Problem Set 2

1. We handled missing attributes in the examples by using the mode of that attribute and inserting it.
2. Insert
3. Insert
4. We implemented pruning with a reduced error algorithm.
5. Insert
6. Insert
7. Insert
8. Insert
9. We think that the pruned tree will perform better because it will reduce overfitting and improve the predictive accuracy.
10. While we all coded/worked together in a room. Jakub worked on ID3, Pruning, and Predictions, Jack worked on Graph and Tait worked on DNF printing. We all contributed to each file, but I have listed the people that contributed the most to that specific file.
11. Bonus: Information gain ratio acts as a way to prevent overfitting as compared to the information gain. Our accuracy would be higher with information gain on the training, but would be lower with the actual data. Also, by only looking at a limited step portion of our data, we increase speed, but sacrifice accuracy. These were good model selections assuming that we were able to sacrifice accuracy for speed.