

Tarea 3: Modelos Supervisados

- **Propósito:** Introducir a los estudiantes a los modelos de aprendizaje supervisado para predicción y clasificación.
- **Herramientas a utilizar:**
 - **Google Colab** para la codificación.
 - **Lenguaje R:** librerías como `caret`, `randomForest`, `e1071`, y `ggplot2`, etc.
- **Descripción de la actividad:**
 - Se entregará un conjunto de datos de predicción (por ejemplo, predicción de precios de casas o clasificación de clientes).
 - Los estudiantes deben:
 1. Dividir los datos en conjuntos de entrenamiento y prueba. Interpretar resultados. **2.5 puntos**
 2. Implementar un modelo de regresión o clasificación para resolver el problema. Interpretar resultados. **2.5 puntos**
 3. Ajustar hiperparámetros utilizando validación cruzada. Interpretar resultados. **2.5 puntos**
 4. Comparar métricas como RMSE, R^2 , o precisión en diferentes modelos aprendidos en clases, matriz de confusión. Interpretar resultados. **2.5 puntos**
 5. Visualizar resultados con gráficos adecuados y establecer las variables mas importantes del modelo. **2.5 puntos**
 6. Agregar controles #@markdown y prueba la predicción de un nuevo registro. **2.5 puntos**
 7. Un video con cámara encendida de máximo 5 minutos. de la explicación del trabajo, propósito, código en colab, base de datos en github, y los resultados obtenidos. **5 puntos**
 - **Resultado esperado:** Un informe que detalle el análisis, los modelos probados y la comparación de resultados.
- **Nota:** Subir un archivo PDF con portada y sólo agregar los enlaces de Google Colab y del video.

Base de datos: El conjunto de datos para la clasificación de fallas de máquinas en entornos industriales incluye 1,000 muestras con mediciones de sensores en tiempo real, como temperatura, vibración, uso de energía, humedad y tipo de máquina, **destinadas a predecir el riesgo de falla (0: normal, 1: en riesgo)**. Diseñado para el mantenimiento predictivo, permite identificar patrones y tendencias en las lecturas de sensores asociados con fallas, ayudando a prevenir costosos tiempos de inactividad mediante intervenciones tempranas.

https://raw.githubusercontent.com/Emax1/DataMining4/refs/heads/main/machine_failure_dataset.xls