**Lab Sheet 6**

**Decision Trees**

Given the car dataset and its attribute description perform the following:

1. Read the data and split data into training and test set.
2. Define a function to calculate the entropy of a dataset, S, based on the target variable.

*Entropy(S) =*

Where *pi* is the probability of class *i*

1. Consider *‘buying’* attribute of car dataset.

Find unique values in the dataset for ‘*buying’* attribute.

Find expected information gain when ‘buying’ attribute becomes known

Where Sv is the subset of dataset with v value in buying attribute.

1. Repeat Q.3 for all attributes and find the attribute with maximum gain.
2. Use the predefined function to do the training using decision tree.

Follow : <http://dataaspirant.com/2017/02/01/decision-tree-algorithm-python-with-scikit-learn/>

1. Compare the results of Decision tree with kNN and Logistic regression.