Tool and Techniques Name

1) Python Libraries :-
    - Numpy
    - Pandas
    - Matplotlib
    - Seaborn
    - Sklearn
    - Imblearn
    - Tensorflow
    - Sklearn.metrics
    - Sklearn.preprocessing
2) Dataset Link :-
    - ✓ Lithuanian airports flight dataset | Kaggle
    - ✓ Here we have used 2 dataset i.e, Arrival and Departure Dataset.
3) Preprocessing Techniques : -
    - Label encoding to encode the categorical data
    - filled the missing values by using mode of that column.
4) Classification Technique

    We have 6 classes in our base3 paper to classify the time deviation i.e,

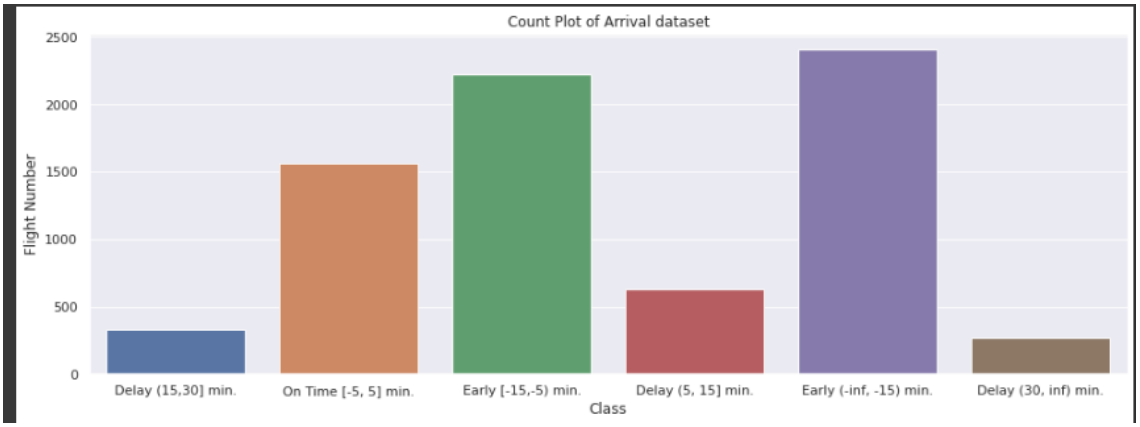
    - Delay(15,30] min
    - On Time [-5,5] min
    - Early [-15,5} min
    - Delay (5,15] min
    - Early (-inf,-15) min
    - Delay (30,inf) min

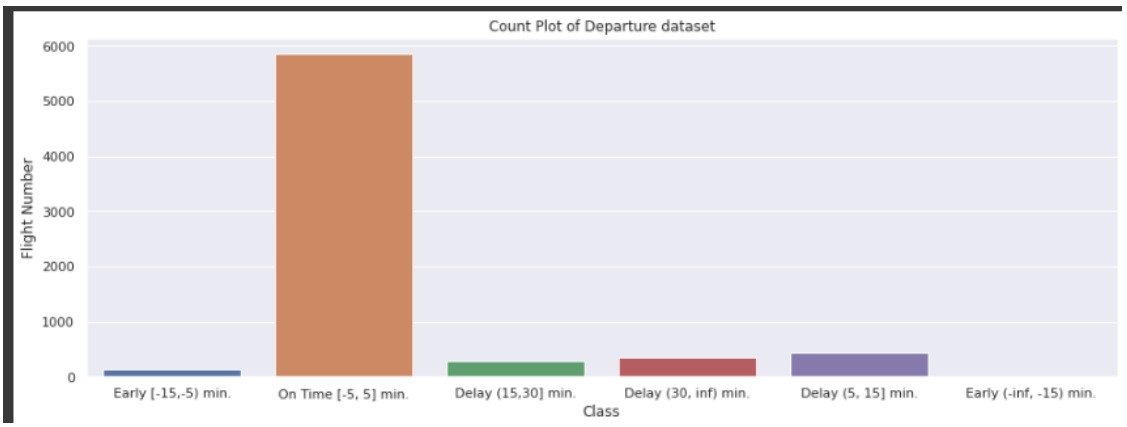
5) Data Split Ratio
    - ✓ Splitted the data into 80:20
6) Base Model
    - Decision tree classifier
    - Random Forest classifier
    - Gradient Boosting Classifier

