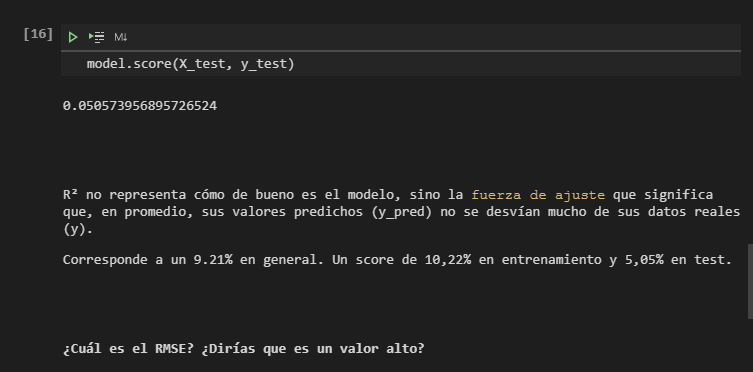
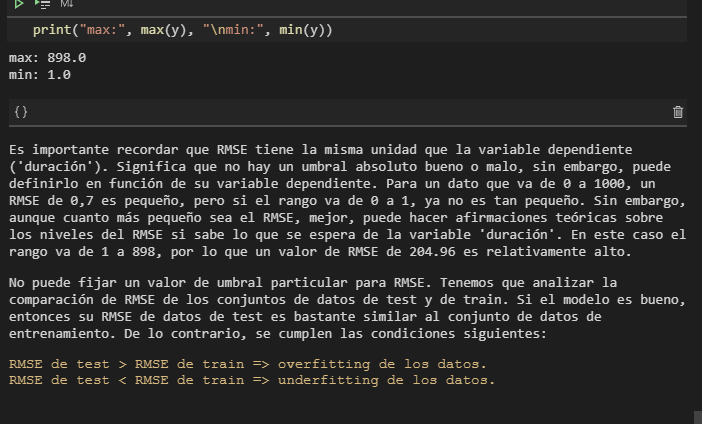
Corrección ejercicios:

-Siempre hay que realizar train y test.

-Imagen:

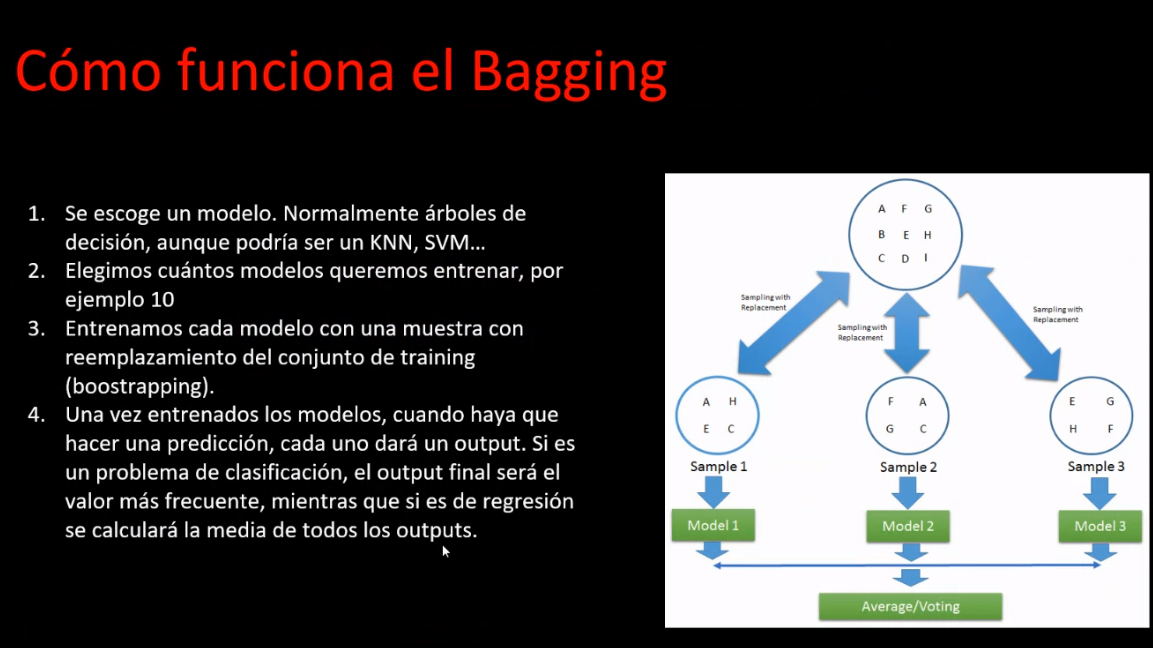


-Imagen:

