Classification Assignment

1.Problem Statement

a. Machine Learning

b.Supervised Learning

c.Classification

2.Total No of Rows: 399

Total No of Columns: 28

3.Preprocessing Method : Get_dummies from Pandas

4.Algorithms Used

1.Logistic Regression

2.Random forest

3.Decision tree

4.Support Vector machine

5.XGBoost

5.Screenshots of results:

1.Logistic Regression: Classification Report

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
              precision    recall  f1-score   support

           0       0.93      1.00      0.97       100
           1       1.00      0.96      0.98       180

    accuracy                           0.97       280
   macro avg       0.97      0.98      0.97       280
weighted avg       0.98      0.97      0.98       280
```

ROC AUC value:

```
Out[11]: 0.9996111111111111
```

2.Random forest : Classification Report & ROC AUC Value

	precision	recall	f1-score	support
0	0.89	0.98	0.93	100
1	0.99	0.93	0.96	180
accuracy			0.95	280
macro avg	0.94	0.96	0.95	280
weighted avg	0.95	0.95	0.95	280

```
Out[12]: 0.9972222222222223
```

3.Decision tree: Classification Report

```
Fitting 5 folds for each of 18 candidates, totalling 90 fits
```

	precision	recall	f1-score	support
0	0.89	0.99	0.94	100
1	0.99	0.93	0.96	180
accuracy			0.95	280
macro avg	0.94	0.96	0.95	280
weighted avg	0.96	0.95	0.95	280

ROC AUC value

```
Out[13]: 0.9616666666666667
```

4.Support Vector machine:Classification Report

```
Fitting 5 folds for each of 54 candidates, totalling 270 fits
```

	precision	recall	f1-score	support
0	0.90	0.99	0.94	100
1	0.99	0.94	0.97	180
accuracy			0.96	280
macro avg	0.95	0.96	0.95	280
weighted avg	0.96	0.96	0.96	280

ROC AUC value:

```
Out[14]: 0.9992222222222222
```

5.XGBoost : Classification Report & ROC AUC Value

```
Fitting 5 folds for each of 210 candidates, totalling 1050 fits
              precision    recall  f1-score   support

           0       0.89      0.93      0.91       100
           1       0.96      0.93      0.95       180

    accuracy                           0.93       280
   macro avg       0.92      0.93      0.93       280
weighted avg       0.93      0.93      0.93       280
```

```
Out[22]: 0.9916111111111111
```

6.Final Model: WRT ROC AUC value

1.Logistic Regression : 99.96%

2.Random forest : 99.72%

3.Decision tree : 96.16%

4.Support Vector machine : 99.92%

5.XGBoost : 99.16%

Reason :

As per evaluation ROC value by the Logistic Regression model is Higher. Hence we choose Logistic Regression for Deployment.