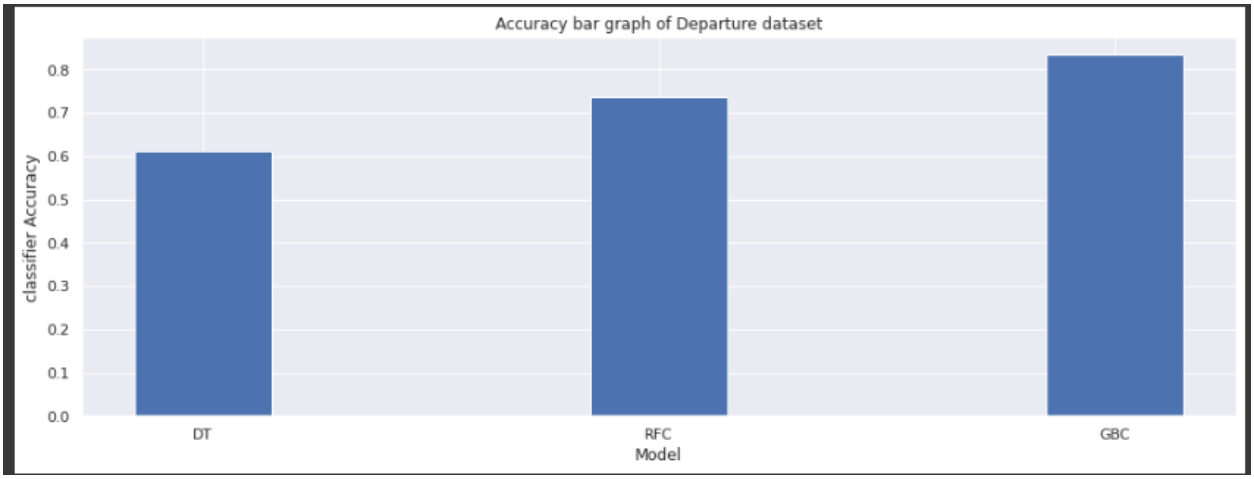
if any another tool and technique used so pls include it and remove it to above points

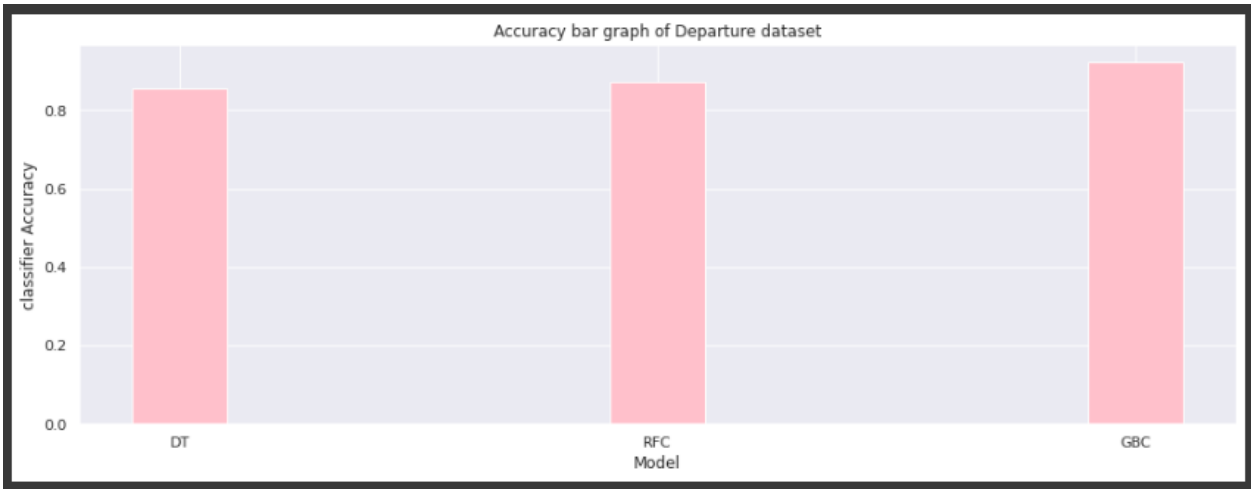
Screenshots of base and proposed results

