**Assignment #1**

Hands-on Programming

1. Decision Tree: Titanic Prediction - Fine Tune your Model

Please read ‘Kaggle instruction.pdf’ before working on this one.

You are going to work on the Titanic survivor prediction problem with Python (‘titanic\_train.csv’ and ‘titanic\_test.csv’ are the same as the two datasets on Kaggle <https://www.kaggle.com/francksylla/titanic-machine-learning-from-disaster/data>). Please try to do something in both data preparation and modeling stage.

In data preparation stage, you need to handle missing values and do one-hot-encoding on categorical variables.

In modeling stage, you can try 3 different combinations of parameters (e.g., criterion, max\_depth, min\_samples\_split) to find a best set of parameter values for this problem among the 3 (best in the sense that it gives you the best ranking on Kaggle). For example, one possibility can be

model = DecisionTreeClassifier(criterion='entropy',max\_depth=3, min\_samples\_split=5)

**Please submit:**

1. A file that specifies the 3 combinations of parameters you tried and the accuracy and ranking of each combination.
2. The ‘Assignment\_1-titanic.ipynb’ file with your commands in (please execute your commands before submission). Please use the model parameter setting that generates your best ranking.
3. The ‘titanic\_tree.pdf’ file (plotted tree) under the best model parameter setting.