Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603 110
(An Autonomous Institution, Affiliated to Anna University, Chennai)

UCS2612 Machine Learning Laboratory

Academic Year: 2023-2024 Even **Batch: 2021-2025**
Faculty In-charges: Y.V. Lokeswari & Nilu R Salim **VI Semester A & B**

Lab Test 1- Q3

Mushroom Dataset

This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family (pp. 500-525). Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like ``leaflets three, let it be'' for Poisonous Oak and Ivy.

Download the dataset from the following link:

https://archive.ics.uci.edu/dataset/73/mushroom

Develop a Python program to classify mushrooms as “poisonous” or “edible” using all the classification models (LR, PLA, MLP, KNN, SVM, Naïve Bayes) you have learnt. Interpret the model that works better for this dataset. Visualize the features from the dataset and interpret the results obtained by the model using the Matplotlib library. **[CO1, K3]**

Use the following steps to do implementation:

1. Loading the dataset.
2. Pre-Processing the data (Encoding, Standardization, Normalization, Handling missing values, Noisy data suitable for the dataset)
3. Exploratory Data Analysis.
4. Feature Engineering techniques.
5. Split the data into training, testing and validation sets.
6. Train the model.
7. Test the model.
8. Measure the performance of the trained model.
9. Represent the training and testing results using ROC curves. Does the model overfit? Comment on your obtained results.
10. Ignore the class label and perform the clustering task. Measure the performance of the model.

Upload the code in GitHub and include the GitHub main branch link in the assignment PDF.