

Name : SHEIKPAREETH

Machine Learning, Supervised, classification, Random forest

In [1]:

```
#Import the libraries and put nicknames
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

In [2]:

```
#Reading the dataset
dataset=pd.read_csv("fraud.csv")
```

In [3]:

```
#It have 16426 rows and 9 columns
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
	16422	743	TRANSFER	6311409.28	6311409.28	0.00	0.00
	16423	743	CASH_OUT	6311409.28	6311409.28	0.00	68488.84
	16424	743	TRANSFER	850002.52	850002.52	0.00	0.00
	16425	743	CASH_OUT	850002.52	850002.52	0.00	6510099.11

16426 rows × 9 columns

In [4]:

```
#Above the value is categorical value so i have used "one hot encoding method" (we cannot a
#I have removed the duplcates or dummy value
dataset=pd.get_dummies(dataset,drop_first=True)
```


In [8]:

```
#Evaluation metrics to use test set
#y_test output of prdicted value
y_pred=classifier.predict(X_test)
```

In [9]:

```
#Calculate confusion matrix to evaluate the accuracy of a classification
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
```

In [10]:

```
print(cm)
```

```
[[4926   0]
 [   1  1]]
```

In [11]:

```
#find clssification report (precision,recall,f1-score,accuracy )
from sklearn.metrics import classification_report
clf_report = classification_report(y_test, y_pred)
```

In [12]:

```
print(clf_report)
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4926
1	1.00	0.50	0.67	2
accuracy			1.00	4928
macro avg	1.00	0.75	0.83	4928
weighted avg	1.00	1.00	1.00	4928