# Assignment 6 - Decision Tree

*The purpose of this assignment is to use R/Python to practice Decision Trees*

This assignment provides you with an opportunity to demonstrate the achievement of the following course learning outcomes:

* Understand and apply Python programming language
* Understand and apply Decision Trees
* Understand how hyper-parameter tuning can be applied to decision trees

## Key Information

* **Type:** *Individual*
* **Weight:** 6.25%
* **Delivery:** Course website upload
* **Due Date:** End of lab session

## Expectations

You are expected to complete this assignment individually.

Respect for academic integrity is crucial to your success. Make sure you understand what constitutes acts of academic dishonesty in the page: [What is Academic Dishonesty?](http://mcmaster.ca/academicintegrity/students/whatis.html)

## Instructions

*Using R/Python, you are to complete the following questions.* ***Please submit your answers (CODE USED AND OUTPUT) as PDF* *files to the course website submission folder. All questions must be answered!***

*Part I*

*Upload Titanic dataset*

*Define Survived column as TARGET variable*

*Select* ***ALL*** *features that can be predictive of the survival status*

*Transform* ***ALL Possible***  *categorical features with Dummy values*

*Import Decision Tree function, train and test function from sklearn library*

*Apply Decision Tree on the split train/test dataset and report the accuracy*

*Part II*

*Use the following code to perform hyper-parameter tuning for the decision tree.*

*from sklearn import tree*

*clf = tree.DecisionTreeClassifier()*

*parameters = {'max\_features': ['log2', 'sqrt','auto'],*

*'criterion': ['entropy', 'gini'],*

*'max\_depth': [2, 3, 5, 10],*

*'min\_samples\_split': [2, 3, 5],*

*'min\_samples\_leaf': [1,5,8]*

*}*

*grid\_obj = GridSearchCV(clf, parameters, cv=5)*

*grid\_obj = grid\_obj.fit(X\_train, y\_train)*

*clf = grid\_obj.best\_estimator\_*

*clf.fit(X\_train, y\_train)*

*Change and expand the possible parameters and the number of cv for the hyper-parameter tuning*

*Report feature importance using “***feature\_importances\_” attribute**

*Report the best features according to the decision tree using* **n\_features\_ attribute**

*Compute your model’s confusion matrix, accuracy, and f1-score.*

## Rubric

To achieve full marks on this assignment, you must have answered all questions above correctly with code submitted that has no errors.