School of Computer Science and Engineering, VIT Chennai.

CSE4020 Machine Learning

**Lab-7 Decision Tree Classifier**

Faculty : Dr. R. Jothi

USE appropriate Scikit-learn libraries

Submit a zip file comprising

* document containing your responses to all of the questions in this assignment.
* your python code (Jupyter notebook)

1. Consider PlayTennis Dataset (see attached csv file) that is used for predicting whether a tennis game is played in the given weather conditions or not. Here the weather conditions are described by features outlook, temperature, humidity, play and wind. The target is play with two class labels *Yes* and *No*.
2. Compute information gain for all the attributes and display them.
3. Find which attribute will become the root node of the decision tree.
4. Scikit DecisionTreeClassifier
   * Train using DecisionTreeClassifier using tennis dataset;
   * Classify the test sample <Rain, Cool, High,Weak>.
   * Draw the decision tree for a max depth 2

graph = Source( tree.export\_graphviz(id3\_model, out\_file = None))

SVG(graph.pipe(format='svg'))

Check whether root node that you identified is same as the one returned by Scikit DecisionTreeClassifier.

1. **Predict heart disease**
2. Apply decision tree classifier on heart disease dataset and report the accuracy.
3. Try to change the following hyperparameters of the decision tree model and report the change in the accuracy
   * 1. Splitting Criterion from entropy to gini
     2. max\_depth
     3. min\_samples\_leaf