarrollado por investigadores de la NASA, introduce el proceso de cómo utilizaron datos históricos y de sensores para modelar el comportamiento de los motores a lo largo del tiempo. Se utilizó estadísticas y machine-learning para identificar patrones de desgaste, y predecir el tiempo restante de vida útil (Remaining Useful Life, RUL) y generar alertas tempranas. Modelos como la regresión lineal, máquinas de soporte vectorial (SVM), redes neuronales artificiales (ANN) y modelos basados en series temporales han sido aplicados con resultados prometedores. Estos enfoques han demostrado ser útiles para transformar grandes volúmenes de datos en información valiosa para la toma de decisiones crear estrategias data-driven.

La aplicación de modelos predictivos en escenarios simulados, como el de este proyecto, permite evaluar la eficacia de los distintos modelos predictivos en condiciones controladas antes de su implementación en entornos reales. Estos estudios apoyan en la mejora continua en las soluciones para la aviación, ya que ayudan a comprender el comportamiento de los modelos al procesar datos complejos y multivariados, como los registrados por los sensores de los motores o identificando tendencias negativas en desempeño de vuelos.

### **4. Definición de Problema**

El problema del proyecto es estimar la vida útil de los motores turbofán, analizando los ciclos de tiempo como también las mediciones de los sensores.

### **5. Análisis Predictivo**

#### **5.1. Determinación de la base de datos**

La base de datos utilizada para este estudio es el dataset CMAPSS de la NASA, centrandonos en el subconjunto FD001. Este dataset simula el comportamiento de motores de aeronaves a lo largo de múltiples ciclos operativos hasta la falla.

## 5.2. Pre-procesamiento y limpieza

En esta fase, se para prepararon los datos para los diferentes análisis. Las tareas realizadas fueron las siguientes:

- Carga de los archivos de datos
  - train\_FD001.txt (train\_file)– Conjunto de datos que contiene la data completa del ciclo 1 hasta que falla el ciclo de falla de cada motor. de 100 motores. Adicional mantiene las grabaciones de sensores.
  - test\_FD001.txt (test\_file) - Conjunto de datos que contiene una parte del historial de cada motor y no llega hasta que el ciclo que falle. Este conjunto se utiliza como parte del modelo para predecir cuándo va a fallar.
  - RUL\_FD001.txt (y\_rul)– Este archivo indica el RUL o vida útil real de cada motor.
- Asignación de nombres a las columnas
  - Le asignamos los nombres de las columnas se describirán en la sección del anexo.
  - Adicional, se agruparon las columnas, esto con el objetivo de poder realizar ciertos filtrados al momento del análisis y modelado.
  - Eng\_info → unit\_number y time\_cycles
  - Eng\_setting → operational\_setting1-3
  - Sensors\_names → s\_01 - 21
- Exploración inicial de los conjuntos de datos
  - Primero, revise con que tipo de datos tenia los diferentes archivos, en donde todos eran enteros o flotantes.
  - Luego, revise si había que realizar imputaciones por valores en nulos, sin embargo la data no tenía nulos.
  - Después de esto, para el conjunto de datos de train\_file utilice la función describe para ver cómo se distribuían mis registros.

train_file.describe().T								
	count	mean	std	min	25%	50%	75%	max
unit_number	20631.0	51.506568	2.922763e+01	1.0000	26.0000	52.0000	77.0000	100.0000
time_cycles	20631.0	108.807862	6.888099e+01	1.0000	52.0000	104.0000	156.0000	362.0000
operational_setting_1	20631.0	-0.000009	2.187313e-03	-0.0087	-0.0015	0.0000	0.0015	0.0087
operational_setting_2	20631.0	0.000002	2.930621e-04	-0.0006	-0.0002	0.0000	0.0003	0.0006
operational_setting_3	20631.0	100.000000	0.000000e+00	100.0000	100.0000	100.0000	100.0000	100.0000
s_1	20631.0	518.670000	0.000000e+00	518.6700	518.6700	518.6700	518.6700	518.6700
s_2	20631.0	642.680934	5.000533e-01	641.2100	642.3250	642.6400	643.0000	644.5300
s_3	20631.0	1590.523119	6.131150e+00	1571.0400	1586.2600	1590.1000	1594.3800	1616.9100
s_4	20631.0	1408.933782	9.000605e+00	1382.2500	1402.3600	1408.0400	1414.5550	1441.4900
s_5	20631.0	14.620000	1.776400e-15	14.6200	14.6200	14.6200	14.6200	14.6200
s_6	20631.0	21.609803	1.388985e-03	21.6000	21.6100	21.6100	21.6100	21.6100
s_7	20631.0	553.367711	8.850923e-01	549.8500	552.8100	553.4400	554.0100	556.0600
s_8	20631.0	2388.096652	7.098548e-02	2387.9000	2388.0500	2388.0900	2388.1400	2388.5600
s_9	20631.0	9065.242941	2.208288e+01	9021.7300	9053.1000	9060.6600	9069.4200	9244.5900
s_10	20631.0	1.300000	0.000000e+00	1.3000	1.3000	1.3000	1.3000	1.3000
s_11	20631.0	47.541168	2.670874e-01	46.8500	47.3500	47.5100	47.7000	48.5300
s_12	20631.0	521.413470	7.375534e-01	518.6900	520.9600	521.4800	521.9500	523.3800
s_13	20631.0	2388.096152	7.191892e-02	2387.8800	2388.0400	2388.0900	2388.1400	2388.5600
s_14	20631.0	8143.752722	1.907618e+01	8099.9400	8133.2450	8140.5400	8148.3100	8293.7200
s_15	20631.0	8.442146	3.750504e-02	8.3249	8.4149	8.4389	8.4656	8.5848
s_16	20631.0	0.030000	1.387812e-17	0.0300	0.0300	0.0300	0.0300	0.0300
s_17	20631.0	393.210654	1.548763e+00	388.0000	392.0000	393.0000	394.0000	400.0000
s_18	20631.0	2388.000000	0.000000e+00	2388.0000	2388.0000	2388.0000	2388.0000	2388.0000
s_19	20631.0	100.000000	0.000000e+00	100.0000	100.0000	100.0000	100.0000	100.0000
s_20	20631.0	38.816271	1.807464e-01	38.1400	38.7000	38.8300	38.9500	39.4300
s_21	20631.0	23.289705	1.082509e-01	22.8942	23.2218	23.2979	23.3668	23.6184

