

Subject: 21AIE312

Lab Session: 01

Notes:

1. Please read the assignment notes carefully and comply to the guidelines provided.
2. A report should be prepared for questions available in “Report Section” of this assignment & submitted to TurnItIn.
3. The report should not have any content (text, figures, photos etc.) copied from any source outside your work. Please provide results obtained from your experiment to support your statements.

Main Section (Mandatory):

Please use the data set provided in attached file “Custom CNN Features.csv”. The dataset contains 3 class labels for 3253 records with 200 attributes / features.

A1. Please perform exploratory data analysis on the dataset to understand the following.

- Data types of attributes
- Missing values if any and how to perform imputations
- Class balance nature of data
- Attribute correlation study with correlation matrix and heatmap plot of the matrix
- Matrix rank study to find the dimensionality of data cloud
- Data range and normalization aspects

A2. Split the set into train & test sets with 75:25 split. Use various classifiers (kNN, Decision Tree, Support Vector Machines, Random Forest, Naïve-Bayes etc.) for classification of the dataset. Study the results of classification with accuracy, Precision, Recall, F1-Score & AUROC metrics.

A3. Check if the models have underfit, regular fit or overfit from the trainset and test set metrics.

A4. Check the timings of training and classification by each model.

Report Assignment:

1. From the performance of the models observed in above experiments, identify the classifier which you’d choose for this job. Justify your statement with data from obtained results. You may use tables, graphs and charts to prepare your report. [4]