

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
dataset=pd.read_csv("fraud.csv")
```

In [3]:

```
dataset
```

Out[3]:

step		type	amount	oldbalanceOrig	newbalanceOrig	oldbalanceDest	newbalar
0	1	PAYMENT	9839.64	170136.00	160296.36	0.00	
1	1	PAYMENT	1864.28	21249.00	19384.72	0.00	
2	1	PAYMENT	11668.14	41554.00	29885.86	0.00	
3	1	PAYMENT	7817.71	53860.00	46042.29	0.00	
4	1	PAYMENT	7107.77	183195.00	176087.23	0.00	
...	
16421	743	CASH_OUT	339682.13	339682.13	0.00	0.00	33
16422	743	TRANSFER	6311409.28	6311409.28	0.00	0.00	
16423	743	CASH_OUT	6311409.28	6311409.28	0.00	68488.84	637
16424	743	TRANSFER	850002.52	850002.52	0.00	0.00	
16425	743	CASH_OUT	850002.52	850002.52	0.00	6510099.11	736

16426 rows × 9 columns



In [4]:

```
dataset=dataset.drop("isFlaggedFraud",axis=1)
```

In [5]:

```
dataset
```

Out[5]:

	step	type	amount	oldbalanceOrig	newbalanceOrig	oldbalanceDest	newbalar
	0	1	PAYMENT	9839.64	170136.00	160296.36	0.00
	1	1	PAYMENT	1864.28	21249.00	19384.72	0.00
	2	1	PAYMENT	11668.14	41554.00	29885.86	0.00
	3	1	PAYMENT	7817.71	53860.00	46042.29	0.00
	4	1	PAYMENT	7107.77	183195.00	176087.23	0.00

16421	743	CASH_OUT	339682.13	339682.13	0.00	0.00	339682.13
16422	743	TRANSFER	6311409.28	6311409.28	0.00	0.00	6311409.28
16423	743	CASH_OUT	6311409.28	6311409.28	0.00	68488.84	6370918.44
16424	743	TRANSFER	850002.52	850002.52	0.00	0.00	850002.52
16425	743	CASH_OUT	850002.52	850002.52	0.00	6510099.11	7350099.64

16426 rows × 8 columns



In [6]:

```
dataset=pd.get_dummies(dataset,drop_first=True)
```

In [7]:

```
dataset
```

Out[7]:

	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFra
0	1	9839.64	170136.00	160296.36	0.00	0.00	
1	1	1864.28	21249.00	19384.72	0.00	0.00	
2	1	11668.14	41554.00	29885.86	0.00	0.00	
3	1	7817.71	53860.00	46042.29	0.00	0.00	
4	1	7107.77	183195.00	176087.23	0.00	0.00	
...
16421	743	339682.13	339682.13	0.00	0.00	339682.13	
16422	743	6311409.28	6311409.28	0.00	0.00	0.00	
16423	743	6311409.28	6311409.28	0.00	68488.84	6379898.11	
16424	743	850002.52	850002.52	0.00	0.00	0.00	
16425	743	850002.52	850002.52	0.00	6510099.11	7360101.63	

16426 rows × 11 columns



In [8]:

```
indep=dataset[["step","amount","oldbalanceOrg","newbalanceOrig","oldbalanceDest","newbalanceDest"]]
dep=dataset[["isFraud"]]
```



In [9]:

```
indep
```

Out[9]:

	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	type
0	1	9839.64	170136.00	160296.36	0.00	0.00	
1	1	1864.28	21249.00	19384.72	0.00	0.00	
2	1	11668.14	41554.00	29885.86	0.00	0.00	
3	1	7817.71	53860.00	46042.29	0.00	0.00	
4	1	7107.77	183195.00	176087.23	0.00	0.00	
...	
16421	743	339682.13	339682.13	0.00	0.00	339682.13	
16422	743	6311409.28	6311409.28	0.00	0.00	0.00	
16423	743	6311409.28	6311409.28	0.00	68488.84	6379898.11	
16424	743	850002.52	850002.52	0.00	0.00	0.00	
16425	743	850002.52	850002.52	0.00	6510099.11	7360101.63	

16426 rows × 10 columns



In [10]:

```
dep
```

Out[10]:

	isFraud
0	0
1	0
2	0
3	0
4	0
...	...
16421	1
16422	1
16423	1
16424	1
16425	1

16426 rows × 1 columns

In [11]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(indep,dep,test_size=1/3,random_state=0)
```

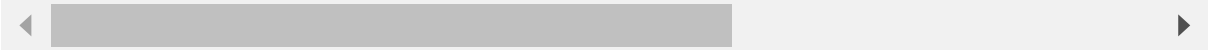
In [12]:

X_train

Out[12]:

