CSCI N317 Spring 2018 Lab 6

**Decision Tree Classification**

Use Weka to perform a decision tree classification. You can use an arff file provided in the Weka dataset or an arff file from a different data source, just not the files used in the demo videos.

Answer the following questions in a WORD document. Use screen catches in each of the following items together with your answers.

1. What is the data about? (10’)
2. Run a classification with an unpruned tree setting. Show the tree structure. What are the other settings you used in the process, e.g. minimum leaf size, cross-validation, etc. (30’)
3. What is the error rate? (10’)
4. Explain the confusion matrix.(10’)
5. Do at least three more runs, such as a pruned version, a version with different setting on leaf size limit, and different setting on percentage split. Explain the differences between these models. (20’)
6. Which model is finally chosen and why? (10’)
7. Create a new record with arbitrary values and use the chosen model to predict the class for this new record. What is the new record and what is the result?(10’)

Submit the WORD file on Canvas. Your file must be in the format of *lastname*\_lab6.docx.