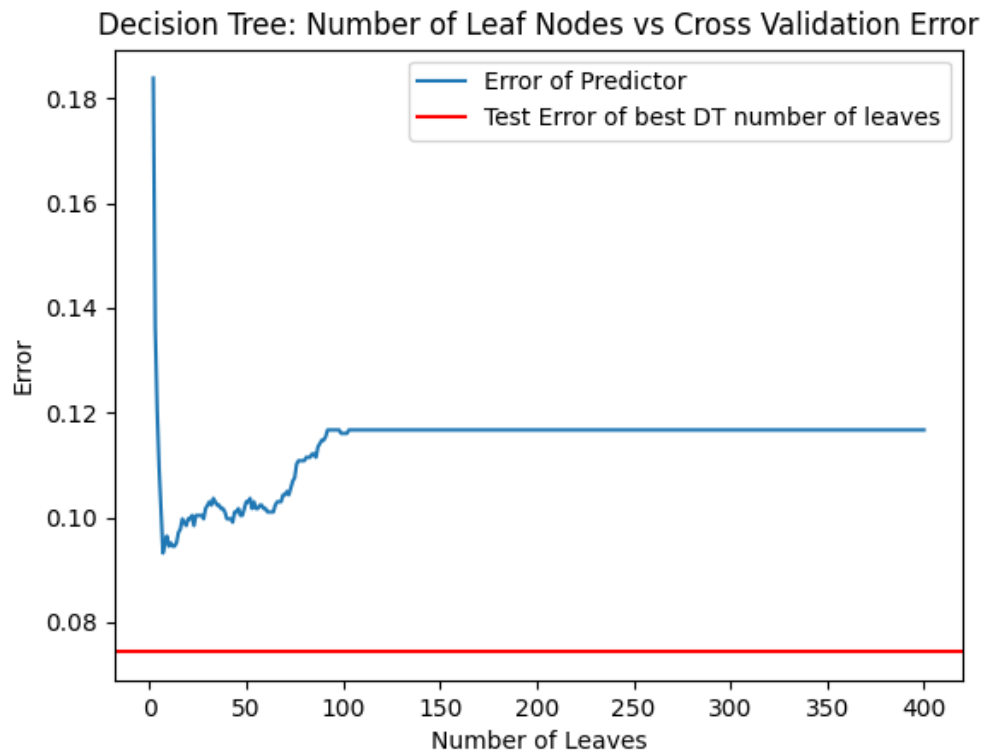
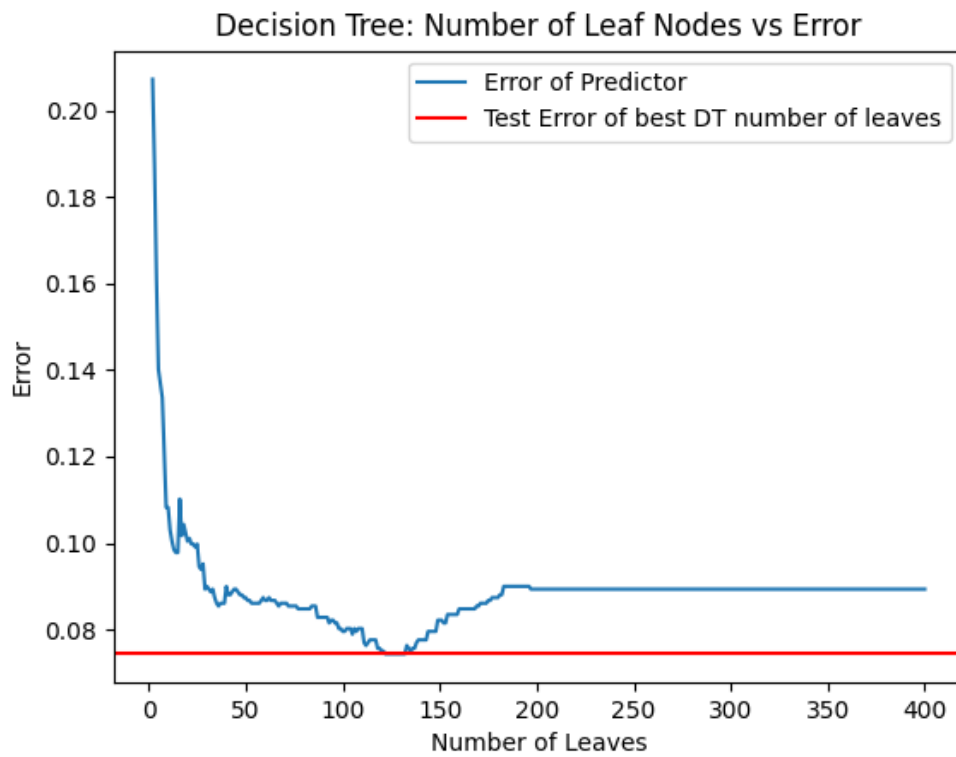