**Fig.1 Resumen de estadísticas de train\_file**

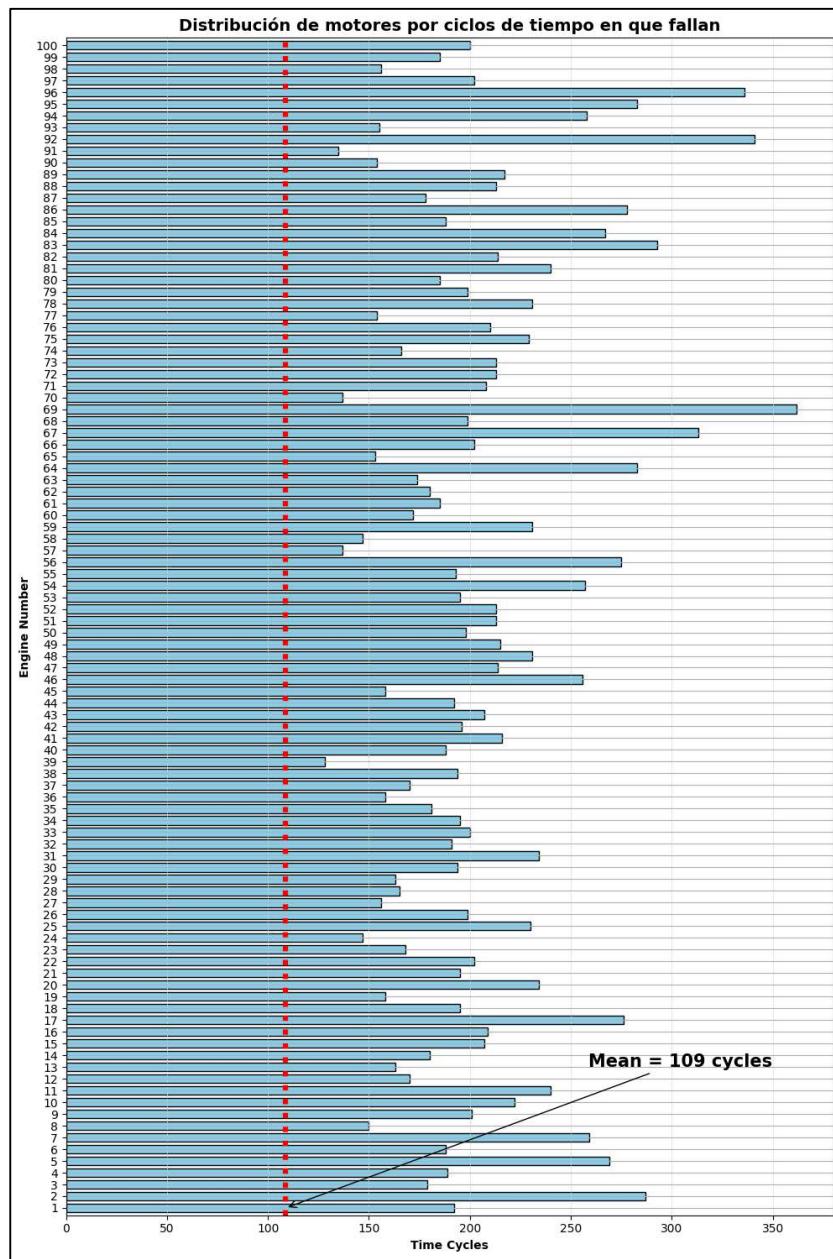
- Identificamos lo siguiente de la Fig.1:
  - Time\_cycles: Presenta una media de 109 ciclos, indicando que los motores operan por aproximadamente 109 ciclos antes de fallar.
  - Sensores: En su mayoría presentan variaciones, sin embargo, sensores (1,5,6,10,16,18,19) se mantienen constantes a través de los ciclos.
- Adicional, en el conjunto de datos de train\_file, agregue una columna de RUL.

### 5.3. Análisis descriptivo

Se realizó un análisis exploratorio de los datos para comprender sus características y patrones iniciales.

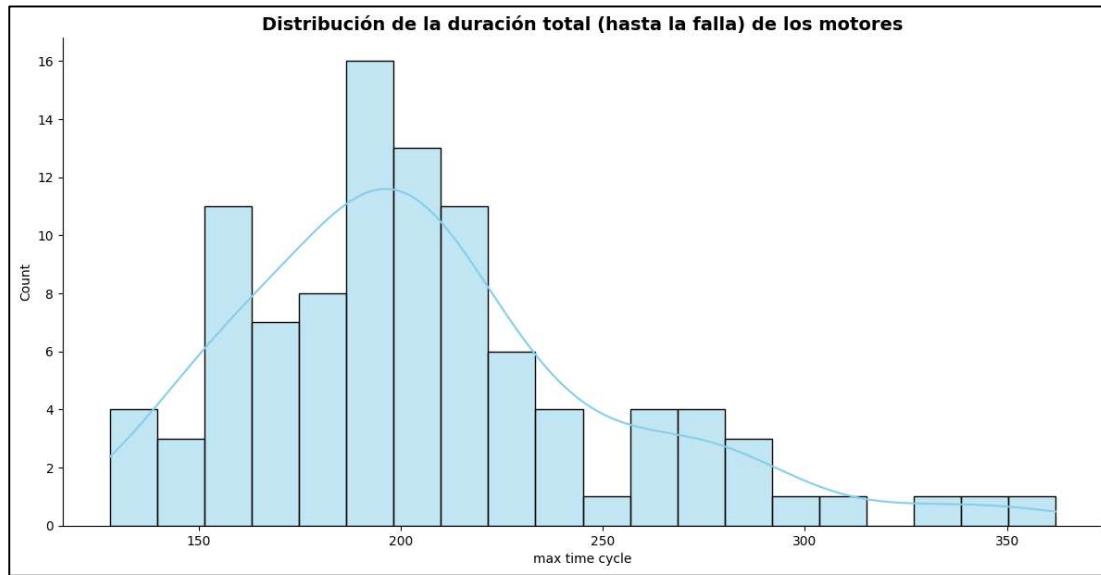
### 5.3.1. Análisis por motores:

