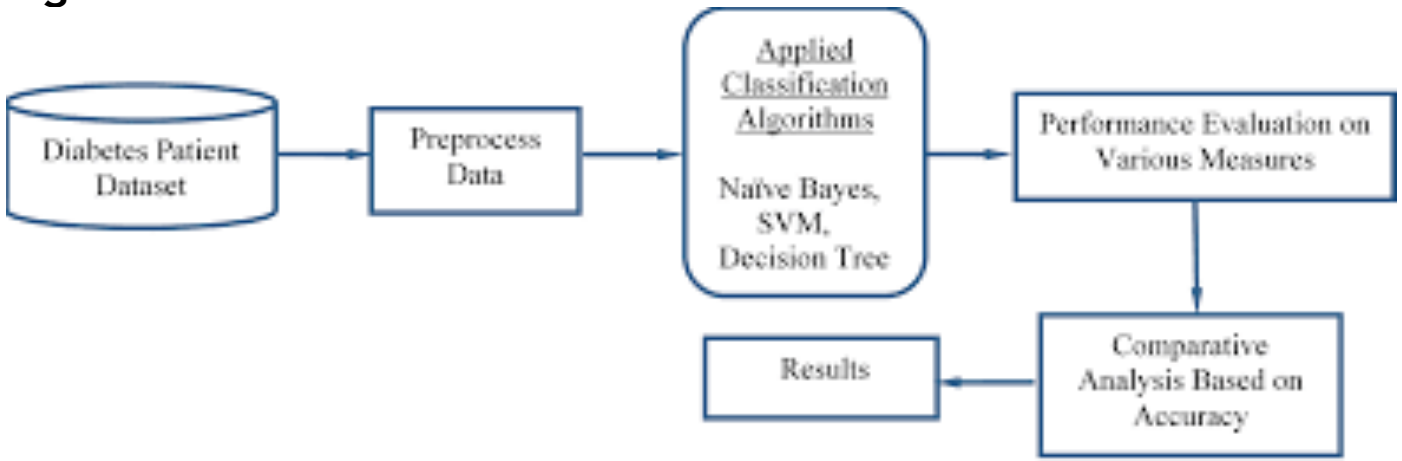
Prediction of diabetes

By:
Dheeravath Dharmaraj

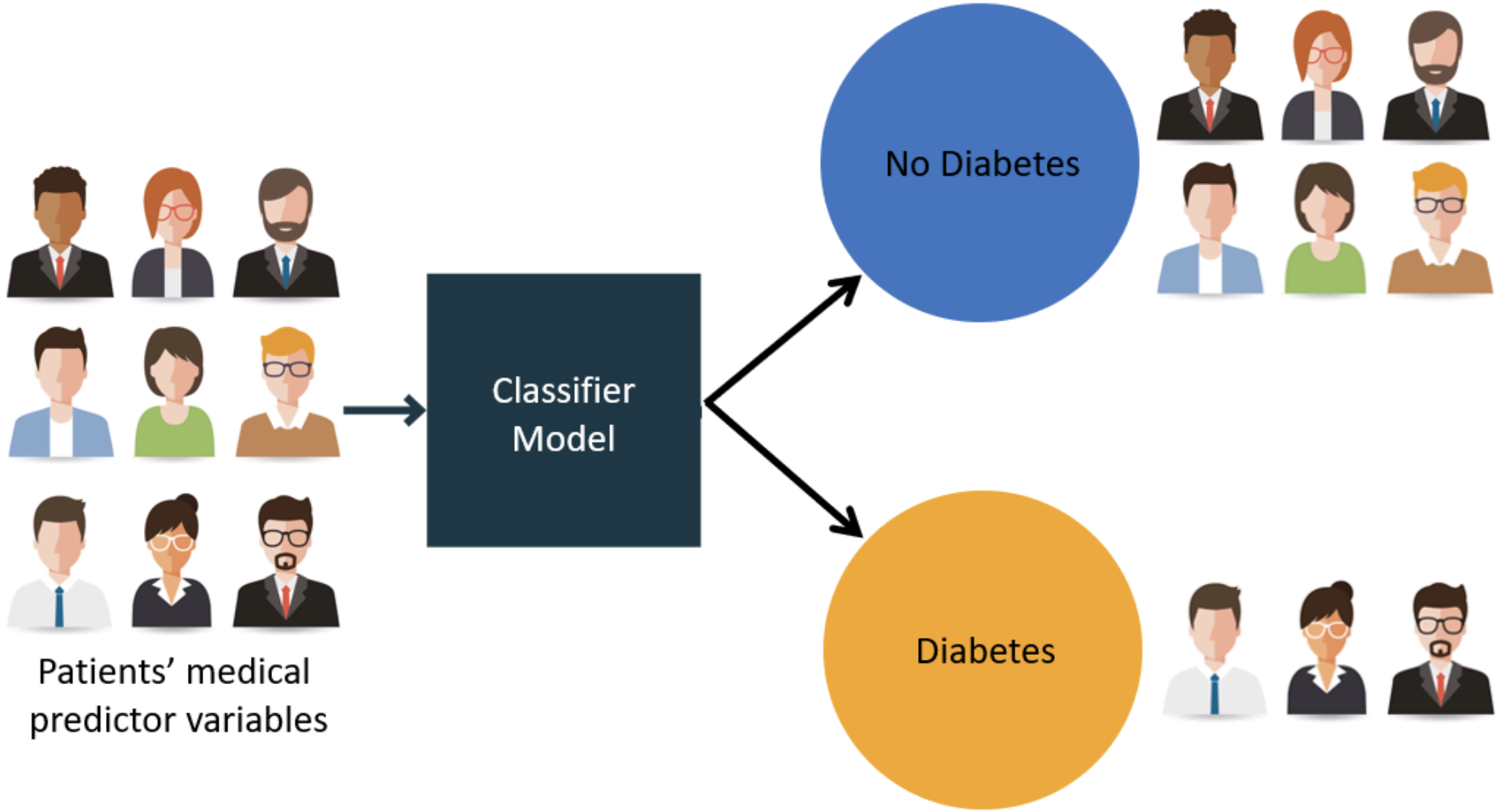
>>To build a predictive machine learning model to predict based on diagnostic measurements whether a patient has diabetes.

>>Building a machine learning classifier model for diabetes.

>>Based on medical diagnostic measurements.