1) Dataset visualization screenshots









2) Preprocessing results

```
<bound method DataFrame.nunique of            DATE           COMPANY.1              TO       FROM  \
0     2019-10-25                SAS        Kopenhaga  PalangaPLQ
1     2019-10-25          AirBaltic        TalinasTLL  VilniusVNO
2     2019-10-25            WizzAir  Londonas/Lutonas  PalangaPLQ
3     2019-10-25      CorendonEurope        HurgadaHRG  VilniusVNO
4     2019-10-25  ScandinavianAirlines      KopenhagaCPH  VilniusVNO
...          ...                 ...               ...         ...
7404  2020-03-14            WizzAir       LondonasLTN  VilniusVNO
7405  2020-03-14            WizzAir    ReikjavikasKEF  VilniusVNO
7406  2020-03-14                NaN        MilanasMXP  VilniusVNO
7407  2020-03-14          AirBaltic           RygaRIX  VilniusVNO
7408  2020-03-14        HOLIDAYEUROPE        HurgadaHRG  VilniusVNO

     TEMPERATURE                SKY INFORMATION WIND SPEED  WIND ANGLE  \
0             13                           Fog.     7 km/h         150
1          11 °C                         Sunny.    22 km/h         180
2          13 °C             Drizzle. Dense fog.     6 km/h         200
3          11 °C                         Sunny.    22 km/h         180
4          11 °C                         Sunny.    22 km/h         180
...          ...                            ...       ...         ...
7404        1 °C   Snow showers. Passing clouds.    20 km/h         220
7405        1 °C  Snow flurries. Passing clouds.    19 km/h         210
7406        2 °C                Passing clouds.    17 km/h         220
7407       -2 °C    Snow flurries. Partly sunny.    41 km/h         350
7408       -2 °C    Snow flurries. Partly sunny.    41 km/h         350

     VISIBILITY SCHEDULED TIME LANDED TIME                  CLASS
0         4 km       00:55:00    01:18:00      Delay (15,30] min.
1        16 km       10:15:00    10:14:00    On Time [-5, 5] min.
2         1 km       10:20:00    10:07:00     Early [-15,-5) min.
3        16 km       10:30:00    10:28:00    On Time [-5, 5] min.
4        16 km       10:35:00    10:43:00      Delay (5, 15] min.
...        ...            ...         ...                   ...
7404      4 km       02:35:00    02:06:00  Early (-inf, -15) min.
7405       NaN       03:15:00    02:54:00  Early (-inf, -15) min.
7406       NaN       03:45:00    04:18:00     Delay (30, inf) min.
7407       NaN       08:30:00    08:34:00    On Time [-5, 5] min.
7408       NaN       08:40:00    08:16:00  Early (-inf, -15) min.

[7409 rows x 12 columns]>
```

