Required libraries are imported and the data set is loaded. Now we pre-process and view data in it’s different categories. The different variables used for diabetes prediction are number of pregnancies, glucose level, age, Blood pressure, skin thickness, Insulin levels, BMI. And the result is given as a binary output as 1 or 0 for detected Diabetes.

See correlation of each variable using a heat map.

Calendar

Description automatically generated with medium confidence

DISTPLOTS

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated Chart, line chart, histogram

Description automatically generated

HISTOGRAMS

Graphical user interface, chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated

SCATTERPLOT

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

PIECHARTS

Chart, pie chart

Description automatically generated

Pairplot

A picture containing diagram

Description automatically generated

A picture containing diagram

Description automatically generated

Chart, scatter chart

Description automatically generated Diagram, histogram

Description automatically generated

Probplot :-

Each data point in y using marker symbols and draws a reference line that represents the theoretical distribution.

Chart, line chart

Description automatically generated

Training the dataset

Splitting the dataset into training and test data

Fitting dataset into different models

Support vector machines

K-nearest neighbours

Decision tree algorithm

Random Forest

# Compare the accuracy\_score on the basis of barplot

Accuracy Scores

####################Logistic Regression####################

0.8051948051948052

####################KNearest Neighbour####################

0.7207792207792207

####################SVM####################

0.8116883116883117

####################Decision Tree Classifier####################

0.7142857142857143

####################Random Forest Classifier####################

0.7987012987012987

Chart, bar chart

Description automatically generated

ROC score:

0.845595545834162

accuracy