* Siempre revisar y y entenderla bien. Si trabajamos categóricos hay que hacer valuecounts para ver si esta desbalanceado.
* A medida vaya modificando X\_train, modificar también X\_test para no tener luego el trabajo de convertirla también.
* Recordar poner class\_weight en balanced siempre que se pueda

Tabla

Descripción generada automáticamente con confianza media

* Siempre establecer una base para empezar que sea un baseline: Quitar columnas categóricas y lanzar un modelo. Preferiblemente un random forest por lo bueno y ligero que es.
* Bueno, bonito y rápido: RandomForest
* Para el Standard Scaler, hacer fit con train y luego transform con train y al final con test.

Flow

Primero hacer un baseline con solo variables numéricas. Luego ir columna por columna categórica viendo si al convertirla mejora el indicador o no.

OJO: Siempre usar el mismo seed (42) para comparar en igualdad de condiciones.

learning\_rate

n\_estimators

max\_depth

min\_samples\_split