# *Advanced Topics in Computer Science II (420-G50-HR)*

# *Assignment 2 – Machine Learning*

Date assigned: April 29, 2022

Date due: May 17, 2022

**Learning Objectives**

Upon successful completion of this assignment, the student will be able to:

* Perform new algorithms for Supervised Machine Learning

To do:

1. The folder Titanic contains two files about actual passengers on the Titanic. The traintest.csv file has one extra column…Survived where 1 indicates the person survived the sinking and a 0 indicates they did not.
2. You are going to create 3 Supervised Machine Learning Classification models to predict if the people listed in the predict.csv dataset survive the sinking or not. These three models MUST use different algorithms. One can be Logistic Regression or KNN; one MUST be Random Forest and the other can be anything you like that is not one of the ones already mentioned.
3. Follow the steps of Machine Learning as discussed in class.
4. Perform a solid data analysis and provide comments in the Python file as to what the analysis shows.
   1. Some of the fields are not relevant and can be deleted.
   2. Some of the fields are missing data and you must decide what to do about that
   3. Some of the fields are very relevant and skew the data
   4. Some of the fields are string and must be converted to numbers
5. Train and test each of the models you create with the traintest data. Then use the predict data to predict whether or not the person survived. The last file (tested.csv) contains the matching records to predict along with the survived column.
6. Make sure each of your models reaches at least the 70% minimal threshold. The higher the better, obviously.

**To submit**

A ZIP format (*firstinitiallastname\_*G50A02.zip) containing all submitted files on Moodle.