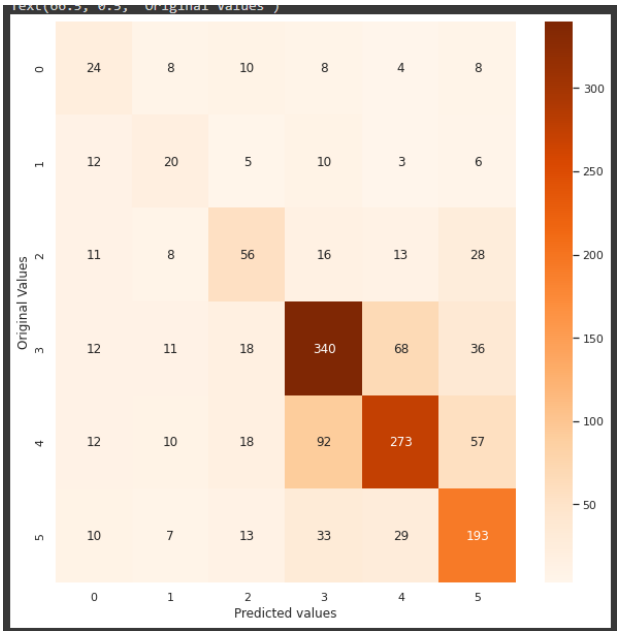
```
<bound method DataFrame.nunique of            DATE                   COMPANY.1        TO            FROM  \
0     2019-10-25                AirBaltic  PalangaPLQ            Ryga
1     2019-10-25                     SAS  PalangaPLQ        Kopenhaga
2     2019-10-25                AirBaltic  VilniusVNO          RygaRIX
3     2019-10-25                  WizzAir  PalangaPLQ  Londonas/Lutonas
4     2019-10-25                 Aeroflot  VilniusVNO         MaskvaSVO
...          ...                      ...         ...              ...
7052  2020-03-16  UkraineInternationalAirlines  VilniusVNO       KijevasKBP
7053  2020-03-16                   Belavia  VilniusVNO       MinskasMSQ
7054  2020-03-16                   Ryanair   KaunasKUN     BristolisBRS
7055  2020-03-16                AirBaltic  PalangaPLQ            Rygа
7056  2020-03-16                AirBaltic  PalangaPLQ            Ryga

     TEMPERATURE          SKY INFORMATION WIND SPEED  WIND ANGLE VISIBILITY  \
0          12 °C                    Fog.     4 km/h         170      2 km
1          12 °C                    Fog.     4 km/h         170      2 km
2          11 °C                  Sunny.    26 km/h         180     16 km
3          13 °C       Drizzle. Dense fog.     6 km/h         200      1 km
4          11 °C                  Sunny.    22 km/h         190     16 km
...          ...                     ...       ...         ...       ...
7052        7 °C                  Sunny.    24 km/h         210     16 km
7053        7 °C                  Sunny.    20 km/h         210     16 km
7054        4 °C  Drizzle. Mostly cloudy.    19 km/h         240      8 km
7055        4 °C                  Clear.    11 km/h         200     20 km
7056        4 °C    Light rain. Overcast.     9 km/h         250     11 km

     SCHEDULED TIME DEPARTURE TIME               CLASS
0          05:40:00       05:34:00   Early [-15,-5) min.
1          07:00:00       06:56:00  On Time [-5, 5] min.
2          10:45:00       10:45:00  On Time [-5, 5] min.
3          10:50:00       10:55:00  On Time [-5, 5] min.
4          11:05:00       11:05:00  On Time [-5, 5] min.
...             ...            ...                 ...
7052       17:05:00       22:53:00  Delay (30, inf) min.
7053       17:40:00       17:40:00  On Time [-5, 5] min.
7054       20:55:00       21:14:00    Delay (15,30] min.
7055       05:50:00       05:47:00  On Time [-5, 5] min.
7056       15:10:00       15:22:00    Delay (5, 15] min.

[7057 rows x 12 columns]>
```

