

# Ejercicio churn

## Machine learning con Python

Autor: José Ángel Carballo Sánchez

El caso práctico consiste en calcular el churn (clientes que abandonan una compañía) de clientes en una operadora de telecomunicaciones. La compañía ha descubierto que entre diciembre de 2019 y enero de 2020 han tenido bastantes bajas de clientes. El objetivo de la práctica es llegar a construir un modelo analítico que sea capaz de predecir los próximos clientes que son potenciales a marcharse de la operadora para poder lanzar una campaña e intentar fidelizarles antes de su portabilidad. Además, los directivos de la compañía quieren conocer las causas por las cuales los clientes se están fugando. Para ello disponemos de varios datasets con datos que aportan información de los clientes.

El procedimiento a seguir será el siguiente:

1. Cargar los datasets y construir un único tablón analítico con todas las variables que consideremos necesarias para cada una de las cosechas que queremos analizar.

Las variables de las que disponemos en los datasets, son las siguientes:

id: es el identificador de cliente. Es común a todas las tablas (la clave para después mergearlas)

#### **TABLA CLIENTES**

edad: edad de los clientes.

facturacion: dinero que pagan los clientes al mes.

antigüedad: fecha de alta del cliente.

provincia: lugar de procedencia del cliente.

num\_líneas: número de líneas móviles contratadas.

num\_dt: número de líneas en impago.

incidencia: es SI, si el cliente ha tenido alguna incidencia/reclamación.

#### **TABLA PRODUCTOS**

conexión: tipo de conexión de internet del cliente.

vel\_conexion: velocidad de conexión de internet.

TV: tipo de paquete de tv contratado por el cliente.

#### **TABLA CONSUMOS**

num\_llamad\_ent: número de llamadas entrantes de todas sus líneas.

num\_llamad\_sal: número de llamadas salientes de todas sus líneas.

mb\_datos: mb de los datos consumidos en todas sus líneas.

seg\_llamad\_ent: segundos consumidos en llamadas entrantes.

seg\_llamad\_sal: segundos consumidos en llamadas salientes.

**TABLA FINANCIACION**

financiación: es SI, si el cliente tiene financiado algún terminal.

imp\_financ: el dinero mensual que paga por los terminales financiados.

descuentos: es SI, si el cliente tiene activo algún descuento (campana).

target: esta variable no la tenemos. Una vez que tenemos los dos tabloncillos analíticos contruירים (diciembre y enero) hay que pensar como podemos contruיר esta variable.

Revisar el funcionamiento de `pd.merge()` para contruיר el tabloncillo.

2. Realizar un pre procesamiento y una limpieza de los datos, siguiendo la metodología vista en clase.
3. Muestrear los datos, contruירendo un dataset de train y test.
4. Contruיר un modelo analítico de clasificación que sea capaz de predecir cuando un cliente se fuga de la empresa, argumentando el tipo de algoritmo utilizado, las variables seleccionadas, obteniendo las métricas oportunas... Explicar lo “bueno” o “malo” que es el modelo a través de las métricas obtenidas.
5. Realizar una mejora del modelo utilizando técnicas vistas en clase (feature engineering, análisis de variables, comprobación de overfitting, validación cruzada, tuneado de hiper parámetros a través de la automatización...)
6. Predecir los clientes de la cosecha de enero que más probabilidad tienen de cambiarse de operadora.
7. Obtener y explicar las claves (a través de las variables...) de la marcha de los clientes en la compañía.