Realice una visualización en donde se observa una consolidación por motor los ciclos recorridos hasta que fallaron. Adicional anote donde presenta una media de 109 ciclos, indicando que los motores operan por aproximadamente 109 ciclos antes de fallar. **Fig.2**



**Fig.2: Distribución de motores por ciclos de tiempo en que fallan**

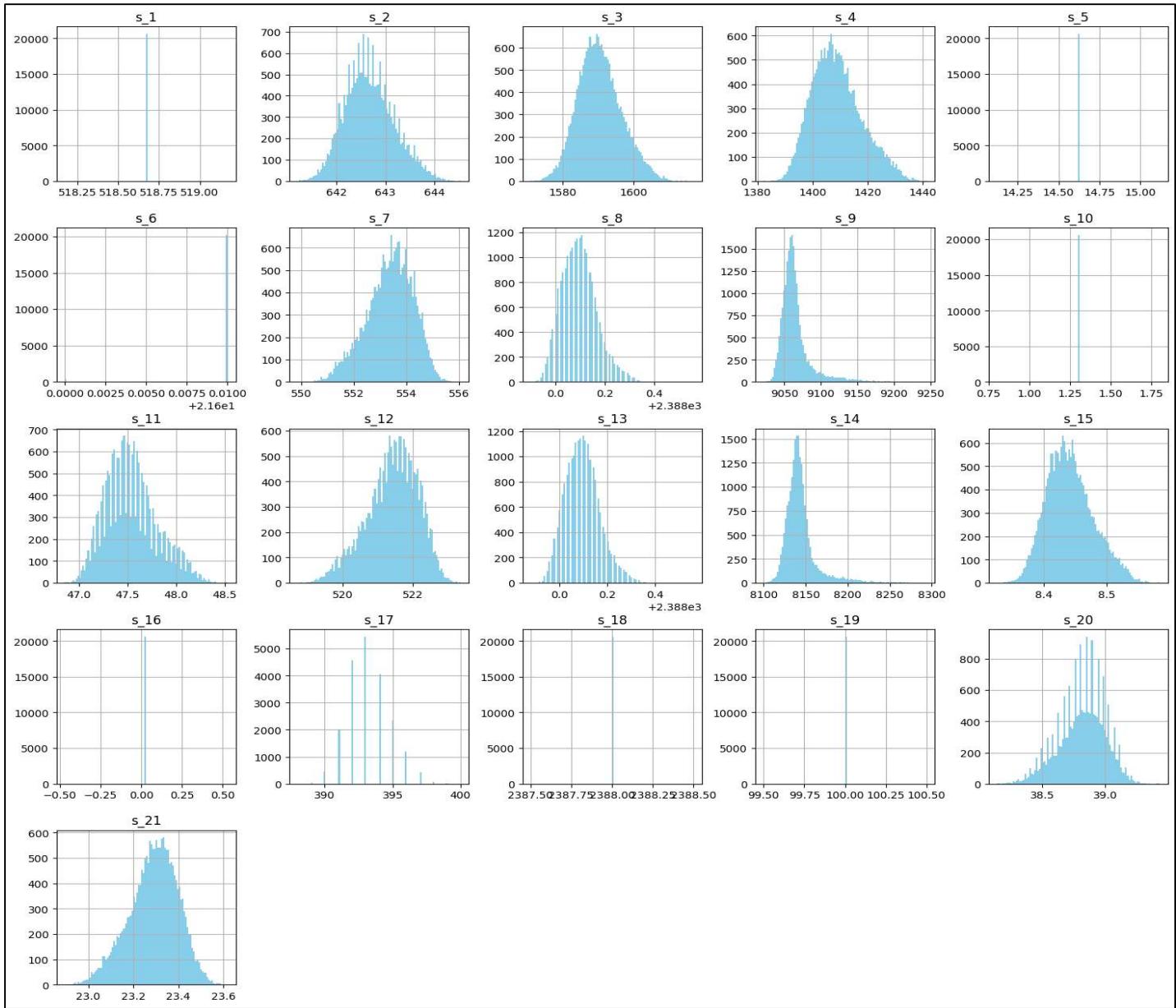
- En la siguiente grafica continuamos observando la distribución del número máximo de ciclos por motor antes de fallar. La distribución no es normal ya que muestra sesgada a la derecha. La distribución también nos indica que la mayoría de los motores fallan entre 190-220. Adicional nos indican que hay ciertos motores que se degradan más lento. **Fig. 3**



**Fig. 3: Distribución de la duración total (hasta la falla) de los motores**

### 5.3.2. Análisis por sensores:

Ahora, evalué los sensores, realizando una distribución de cada sensor para entender lo que había preliminarmente visto en el resumen estadístico. Estos sensores tendrán un rol importante en el modelo predictivo, ya que las variaciones que presentan contribuyen a la estimación de cuando fallaría el motor así determinando la vida útil.



