# 457\_Machine\_Learning

# Obligatorio final

Jessica Menendez

Ignacio Amarilla

# Informe del Modelo: "Mi Cliente Comprador"

## Introducción

El objetivo del proyecto es desarrollar un modelo de Machine Learning para predecir si un cliente potencial realizará una compra, basado en diversas características sociodemográficas y transaccionales. Para este propósito, se empleó un enfoque supervisado utilizando la biblioteca scikit-learn en Python.

## Estructura del Código

### 1. Preprocesamiento de Datos

En esta etapa, se manejaron valores nulos y se prepararon los datos para su uso en el modelo:  
1. Valores nulos en las columnas 'Género’ se reemplazaron por 'Desconocido' y 'Mes de contacto' se reemplazaron por la moda.  
2. Las variables categóricas fueron codificadas usando get\_dummies.  
3. Las variables numéricas fueron estandarizadas utilizando StandardScaler.

### 2. Separación de Datos

Los datos se dividieron en independientes (X) y dependientes (y). La variable objetivo 'Compra' fue mapeada a valores binarios: 'si' a 1 y 'no' a 0. Posteriormente, se realizó una división en conjuntos de entrenamiento y prueba con un 20% de datos para prueba.

### 3. Creación del Modelo y Optimización de Hiperparámetros

Se utilizó el modelo Random Forest Classifier.

GridSearchCV se empleó para encontrar los mejores hiperparámetros del modelo. El parámetro de evaluación fue 'accuracy', y se probaron diferentes combinaciones de hiperparámetros como el número de árboles (n\_estimators), la profundidad máxima (max\_depth), y los criterios de división y hojas mínimas.

### 4. Validación Cruzada

Se realizó una validación cruzada con 5 particiones en el conjunto de entrenamiento para evaluar la estabilidad del modelo. Se reportó el promedio de las métricas obtenidas.

### 5. Entrenamiento y Evaluación

El modelo fue entrenado con los datos de entrenamiento y evaluado con el conjunto de prueba. Se calcularon métricas de desempeño como la precisión (accuracy), el informe de clasificación y la matriz de confusión.

## Resultados

El modelo alcanzó un nivel de precisión satisfactorio en la predicción de clientes compradores. El informe de clasificación y la matriz de confusión proporcionaron una visión detallada del desempeño para cada clase.

## Sugerencias de Mejora

1. Probar con otros modelos como Gradient Boosting, XGBoost o LightGBM.  
2. Incrementar el tamaño del dataset para mejorar la capacidad de generalización del modelo.  
3. Ajustar los hiperparámetros con mayor profundidad.  
4. Realizar una selección de características basada en análisis de importancia.