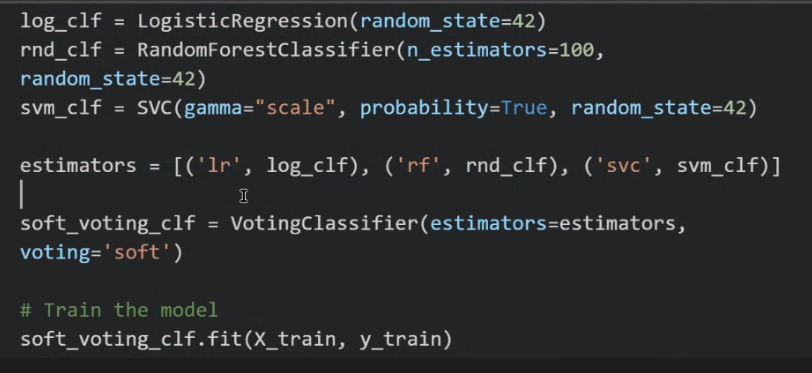


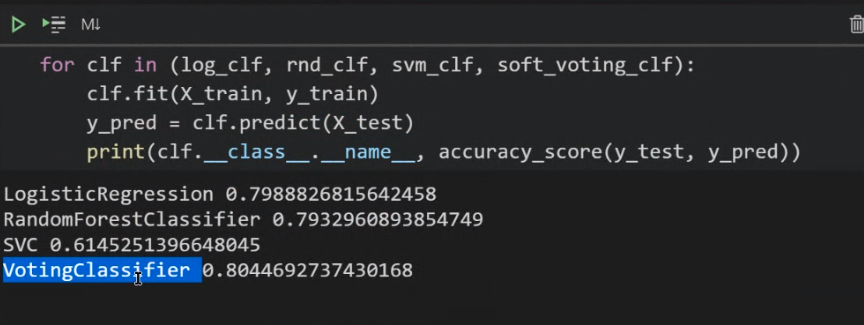
Clase



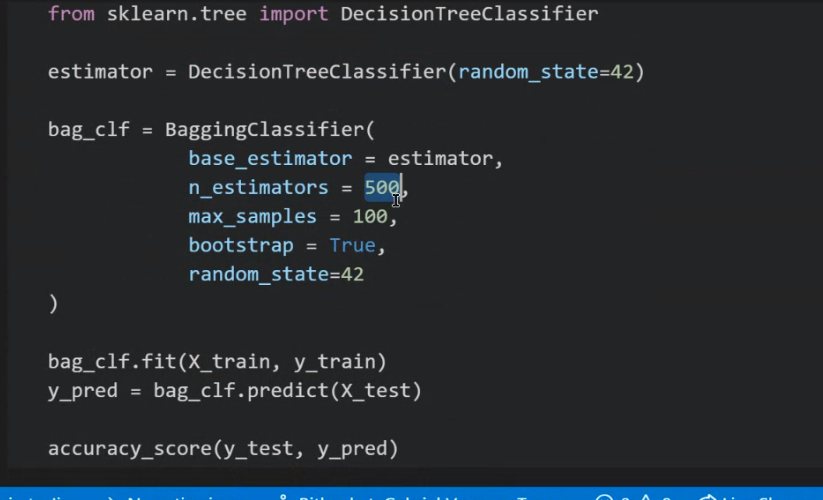


A clasify le das los algoritmos que quieres y te calcula todos los scores con hard o softvoting según quieras:

