UNIVERSITY OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

Data Science & Data Analytics Lab Project
CS695A

Analyze the Assigned dataset using any three of the following Machine Learning models:

1. Multilayer Perceptron Feed-Forward Network
2. Random Forest
3. Support Vector Machine
4. Naïve Bayes Classifier
5. K-Nearest Neighbour

Datasets: (source: https://archive.ics.uci.edu/ml/datasets.html)

1. ILPD (Indian Liver Patient Dataset) Data Set
2. Ozone Level Detection Data Set
3. Banknote authentication Data Set
4. Occupancy Detection Data Set
5. SPECT Heart Data Set

Method:

1. Importing Dataset: Import the dataset using 'pandas' package
2. Pre-processing: Check for missing values or any other discrepancies in the dataset. Use 'pandas' package or 'SimpleImputer' of 'sklearn.impute' module to tackle the missing values. Hint: Replace the missing values by the average of the existing values of the attribute. Perform any other necessary pre-processing.
3. Build Classifier: Build the classifier using the Scikit-learn package (Python) or any other similar package.
4. Split Dataset: Split the dataset into training and testing sets. Use 70% data for training. Hint: Use 'train_test_split' function of 'sklearn.model_selection' module.
5. Training Model: Train the classifier using the training set
6. Testing Model: Test the classifier using the test set
7. Performance Analysis: Find Accuracy, Precision, Recall of the model. Hint: sklearn.metrics can be used to find the metric values.

8. Comparative Analysis: Do Step 2 to 7 for all three classifiers of your choice. Compare the results and comment on the best classifier.

Dataset Allocation:

- Each group consists of two members (Strict).
- Groups are formed as follows;
  - Roll 1 & 2 will form group no. 1
  - Roll 3 & 4 will form group no. 2,
    and so on.
- Allotment List:

Dataset No.	Group No.
1	1 – 10
2	11 – 20
3	21 – 30
4	31 – 40
5	41 – 50

Submission:

1. The lab project carries 20 marks of the Final semester lab exam.
2. The project is needed to be submitted on or before Last week of March' **20**20 (24/03/2020 – 28/03/2020) on respective Lab days.
3. There will be no extension of the date of submission.
4. A Project Report (including source code) is needed to be submitted during project submission. On the day of submission, the project is needed to be demonstrated.