

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

**TEMAS EN INGENIERÍA MECATRÓNICA H (1MTR19)**

**Examen 1  
(Ciclo de verano 2024)**

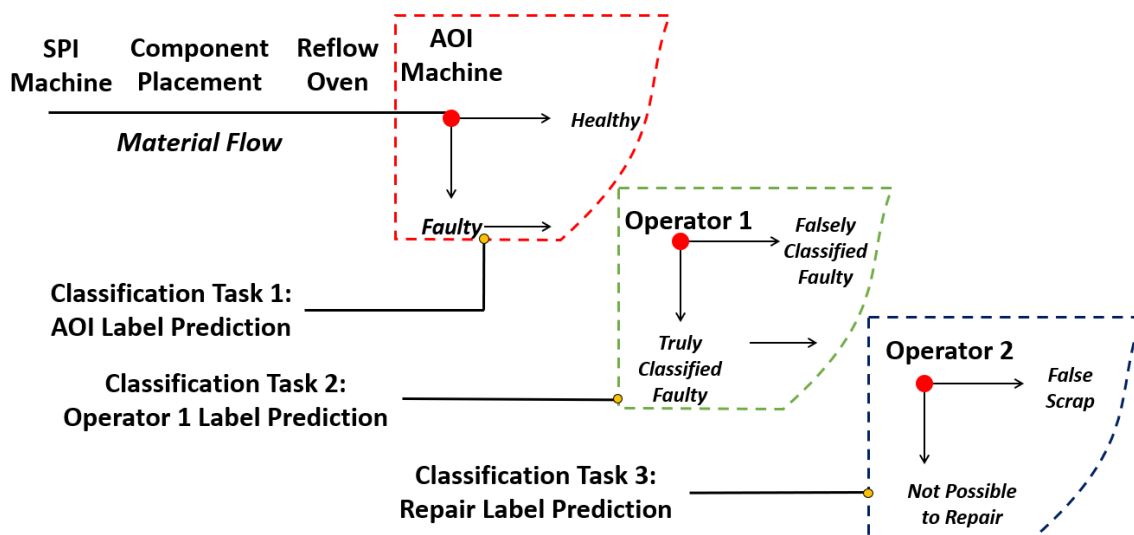
**Indicaciones generales:**

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

Puntaje total: 20 puntos

---

**Caso de estudio:** Detección de fallas de circuitos electricos.



Para detalles del caso de estudio en general: <https://phm-europe.org/data-challenge>

Data:

**SPI:** 'SPI\_training\_0.csv.zip', 'SPI\_training\_1.csv.zip', 'SPI\_training\_2.csv.zip', 'SPI\_training\_3.csv.zip'

**AOI:** 'AOI\_training.csv.zip'

Descripción de la data: Ver documento 'descripcion de la data'

Para este examen, la tarea consiste en realizar la clasificación llevada a cabo por el 'Operador 2' (task 3, 'RepairLabel'), la cual es binaria y se divide en dos categorías: 'False Scrap' y 'Not Possible to Repair'. La predicción se hará al nivel de [PanelID, FigureID, ComponentID].

Nota: En la data, la variable 'RepairLabel' tiene tres categorías : 'FalseScrap', 'NotPossibleToRepair', 'NotYetClassified' y datos nulos. Para el presente examen, **no tomar en cuenta los datos con categoría 'NotYetClassified' ni los nulos de la columna 'RepairLabel'**. Tomar acciones necesarias.

Los estudiantes deberán desarrollar un modelo base utilizando técnicas de aprendizaje automático. Se recomienda emplear un modelo simple como XGBoost o algún modelo comparable, aplicado específicamente a este caso de estudio. El desarrollo del modelo base debe abarcar los siguientes pasos:

1. **Análisis Exploratorio de Datos (EDA):** 2 puntos
  - Identificación de tendencias, patrones y valores nulos para los datos AOI y SPI. Mostrar gráficas y comentarios.
2. **Ingeniería de características (feature engineering):** 4 puntos
  - Obtener feature matrix para data SPI. Considerar todas las variables numéricas (Volumen, Size, etc) con respecto a PanelID, FigureID, ComponentID. Usar todas las agregaciones posibles (mean, min, max, std, median, etc)
  - Obtener feature matrix para data AOI. Considerar columna RepairLabel con respecto a PanelID, FigureID, ComponentID.
3. **Integración de datos:** 2 puntos
  - Combinación efectiva de diferentes conjuntos de datos. (Integración de data AOI y SPI luego de la agregación). Se usa los feature matrix obtenidos en el punto 2.
4. **Definición de variables independientes y dependientes:** 2 puntos
  - Seleccionar las variables independientes (X)
  - Definir la variable dependiente según lo requerido en el problema de clasificación (y)
5. **Separación de los datos en conjuntos de entrenamiento y validación:** 2 puntos
  - División adecuada de los datos para entrenamiento y validación del modelo. (70% entrenamiento - 30% validación)
6. **Realización de predicciones:** 2 puntos
  - Uso del modelo para realizar predicciones claras y precisas
7. **Evaluación del modelo, incluyendo F1 score, accuracy y la matriz de confusión:** 4 puntos
  - Análisis detallado del rendimiento del modelo utilizando F1 score, accuracy y matriz de confusión (confusion matrix).

8. **Análisis de la importancia de las características (feature importance):** 2 puntos

- Evaluación y explicación de qué características tienen mayor influencia en las predicciones del modelo. Generar grafica de ranking de features (feature importance plot).

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

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](mailto:john.taco@pucp.edu.pe) hasta las 12:00 pm del viernes 2 de febrero del 2024.

- **Asunto del Correo:** Examen1\_1MTR19\_[su\_código\_PUCP]
- **Nombre del Archivo Jupyter (ipynb):**  
Examen1\_1MTR19\_[su\_código\_PUCP]\_codigo.ipynb
- **Nombre del Archivo PDF:** Examen1\_1MTR19\_[su\_código\_PUCP]\_codigo.pdf

Profesor del curso: John Oliver Taco López

San Miguel, 02 de febrero de 2024