Project-Report #5

Here are detailed instruction for Project-Report #5:

***Note: You will need to report the running time for each classifier for LOOCV and 5-fold CV. If your dataset is large, pay extra attention to the running time for LOOCV as it may take up to hours to complete.***

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Part-1**: Use dataset from Project #3 for gender classification:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

First, consider Decision Tree for classification for your dataset.

Next, consider all three classifiers we learned from Chapter 8 so far: Bagging, RandomForest, and Boosting.

Step1, apply 5-fold cross validation to all three classifiers of Bagging, RandomForest, and Boosting. Find the overall prediction accuracy for each classifier.

**Note: you should KEEP the same random seed for all partition of 5-fold CV for all classifiers forthmentioned, so you will be able to compare *apple to apple* later on.**

Step2: Make a side-to-side Boxplot for the prediction accuracy for all seven classifiers on 5-fold CV: Logistic Regression, LDA, QDA, K-NN, Bagging, RandomForest, and Boosting.

Step3: Summarize your outputs into a large table for all prediction accuracies with standard errors for 5-fold CV, for all seven classifiers. **Summarize your outputs into a large table for all prediction accuracies & standard errors for 5-fold CV, for all forth-mentioned classifiers.**

Step4: Apply LOOCV for all seven classifiers of Logistic Regression, LDA, QDA, K-NN, Bagging, RandomForest, and Boosting.

Step5: Summarize your outputs into a large table for all prediction accuracies with standard errors for LOOCV, for all seven classifiers.

Step6: Make a side-to-side Boxplot for the prediction accuracy for all four classifier on LOOCV.

**Summarize your outputs into a large table for all prediction accuracies & standard errors for LOOCV, for all forth-mentioned classifiers.**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Part 2: Use dataset from Project #3 for age estimation:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Please consider all possible methods of Decision Tree, Bagging, RandomForest, and Boosting for age estimation.

Find you MAE (Mean Absolute Estimate).

Draw conclusions respectively.