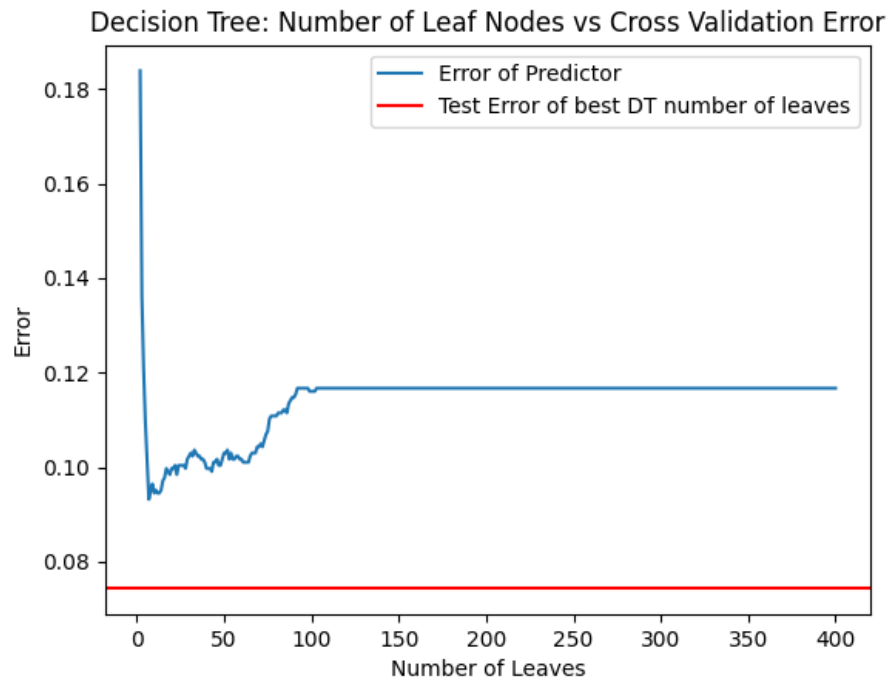
McMaster University

4SL3 Assignment 4

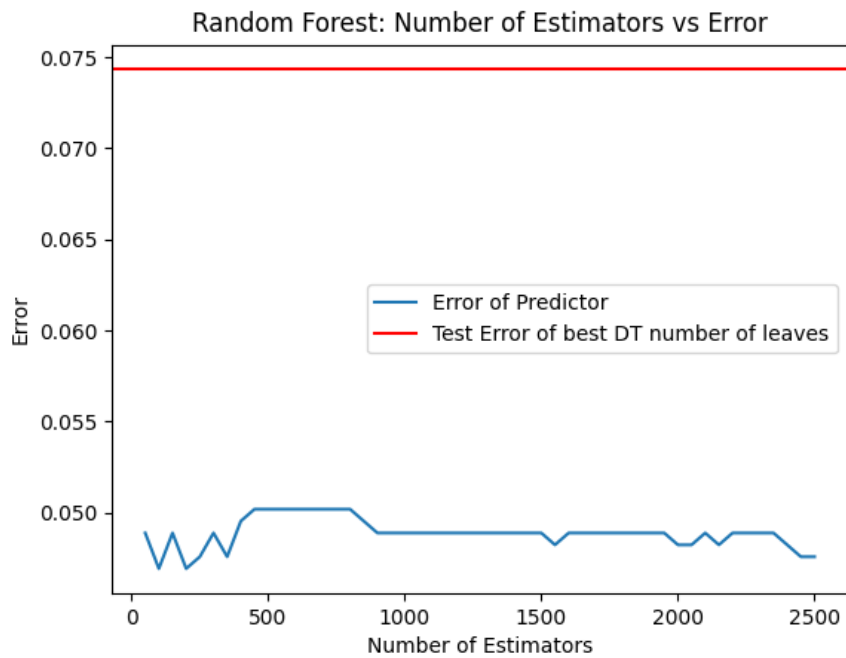
Uday Sharma - 400139248
11-24-2021



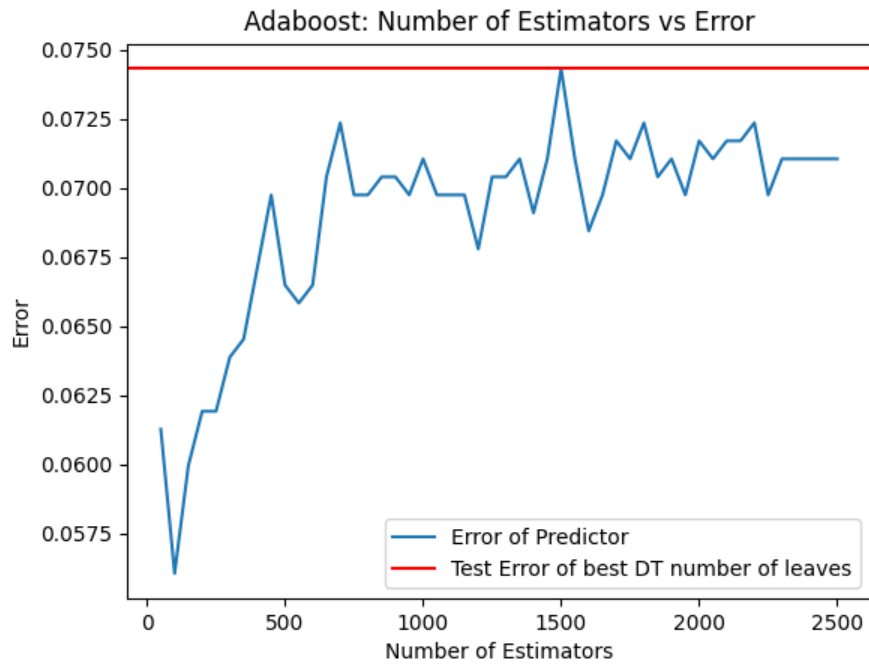