3) Feature extraction/selection results
4) Classification Results
5) Base model results
    - Arrival Dataset
        - Decision Tree Classifier

	precision	recall	f1-score	support
0	0.30	0.39	0.34	62
1	0.31	0.36	0.33	56
2	0.47	0.42	0.44	132
3	0.68	0.70	0.69	485
4	0.70	0.59	0.64	462
5	0.59	0.68	0.63	285
accuracy			0.61	1482
macro avg	0.51	0.52	0.51	1482
weighted avg	0.62	0.61	0.61	1482

Accuracy: 0.611336032388664

```
Precision by DT of testing data is: 0.611
Recall by DT of testing data is: 0.611
F1 score by DT of testing data is: 0.611
```

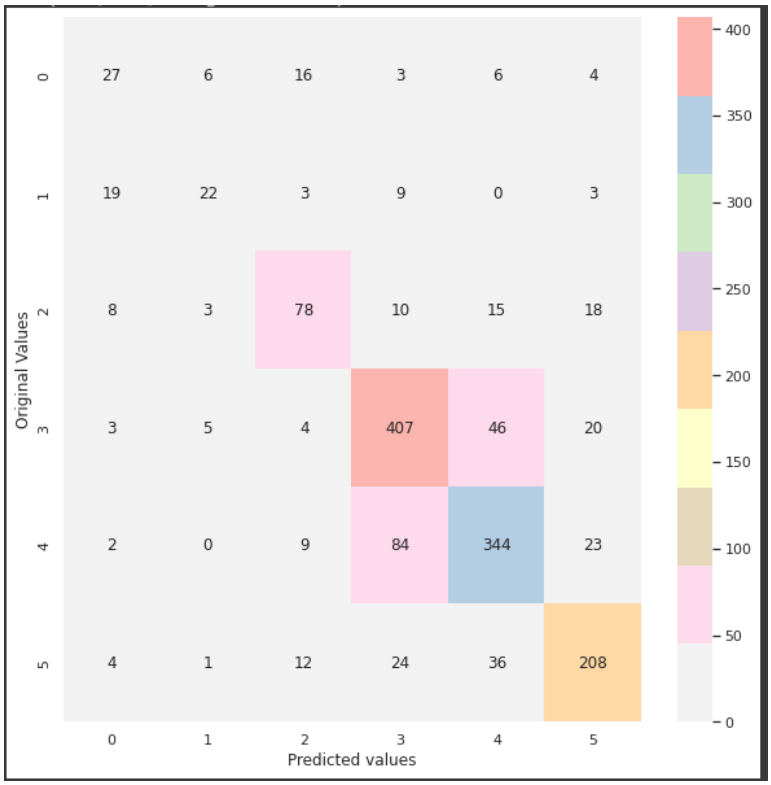


- Random Forest:-

	precision	recall	f1-score	support
0	0.43	0.47	0.45	62
1	0.63	0.43	0.51	56
2	0.64	0.58	0.61	132
3	0.76	0.85	0.80	485
4	0.77	0.75	0.76	462
5	0.75	0.72	0.74	285
accuracy			0.74	1482
macro avg	0.67	0.63	0.64	1482
weighted avg	0.74	0.74	0.73	1482

Accuracy: 0.7368421052631579

```
Precision by RF of testing data is: 0.737
Recall by RF of testing data is: 0.737
F1 score by RF of testing data is: 0.737
```

