


```

In [23]: import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import Normalizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import *
from sklearn.preprocessing import *

# Reading Train Set Data File
dataset = pd.read_csv('Data/Preprocessed/P_Train.csv')
testdata = pd.read_csv('Data/Preprocessed/P_Test.csv')

dataset.drop(['Jul', 'TrafficType', 'BounceRates'], axis=1, inplace=True)
testdata.drop(['Jul', 'TrafficType', 'BounceRates'], axis=1, inplace=True)

X = dataset.iloc[:, :-1]
y = dataset.iloc[:, -1]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 42, shuffle = True, stratify = None)

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.fit_transform(X_test)
X = sc.fit_transform(X)
testdata = sc.fit_transform(testdata)

rf = RandomForestClassifier(bootstrap=True, class_weight=None, max_depth=8, max_features=10, max_leaf_nodes=None,
                           min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=100,
                           n_jobs=1, oob_score=False, random_state=None, verbose=0, warm_start=False)

rf.fit(X_train, y_train)
y_pred = rf.predict(X_train)
y_pred1 = rf.predict(X_test)
y_pred2 = rf.predict(X)
y_pred3 = rf.predict(testdata)

conf1 = confusion_matrix(y_train, y_pred)
conf2 = confusion_matrix(y_test, y_pred1)
conf3 = confusion_matrix(y, y_pred2)

roc_auc1 = roc_auc_score(y_train, y_pred)
roc_auc2 = roc_auc_score(y_test, y_pred1)
roc_auc3 = roc_auc_score(y, y_pred2)

print("Train Accuracy: {:.3f}%      AUC: {:.4f}".format(((conf1[0][0]+conf1[1][1])/sum(sum(conf1)))*100, roc_auc1))

```

```
print("Test Accuracy: {:.3f}%      AUC: {:.4f} ".format(((conf2[0][0]+conf2[1][1])/sum(sum(conf2)))*100,roc_auc2))
print("Total Set Accuracy: {:.3f}%  AUC: {:.4f}".format(((conf3[0][0]+conf3[1][1])/sum(sum(conf3)))*100,roc_auc3))
print("Confusion Matrix for Test Set :\n",str(conf2))
print("Number of Ones for Test Data :",np.sum(y_pred3))
```

Train Accuracy: 93.404% AUC: 0.8348
Test Accuracy: 91.460% AUC: 0.8026
Total Set Accuracy: 92.758% AUC: 0.8239
Confusion Matrix for Test Set :
[[1705 59]
 [120 212]]
Number of Ones for Test Data : 223

```
In [19]: df=pd.DataFrame(Y,dtype=int)
df.index += 1
df.to_csv('Data/Predict_w.csv', sep=',', encoding='utf-8', header=['Revenue'], index_label='ID')
```

```
In [ ]:
```