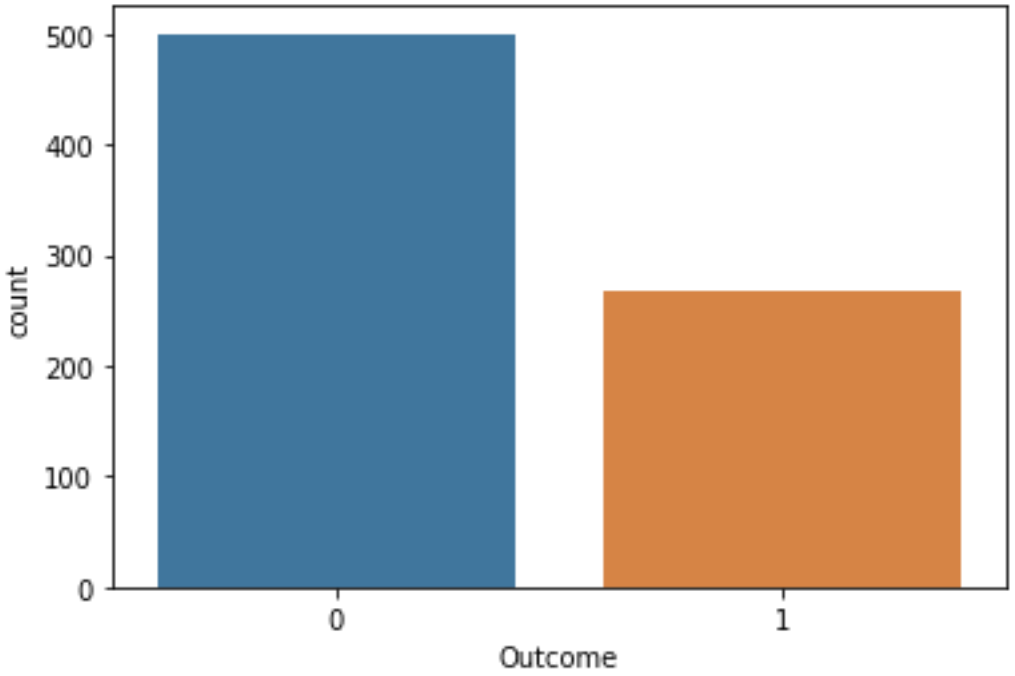


Objective


No Diabetes
Classifier Model
Diabetes
Patients' medical predictor variables

Dataset

downloaded from Kaggle

