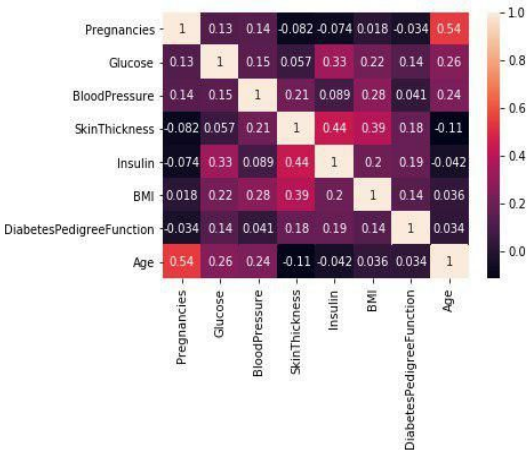
Diabetes classification using SVM

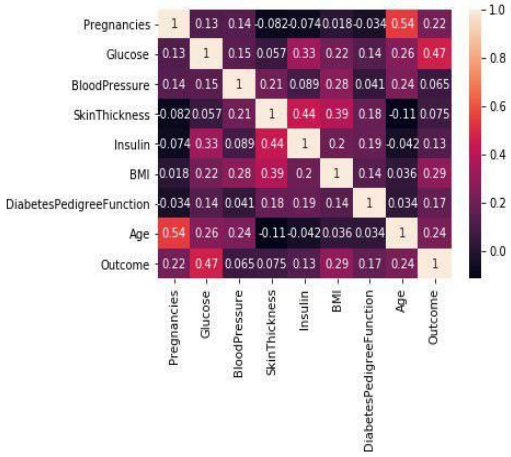
Siva Kailash S - 2016103593
Gokul R - 2016103523

Aim:
To predict the presence of Diabetes using SVM and apply dimensionality reduction using PCA.

Pima Indians Diabetes Data Set :
https://www.kaggle.com/uciml/pima-indians-diabetes-database

Correlation Matrices





Algorithms : SVM Kernel : RBF

RBF kernel is used for nonlinear datasets

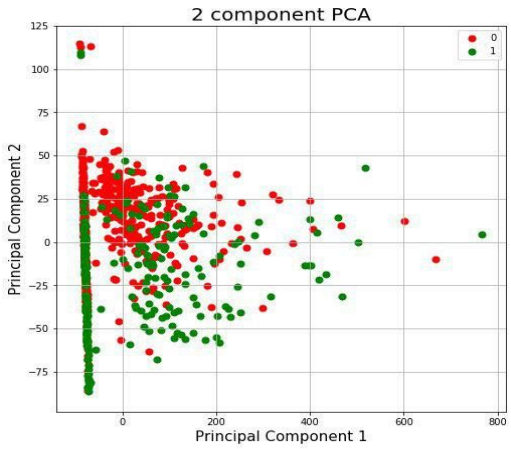
Confusion Matrix
[[117 13]
 [30 32]]

Accuracy
0.7760416666666666

We now Try to reduce the dimensions of the model.

Two Components

PCA - 2 Components



Confusion Matrix
[[118 12]
[37 25]]

Accuracy
0.7447916666666666

PCA- Three Components

Confusion Matrix
[[115 15]
[30 32]]

Accuracy
0.765625

Conclusion :

The reduction of dimensions may have reduced the complexity of the code but also reduced the accuracy as well. This can be reduced by doing some preprocessing or other dimensionality reduction algorithm that produces better accuracy.

Code : https://github.com/SivaK18/MachineLearning