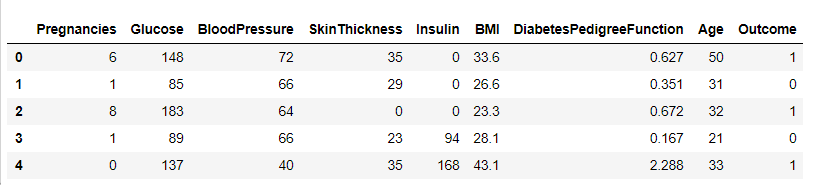
**DIABETES DATA**

I have used the Logistic Regression to solve this problem, because the outcome is to predict a Boolean result i.e., whether the patient has diabetes or not. Not just using the Logistic Regression but also using the Standard Scaling, Normalisation, MinMax Scaling, and also Hyper parameter tuning is done. The accuracies are compared and the best suitable model is chosen to predict the data.

**Dataset knowledge:**

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.



**Columns:**

Pregnancies: Number of times pregnant

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 (

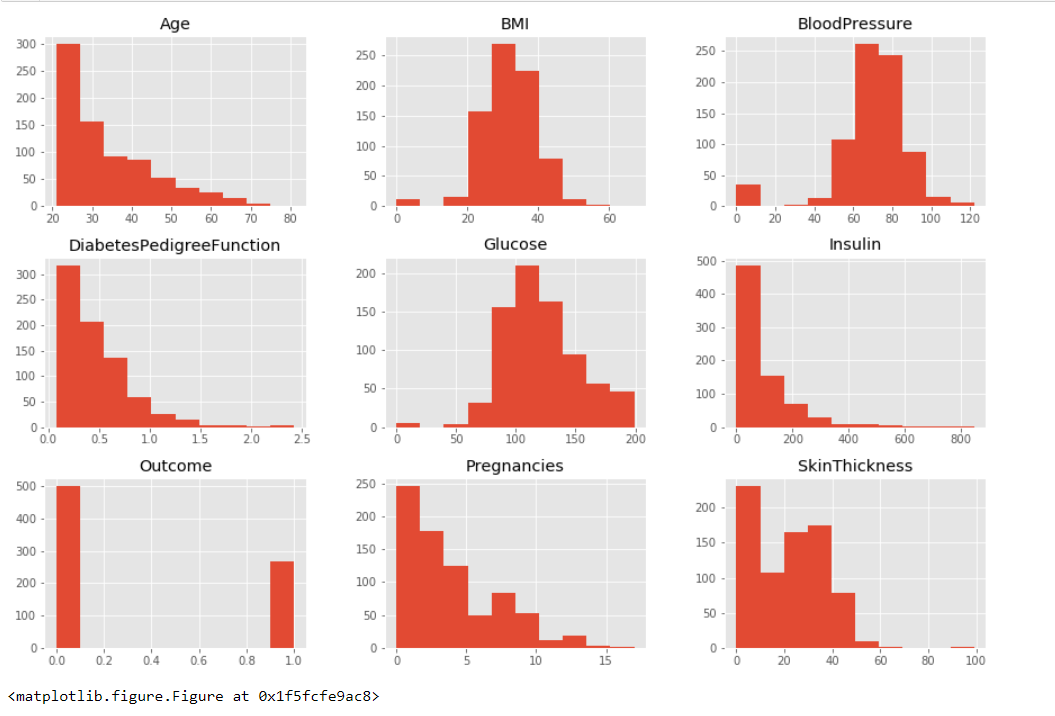
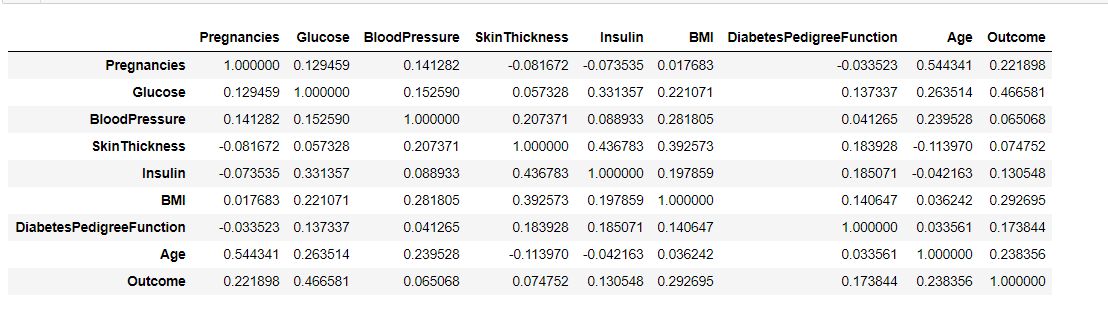
BMI: Body mass index (Weight (kg)/(Height (m)^2))

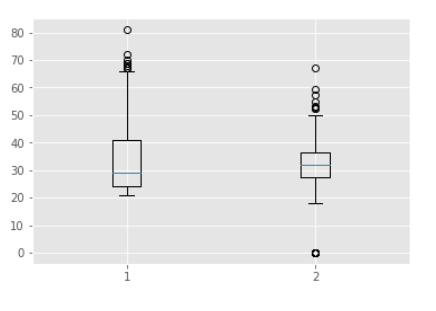
DiabetesPedigreeFunction: Diabetes pedigree function

Age: Age (years)

Outcome: Class variable (0 or 1)

**Pre-Processing:**

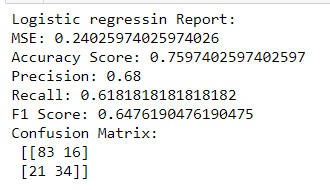
The correlation is calculated to see whether the data has outliers or noise. And also to check how well the data has been correlated. The histograms are plotted so that to see how well the data has been distributed and to choose the dependent variables for the model. Here, the data has been well distributed among age and BMI, so that is chosen. 



**ML models:**

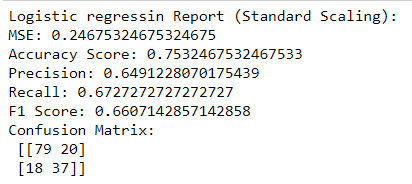
Logistic Regression:

The Logistic Regression is applied without scaling and normalization. The accuracy for this model is calculated and it turns out to be 74.7%



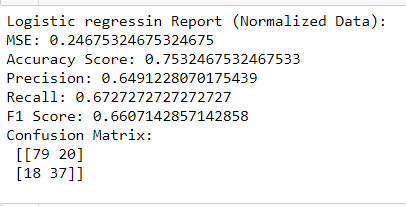
Logistic Regression (Standard Scaling):

Accuracy:75.3%

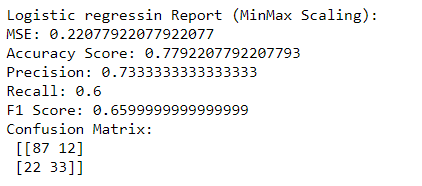


Logistic Regression (Normalized data):

Accuracy : 75.3%



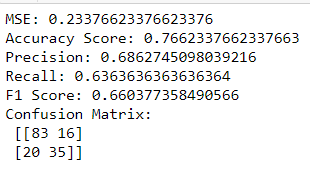
Logistic Regression (MinMax Scaling):

Accuracy : 77.9%

Hyperparameter tuning:

Grid search is arguably the most basic hyperparameter tuning method. With this technique, we simply build a model for each possible combination of all of the hyperparameter values provided, evaluating each model, and selecting the architecture which produces the best results.

Accuracy : 76%



Unarguably , Logistic Regression with MinMax Scaling Model is the best suitable for this data for prediction.