

**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**

**Faculty In-charges: Y.V. Lokeswari & Nilu R Salim**

**Lab Test 1**

**Batch: 2021-2025**

**VI Semester A & B**

**Pima Indians Diabetes Database**

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

**Target is Outcome**

<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

Develop a python program to predict the diabetes 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.