Indian Institute of Technology Delhi
Department of Mathematics
II Semester 2022-2023
Assignment
Weightage 25% Due Date 28th Match 2023

You need to form a group of 3 members and jointly complete the assignment using Python Jupyter Notebook.

Q1. Choose a data set from UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=list) (or any other Source) for Multi class classification problems.

(i) Your first task is characterize the data set. Answer the following questions about the data: [3]

1) What the data is about.
2) What type of benefit you might hope to get from data mining.
3) Discuss data quality issues: For each attribute,
   a) Are there problems with the data?
   b) What might be an appropriate response to the quality issues.

(ii) Implement (1) Decision Tree, (2) Random Forest, (3) Naïve Bayes Classifier and (4) KNN classifier, (5) SVM, and (6) ANN and compare the performances using k-fold cross validation and other tuning techniques (grid search and parameter search, where ever applicable) [10]

Q2. Use MNIST DATASET

(a) Use the above classifiers to do multi-class classification where the idea is to classify the image to one of the ten digits (0-9). [9]

(b) Exploration of Different Evaluation Metrics

Evaluate your methods using different evaluation metrics. Tune the parameters using two powerful techniques of grid search and parameter search. [3]