#### SANZ ESPADA, ELENA

EJERCICIO 1:

Texto

Descripción generada automáticamente

Hay algunas variables que has representado con countplot y deberías representarlas con histplot

Texto

Descripción generada automáticamente

Hay una variable menos correlacionada que delinq.2yrs, puedes quitar más de dos, pero siempre las menos correlacionadas.

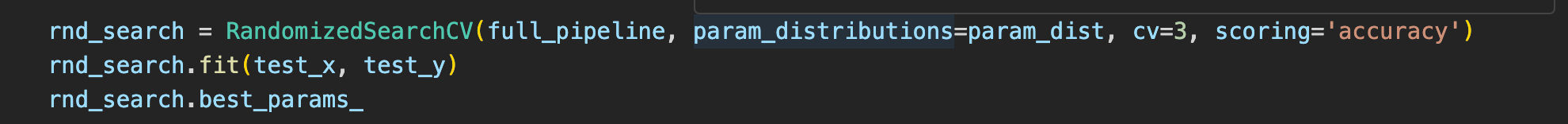
A num\_pipeline le falta por lo menos estandarizar los datos.

Texto

Descripción generada automáticamente

Estás entrenando dos veces, primero con fit y luego con cross\_val\_score y encima en scoring de cross\_val\_cores estas usando como métrica accuracy, que no es nada adecuada para la clasificación binaria y menos todavía si no están bien balanceados los datos.

Para hallar las diferentes métricas no tienes que usar ninguno de estos, tienes que usar cross\_val\_predict



En este caso accuracy no es aprobado

Texto

Descripción generada automáticamente

Para calcular la curva de roc hay que usar para predecir predict\_proba (pero no hace falta que calcules la curva de roc)

Texto

Descripción generada automáticamente

Estos bucles son un desastre.

No pides todos los datos necesarios y a algunos el nombre de la columna no es correcto, en resumen, no funciona

EJERCICIO 2:

Sin hacer.

Hay un ejemplo parecido en el documento 2\_7, básate en él (ejemplo 2, aunque la parte de cargar los datos está en ejemplo1).

EJERCICIO 3

Sin hacer

Explora y divides los datos como siempre

Luego crea el modelo y entrenalos sobre las transformacione hechas:

Texto

Descripción generada automáticamente