MAS DSE 220

Assignment #3

1. Perform and plot the decision surface of the decision tree, forests of randomized trees and Boosting trained on the Titanic Data Set. Spend a little bit time on data prep. (Hint: Kaggle has a nice description). Perform and plot histogram of feature importance analysis before the training. Several different configuration of each of the models/parameters should be explored, analyzed and plotted. Demonstrate how increase in parameters influences accuracy. Describe your process of parameter tuning and provide in detailed discussion of the results.

2. Perform Stacking Ensemble analysis on the Boston Housing Data set. You can use Weka or other tool to perform Stacking or write your own Python code. Include interesting plots and attribute importance analysis to support the choice of the final model configuration chosen. Any combination of any of the Machine learning algorithms we have covered in the class so far or you are already familiar with is acceptable to be used in the Stacking Ensemble.