	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	typ
14402	561	2763025.35	2763025.35	0.00	0.00	0.00	
11624	302	633384.46	633384.46	0.00	0.00	0.00	
6422	6	9396.00	936212.00	926816.00	0.00	0.00	
9427	103	710385.55	710385.55	0.00	0.00	0.00	
5126	5	59691.76	0.00	0.00	4254247.97	4313939.72	
...	
9225	84	152331.26	152331.26	0.00	0.00	0.00	
13123	438	4626269.69	4626269.69	0.00	4501183.12	9127452.81	
9845	145	1767886.02	1767886.02	0.00	0.00	0.00	
10799	230	10000000.00	13774746.08	3774746.08	0.00	0.00	
2732	2	147141.81	1398839.54	1545981.34	221285.85	74144.04	

10950 rows × 10 columns



In [13]:

X_test

Out[13]:

	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	type
11741	313	5818863.13	5818863.13	0.00	227196.79	6046059.92	
13248	449	133632.24	133632.24	0.00	0.00	0.00	
7082	6	37887.08	344.00	0.00	0.00	0.00	
12419	378	187761.76	187761.76	0.00	360353.54	687445.20	
6725	6	22923.28	0.00	0.00	0.00	0.00	
...
10870	235	1369544.19	1369544.19	0.00	0.00	0.00	
6739	6	13916.14	0.00	0.00	0.00	0.00	
11628	303	447009.83	447009.83	0.00	0.00	0.00	
6178	6	8793.16	836517.00	827723.84	0.00	0.00	
15786	689	276253.24	276253.24	0.00	0.00	0.00	

5476 rows × 10 columns



In [16]:

```

from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestClassifier
param_grid={'criterion':['gini', 'entropy'], 'oob_score':[True,False]}
grid=GridSearchCV(RandomForestClassifier(),param_grid,refit=True,verbose=3,n_jobs=-1)
grid.fit(X_train,Y_train)

```

C:\Users\athis\Anaconda3\lib\site-packages\sklearn\model_selection_split.py:1978: FutureWarning: The default value of cv will change from 3 to 5 in version 0.22. Specify it explicitly to silence this warning.

warnings.warn(CV_WARNING, FutureWarning)
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.

Fitting 3 folds for each of 4 candidates, totalling 12 fits

[Parallel(n_jobs=-1)]: Done 2 out of 12 | elapsed: 2.0s remaining: 10.2s

[Parallel(n_jobs=-1)]: Done 7 out of 12 | elapsed: 2.0s remaining: 1.4s

[Parallel(n_jobs=-1)]: Done 12 out of 12 | elapsed: 2.1s remaining: 0.0s

[Parallel(n_jobs=-1)]: Done 12 out of 12 | elapsed: 2.1s finished
C:\Users\athis\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
C:\Users\athis\Anaconda3\lib\site-packages\sklearn\model_selection_search.py:715: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

self.best_estimator_.fit(X, y, **fit_params)
C:\Users\athis\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:460: UserWarning: Some inputs do not have OOB scores. This probably means too few trees were used to compute any reliable oob estimates.

warn("Some inputs do not have OOB scores. ")
C:\Users\athis\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:465: RuntimeWarning: invalid value encountered in true_divide
predictions[k].sum(axis=1)[: , np.newaxis])

Out[16]:

```

GridSearchCV(cv='warn', error_score='raise-deprecating',
             estimator=RandomForestClassifier(bootstrap=True, class_weight=None,
                                               criterion='gini', max_depth=None,
                                               max_features='auto',
                                               max_leaf_nodes=None,
                                               min_impurity_decrease=0.0,
                                               min_impurity_split=None,
                                               min_samples_leaf=1,
                                               min_samples_split=2,
                                               min_weight_fraction_leaf=0.0,
                                               n_estimators='warn', n_jobs=None,
                                               oob_score=False,
                                               random_state=None, verbose=0,
                                               warm_start=False),

```

```

iid='warn', n_jobs=-1,
param_grid={'criterion': ['gini', 'entropy'],
            'oob_score': [True, False]},
pre_dispatch='2*n_jobs', refit=True, return_train_score=False,
e,
scoring=None, verbose=3)

```

In [17]:

```
Y_pred=grid.predict(X_test)
```

In [18]:

```
Y_pred
```

Out[18]:

```
array([1, 1, 0, ..., 1, 0, 1], dtype=int64)
```

In [19]:

```

from sklearn.metrics import confusion_matrix
cm=confusion_matrix(Y_test,Y_pred)
print(cm)

```

```

[[2721   1]
 [   9 2745]]

```

In [20]:

```

from sklearn.metrics import classification_report
cr=classification_report(Y_test,Y_pred)
print(cr)

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2722
1	1.00	1.00	1.00	2754
accuracy			1.00	5476
macro avg	1.00	1.00	1.00	5476
weighted avg	1.00	1.00	1.00	5476

Accuracy for RandomForestClassification Algorithm = 1.00

In [23]:

```
result=grid.predict([[743,339682.13,339682.13,0.00,0.00,339682.13,1,0,0,0]])
```


In [24]:

```
result
```

Out[24]:

```
array([1], dtype=int64)
```

In []: