**COMP 8901 – Assignment #5 Due date: Nov 12 (midnight)**

**Learning Decision Trees from Training Examples**

In this assignment, you will implement the ID3 algorithm for learning decision trees from data. On D2L, you can find a dataset for your use. It involves predicting who survived the sinking of the Titanic with only a few features. It has been divided into a training set and a testing set.

**Your task**

You must write a program that takes two files as input: training data and testing data. This can be a command line application launched as follows:

>BuildAndTestDecisionTree <trainsetFilename> <testsetFilename>

Or it can have some other suitable interface.

When the program is run, it should do the following:

1. Use the training set to build a decision tree using ID3.
2. Print out this decision tree (using simple ASCII text, or something prettier if you like)
3. Categorize the TESTING SET using the induced tree, reporting which examples were INCORRECTLY classified, as well as the FRACTION that were incorrectly classified. Just print out the NAMES of the examples incorrectly classified.

Note… you should use the provided data for testing, and as a template. I will also run your program with an alternative data set (the format will be the same, but the feature set will be totally different). The important point is that blank lines and commented lines are ignored. But there can be more features with more values, as long as they are provided in the same format.

**Submission**

You must submit all code, as well as an executable version of your program. In addition, you must submit a 1-2 page written document that addresses the following questions:

1. Does the learned tree for the test data make sense? i.e. does it demonstrate a general pattern, or does it just memorize examples?
2. If there are any errors on the test data, explain why they have occurred.
3. Give a ***specific suggestion***for an extra processing step that could be added to improve the learned tree. (i.e. just saying “pruning” is not enough… you need to suggest an approach or method to improve the tree in an automated way)

Zip all files up and put on share in by the due date.