



Universidad Anáhuac Machine Learning

Práctica 1: Técnica Supervisada “Árboles de Decisión”

Instrucciones. Lea cuidadosamente cada inciso y desarrolle el programa que se le solicita en Jupyter Notebook. **Al finalizar, guarde el archivo con el nombre Práctica1_SuNombre y subirlo a Brighspace**

Una organización política está interesada en predecir el nivel de apoyo ciudadano que puede alcanzar una campaña antes de asignar recursos adicionales.

El nivel de apoyo ciudadano se clasifica en tres categorías:

High → Alto apoyo ciudadano

Medium → Apoyo ciudadano moderado

Low → Bajo apoyo ciudadano

Con base en información socioeconómica, de exposición mediática y de esfuerzo de campaña, se busca construir un modelo de clasificación que permita estimar el nivel de apoyo esperado para nuevas campañas.

Importe las librerías necesarias (pandas, matplotlib, DecisionTreeClassifier, train_test_split , etc).

- 1) Lea el archivo de excel “Politics_BD.xlsx” y cárguelo en un *Dataframe* llamado *data*, imprima los primeros 5 registros para revisar que se haya cargado de manera correcta.
- 2) Obtenga el tamaño de la BD y obtenga la descripción general de la base de datos
- 3) Asigne en un *Dataframe* llamado X las variables predictoras (District_Type, Region_ID, Education_Index, Income_Level, Media_Exposure, Campaign_Intensity, Citizen_Engagement) y en un *Dataframe* llamado Y la variable objetivo (Support_Level). Imprima los 5 primeros registros de X y Y para asegurar que se cargaron correctamente.
- 4) Divida la base de datos en un conjunto de prueba (*test*) y de entrenamiento (*train*). Tomando 20 por ciento de la base total para el conjunto de prueba (*test*).
- 5) Construya su modelo utilizando árboles de decisión y ajústelos al conjunto de entrenamiento (*train*).
- 6) Obtenga el *score* obtenido para el conjunto de prueba *test*.
- 7) Genere la matriz de confusión.
- 8) Tomando en cuenta el inciso anterior, indique
 - A) Cuántos registros se clasificaron correctamente como Medium (1)
 - B) Cuántos registros eran Medium pero se clasificaron incorrectamente como High (2)
 - C) Cuántos registros se clasificaron correctamente como High (2)
- 9) Si un ciudadano tiene las siguientes características District_Type =1, Region_ID=1, Education_Index=5, Income_Level=5, Media_Exposure=4, Campaign_Intensity=5, Citizen_Engagement=88. ¿Qué categoría le asigna automáticamente el árbol de decisión 0-Low, 1-Medium, 2- High?
- 10) Grafique el árbol. Interpretelo y explíquelo de manera breve.