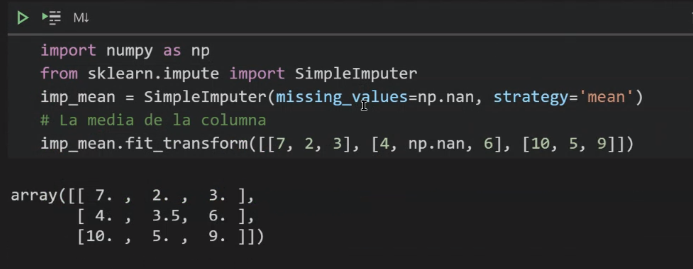


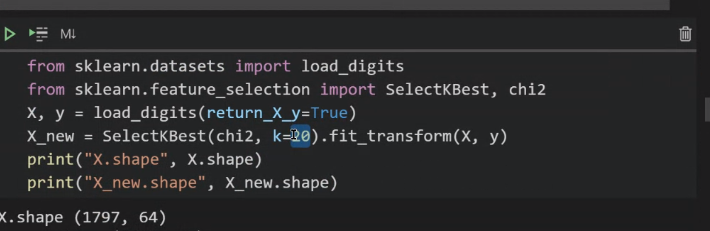


-Bagging se usa para el decisión tree



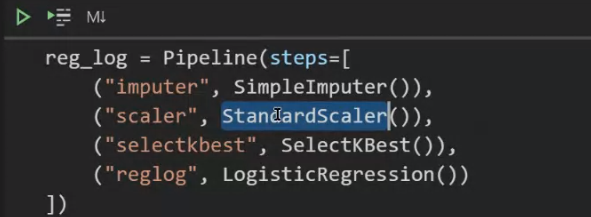
-Adaboost hace que cada uno de los arboles, le otorga algo al siguiente



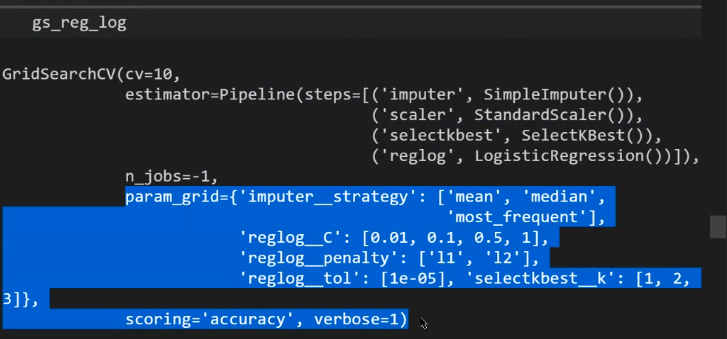
-Chi2 te coje las columnas que mejor se correlacionen en un dataset. K es el numero de columnas que quieres.

GRIDSEARCH

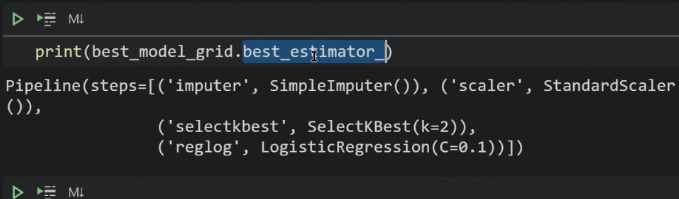
-Pipeline: te hace una secuencia de cambios de features sobre una tabla. Imputer, por ejemplo, transforma los value NaN a la media. Además, podemos ponerle diferentos argumentos para los parámetros de imputer para que haga una secuencia de entrenamientos con distintos argumentos en esos parámetros. En este caso práctico, ha hecho 720 entrenamientos. Con el parámetro verbose, puedo darle especificidad en el mostreo de resultados. Contra más alto sea el número, en más líneas dividirá el resultado, el máximo en este caso es 720. De base es 1.



Con el comando que muestro a continuación, puedo ver como se ha generado el gridsearch

