ML Decision Tree Report

Implementation part:

The top3 is (1) WaitEstimate:16.0 -> (2) Patrons:1.5 -> (3) Hungry: 0.5
I use DFS to construct my tree. First, traverse the left subtree and record the counterpart of threshold. While we reach an end of path, we could decide whether T or F. Then jump out pf current path and we could traverse the other side.

Choose the best split among all data recursively until reach leaf or other constraint!

I think the most difficult part is the concept of recursion, while we are using recursion, we must consider about the structure carefully! In this project, we may also need take the zero sample under any circumstance! I tried to fix my bug in code and check out it's eventually the zero element which is involve in entropy and the decision of T or F!

Prediction part:

At this part, we are asked to predict the death of subject_id! As you see, there is 83 attributes, which is used to classify the data! With such a big data, we might face overfitting issue. So, we need to pick out the most important data and use it to classify! By observation, we found out that some attributes such as subject_id is useless in decision branching! If you use that kind of attribute to classify data, it might get not bad entropy, however it turned out to be a bad performance.

After we did such preprocessing of data, we might get better result!