

## Desafío - Clasificación desde Machine Learning

- Para realizar este desafío debes haber estudiado previamente todo el material disponibilizado correspondiente a la unidad.
- Una vez terminado el desafío, comprime la carpeta que contiene el desarrollo de los requerimientos solicitados y sube el `.zip` en el LMS.
- Desarrollo desafío:
  - El desafío se debe desarrollar de manera Individual.
  - Para la realización del desafío necesitarás apoyarte del archivo Apoyo Desafío - Clasificación desde Machine Learning.

### 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 `lmpplot` 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 `lmpplot` 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:

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

- 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:

- Genere subconjuntos de entrenamiento y pruebas, separando el 33% de la muestra en el conjunto de pruebas.
- Estandarice las matrices de atributos de entrenamiento y pruebas usando `StandardScaler()` ajustado solo sobre el conjunto de entrenamiento.
- Implemente el modelo sin alterar los hiper parámetros.
- Genere la predicción de las clases y de la probabilidad.
- Reporte las métricas de precisión, recall y f1 para ambas clases.
- Grafique la curva Receiver Operator Characteristic.
- En base a los pasos, comente sobre el desempeño predictivo del modelo haciendo mención a todas las métricas calculadas y a la curva ROC.