# Department of Computing

# CS370: Artificial Intelligence

# Class: BSCS-8C

# Lab 07: Random Forest Classifier

# Date: 09-04-2021

# Time: 10:00-1:00

# Instructor: Dr. Hashir Kiani

# Lab Engineer: Ms Rabbia Hassan

# Lab 07: Random Forests

**Introduction:**

Bagging identification trees often ends up giving us K trees that are highly correlated. If there is one attribute with a very low disorder, it is likely to be the root of most of the trees. The random forest model is a form of bagging in which we take extra steps to make the ensemble of trees more diverse. The key idea is to randomly vary the attribute choices rather than the training examples. At each split point in constructing the tree we select a random sampling of attributes and then compute which of those have the lowest disorder. If there are 𝑑 attributes, a common default choice is that each split randomly picks √𝑑 attributes to consider for classification problems. Random forests reduce the problem of overfitting.

**Lab Task:**

In this lab, you are supposed to implement the Random Forest Classifier. Data in excel files (both the training and test sets) are uploaded on LMS. In the said training and test data files, each row contains data about one instance of a plant category where four predictors/attributes are recorded for each plant (namely, leaf length, leaf width, flower length, and flower width), while “plant” is the target class which could be any one of the following at a time: “Arctica” or “Harlequin” or “Caroliniana”.

You are supposed to develop random forests with

a. 100 trees

b. 300 trees

c. 500 trees

Figure out how the results vary on the test data (How results vary by increasing the number of trees). Please submit your code and screenshots of output.

Note: Scikit-learn might be a useful resource during implementation, feel free to use this-or any other suitable library.

This link can be very useful.

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>