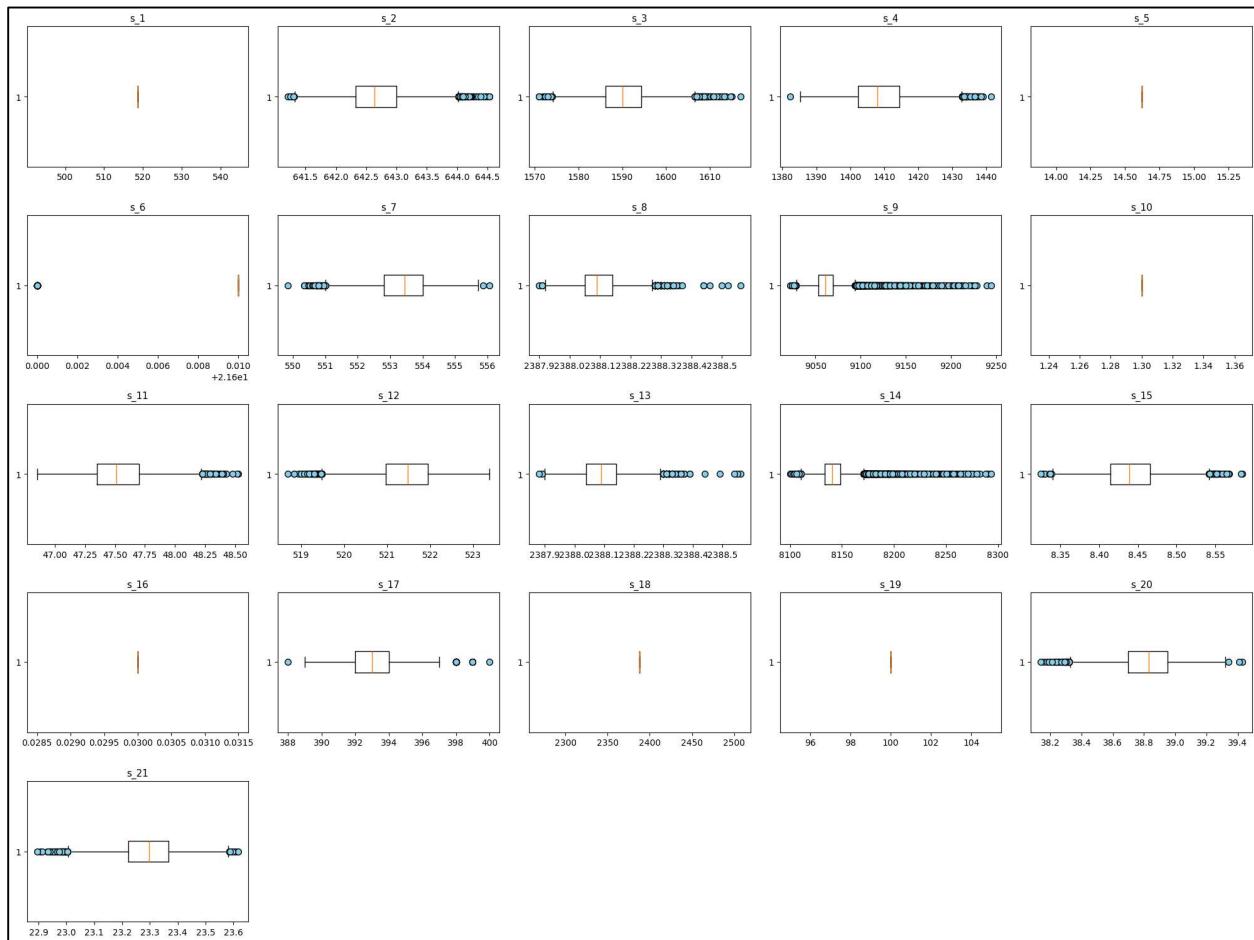
**Fig. 4: Distribución de los sensores de los motores**

Observamos lo siguiente:

- Los sensores (2,3,4,7,8,9,12,13,14,15,20,21), presentan una distribución normal, siendo su forma de campana, simétricas alrededor de la media. Están serán las variables candidatas para los modelos.
- Sin embargo, los sensores (1,5,10, 16,18 19) se mantienen constantes a través de los ciclos, como no presentan variabilidad quizás no sean aplicables a los modelos.

