Efficiency Report: Group 12

Background:

Within random forest algorithms, one of the major aspects which decreases efficiency is the time it takes to develop trees. Each tree takes a set amount of time to develop, the more trees the more time it takes. Unfortunately, the current barrier is that fewer trees means less accuracy, and thus a compromise must be struck between the two, decreasing efficiency.

Our Proposed Solution:

The solution proposed in the previous paper, “Accuracy Report: Group 12,” presents our proposed change to the process of random forest algorithms. Beyond having a benefit to accuracy, this method can also benefit efficiency. Because each decision tree takes time to develop, a random forest algorithm takes longer and is less efficient the more trees are required for good accuracy and to utilize a significant portion of the data set. Therefore:

* + - * + Because our proposed solution uses all data within a data set, accuracy is significantly increased.
        + Therefore, in theory, smaller data sets can be used in order to gain the same level of accuracy.
        + This allows for a balance to be struck between accuracy and efficiency
        + Smaller data sets/fewer trees can be used while maintaining a slightly higher level of accuracy than could previously be achieved without our modification.

Conclusions:

The process defined within “Accuracy Report: Group 12” has the potential to allow for a balance to be struck between efficiency and accuracy. With the use of all data and thus the ability for greater accuracy, machine learning programmers could decrease the number of trees produced and aggregated while still maintaining similar or higher levels of accuracy than they would of had without this method. This opens up this field of research to more possibilities without huge data sets as well as streamlining the process overall for better use of data.