CS4092D Machine Learning Lab
Module6 (DT): Exercise
Decision Trees
S4 MCA: Winter 2022-23

Date: 28-Mar-2023

Q. Implement the **CART Decision Tree(DT)** classification algorithm from scratch using Python. You must use only the basic libraries in Python to implement this exercise. We have shared 'BankNote_Authentication.csv' file for implementing this problem. Data were extracted from images that were taken from genuine and forged banknote-like specimens. The dependent variable in the dataset is the feature 'class', which is a binary variable (0 for genuine notes and 1 for forged notes). All the necessary feature descriptions can be found in the 'BankNote_Authentication .txt' file.

For building the DT model, **divide your dataset into train and test sets**: you may **randomly** take 70% of the datapoints(rows) in the dataset as the train-dataset and the remaining 30% of the datapoints goes to the test-dataset.

Run the performance evaluation metrics of your classification model on the given data set. Compare your **model's performance** with respect to Python's DecisionTreeClassifier() function defined in sklearn.tree.

Note: There are four main error metrics that are commonly used for evaluating and reporting the performance of a classification model in machine learning:

- Precision
- Recall
- F1-Score
- AUC & ROC Curve