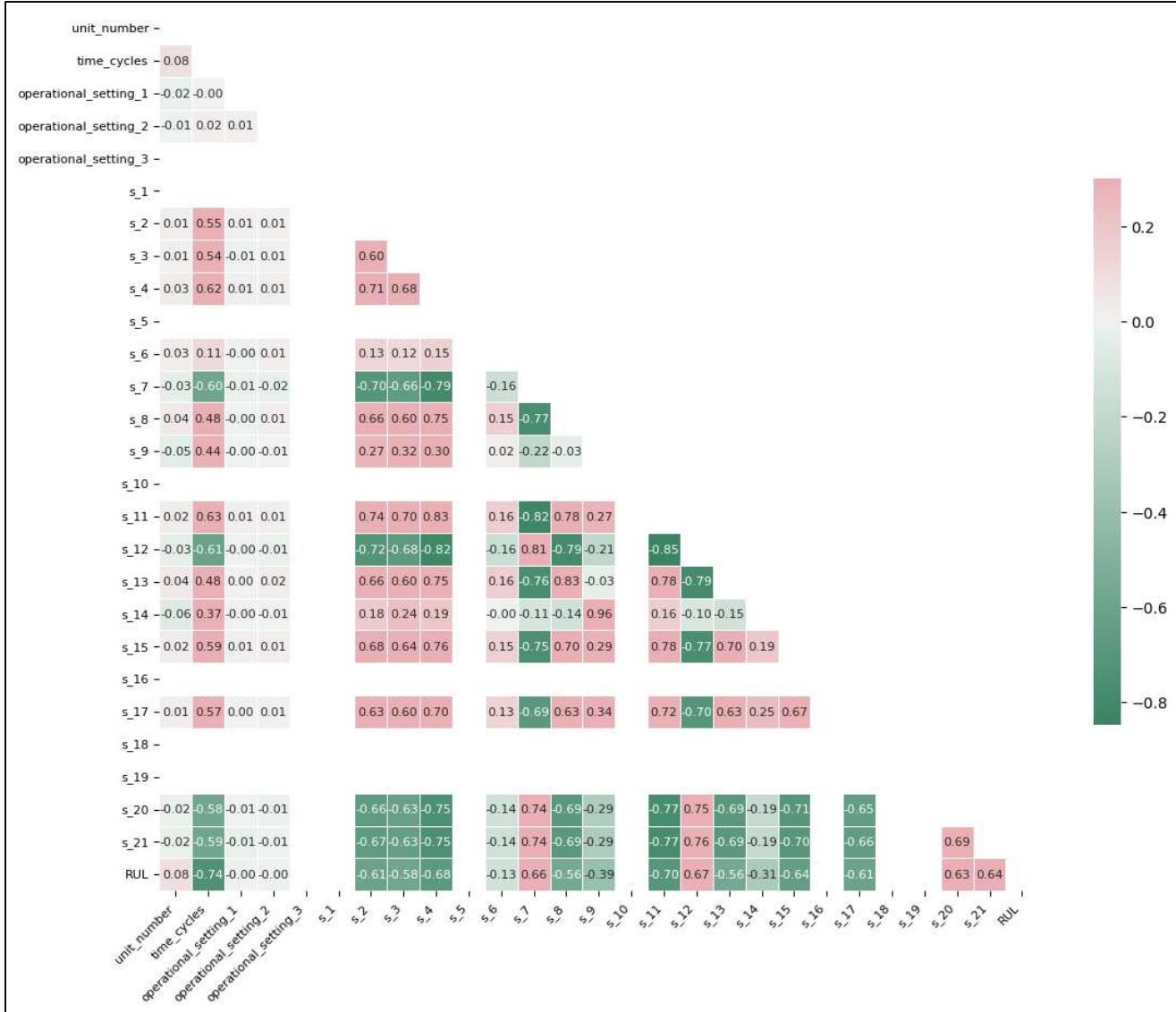
- Los sensores (9,14), son variables sesgadas, con una cola larga hacia valores altos, que pudiese ser outliers o valores atípicos.
- El sensor (17), muestra una distribución rara, con ruido. Esto pudiese sensor que quizás este grabando valores erróneos o una variable categórica codificada numéricamente. La información no brindo mas detalle sobre este parámetro.

Realice una visualización de box plot por sensor y confirma los que observamos en la distribución:



**Fig. 5: Box-plot de los sensores de los motores**

Luego, procedí a hacer un análisis de correlación entre RUL y sensores, con el objetivo de identificar aquellas variables que muestran una relación fuerte con la vida útil restante del motor, ya que estas son candidatas ideales para alimentar el modelo predictivo.



**Fig. 5: Análisis de correlación entre RUL y variables**

Identificamos lo siguiente:

- En el caso de esta data, correlación negativa sería lo esperado ya que las variables tienden a disminuir a medida que se acerca el fallo del motor, por lo tanto, su comportamiento está fuertemente relacionado con la degradación del sistema.
  - **time\_cycles**: Muestra una correlación negativa muy fuerte (-0.9), lo cual es esperable, ya que a mayor número de ciclos transcurridos, menor será la vida útil restante.

- **Sensores (9 y 14):** Presentan correlaciones entre -0.7 y -0.8 con RUL, lo que indica que la velocidad del núcleo del motor es un fuerte indicador de desgaste.
- **Sensores (7 y 10):** También muestran correlación negativa significativa, lo que sugiere que la presión de salida del compresor de alta presión y la relación de presiones internas se degradan con el tiempo.

#### **5.4. Selección de variables**

La selección de variables se realizó considerando el conocimiento del funcionamiento de los motores, así como los resultados del análisis descriptivo y de correlación realizados. Selecciona las variables de (time\_cycles) y lecturas de sensores, porque demuestran una alta correlación con la vida útil remanente de los motores (RUL).

Estas variables aportan información crítica sobre el desgaste progresivo del motor y permiten capturar patrones de degradación relevantes para la predicción.

#### **5.5. Selección de Modelos**

##### **5.5.1. Preparación del modelo.**

Antes de iniciar el modelo procedo a utilizar la función train\_test\_split para separar los datos del test\_file para entrenar y probar el modelo. Utilice un tamaño de prueba del 30% para evaluar el rendimiento y el resto para el entrenamiento.

Luego de esto aplique una función de MinMax scalar esto para normalizar la data de los sensores ya que las lecturas tienen escalas distintas por ejemplo, presión puede ir de 0 a 1000, mientras que temperatura puede ir de 0 a 300. Adicional, esto ayuda al modelado porque evita que variables de valores grandes sesguen el modelo.

### 5.5.2. Modelo Regresión Lineal

Utilice el modelo de regresión lineal por ser una herramienta sencilla, eficiente e interpretable, ideal como punto de partida para este modelo

Asimismo, dado que estos datos muestran la degradación de los motores e indica que es progresiva, una regresión lineal tiene el potencial de capturar adecuadamente esa tendencia.

