

Actividad Integradora (Módulos 1 y 2)

Proyecto: Predicción de una Variable

Objetivo: Aplicar el ciclo de vida de la Ciencia de Datos a un conjunto de datos real, integrando los conocimientos de Python, manipulación de datos y modelado básico.

Instrucciones para los estudiantes:

1. **Formulación:** Elijan un conjunto de datos público (por ejemplo, de Kaggle o un repositorio conocido) con al menos 1,000 filas y una variable objetivo (numérica para Regresión o binaria para Clasificación).
2. **Adquisición y Exploración (Módulo 1 y 2):** Cargar el dataset en un *Jupyter Notebook*. Realizar un **Análisis Exploratorio de Datos (EDA)**: calcular estadísticas descriptivas, visualizar la distribución de las variables clave y las correlaciones, e identificar problemas de calidad de datos.
3. **Preprocesamiento (Módulo 2):** Limpiar y transformar los datos: tratar valores perdidos, manejar *outliers* y codificar variables categóricas. Escalar las variables numéricas.
4. **Modelado Básico (Módulo 3):**
 - Dividir el conjunto de datos en entrenamiento y prueba.
 - Entrenar al menos dos modelos vistos (por ejemplo, Regresión Lineal/Logística y Árbol de Decisión).
5. **Entrega:** Documento (*Notebook*) con el código, las visualizaciones y un breve reporte de las métricas de rendimiento de cada modelo (p. ej., R^2 para regresión o *Accuracy* para clasificación).