Performance Evaluation and Model Selection

| Date	12 November 2022
Team ID	PNT2022TMID13084
Project Name	Flight Delay Prediction Using Machine Learning

Model Buliding:

Decision Tree with 0.6875834445927904

Random Forest with 0.7903871829105474

We will explore Random Forest and Decision Tree
Random Forest:

```
Testing Sensitivity for Random Forest 0.9169732002101945
Testing Specificity for Random Forest 0.23255813953488372
Testing Precision for Random Forest 0.9814398200224972
Testing accuracy for Random Forest 0.7903871829105474
```
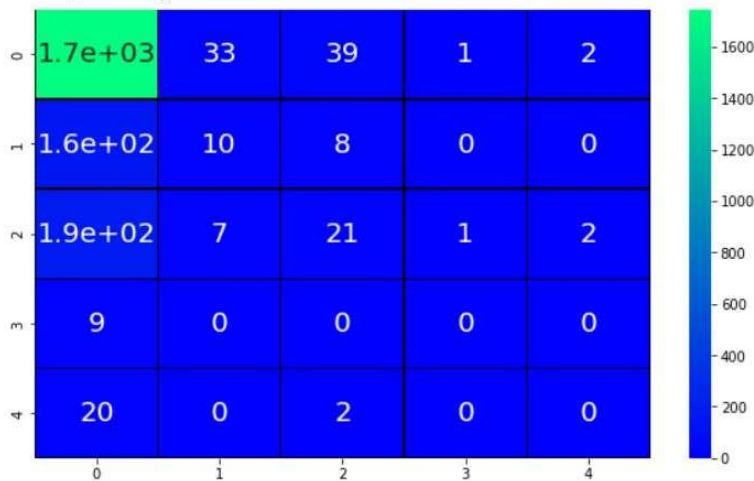

Decision Tree:

```
Testing Accuracy for Decision Tree 0.8320355951056729
Testing Sensitivity for Decision Tree 0.9201497192763568
Testing Specificity for Decision Tree 0.1076923076923077
Testing Precision for Decision Tree 0.8944815039417829
Testing accuracy for Decision Tree 0.6875834445927904
```
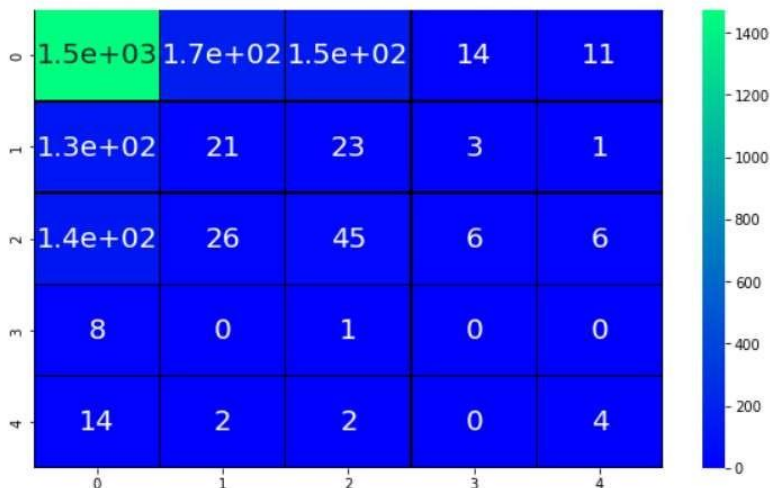
Model: Random Forest Classification performance values

```
[90] accuracy_score(y_test,pred)

     0.8095238095238095
```

```
    print(classification_report(y_test,pred))#RandomForest

                  precision    recall  f1-score   support

             0.0       0.82      0.98      0.89      1797
             1.0       0.22      0.03      0.06       181
             2.0       0.38      0.07      0.12       233
             3.0       1.00      1.00      1.00        15
             4.0       1.00      1.00      1.00        21

        accuracy                           0.81      2247
       macro avg       0.68      0.62      0.61      2247
    weighted avg       0.73      0.81      0.75      2247
```

```
    conf_mat=confusion_matrix(y_test,pred)
    conf_mat

    array([[1760,   18,   19,    0,    0],
           [ 166,    6,    9,    0,    0],
           [ 213,    3,   17,    0,    0],
           [   0,    0,    0,   15,    0],
           [   0,    0,    0,    0,   21]])
```

Random forest classification metrics

Model: Decision Tree performance values

```
[97] print(classification_report(y_test,pred1))

                  precision    recall  f1-score   support

             0.0       0.83      0.80      0.82      1797
             1.0       0.12      0.14      0.13       181
             2.0       0.16      0.18      0.17       233
             3.0       1.00      1.00      1.00        15
             4.0       1.00      1.00      1.00        21

        accuracy                           0.69      2247
       macro avg       0.62      0.62      0.62      2247
    weighted avg       0.70      0.69      0.70      2247
```

```
[98] accuracy_score(y_test,pred1)

     0.6884735202492211
```

```
[99] conf_mat=confusion_matrix(y_test,pred1)
     conf_mat

     array([[1445,  162,  190,    0,    0],
            [ 137,   25,   19,    0,    0],
            [ 162,   30,   41,    0,    0],
            [   0,    0,    0,   15,    0],
            [   0,    0,    0,    0,   21]])
```

Decision tree metrics

Model Saving:Random Forest gives the best accuracy then others , so we save random
forest model using pickle.

```
[101] import pickle

     pickle.dump(dc,open("dcmodel.pkl",'wb'))
```