**Assignment Problems: Day 9**

The following problems require access to the file named ‘Dataset\_Day9.csv’, provided with this assignment. (**Use Random State = 1234**)

This is the same assignment data as Day 8.

The datasets include data from 768 women with several medical predictor variables and one target variable(‘Outcome’). The classification goal is to predict whether or not the patients in the dataset have diabetes or not.

A screenshot of a computer code

Description automatically generated

For columns:

*Glucose,   
BloodPressure,   
BMI,   
DiabetesPedigreeFunction,  
Age*

If the column value is 0, then they should be considered as **missing data.**

Problems to solve –

1. Firstly, replace all Missing values with relevant figures.
2. Then remove all existing outliers and get the final data for classification.
3. Split the data into 75% training and 25% testing data. Then, use a SVM classifier algorithm with target variable as ‘Outcome’.
   1. Print the default model performance metrics: Accuracy, Precision, Recall, F1Score
   2. Print Precision & Recall & F1-Score vs kernel type(**'linear', 'poly', 'rbf', 'sigmoid'**) curve (All metrics on the same graph). Find the kernel type for which F1-score is the highest. Take (C = 0.001,0.01,0.1,1,10)
   3. Plot a curve on Precision & Recall & F1-Score vs **appropriate range of C** using the best kernel type you obtained in question(3b), (All metrics on the same graph). Find the C for which F1-score is the highest for the given kernel type.  
      Example: Take C between (0,10) in small increments of 0.05