

# Assignment: Decisions Trees and Classification

**Due Date:** Thursday July 31<sup>st</sup>

## 1. Targeted Marketing Campaign

In this problem we will use historical data from past customer responses to build a classification model. In class next week we will apply the trained model to a new set of prospects to whom we may want extend an offer for a PEP. Rather than doing a mass marketing campaign to all new prospects, we would like to target those that are likely to respond positively to our offer (according to our classification model).

There is one data sets available, comma delimited, and the first row contains the field names):

- [bank-data.csv](#) - Preclassified training data Set for Building a Model
- A. Open the .CSV file and inspect the attributes, is there one or more that you feel should be dropped? If so, do it now.
- B. Save the file in .ARFF format
- C. Load the .ARFF formatted data file
- D. Use **WEKA** or software package of your choice to create a "C4.5" classification model based on the pre-classified training data. In WEKA, the C4.5 algorithm is implemented by "**weka.classifiers.trees.J48**".
  1. Use 10-fold cross-validation to evaluate your model accuracy.
  2. Spend time experimenting with different setting for the tree, such as pruning, binary branching, etc. to see if you can improve the performance of your model.
  3. Review the structure of the tree, note the more important features and information that Weka is providing about the structure of the tree.
  4. Examine the performance measures that Weka provides, review lecture notes if you are unsure what these measure are (we did not discuss Kappa).
  5. Generate the ROC curve for your final model, left click your best model, select "View Threshold Curve".
  6. Record the final decision tree and model accuracy statistics obtained from your model. Be sure to indicate the parameters you use in building your classification model. You can save the statistics and results by right-clicking the last result set in the "Result list" window and selecting "Save result buffer."

7. You should also generate and create a screen shot of your tree by selecting the "Visualize tree" command from the same menu.
- E. And for you overachievers, you know that you want to do this, can you improve the performance of your model by using a Random Forrest? If so, how many trees in the Random Forrest?

You should provide the decision tree together with the accuracy results from the cross-validation as part of your submission.

