**Lab Instructions – 2**

Objective: Building Univariate Decision Tree and understanding the effect of various choices.

Download the “Seeds” dataset from the following link: <https://archive.ics.uci.edu/ml/datasets/seeds>

It is a tab separated text file that contains numeric values for seeds of three different varieties of wheat with 7 features. Thus we have three classes and seven features in our dataset.

The purpose of this exercise is that you should be able to observe the effect of the choices that are available when using ID3 for building univariate trees.

1. Choice regarding training set: Choose 70% of samples randomly for training. Build the tree and write down the corresponding decision rules. Use Information Gain for finding the split point. Now choose some other 70% samples for training and again build the tree (using IG) and the corresponding rules. Did the rules change? Why? What was the accuracy for the two cases? You can check the accuracy by predicting the class of the remaining 30% of samples and then checking versus the actual class.
2. Choice regarding impurity measure: Repeat the task of (a) above but use Ginni Index for finding the split point instead of IG. Did the rules change from (a)? Why? Which impurity measure gave a higher accuracy?

Prepare a report with the rules, the accuracies and your explanation regarding these in a word file.