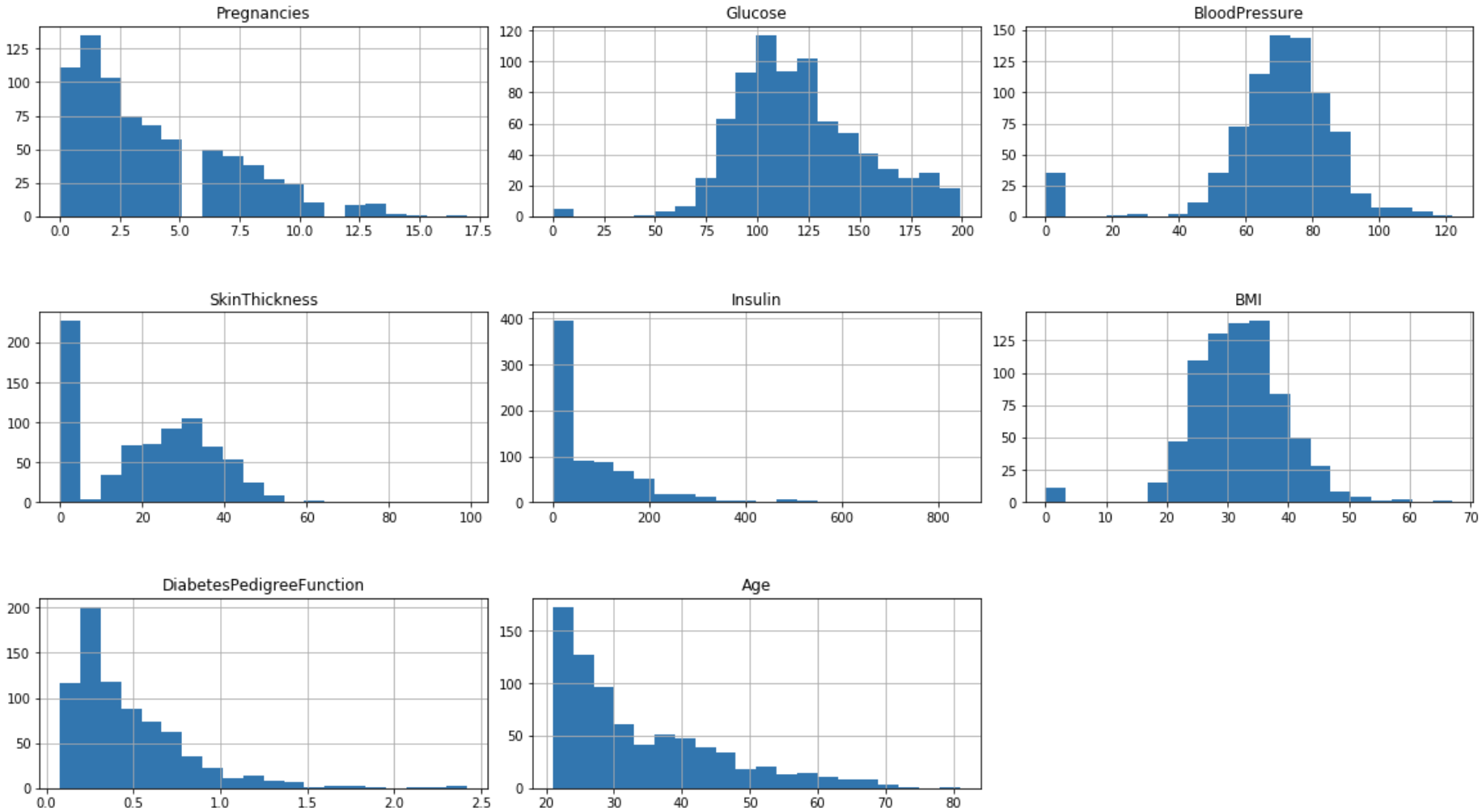


There are 768 observations with 8 medical predictor features (input) and 1 target variable (output 0 for ”no” or 1 for ”yes”)

