How to Improve the Efficiency of the Random Forest Method

Efficiency refers to how complex the system is. The complexity can be improved in several ways. The efficiency of the aggregate can be improved through reducing overlap of questions conducted by each tree i.e. become more independent. This causes each decision tree to be more uniquely defined, and removes redundant sections of code. Rearranging the order of the decision trees will also allow for identification of surplus trees, which can also be removed to reduce complexity.

The efficiency of the system can be improved by adding parameters that limit the number of branches and leaves allowed, so we decided that if we were able to use python/a similar program, that we would improve the method by changing from the decision tree method to the k-neighbour method. A drawback on this modification would be that the accuracy of the aggregate would decrease. This could be counteracted however through increasing the number of neighbours, to take a larger sample, however only to a certain extent. Eventually the number of neighbours becomes so large that it takes values that are irrelevant, and will increase the amount of data being used. Thus, the value of “k” in this instance must be such that it has improved accuracy without compromising the gained efficiency.