



## Assignment #3

Solve the following exercises and report your answers.

---

### Programming Problem (Code + Report)

Attached is the "Concrete\_Data.xlsx" dataset file. Given 8 numerical attributes, we are required to predict the concrete's compressive strength which is given in the last column. We will treat this problem as a classification task to predict if the concrete sample is strong or not. We assume that the sample is strong if the compressive strength is  $>40$ .

You are required to implement by yourself two ensemble learning approaches (bagging and boosting) and explore the effect of increasing the number of estimators on training and test errors. You will develop two approaches:

1. **Bagging Generic**, and try it on a built-in Decision Tree model.

Compare between Out-of-Bag error, and test accuracy. Plot them versus a different number of estimators for each approach (5, 10, 50, 100, 200, 500), and comment on your results.

2. **AdaBoost Generic**, and try it on decision stump as a weak learner.

You need to try a different number of estimators for each approach (5, 10, 50, 100, 200, 500). Plot a graph of the training and test accuracy as a function of the number of weak estimators, and comment on your results.

#### Rules:

- This is an individual assignment.
- Any evidence of plagiarism will be penalized by all the grades of the assignment.
- You should **only** submit the updated notebooks (with solutions displayed clearly) with your report.

**Deliverables:** The python notebook with your name, section & bench number included.

**Deadline:** Monday, May 22th, 2023, 23:59