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

Predict Census Income

This data was extracted from the 1994 Census bureau database by Ronny Kohavi and Barry Becker (Data Mining and Visualization, Silicon Graphics). A set of reasonably clean records was extracted using the following conditions: ((AAGE>16) && (AGI>100) && (AFNLWGT>1) && (HRSWK>0)).
Income is the target label.
***The prediction task is to determine whether a person makes over $50K a year*.**

https://www.kaggle.com/datasets/uciml/adult-census-income?resource=download

Develop a python program to predict the income of a person using all the classification models (LR, PLA, MLP, KNN, SVM, Naïve Bayes) you have learnt. Interpret the model which works better for this dataset. Visualize the features from the dataset and interpret the results obtained by the model using 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)
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 clustering task. Measure the performance of the model.

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