Testing and Optimising RandomForest and GradientBoosting Classifiers on diabetes dataset

RandomForest

-Change the max depth (optimum appears to be around max depth = 8)

GradientBoosting

-Change the max depth (optimum around max depth = 2 or 4)

-Change the number of estimators (default appears to be around 100)

-Change the learning rate (default=0.1, optimum=0.18 for max depth =2)

GridSearchCV: model selection tool used to run through a range of given hyperparameters of the model and give the best suited, by ways of cross validation scoring (cross\_val\_score tool). Although this process is computationally expensive and takes time to process.

RandomizedSearchCV: another model selection tool similar to GridSearchCV but with the aim of improving the computational time. Can be used for a larger range of parameters, initially, for the model on the dataset. It randomly selects combinations to run rather than going through all possible combinations. This method saves time and can be used prior to GridSearchCV to find a rough location of the optimum hyperparameters.