

Aplicar la metodología CRISP-DM para realizar una tarea de regresión, seleccionando características mediante Recursive Feature Elimination (RFE) y árboles de decisión, y desplegar el modelo utilizando Flask y Render. El alumno entregará tres productos: un documento en Word, una libreta en Jupyter y la publicación del modelo en Render.

Selección del DataSet

Debes buscar un DataSet para regresión que cumpla con los siguientes requisitos:

- El atributo dependiente debe ser una variable numérica continua.
- Las variables independientes (X) deben contener de 8 a 100 columnas.
- Registrar el nombre y la URL del DataSet en el archivo proporcionado, asegurándose de no repetir DataSets con otros alumnos.

Productos a Entregar

1. Documento en Word:

- **Portada:** Incluir datos personales (Nombre, materia, profesor, grupo, etc.).
- **Entendimiento del Negocio:**
 - **Introducción:** Describir el DataSet, número de registros y atributos, y objetivo de la predicción. Explicar el contexto y relevancia del problema.
- **Entendimiento de los Datos:**
 - **Descripción de Atributos:** Describir cada atributo, su tipo (Nominal, Ordinal, Interval, Ratio), tipo de dato (object, int64, float64, etc.) y si contiene valores nulos o faltantes.
 - **Análisis exploratorio de los datos:** presentar gráficas o estadísticas para el análisis y explicar cada caso.
- **Preparación de los Datos:**
 - **Tratamiento de Valores Faltantes:** Generar y tratar valores nulos si no existen en el DataSet. Explicar y justificar la técnica utilizada.
 - **Conversión de Atributos:** Convertir atributos de texto a numéricos si es necesario. Explicar la necesidad de esta conversión.
 - **Escalado de Atributos:** Normalizar los valores de X si es necesario, justificando la decisión.
 - **Selección de Atributos:** Seleccionar entre 4 y 6 atributos utilizando RFE, matriz de correlación y árboles de decisión. Justificar, con base en los tres métodos la elección y exclusión de atributos.
- **Modelado:**
 - **Entrenamiento de Modelos:** Entrenar dos modelos de regresión, uno de los cuales debe ser una red neuronal (el otro modelo no debe ser red neuronal). Utilizar k-fold cross-validation.
- **Evaluación de Modelos:**
 - **Métricas de Evaluación:** Evaluar los modelos utilizando R^2 y alguna otra métrica adecuada (RMSE, MAE), realizar validación cruzada y reportar la media, desviación estándar y varianza.

- **Gráfica de Dispersión:** Incluir y analizar gráficas de dispersión de los resultados de ambos modelos.
 - **Gráfica de Residuos:** Incluir y analizar gráficas de residuos para cada modelo (Para las gráficas realiza una ejecución empleando Train Test Split).
 - **Evaluación:**
 - **Conclusión:** Indicar cuál modelo fue mejor y explicar el motivo.
 - **Despliegue:**
 - **Implementación en Render:** Describir brevemente la implementación y despliegue del modelo utilizando Flask y Render.
 - **Bibliografía:** Incluir referencias en formato APA o IEEE.
 - **Generalidades:** Incluir explicación, código y evidencia de ejecución del código copiados de la libreta de Jupyter.
2. El documento debe ser entregado impreso en un folder (color azul para el grupo A, beige para el grupo B) con las hojas grapadas en la esquina superior izquierda.
 3. **Libreta de Jupyter:**
 - **Contenido:** Incluir todo el código del ejercicio desde la carga del DataSet, selección de atributos, tratamiento de valores faltantes, generación y evaluación de los modelos.
 - **Mostrar la Ejecución:** Presentar la ejecución de la libreta al inicio de la clase.
 4. **Despliegue del Modelo en Render:**
 - **Servicio Web:** Crear un servicio web utilizando Flask.
 - **Página Web:** Desarrollar una página web que muestre el modelo de regresión con un formulario para ingresar valores.
 - **Despliegue en Render:** Desplegar el modelo en la plataforma Render.
 - **Compartir URL:** Compartir la URL del formulario y colocar el link en la libreta de Jupyter.

Entrega y Evaluación

- **Documento en Word:** Entregar impreso en un folder, con las hojas grapadas en la esquina superior izquierda.
- **Libreta de Jupyter:** Mostrar la ejecución al inicio de la clase.
- **URL del Formulario:** Compartir la URL donde se puede interactuar con el modelo desplegado en Render.

Esta tarea tiene una ponderación del 30% de la evaluación mensual.