5. RF Results

This section was written by Matthew Adler:

Random Forest Results:

  If the dataset is large enough, a random forest is a good choice. But, there is a caveat. In our analysis of time taken, the more trees used to get a more accurate model(more trees = better accuracy), then it takes much longer. Furthermore, if you don't know the ideal number of trees i.e. the limit reached for accuracy (much like the minimum value reached in gradient descent), it can be quite time consuming. Also, in our findings, it can be argued that if you're under a time constraint and improvement in accuracy isn't that necessary then it might be better to simply leave the model with less trees because our results in going from say 40 trees to 80 trees was just a two percent accuracy improvement.

Accuracies given different number of trees:

```
1 tree score 0.4071388733965421
10 trees score 0.673452314556609
20 trees score 0.7438650306748467
40 trees score 0.7858337981037368
80 trees score 0.8074456218627998
160 trees score 0.8216675962074735
320 trees score 0.8234802007808143
640 trees score 0.8290574456218628
```