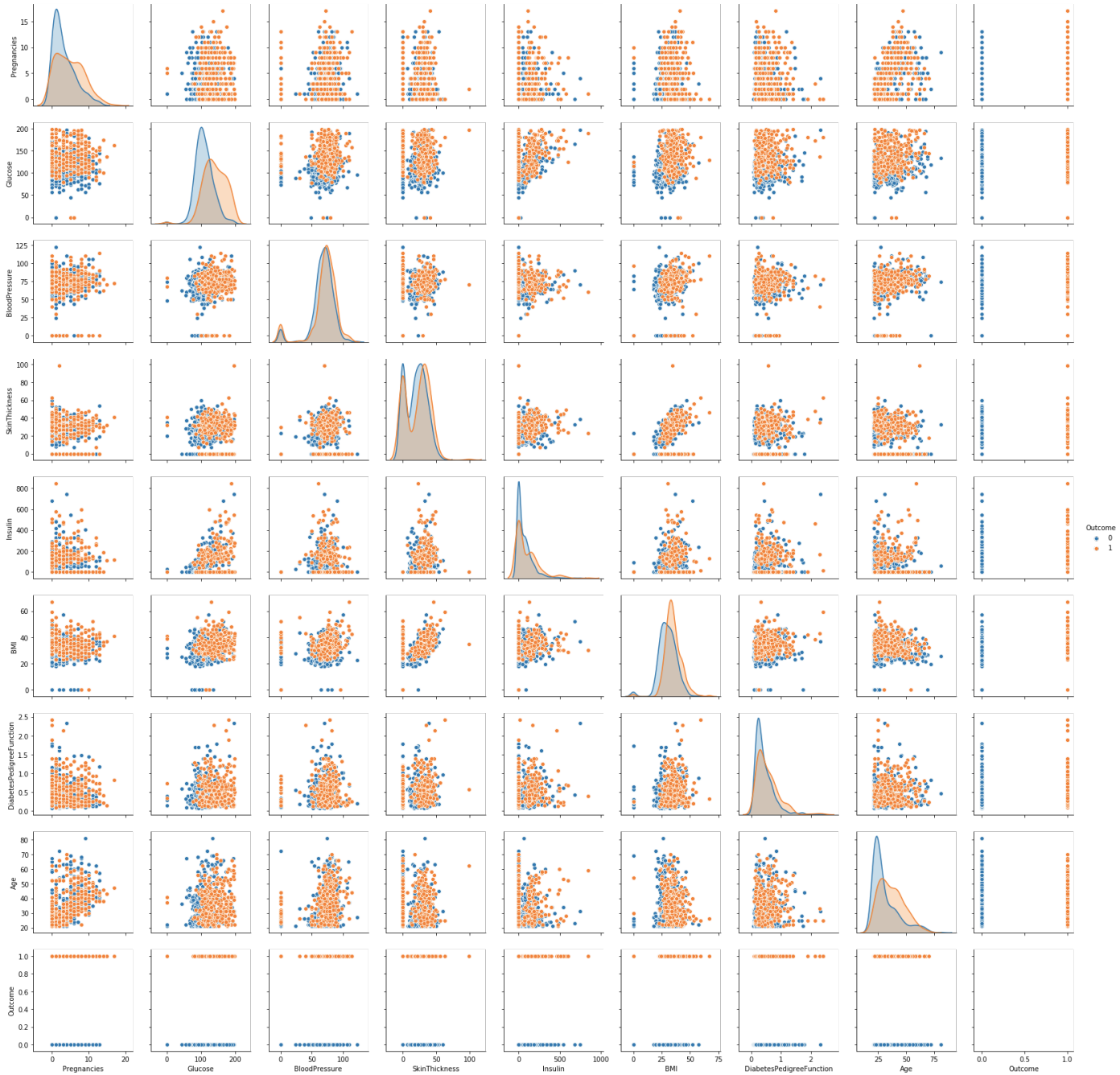
The 8 medical predictor features are:

