

Build your own dataset (with unique data-points) using Python libraries. The dataset must consist of at least 1000 data-points in the form (ID, A, B, C, D, E, F, Class) where

$$0 \leq A, B, C, D, E, F \leq 100 \quad (1)$$

and ID is the sequentially generated number of that data-point. The value of Class is 0 or 1 (randomly assigned). Create this dataset just once since you will be using it for the next steps.

Create a random 80-20 split of the data into train and test respectively. Store them in two different csv\_files. Plot the train data and test data with different labels on the same plot. (Note : You will need to submit the csv\_files. Adhere to the format mentioned above)

Write your own implementation of K-Nearest Neighbors. Input to the program is the value K. Run this implementation for different values of K (1 to 21 inclusive). Do the same using the library (sklearn) implementation of K-NN.

Compare and plot the accuracy and runtime for different values of K between your implementation and sklearn. What do you observe? Mention in the report. Mention any other inferences you can make.

If the input to the K-NN is an even number, what do you expect, and what do you observe? Mention it in the report.