Findings & Observation documentation

Name: Preeti Yadav

PRN: 230340325039

Objective:

The main objective is to generate a predictive model which can accurately classify transactions as either legitimate or fraudulent. This model will be based on the available data and will be used to make predictions on new data.

For training data = 0.80 and testing data = 0.20

MODEL COMPARISON TABLE:

S.No	Model	Accuracy	Precision	Recall	f1-score
1	Logistic Regression	0.90	0.85	0.96	0.90
2	KNN	0.85	0.78	0.96	0.86
3	Decision Tree Classifier	0.93	0.96	0.91	0.93
4	Random Forest	0.79	0.74	0.91	0.8

1) Logistic Regression:

confusion_matrix

```
[[1557   73]
 [ 267 1389]]
```

accuracy_score

```
0.8965307364576993
```

classification_report

	precision	recall	f1-score	support
0	0.85	0.96	0.90	1630
1	0.95	0.84	0.89	1656
accuracy			0.90	3286
macro avg	0.90	0.90	0.90	3286
weighted avg	0.90	0.90	0.90	3286

2) **K Nearest Neighbours:**

```
confusion_matrix
[[1572   58]
 [ 441 1215]]

accuracy_score
0.8481436396835058

classification_report:
```

	precision	recall	f1-score	support
0	0.78	0.96	0.86	1630
1	0.95	0.73	0.83	1656
accuracy			0.85	3286
macro avg	0.87	0.85	0.85	3286
weighted avg	0.87	0.85	0.85	3286

3) Decision Tree Classifier:

Accuracy = 0.9324406573341448

Precision = 0.96

Recall = 0.91

Confusion matrix:

```
[[1477  153]
 [  69 1587]]
```

Classification report:

	precision	recall	f1-score	support
0	0.96	0.91	0.93	1630
1	0.91	0.96	0.93	1656
accuracy			0.93	3286
macro avg	0.93	0.93	0.93	3286
weighted avg	0.93	0.93	0.93	3286

4) Random Forest :

```
confusion_matrix
[[1478  152]
 [ 524 1132]]


accuracy_score
0.7942787583688375


classification_report
              precision    recall  f1-score   support

           0       0.74      0.91      0.81      1630
           1       0.88      0.68      0.77      1656

    accuracy                           0.79      3286
   macro avg       0.81      0.80      0.79      3286
weighted avg       0.81      0.79      0.79      3286
```

Conclusion:

	Model	Accuracy Score
0	LogisticRegression	0.90
1	KNN	0.85
2	Decision Tree	0.93
3	Random Forest	0.80

From the above table we can conclude that Decision Tree has highest Accuracy Score i.e., 93% followed by Logistic Regression which is 90%.