



Lab Model Selection

Summary

In this lab you will:

- Try different estimators in the notebook and observe changes in the pipeline performance

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

Instructions

1. Open the original notebook version for the [banknote_auth](#) experiment
 - a. Run the notebook and take note of the holdout and cross-validation scores
2. Review the [Compose Pipeline](#) cell of the notebook
 - a. Find the code block under the comment `# assembling cognito_Pipeline`
 - b. Find the line that appends the `estimator` to the pipeline. Notice the `XGBClassifier` estimator is appended here.
 - c. Recall from the Pipeline leaderboard for this AutoAI experiment that the Random Forest Classifier also performed well for this dataset. Replace the `XGBClassifier` with the `RandomForestClassifier` from scikit-learn. Use `sklearn.ensemble.RandomForestClassifier()` with the default parameters.
 - d. Re-run the notebook and notice the output for the holdout and cross-validation scores. Save this version of the notebook.
 - e. Replace the Random Forest classifier with `sklearn.svm.SVC()` and re-run the notebook to observe any changes.
3. Save this version of the notebook, but revert back to the original notebook version.
4. Optional: Open the original notebook version for the [parkinsons_updrs](#) experiment. Try the following models using the same steps as above and compare performance. We'll examine hyperparameter optimization in later labs. You can use the defaults for the estimators here. If you have time, you can try tuning some parameters.
 - a. `sklearn.tree.DecisionTreeRegressor()`
 - i. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html#sklearn.tree.DecisionTreeRegressor>
 - b. `sklearn.linear_model.LinearRegression()`
 - i. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html