

Hoja de Trabajo – Clasificación

Para cada uno de los ejercicios:

- Suba los archivos de .ipynb indicados
- Exporte los resultados de su notebook en formato HTML y adjúntelo también
- Responda las preguntas indicadas

1. Clasificación de Ordenes

- a. Utilice el siguiente script para crear una vista llamada “VW_OrdenesEncimaPromedio” en la base de datos de [RepuestosWeb] utilizada en clases anteriores

```
CREATE VIEW VW_OrdenesEncimaPromedio
as
    SELECT ci.Nombre AS NombreCiudad,
           c.Genero,
           p.ID_Parte,
           p.ID_Categoria,
           o.Total_Orden,
           TotalEncimaPromedio = CASE WHEN (o.Total_Orden/
                                           (SELECT AVG(o2.Total_Orden)
                                            FROM dbo.Orden o2)) > 1
                                     THEN 1 ELSE 0 end
FROM      dbo.Clientes c
INNER JOIN dbo.Orden o
    ON o.ID_Cliente = c.ID_Cliente
INNER JOIN dbo.Detalle_orden do
    ON do.ID_Orden = o.ID_Orden
INNER JOIN dbo.Partes p
    ON p.ID_Parte = do.ID_Parte
INNER JOIN dbo.Ciudad ci
    ON ci.ID_Ciudad = o.ID_Ciudad
INNER JOIN dbo.Categoria cat
    ON cat.ID_Categoria = p.ID_Categoria
```

La vista anterior agrupa atributos de ordenes históricas según género del cliente, parte que adquirió, categoría de la parte y la ciudad donde se realizó la orden, luego muestra en la columna “TotalencimaPromedio” los valores 1 o 0, donde 1 indica que el monto total de la orden supera el promedio de las órdenes y 0 que no.

- b. Cree un algoritmo de Python en VS code y desarrolle un modelo de Naive Bayes, asegurándose que atributos no numéricos se clasifiquen en diferentes categorías con “Encode” o que los atributos numéricos sigan la misma escala
- c. Haga un Split de 80-20 donde el 80% sirva para crear el modelo y el 20% restante para realizar pruebas usando la función “predict()”
- d. Cree la matriz de confusión y calcule el nivel de precisión de su modelo.
- e. Grafique la matriz de confusión usando la librería “plotly” vista en clase
- f. Compare ambos modelos (Naive Bayes y Árboles de decisión) según la cantidad de precisión en base a la matriz de confusión y en base al AUC
 - i. ¿Cuál modelo es mejor y por qué?

