OBSERVATION REPORT

1. EXPLORATORY DATA ANALYSIS

1) The histograms and barplots are in the separate folder Graphs in ExploratoryDataAnalysis Folder. From the histograms, we can observe that BMI, BP, Diabetes and Plasma have a Bell shaped curve. Hence, those attributes are normally distributed.

2) Plasma has the maximum correlation with the class variable=0.466581398306874

3)

	Pregnant	Plasma	BP	Triceps	Insulin	BMI
Pregnant	1.00000000	0.12945867	0.14128198	-0.08167177	-0.07353461	0.01768309
Plasma	0.12945867	1.00000000	0.15258959	0.05732789	0.33135711	0.22107107
BP	0.14128198	0.15258959	1.00000000	0.20737054	0.08893338	0.28180529
Triceps	-0.08167177	0.05732789	0.20737054	1.00000000	0.43678257	0.39257320
Insulin	-0.07353461	0.33135711	0.08893338	0.43678257	1.00000000	0.19785906
BMI	0.01768309	0.22107107	0.28180529	0.39257320	0.19785906	1.00000000
Diabetes	-0.03352267	0.13733730	0.04126495	0.18392757	0.18507093	0.14064695
Age	0.54434123	0.26351432	0.23952795	-0.11397026	-0.04216295	0.03624187

	Diabetes	Age
Pregnant	-0.03352267	0.54434123
Plasma	0.13733730	0.26351432
BP	0.04126495	0.23952795
Triceps	0.18392757	-0.11397026
Insulin	0.18507093	-0.04216295
BMI	0.14064695	0.03624187
Diabetes	1.00000000	0.03356131
Age	0.03356131	1.00000000

Pregnant and Age attributes have the highest mutual correlation=0.544341228402339

2. NAIVE BAYESIAN CLASSIFIER

1) R package "e1071" has been used to perform Naives Bayes Classification.

2)

	Experiment	Accuracy
[1,]	1	76.62338
[2,]	2	76.62338
[3,]	3	83.11688
[4,]	4	72.72727
[5,]	5	81.81818
[6,]	6	77.92208
[7,]	7	74.02597
[8,]	8	72.72727
[9,]	9	80.51948
[10,]	10	77.92208

3) Overall Accuracy= 77.4026%

3. SVM CLASSIFIER

From the output of the SVMClassifier.R program, we observe that by default, the kernal type is Radial.

The report containing average accuracy of the different types of Kernal for SVM shown bellow.

```
     Kernal         Average Accuracy of 10 experiments
[1,] "Linear"     "79.0909090909091"
[2,] "Polynomial" "76.7532467532468"
[3,] "Radial"     "78.3116883116883"
[4,] "Sigmoid"    "71.1688311688312"
```

From the above report, we can say that the accuracy is higher for a Linear type of kernal SVM.

4. kNN CLASSIFIER

The Report results containing the Accuracy evaluation of the kNN Classifier is as follows:

```
      k       Average Accuracy of 10 experiments
[1,] "3"  "40.7597402597403"
[2,] "5"  "43.8636363636364"
[3,] "7"  "43.3688311688312"
[4,] "9"  "44.2662337662338"
[5,] "11" "44.2727272727273"
```

From the report, we can say that when k=11, that is when a large amount of neighbors are considered for the Pima dataset, Accuracy is higher.

After a complete analysis, we can say that A LINEAR SVM CLASSIFIER is the best Classifier of all.