



## Lab Automated HPO

### Summary

In this lab you will:

- Try different hyperparameter values and observe the effects on the pipelines

Refer to the demo videos from this lesson for a step-by-step demonstration of how to complete the lab.

### Instructions

1. Examine hyperparameter settings for the [banknote\\_auth](#) experiment
  - a. Open the original version of the notebook which uses the [XGBClassifier](#)
  - b. Find the line of code in the cell titled [Compose](#) Pipeline which appends the [estimator](#).
    - i. You may want to add some line breaks here to make it easier to read this line of code.
    - ii. The RBFOpt procedure is not exposed to the notebook. The values that are set for the hyperparameters for the [XGBClassifier](#) in this notebook are the result of the optimization.
  - c. Compare the hyperparameters set by the RBFOpt procedure to the default hyperparameters for the [XGBClassifier](#). View the documentation here: <https://xgboost.readthedocs.io/en/latest/parameter.html>
  - d. Notice that the hyperparameter [max\\_depth](#) has a default value of [6](#), but has been set to [3](#). Change the [max\\_depth](#) to [6](#) and re-run the notebook to observe changes in the pipeline performance.
  - e. The default [learning\\_rate](#) is [0.3](#), but the hyperparameter has been set to [0.1](#). Try the value [0.3](#) instead and observe how the performance changes.
  - f. If you have time, tune some of the other hyperparameters as you would like and continue to observe how that affects pipeline performance.
2. Optional: Examine the hyperparameter settings for the [parkinsons\\_updrs](#) experiment
  - a. Open the original notebook which uses the [RandomForestRegressor](#) estimator
  - b. See below or view the documentation to compare the hyperparameter settings to the default values: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
  - c. Notice that the default for [n\\_estimators=100](#) and the value is set to [10](#) for the pipeline. Change the value and observe the difference in the pipeline performance. You might have noticed a warning that too few trees were used in the model in order to have reliable [oob\\_score](#) with the setting at [10](#). Notice the warning is gone now
  - d. Also, trying setting [oob\\_score](#) to [False](#) and observe performance.
  - e. If you have time, tune some of the other hyperparameters as you would like and continue to observe how that affects pipeline performance



```
class sklearn.ensemble. RandomForestRegressor(n_estimators=100,  
criterion='mse', max_depth=None, min_samples_split=2, min_samples_leaf=1,  
min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None,  
min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True,  
oob_score=False, n_jobs=None, random_state=None, verbose=0,  
warm_start=False, ccp_alpha=0.0, max_samples=None)
```

[\[source\]](#)

3. Optional: If you have time, explore the capabilities of the RBFOpt library
  - a. <https://github.com/coin-or/rbfopt/blob/master/manual.pdf>
  - b. <https://rbfopt.readthedocs.io/en/latest/>