

## IT 448 Programming Assignment 4

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) using Classification (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 formalize the original regression task to a classification task and develop a predictor for “TKN” using the SVM classification method. The following shows the detail requirements.

To formalize the regression task to a classification task, you need to (1) find the minimum (min) and maximum (max) values of TKN; (2) based on the number of classes (k) you want to formalize in the problem, you need to convert the original TKN values to their corresponding classes.

For example, if the min = 0 and max = 49.99, and you want to define 5 classes (k = 5), then you can have the following conversion table:

Class ID	Numerical Value Range
0	0 ~ 9.99
1	10 ~ 19.99
2	20 ~ 29.99
3	30 ~ 39.99
4	40 ~ 49.99

For instance, given an original value is 23, which is  $20 \leq 23 \leq 29.99$ , you convert to class 2.

For the predicted target (“TKN”), you are asked to use **20, 40, and 80** classes, respectively, to create three SVM classifiers. You can base on the results from the previous assignment 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 classifier, evaluate the model with the following performance metrics: **accuracy, precision, recall, and F1 score.**

Please finish this assignment using Colab Jupyter Notebook. Please name the file: **yourFirstName\_yourLastName\_PA4.jpynb.**