Argumentos para optimizar Random Forest con Gridsearch

N\_estimators [50,100,1000]

Max\_depth [3,5,7,9]

Max\_features [0.2,0.4,0.6]

Max\_samples [0.2,0.6]

Min\_sample\_leaf [5,10,15,20]

GridSearch & Pipelines

Gridsearch te ahorra hacer los bucles de forma mas eficiente. Busca los mejores parámetros.

Standard Scaler (SS) = Media y desviación estándar

Pipeline: Serie de pasos que establece funciones en el mismo modelo. Te ahorra transformar toda nueva data que uses para predecir.

Kernels: linear (logistic regression), poly (polynomial), rbf (El mejor, es el de los círculos),

Texto

Descripción generada automáticamente

GridSearch

En regresión utiliza predeterminadamente el negativo MSE. Esto porque en la lógica de GridSearch, mayor valor es mejor. Por lo tanto, las métricas de regresión todas negativas con la excepción de R2, la cual es la única métrica de regresión positiva.

En las categóricas, todas las métricas son positivas.

Pipeline

Introduce funciones junto con los modelos.

RandomSearch

En vez de utilizar todas las muestras como lo haría gridsearch, le das un argumento con numero de iteraciones y toma muestras aleatorias de los rangos dados.

IDEA

Puedo usar el RandomSearch para ver por donde van los tiros respecto a las variables y luego usar un gridsearch para buscar el optimo en esos rangos.

NOTA

En RandomForestClassifier SIEMPRE usar el argumento class\_weight=”balanced”