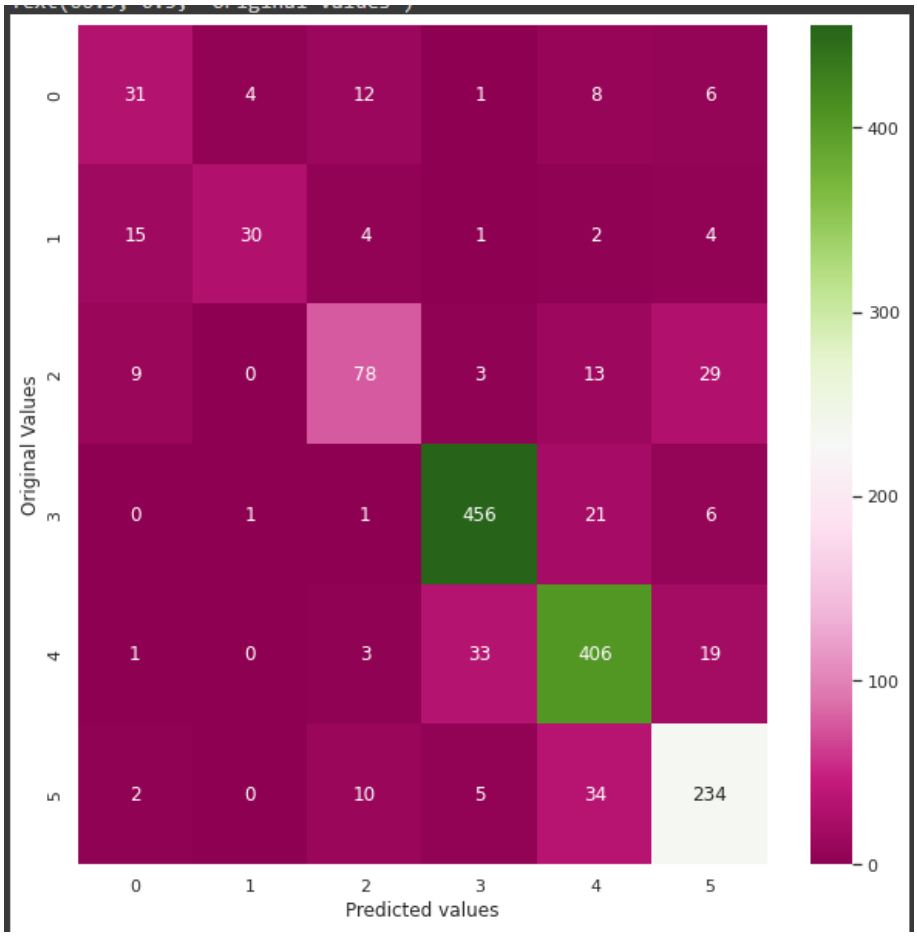


- **Gradient Boosting: -**

	precision	recall	f1-score	support
0	0.53	0.50	0.52	62
1	0.86	0.54	0.66	56
2	0.72	0.59	0.65	132
3	0.91	0.94	0.93	485
4	0.84	0.88	0.86	462
5	0.79	0.82	0.80	285
accuracy			0.83	1482
macro avg	0.78	0.71	0.74	1482
weighted avg	0.83	0.83	0.83	1482

Accuracy: 0.8333333333333334

```
Precision by GBC of testing data is: 0.833
Recall by GBC of testing data is: 0.833
F1 score by GBC of testing data is: 0.833
```

