HW 5: Decision Trees

In this exercise, we will be using Weka to build a prediction model using a telecommunications data set looking at customer churn.

Import telecoms_churn.arff into Weka, and answer the following questions:

1. Data Questions
   a. Which attributes should be included in the model, and which (if any) attributes should be eliminated from the model and why?
   b. Some of the might need to be transformed. Which attributes do you feel should be transformation and why? Which transformation do you suggest and why?

Next, construct a Decision Trees Model using the J48 Algorithm. Remove any attributes you believe should be removed, and transform the required attributes using the techniques outlined in 1.B

Question:

2. Data Model Questions
   a. With no tuning, report the Correctly Classified Instances and the % of instances
   b. Please describe what each element in the Confusion Matric means in this model
   c. Include screen shot of model output.

3. Data Model Tuning
   a. Experiment with your transformations. Please try 5 different bin sizes, and rerun the model for each. Do different bin sizes effect the model? If so, describe how.
      i. Hint: consider accuracy and tree construction
   b. Experiment with Pruning. Using the decision-tree-Weka.pptx as a guide re-run you model adjusting each of the parameters below:
      i. BinarySplit : True or False
      ii. **unpruned**": True or False
      iii. **ConfidenceFactor** range
   c. Describe what impact, if any, these tuning factors had on model accuracy and tree construction
4. Optimal Model
   a. Using all of the variables defined above; please define the best model accuracy you can achieve.
   b. Provide screen shot for optimal model