

## IT 448 Programming Assignment 3

Answers to this assignment are due by **the time specified on ReggieNet.**

You need submit your solution as required below to the ReggieNet.

**No late submission!!!**

### Water Characteristics Predictor (WCP) (35%)

Given a water dataset (WaterData.csv), which contains nine water characteristics (i.e., Total Solids, SS, BOD5, NH3, Org-N, P-TOT, SO4, TKN, PRCP\_NOOA) between 2001 and 2018. Please note that some data are missing in the provided dataset.

In this programming assignment, you are asked to develop two predictors for “BOD5” and “TKN” using Random Forest. The detailed requirements are listed below:

For each predicted target (“BOD5” and “TKN”), you need to identify the best set of features (e.g., selected water characteristics), and prepare the related training dataset and test dataset. You should use 80% data for training and 20% for testing. For each Random Forest based predictor, optimize the hyperparameters (n\_estimators) to achieve the best prediction performance.

You are required to use r2\_score to evaluate your model and mean\_squared\_error to present a numerical performance result. Both are available in sklearn.metrics

Please finish this assignment using Colab Jupyter Notebook. Please name the file as yourFirstName\_yourLastName.jpynb.