



Desafío - Clasificación desde Machine Learning

- Para poder realizar este desafío debes haber revisado la lectura y videos correspondiente a la unidad.
- Crea una carpeta de trabajo y guarda todos los archivos correspondientes (notebook y csv).
- Una vez terminado el desafío, comprime la carpeta y sube el `.zip`

Descripción

En esta sesión trabajaremos con una base de datos sobre clientes morosos de un banco. Dentro de ésta se registran las siguientes observaciones:

- `default` : Variable binaria. Registra si el cliente entró en morosidad o no.
- `income` : Ingreso promedio declarado por el cliente.
- `balance` : Total de saldo en la cuenta de crédito.
- `student` : Variable binaria. Registra si el cliente es estudiante o no.

Desafío 1: Preparación del ambiente de trabajo

- Importe los módulos básicos para el análisis de datos
- Importe `seaborn` .
- Importe `statsmodels`
- Importe `LogisticRegression` , `train_test_split` , `StandardScaler` , `classification_report` y `roc_curve` de `sklearn`

Desafío 2: Análisis exploratorio

- Explore las principales relaciones entre las variables. Utilizando los métodos `boxplot` y `lplot` de `seaborn` implemente los siguientes gráficos.
 - Un boxplot del balance total de las cuentas, segmentado por el status de estudiante.
 - Un boxplot del ingreso total de cada cliente, segmentado por el status de estudiante.
 - Un boxplot del balance total de las cuentas, segmentado por el status de default.
 - Un boxplot del ingreso total de cada cliente, segmentado por el status de default.
 - Un diagrama de dispersión entre `balance` e `income`, segmentado por el status de estudiante.
 - Un diagrama de dispersión entre `balance` e `income`, segmentado por el status de default.
- Comente brevemente sobre los principales patrones de comportamiento.
- *tip*: Puede utilizar el argumento `hue` en `lplot` para identificar las observaciones por status.

Desafío 3: Modelación econométrica

- Recodifique las variables `student` y `default` para que adquieran atributos numéricos.
- Implemente el siguiente modelo econométrico:

$$\log\left(\frac{\Pr(\text{default} = 1)}{1 - \Pr(\text{default} = 1)}\right) = \beta_0 + \beta_1 \cdot \text{balance} + \beta_2 \cdot \text{income} + \beta_3 \cdot \text{student}$$

- A partir de este modelo, depúrelo y mantenga las principales características. Ejecute de nuevo el modelo y comente brevemente sobre el comportamiento de los atributos.

Desafío 4: Entrenar

A partir del modelo depurado, genere un modelo predictivo con las siguientes características:

- Estandarize la matriz de atributos.
- Genere subconjuntos de entrenamiento y validación, separando el 33% de la muestra en el conjunto de validación.
- Implemente el modelo sin alterar los hiperparámetros.
- Genere la predicción de las clases y de la probabilidad.
- Reporte las métricas de precision, recall y f1 para ambas clases.
- Grafique la curva Receiver Operator Characteristic.
- En base a los pasos, comente sobre el desempeño predictivo del modelo.