Assignment 2: Machine Learning (MSC527): Classification Problem

With the attached dataset perform the following classification algorithms in Python**.**

1. **Logistic Regression**
2. **kNN**
3. **Decision Tree**
4. **Naïve Bayes**
5. **Support Vector Machine**

Split the dataset into Training and Test data to check for the performance, like accuracy, confusion matrix, AUC, etc. and finally write a report comparing the results obtained from the above techniques.

The description of the dataset is also provided.

Last date of submission: **April 12, 2024**

Upload your assignment as a single pdf file in Google classroom.

Context: This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes.

Content : 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.

Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test

BloodPressure: Diastolic blood pressure (mm Hg)

SkinThickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mu U/ml)

BMI: Body mass index (weight in kg/(height in m)^2)

DiabetesPedigreeFunction: Diabetes pedigree function

Age: Age (years)

Outcome: Class variable (0 or 1)