

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA**

**INTELIGENCIA ARTIFICIAL INDUSTRIAL (1MTR19)
Examen 2
(Ciclo de verano 2024)**

Indicaciones generales:

- Duración: 3 horas.
- **La presentación influirá en la calificación.**

Puntaje total: 20 puntos

Caso de estudio: Remaining useful life estimation of Aircraft Engine

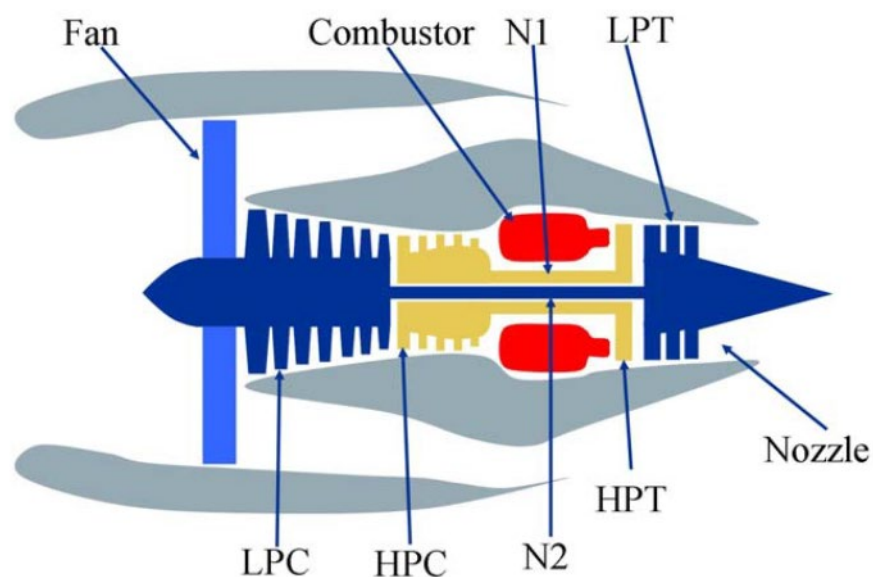


Image [[Source](#)]

Para este examen, la tarea consiste en crear un algoritmo que estime la vida útil remanente de motores de avión. Los estudiantes deberán desarrollar una solución completa al problema usando machine learning. Leer el archivo 'ReadMe' en paideia 'Examen 2' para detalles de la data. La data esta con el nombre 'train_FD001' en paideia 'Examen 2'.

1. EDA & Preprocesamiento de datos (5 puntos)

- Gráficos de línea (comentar)
- Tratamiento de outliers, valores faltantes, etc.
- Eliminación de sensores innecesarios.

- Crear columna Remaining Useful Life apartir de la información de la data.

2. **Modeling** (5 puntos)

- Considerar unit_number [1-20] como test y [21-100] como training.
- Seleccionar variables dependientes e independientes.
- Entrenar modelo.
- Evaluar modelo (RMSE) y compararlo con baseline de valor constante (media). Hacer comentarios.
- Graficar Feature Importance Plot (comentar).

3. **Optimización de hiperparámetros** (5 puntos)

- Optimizar los hiperparámetros del modelo usado en 'Modeling'.
- Para XGBoost o modelo similar, optimizar 'n_estimators', 'learning_rate' y 'max_depth'.
- Entrenar el modelo con hiperparámetros optimizados y compararlo con modelo entrenado en '3. Modeling'.

4. **Cross validation** (5 puntos)

Implementar cross validation en la solución para 5 combinaciones distintas de grupos de motores ('unit_number') para dividirlos en datos de test y entrenamiento. En cada caso de cross validation, un conjunto de veinte motores ('unit_number') se asignará como datos de prueba (testing data), y los motores restantes como datos de entrenamiento (training data).

Por ejemplo:

- Caso 1: Datos de Prueba ('unit_number' [1-20]), Datos de Entrenamiento ('unit_number' [21-100]).
- Caso 2: Datos de Prueba ('unit_number' [21-40]), Datos de Entrenamiento: los motores restantes.
- Y así continúan sucesivamente ...

Para cada caso se debe:

1. Realizar predicciones para los datos de prueba.
2. Calcular y registrar el puntaje RMSE (Root Mean Square Error) para las predicciones.

Al finalizar todos los casos, se debe calcular y mostrar el promedio de los puntajes RMSE.

Los estudiantes pueden incluir cualquier otro paso que consideren necesario para mejorar o complementar el análisis.

Link de competencia (referencial): [PHM 2008 Challenge | NASA Open Data Portal](#)

Entregables: Código (.ipynb) con comentarios e impresión del código en pdf. Los comentarios son importantes para la calificación.

Enviar los entregables al correo: john.taco@pucp.edu.pe hasta las 10:00 pm del jueves 29 de febrero del 2024.

- Asunto del Correo: Examen2_1MTR19_[su_código_PUCP]
- Nombre del Archivo Jupyter (ipynb):
Examen2_1MTR19_[su_código_PUCP]_codigo.ipynb
- Nombre del Archivo PDF:
Examen2_1MTR19_[su_código_PUCP]_codigo.pdf

Profesor del curso: Ing. John Oliver Taco López

San Miguel, 29 de febrero de 2024