- **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 (mu U/ml)
- **BMI**: Body mass index (weight in kg/(height in $m)^2$)
- **DiabetesPedigreeFunction**: Diabetes pedigree function
- **Age**: Age (years)

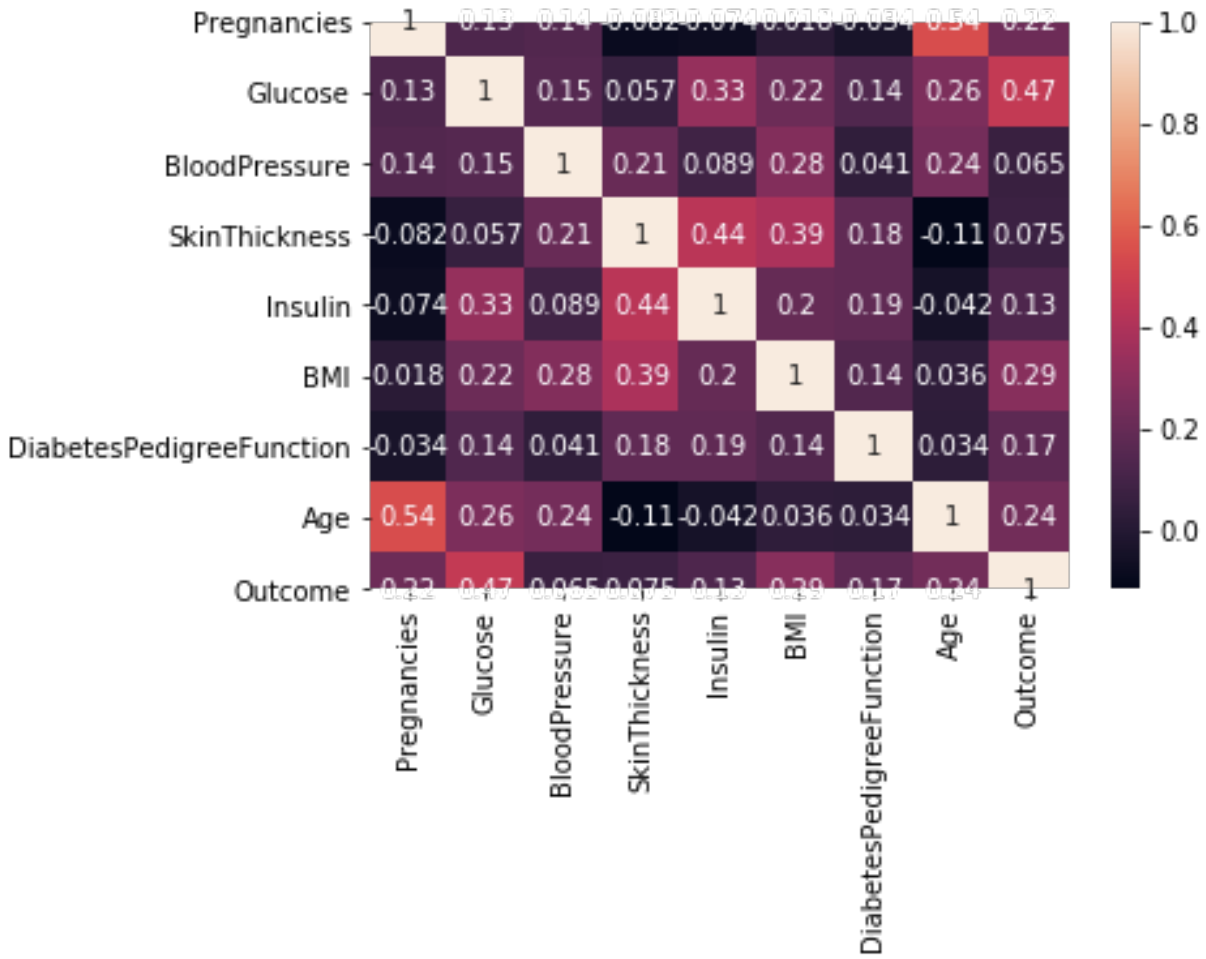
Histogram of each feature



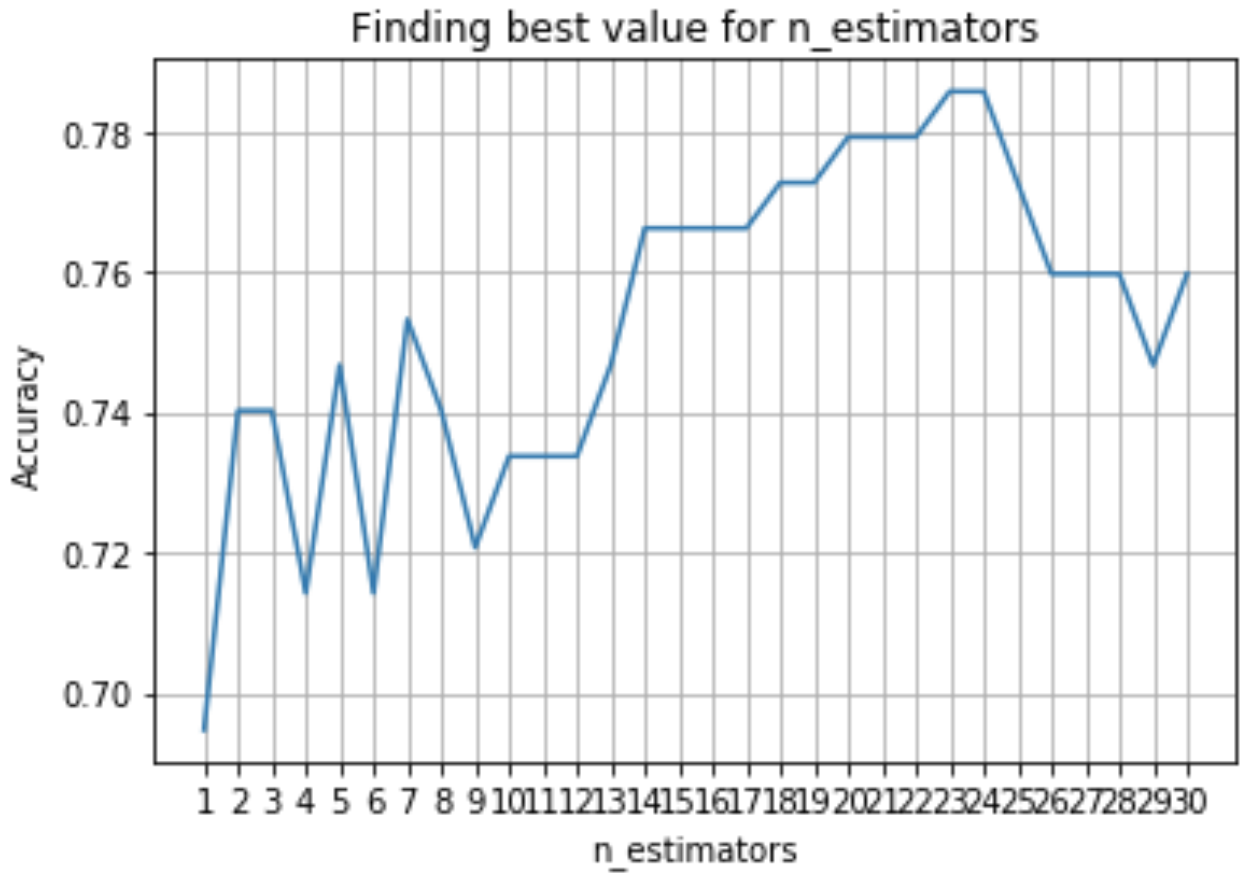
Pair Plots


Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome
Outcome
0
1

Heatmap


