

Tangelo

Business case



Tangelo

Company

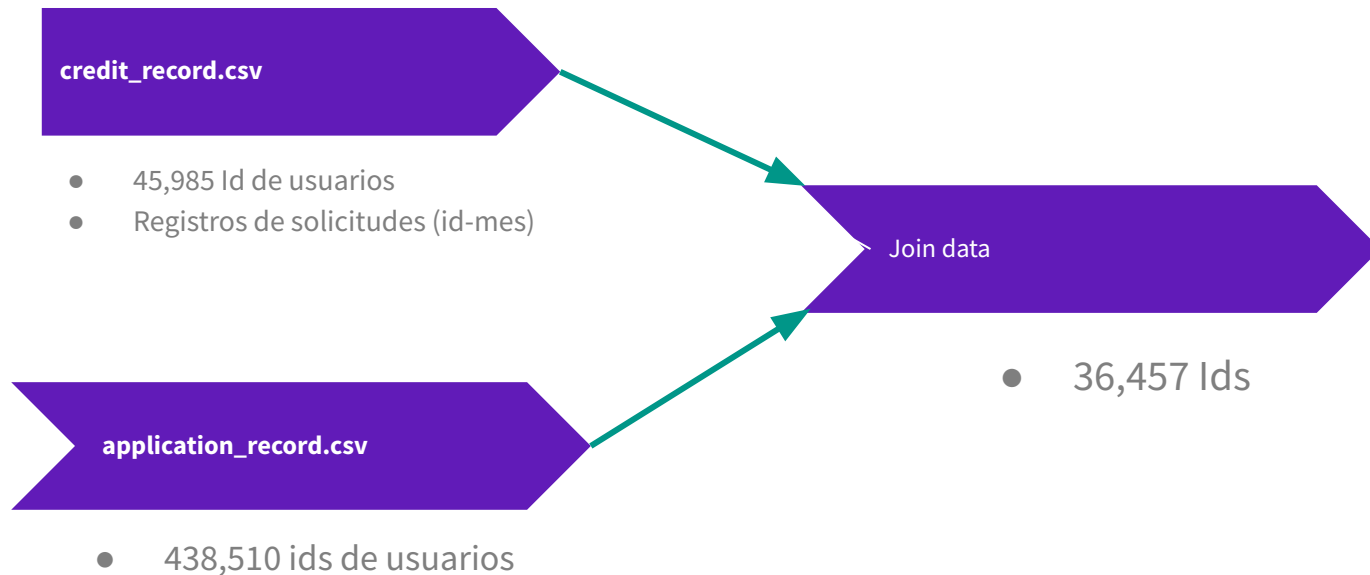
Empresa especializada en crédito digital en productos:

- Crédito personal
- Puntos de venta
- Líneas de crédito

Problem statement

Busca identificar posibles deudores (usuarios que caen en fault) de créditos personales de un DB

Compresión de los datos

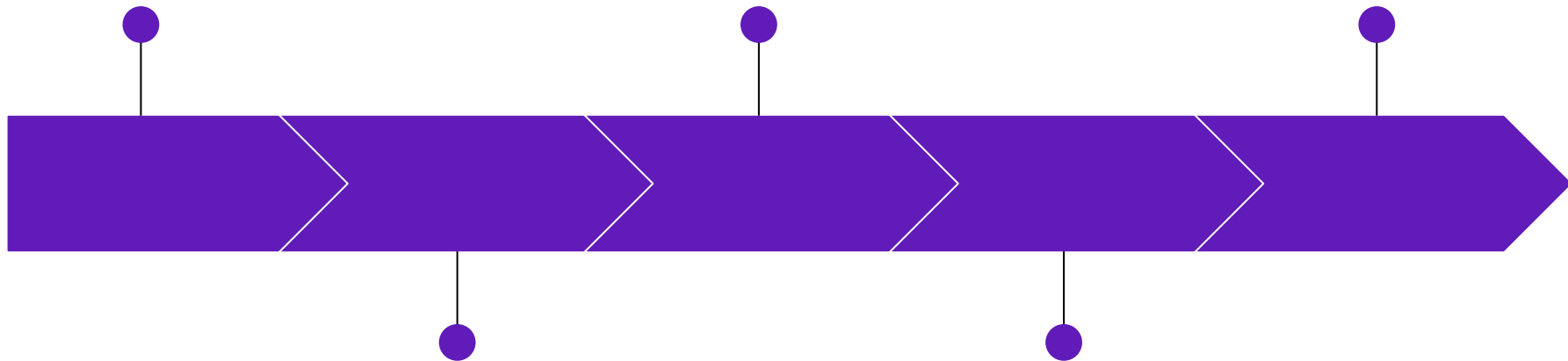


Preparación de los datos

Se elimino la variable *OCCUPATION_TYPE* por tener más de 30% de nulos

Los valores 'X' y 'C' se consideran como buenos (**no fault**) además para los valores *STATUS* >0 se considera fault

Se modifican las variables categóricas y se estandarizan las numéricas



CNT_CHILDREN presenta registros con valores 19 y 14 sin embargo no los consideramos como valores atípicos para quitarlos de la muestra.

Se sobremuestra para balancear el dataset

Modelado

Se elige una maquina de bosques aleatorios: por ser rápida de entrenar y fácil de interpretar

Los resultados de accuracy son: .62 y recall de .58

Interpretación

- Los meses vencidos
- El saldo en el cajero
- La edad del usuario
- Antigüedad en el trabajo

Son las variables que más información aportan al modelo