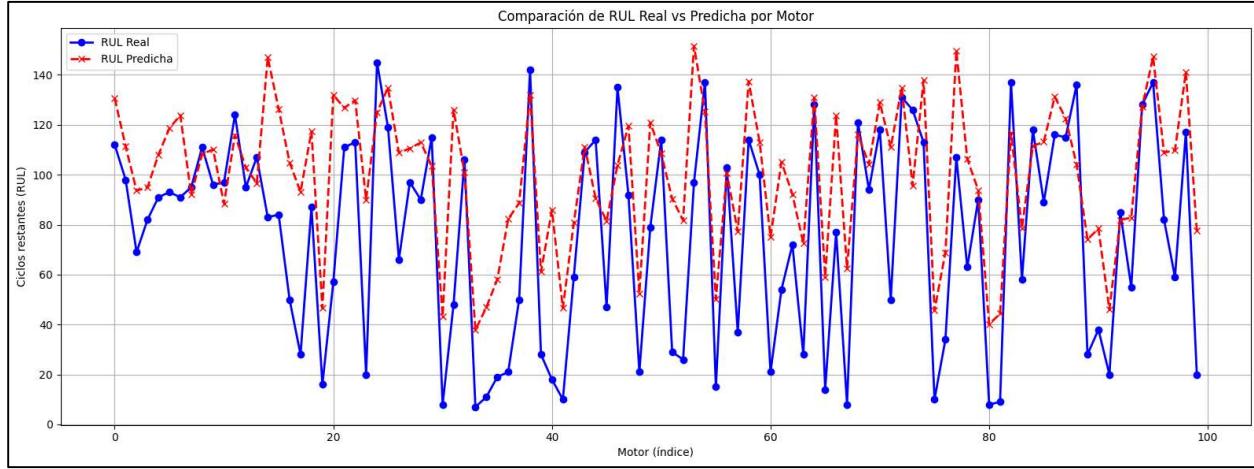
Esto permite establecer una línea base sobre la cual comparar posteriormente con otros modelos.

Escenario	Set	MAE	RMSE	R <sup>2</sup>
Todos los sensores y una muestra de los datos (Fig.6)	Train	45.65	56.20	0.3439
	Test	43.52	54.26	0.3571
	Valid	29.13	34.85	0.2967
Excluyendo sensores constantes y usando muestra de datos (Fig.7)	Train	41.68	47.86	0.3793
	Test	44.36	55.63	0.3245
	Valid	29.13	34.85	0.2967
Excluyendo sensores constantes e incluyendo todos los datos. (Fig. 8)	Train	34.33	40.53	0.5545
	Test	33.56	40.10	0.5545
	Valid	30.47	38.39	0.1465

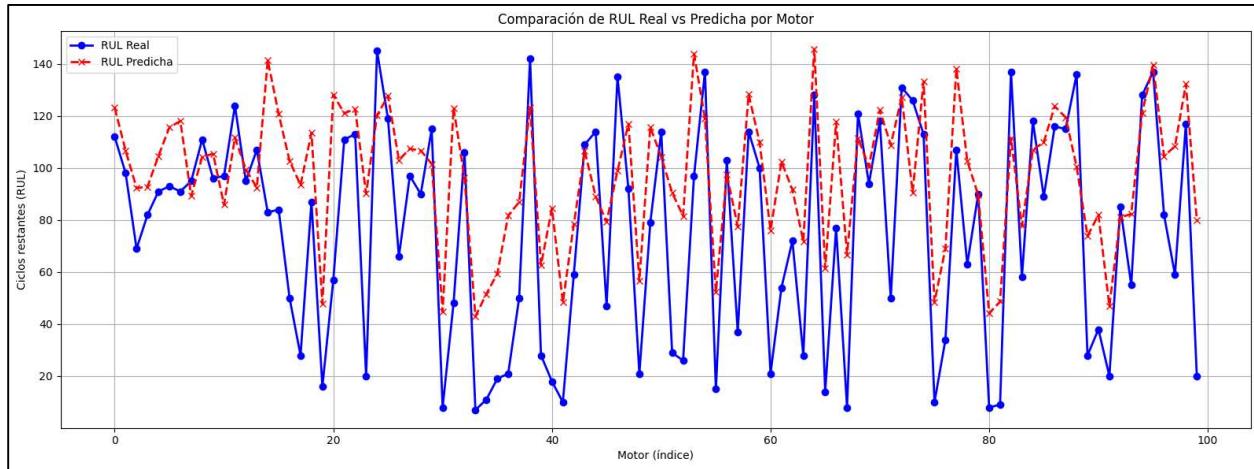
Luego de analizar los tres escenarios, se observa que ninguno logra aproximarse con precisión al RUL real, ya que todos tienden a sobreestimar.

El mejor desempeño relativo se obtiene en el Escenario 3, sin embargo, el valor bajo del coeficiente de determinación ( $R^2$ ) en el conjunto de validación indica la presencia de sobreajuste. Aunque las métricas en el conjunto de entrenamiento y prueba son buenas, al compararlas con el comportamiento en datos reales (validación), se evidencia que el modelo no generaliza adecuadamente.

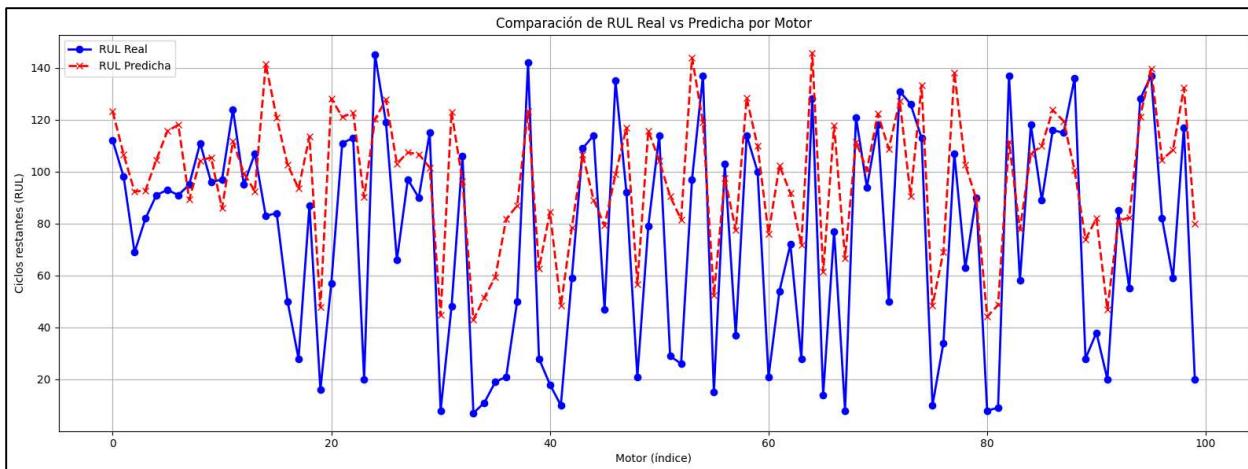
Esto sugiere que es necesario seguir afinando el modelo, ya sea eliminando más sensores irrelevantes, incorporando más datos o incluso evaluando algoritmos más robustos para mejorar la capacidad predictiva.



