

Proyecto Integrador

Módulo 2

En este proyecto integrador aplicarás los conocimientos adquiridos en el módulo 2, donde revisamos técnicas de **regresión**. El objetivo es seleccionar un conjunto de datos de interés, realizar un análisis completo y poner en práctica un modelo predictivo o exploratorio, según corresponda.

1. Selección de Datos

Explora fuentes de datos abiertas o utiliza un conjunto propio en un área de interés. Puedes considerar:

- Kaggle: <https://www.kaggle.com/datasets>
- UCI Machine Learning Repository: <https://archive.ics.uci.edu/>
- DataHub: <https://datahub.io/>
- Datos Abiertos México: <https://datos.gob.mx/>

Elige un conjunto de datos que te permita explorar un problema de **regresión**.

2. Entregable

El trabajo se presentará en una libreta de código (Jupyter Notebook, Google Colab o similar), organizada en secciones con celdas de texto y código.

3. Secciones

a) Introducción

- Describe brevemente el origen del conjunto de datos y el área de interés.
- Explica el problema que deseas resolver o explorar.
- Justifica la técnica elegida (clustering, regresión o clasificación) en función del problema.

b) Análisis Exploratorio de Datos (EDA) y Procesamiento

- Realiza un análisis exploratorio que incluya:
 - Distribuciones univariadas: histogramas para variables numéricas, gráficas de barras para variables categóricas.
 - Análisis bivariado o multivariado (ej. correlaciones, diagramas de dispersión, heatmaps, boxplots) según corresponda.
- Aplica técnicas de procesamiento según sea necesario:
 - Limpieza de valores nulos o atípicos.
 - Codificación de variables categóricas.
 - Normalización o estandarización de variables numéricas.
 - Selección o reducción de variables (si aplica).

c) Modelado

- Implementa al menos un modelo de regresión, clasificación o clustering.
- Explica brevemente la técnica seleccionada.
- Define y aplica métodos de evaluación adecuados según el tipo de problema:
 - Regresión: RMSE, MAE, R².
 - Clasificación: accuracy, F1-score, matriz de confusión, ROC-AUC.
 - Clustering: métricas internas (silhouette score, inercia).

d) Resultados y Conclusiones

- Presenta e interpreta los resultados obtenidos.
- Resume los hallazgos más relevantes.
- Reflexiona sobre posibles mejoras, limitaciones y aplicaciones prácticas del modelo.

e) Presentaciones

- Cada equipo tendrá 15 minutos (máximo) para presentar sus resultados. En la presentación deberán incluir las 4 secciones previamente indicadas.

4. Recomendaciones Generales

- Documenta tu libreta de manera clara con celdas de texto explicativas.
- Incluye visualizaciones que apoyen la interpretación de los datos y resultados.
- Sé conciso en la redacción, pero suficiente en la explicación técnica.
- El proyecto debe demostrar tanto el dominio técnico de las herramientas como la capacidad analítica para interpretar resultados.

5. Lista de Cotejo

- Explica origen de los datos y problema a resolver.
- Realiza análisis exploratorio (gráficas y descripción).
- Aplica técnicas de procesamiento de datos.
- Implementa al menos un modelo (clustering, regresión o clasificación).
- Evalúa el modelo con métricas adecuadas.
- Presenta conclusiones claras.