**Assignment Problems: Day 12**

The following problems require access to the file named ‘Dataset\_Day12.csv’, provided with this assignment.

This is the same assignment data as Day 11.

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*

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 80% training and 20% testing data. Use target variable as ‘Outcome’.
   1. Use Naïve Bayes Classifier algorithm to classify *Outcome* and print the default model performance metrics: Accuracy, Precision, Recall, F1Score.
   2. Use k-Fold Cross Validation to find the best model performance metrics .
   3. Try out Adaboost algorithm for Naïve Bayes Classifier improvement. [Optional]