**Fig. 6: Escenario 1**



**Fig. 7: Escenario 2**



**Fig. 8: Escenario 3**

### 5.5.3. Modelo Random Forest

Ahora aplique un modelo de Random Forest Regressor con el objetivo de mejorar la precisión en la predicción de la RUL.

A diferencia de la regresión lineal, Random Forest permite capturar relaciones no lineales e interacciones entre los sensores, sin necesidad de suposiciones sobre la forma de la función subyacente.

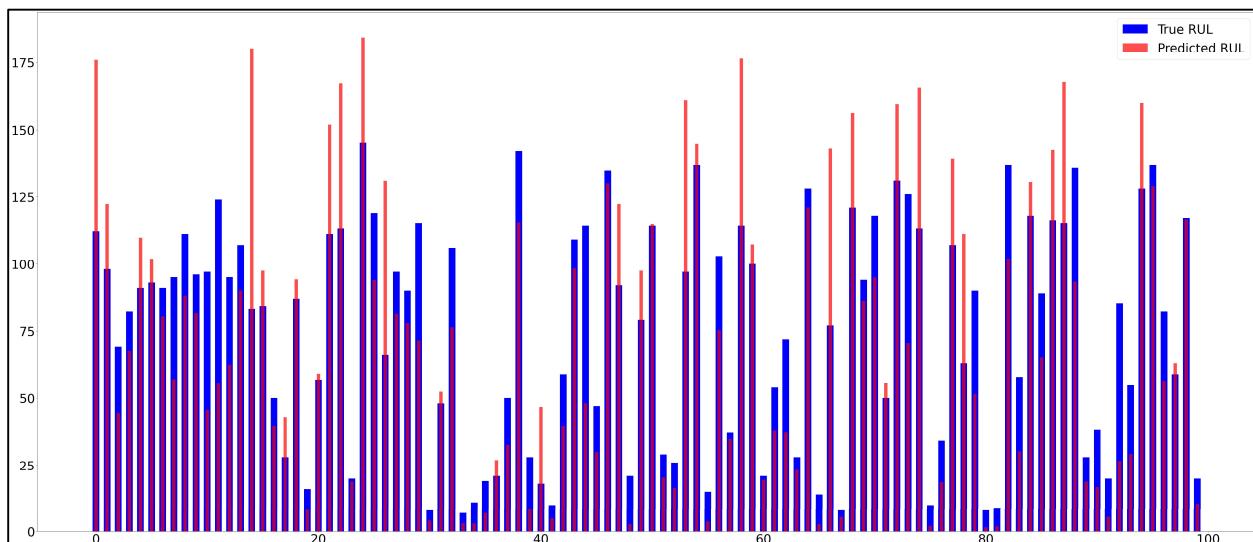
Además, este tipo de modelo es resistente al sobreajuste y maneja bien datos ruidosos o con valores atípicos, lo cual es común en sistemas reales de mantenimiento predictivo.

Escenario	Set	MAE	RMSE	R <sup>2</sup>
Todos los sensores y una muestra de los datos (Fig.9)	Train	10.83	15.41	0.9507
	Test	30.99	44.36	0.5703
	Valid	23.50	30.65	0.4559
Excluyendo sensores constantes y usando muestra de datos (Fig.10)	Train	9.12	12.04	0.9607
	Test	31.49	45.82	0.5417
	Valid	20.49	26.55	0.5919
Excluyendo sensores constantes e incluyendo todos los datos. (Fig. 11)	Train	7.21	9.79	0.9740
	Test	22.50	29.96	0.7514
	Valid	28.72	37.19	0.1990

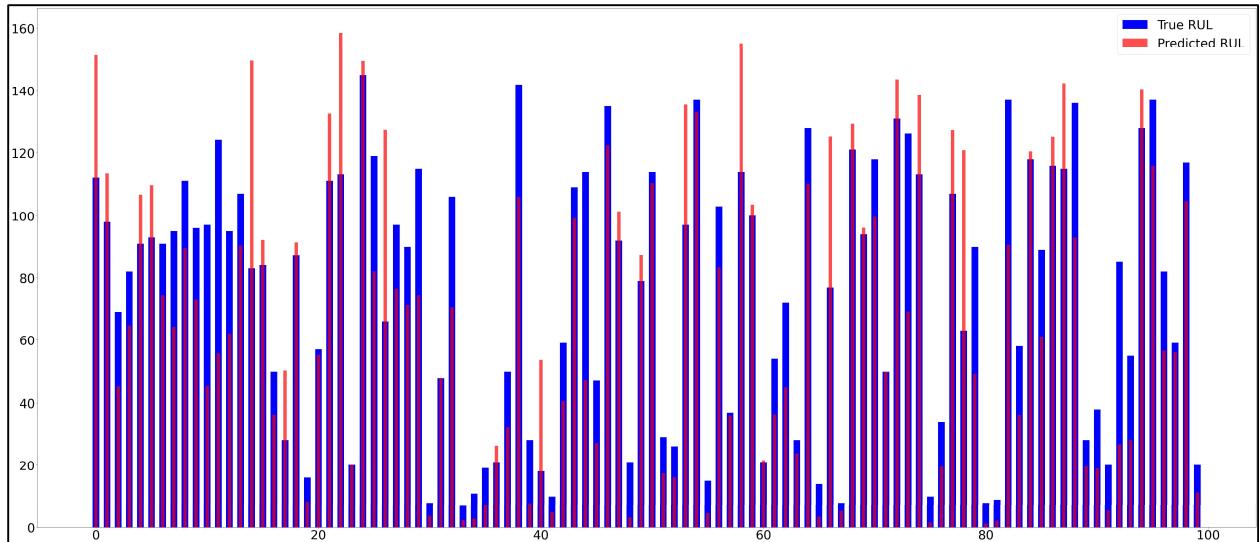
Inicialmente, el modelo Random Forest proyecta mejores resultados en comparación con la regresión lineal. Sin embargo, sigue presentando indicios de sobreajuste.