(The best number of leaf nodes were determined as 132, where we see the smallest test error.)



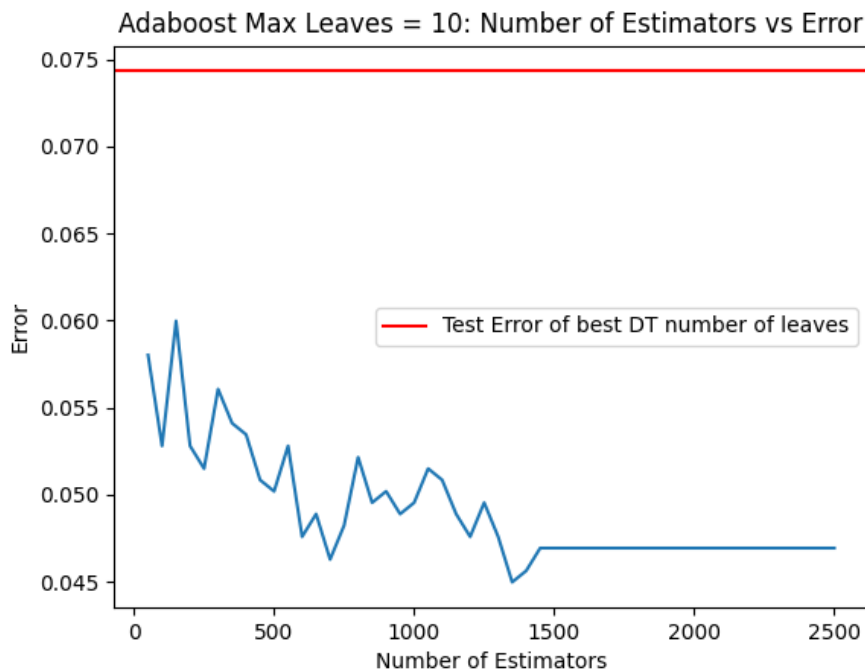
(The best number of estimators were determined as 50, where we see the minimum test error.)