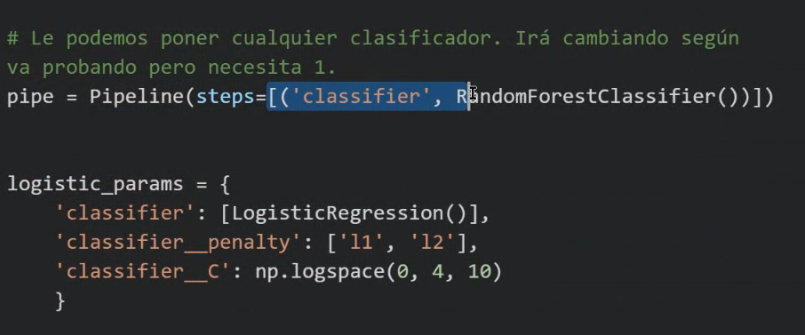


-Ojo! Puedo saber cual es el parámetro que mejor resultado da al ser cambiado

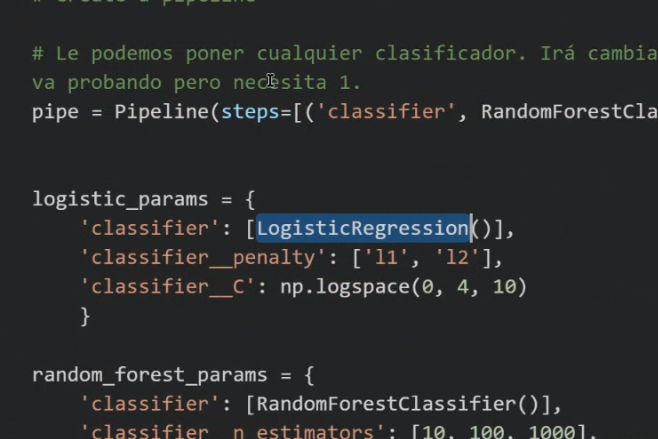


Resumen GRIDSEARCH

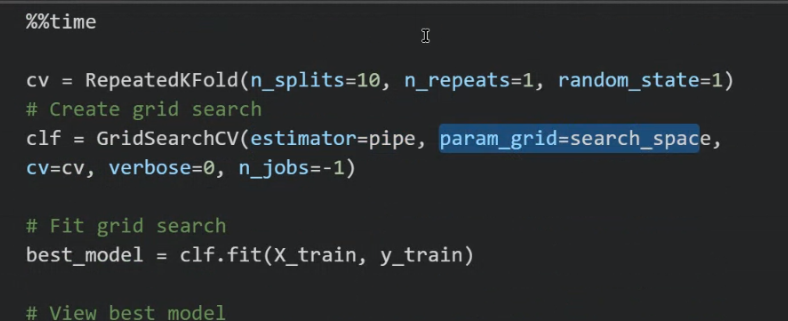
Primero creamos el pipeline



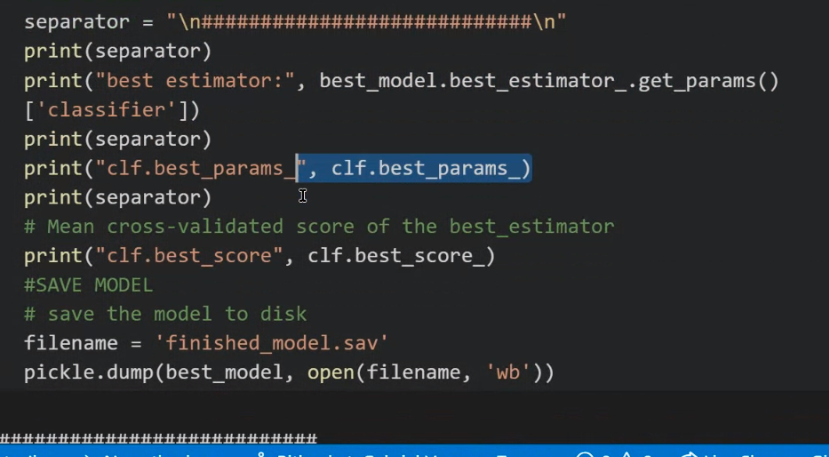
Luego creamos los parámetros del gridsearch



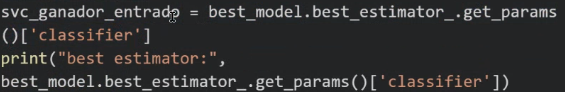
Creamos el gridshearch y le pasámos los parámetros.



Resultados: mejor estimador, mejores parámetros y mejor score. Te guarda un .sav con pickle, que es un modelo entrenado.



Para coger el modelo ganador, es el best estimator, para este caso, es el svc, es decir:



Ese lo puedo volver a entrenar o trabajar con el etc…