El Escenario 2 muestra el mejor equilibrio entre precisión y generalización, con un sobreajuste más controlado. Aunque el Escenario 3 obtiene las mejores métricas en entrenamiento y prueba, su bajo valor de  $R^2$  en el conjunto de validación indica que no es tan confiable al incorporar nuevos motores.

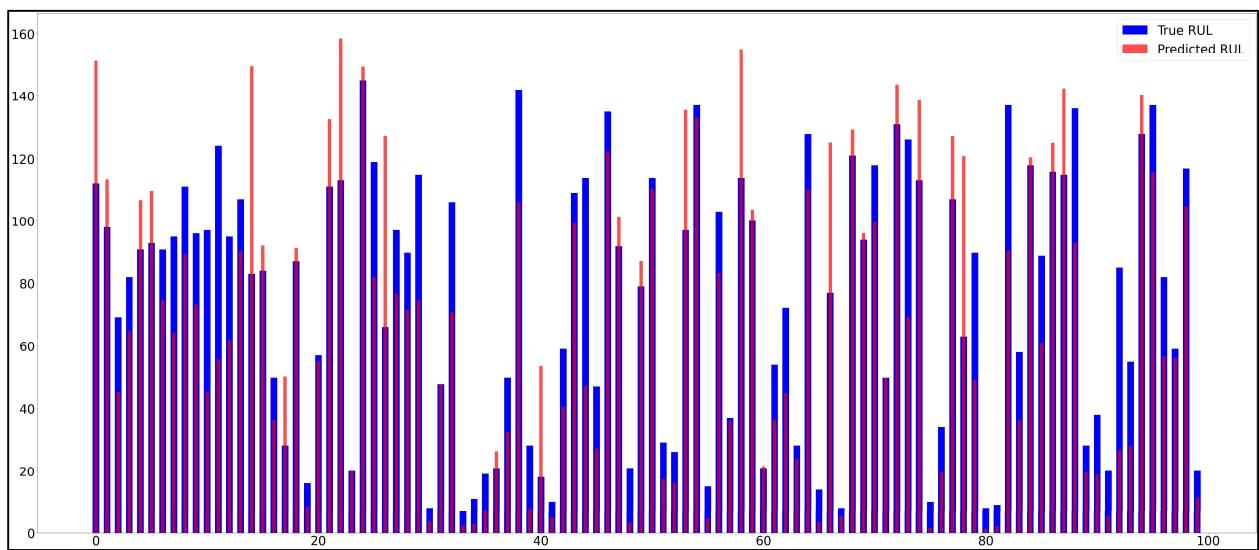
Además, se confirma que eliminar sensores con valores constantes contribuye a mejorar el rendimiento del modelo.



**Fig. 9: Escenario 1**



*Fig.10: Escenario 2*



*Fig.11: Escenario 3*

## 6. Conclusión

En conclusión, este estudio resultó ser interesante, ya que me permitió aplicar diferentes tipos de modelos predictivos sobre un caso práctico y relevante en el ámbito aeronáutico. A través del uso de regresión lineal y Random Forest, pude comparar enfoques simples y avanzados para estimar la vida útil remanente (RUL) de motores simulados.

Si bien el modelo de Random Forest mostró un mejor desempeño general que la regresión lineal, los resultados en el conjunto de validación evidenciaron la necesidad de seguir afinando el modelo, ya que las predicciones aún no se aproximan de forma óptima al RUL real.

Este ejercicio permitió evidenciar la importancia de la selección de variables, la eliminación de sensores constantes y el uso adecuado de técnicas de preprocesamiento como la normalización. Además, destacó cómo la correcta división de los datos y la validación sobre motores no vistos es esencial para evitar el sobreajuste y garantizar la confiabilidad del modelo en un entorno operativo.

Considero que este proyecto representa una base sólida sobre la cual continuar explorando técnicas más complejas.

## 7. Recomendaciones

La recomendación que haría para este estudio sería seguir afianzando estos modelos y realizar diversas pruebas con diferentes sensores, ya que considero que modelos predictivos tiene una gran parte de prueba y error.

Adicional, luego de realizar este análisis, considero que el enfoque podría haberse orientado también hacia la definición de rangos críticos de RUL, que permitan anticipar de forma más clara si un motor está próximo a fallar. Establecer umbrales de alerta podría aportar mayor valor operativo al modelo, facilitando la toma de decisiones proactivas en el contexto del mantenimiento predictivo.

## 8. Bibliografía

- NASA. (s.f.). *NASA Open Data Portal*. <https://data.nasa.gov/dataset/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., ... & Duchesnay, É. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830. <https://scikit-learn.org/stable/index.html>
- Derbel, W. (2022). *NASA - Predictive Maintenance RUL*. Kaggle. <https://www.kaggle.com/code/wassimderbel/nasa-predictive-maintenance-rul/>
- GeeksforGeeks. (2023, abril 25). *Random Forest Algorithm in Machine Learning*. <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>