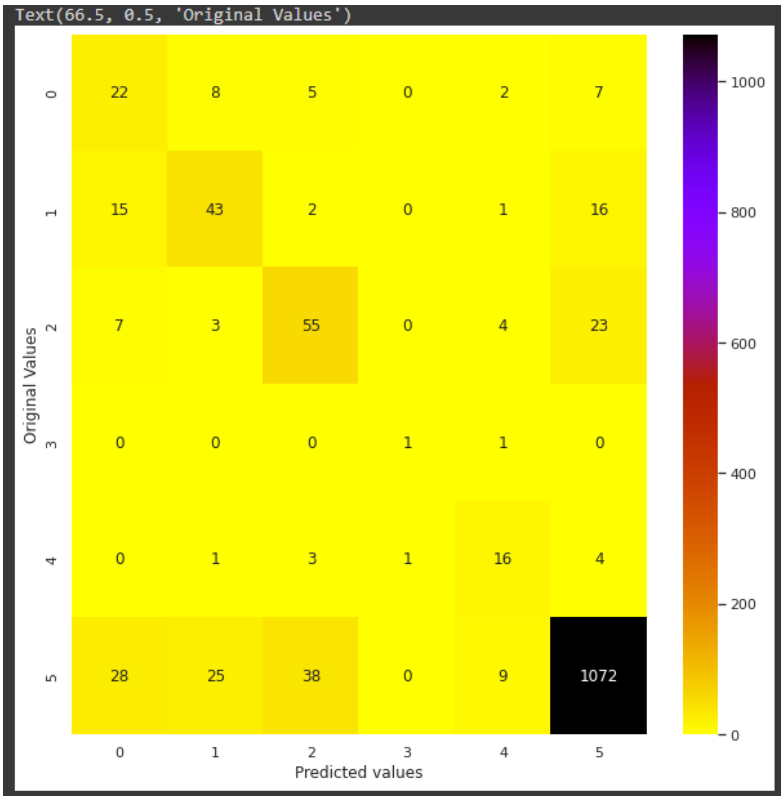


- **Departure Dataset**
  - **Decision Tree**

	precision	recall	f1-score	support
0	0.31	0.50	0.38	44
1	0.54	0.56	0.55	77
2	0.53	0.60	0.56	92
3	0.50	0.50	0.50	2
4	0.48	0.64	0.55	25
5	0.96	0.91	0.93	1172
accuracy			0.86	1412
macro avg	0.55	0.62	0.58	1412
weighted avg	0.88	0.86	0.86	1412

Accuracy: 0.8562322946175638

```
Precision by DT of testing data is: 0.856
Recall by DT of testing data is: 0.856
F1 score by DT of testing data is: 0.856
```

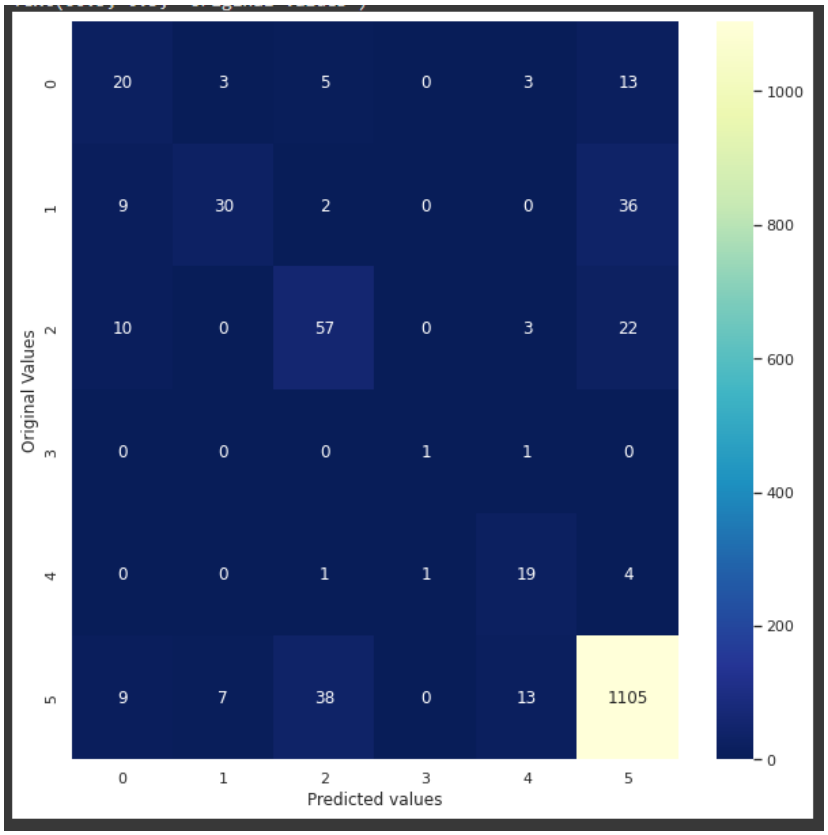


- **Random Forest:-**

	precision	recall	f1-score	support
0	0.42	0.45	0.43	44
1	0.75	0.39	0.51	77
2	0.55	0.62	0.58	92
3	0.50	0.50	0.50	2
4	0.49	0.76	0.59	25
5	0.94	0.94	0.94	1172
accuracy			0.87	1412
macro avg	0.61	0.61	0.59	1412
weighted avg	0.88	0.87	0.87	1412

```
Accuracy:  0.8725212464589235
```

```
Precision by RF of testing data is: 0.873
Recall by RF of testing data is: 0.873
F1 score by RF of testing data is: 0.873
```