Pregnancies 1 0.13 0.14 -0.082 -0.074 0.018 -0.034 0.54 0.22
Glucose 0.13 1 0.15 0.057 0.33 0.22 0.14 0.26 0.47
BloodPressure 0.14 0.15 1 0.21 0.089 0.28 0.041 0.24 0.065
SkinThickness -0.082 0.057 0.21 1 0.44 0.39 0.18 -0.11 0.075
Insulin -0.074 0.33 0.089 0.44 1 0.2 0.19 -0.042 0.13
BMI 0.018 0.22 0.28 0.39 0.2 1 0.14 0.036 0.29
DiabetesPedigreeFunction -0.034 0.14 0.041 0.18 0.19 0.14 1 0.034 0.17
Age 0.54 0.26 0.24 -0.11 -0.042 0.036 0.034 1 0.24
Outcome 0.22 0.47 0.065 0.075 0.13 0.29 0.17 0.24 1
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome
1.0
0.8
0.6
0.4
0.2
0.0

graph for n_neighbors

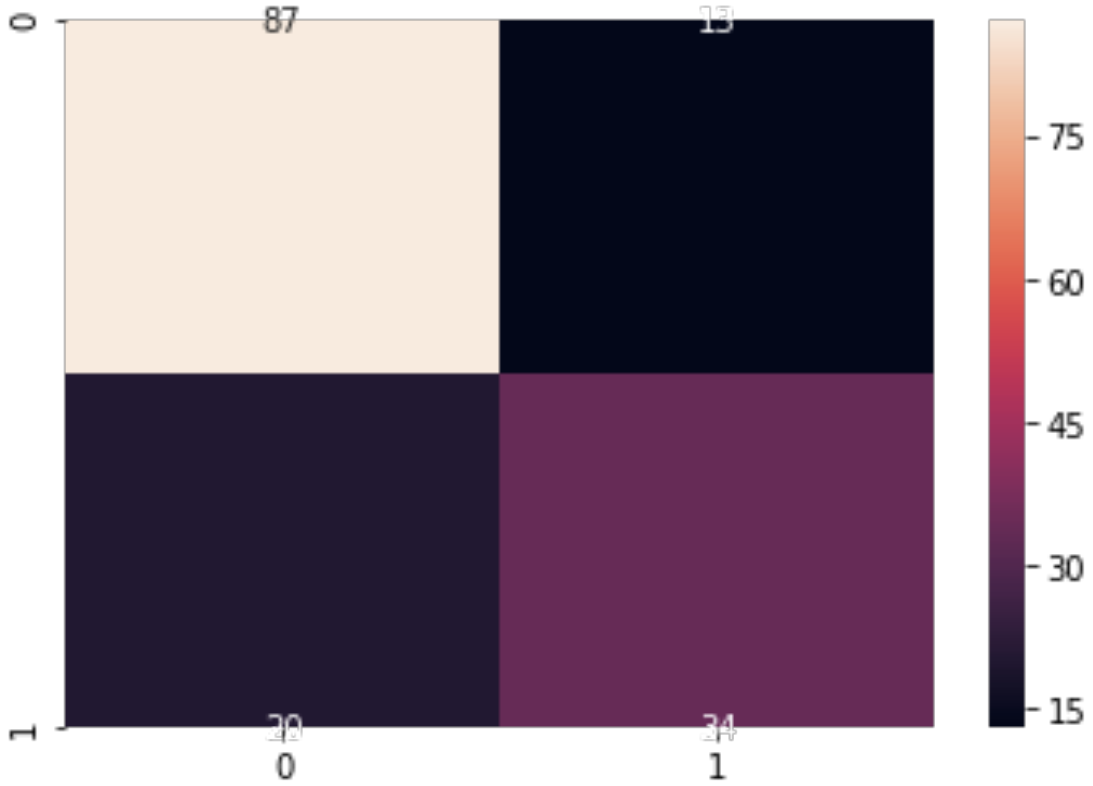


Model Evaluation

Below 6 models have been evaluated:

- Naive Bayes
- Logistic Regression
- K Nearest Neighbour
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Classification (SVC)

Heatmap of Confusion matrix



THANK YOU