**Instructions Lab 8**

The task for today’s lab is to perform feature section using Mutual Information. In today’s lab you will work with the Raisin dataset and you can download it from <https://archive.ics.uci.edu/ml/datasets/Raisin+Dataset>

1. Download the dataset, check for missing values and impute them using K nearest neighbors.

2. For each feature, find the MI of the feature with the class label. Arrange the features in a sorted order. Say, the sorted order is {f1, f2 … fD} where f1 has the highest MI and fD has the lowest MI with class label.

3. Build a classifier with all features and measure the error rate. Say the error rate is ED.

4. Now build a classifier with all features except fD and measure the error rate. Sa y the error rate is ED-1.

5. If ED-1 < ED. then build a classifier with all features except fD and fD-1. Again measure the error rate. Say the error rate is ED-2.

6. Repeat the process till ED-i < ED-i-1.

7. As a precaution, use 10 fold cross validation each time you build a classifier and also ensure that proper hyperparameter tuning is done.

8. Use decision trees classification algorithm.

**Show your work, even if it is partial, during the lab hours. What you show during the lab will contribute toward your final grade.**