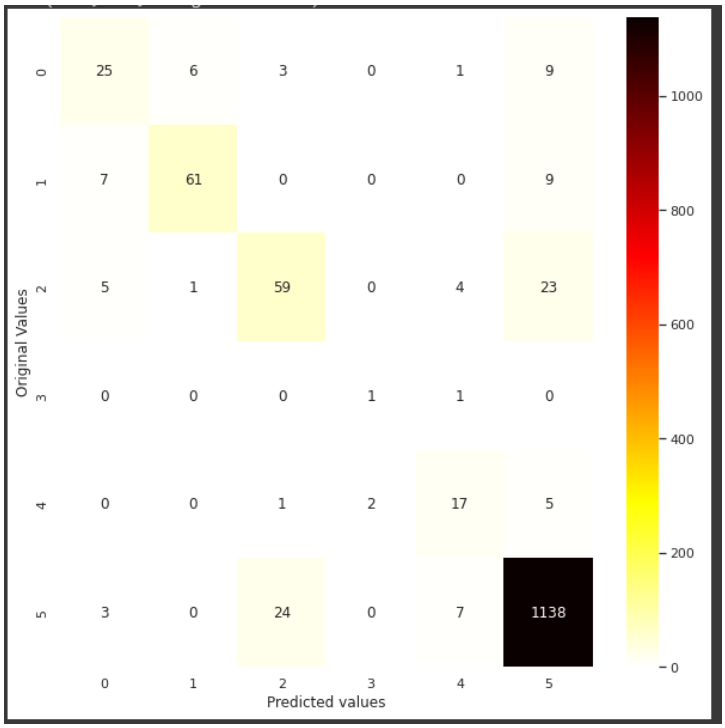


- **Graident Boosting:-**

	precision	recall	f1-score	support
0	0.62	0.57	0.60	44
1	0.90	0.79	0.84	77
2	0.68	0.64	0.66	92
3	0.33	0.50	0.40	2
4	0.57	0.68	0.62	25
5	0.96	0.97	0.97	1172
accuracy			0.92	1412
macro avg	0.68	0.69	0.68	1412
weighted avg	0.92	0.92	0.92	1412

```
Accuracy:  0.9213881019830028
```

```
Precision by GBC of testing data is: 0.921
Recall by GBC of testing data is: 0.921
F1 score by GBC of testing data is: 0.921
```



Base Result table

ARRIVAL DATASET

MODEL	Accuracy	Precision	Recall	F1 Score
DT	61.13%	0.61	0.61	0.61
RFC	73.68%	0.73	0.73	0;.73
GBC	83.33%	0.83	0.83	0.83

`Q`

DEPARTURE DATASET

MODEL	Accuracy	Precision	Recall	F1 Score
DT	85.62%	0.85	0.85	0.85
RFC	87.25%	0.87	0.87	0.87

GBC	92.13%	0.92	0.92	0.92

Base and proposed results comparison table

ARRIVAL DATASET

MODEL	Accuracy	Precision	Recall	F1 Score
Proposed	91.02	0.91	0.91	0.91
Base	83.33%	0.83	0.83	0.83

DEPARTURE DATASET

MODEL	Accuracy	Precision	Recall	F1 Score
Proposed	98.94	0.98	0.98	0.98
Base	92.13%	0.92	0.92	0.92