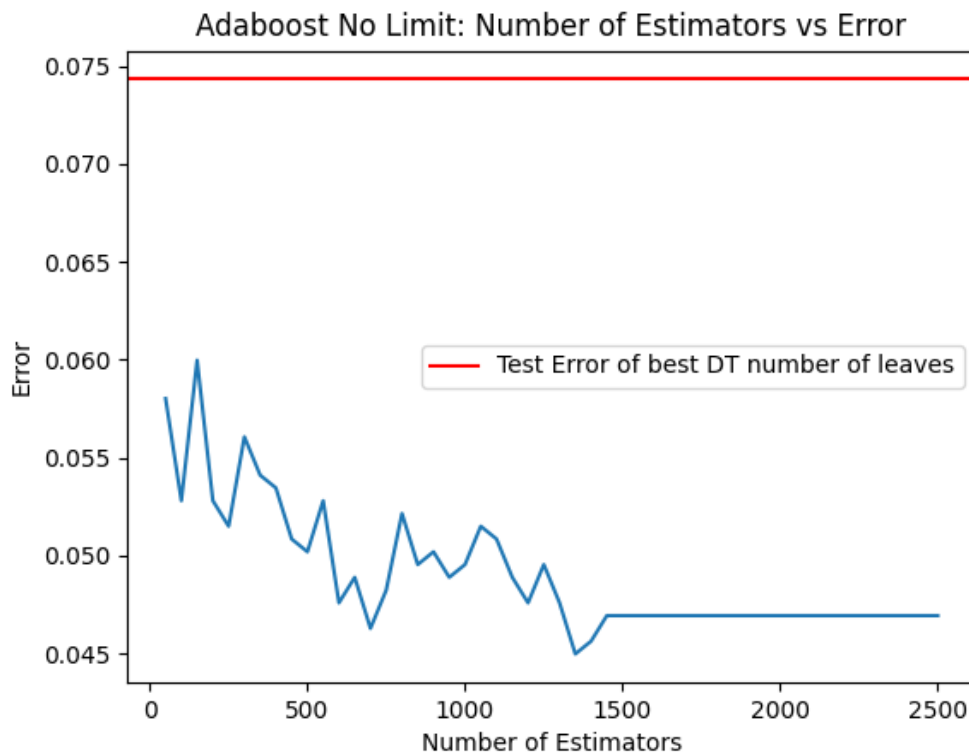
(The best number of estimators were determined as 100, where we see the minimum test error.)



(The best number of estimators were determined as 100, where we see the minimum test error.)



(The best number of estimators were determined as 1350, where we see the minimum test error.)



(The best number of estimators were determined as 1350, where we see the minimum test error.)

In general, the models are relatively close in performance, except for the decision tree classifier being the least accurate (however, it is the most performant). It is straightforward to determine the best type of classifier given certain constraints, however the 2 final Adaboost classifiers are a little difficult to differentiate between.

As we can see with the Adaboost classifiers, whether a limit is imposed or not, the result stays the same, and the optimal number of classifiers remains the same, which goes to show that the optimal number of classifiers for the base Decision Tree classifiers in the Adaboost implementation exist before you get to a maximum of 10 leaf nodes. However, since we have determined that, we can also deduce that the Adaboost with a max of 10 leaf nodes, is the better one between the two, by imposing constraints, we will be benefitting the asymptotic runtime of the algorithm as each iteration would have a stop condition.

Another key takeaway is the performance of the Random Forest classifier. We notice that the test error is very consistent among the 5 models. The best choice of model would really depend on the constraints, how much overhead are we okay with? At 1350 Adaboost estimators, we see the true minimum test error across all our models, however, we needed to test up until 500 estimators to reach a range that is lower than what is present in our Random Forrest.

If we wanted to optimize for delivery, the best choice would probably be the Random Forest, if we were optimizing for accuracy, 1350 Estimators on an Adaboost model with Decision trees as classifiers would be the best choice.