CS 412 Assignment 3 Report

By: Joshua Harris

NetID: jharri39

UIN: 661654036

Brief:

For the Assignment I took the Decision Tree route. I unfortunately did not have time to do the Random Forrest implementation and would ask that all missing data from the Random Forrest not count against me when I do calculations on the data as a whole.

The purpose behind my classification framework is to attempt to sort the data by attribute values, one attribute at a time, such that the attribute that is splitting the data yields the most gain.

Model Evaluations:

Accuracy:

Chess: 70.74%

Led24: 88.13%

Mushroom: 62.71%

Error Rate:

Chess: 29.26%

Led24: 11.87%

Mushroom: 37.29%

Sensitivity:

Chess: 45.89%

Led24:88.38%

Mushroom:99.35%

Specificity

Chess: 97.83%

Led24: 87.89%

Mushroom: 21.73%

Precision:

Chess: 95.83%

Led24: 87.30%

Mushroom: 58.67%

F-1 Score:

Chess: 62.06%

Led24: 87 84%.

Mushroom: 73.78%

F\beta score (\beta = .5):

Chess: 78.70%

Led24: 87.51%

Mushroom: 63.90%

F\beta score (\beta = 2):

Chess: 51.22%

Led24: 88.16%

Mushroom: 87.25%

Parameters:

I never encountered many parameters in my implementation. My decision tree is filled out until it meets one of three base cases:

1. All the elements fall into the new classification
2. There is no Gain Ratio change(ie gain ratio = 0)
3. There are no more attributes to calculate the gain ratio on since they have all been chosen to be the splitting attribute.

I cannot talk about parameters for Ensemble method as I did not have time to complete the implementation.

Conclusion:

The decision tree implementation (assuming that it is correct) seems to be a decent choice for classification frameworks. The idea is simple by splitting on attributes that are going to have the greatest effect. It seems to work better the larger the set of data is. Also it seems like it would be a good fit with an Ensemble method. This is mainly because as the data is split and classified it is easy to run an ensemble method on the different levels to further classify the data. Although once again I am not positive since I have no data on an Ensemble method.