Flight Ticket Price Predicting

Submitted by:

Smita More

Goals

- Forecast flight prices
- Selecting optimum time for travel
- Selecting the cheapest flight to the desired destination

Scraping – Scraping data from Kayak.com. I scrape following information about flight.

- Source & Destination
- Date (2nd June to 10th June, 2022)
- Price
- Duration
- Total Stops
- Airline

Scraping Routes – I scrape following routes

Pune – PNQ

Mumbai – BOM

Delhi – DEL

Chennai – MAA

Bangalore – BLR

Kolkata – CCU

Ahmedabad – AMD

Merge all data in single excel file

Model Building - *The Random Forest model performed the best.*

424]:
```
test_df = pd.DataFrame({
    "Predicted Price" : rf.predict(X_test),
    "Actual Price" : y_test,
}).reset_index(drop = True)

test_df
```

424]:

	Predicted Price	Actual Price
0	437.210000	344
1	435.390083	475
2	455.840000	410
3	1400.370000	443
4	1245.050000	1445
...	...	...
339	487.985079	527
340	733.443333	1741
341	566.526904	434
342	564.270000	529
343	392.280000	402

344 rows × 2 columns

```
# saving the model
import pickle as pkl
file = open('rf_flightprice_prediction.pkl', 'wb')
pkl.dump(rf, file)
```

```
# open the model
model = open('rf_flight_prediction.pkl','rb')
rf_flight_prediction = pkl.load(model)
```

```
print(f'R2 score {metrics.r2_score(y_test,rf_flight_prediction.predict(X_test))}')
print("MAE:" , metrics.mean_absolute_error(y_test,rf_flight_prediction.predict(X_test)))
print("MSE:" , metrics.mean_squared_error(y_test,rf_flight_prediction.predict(X_test)))
print("RMSE:" , np.sqrt(metrics.mean_squared_error(y_test,rf_flight_prediction.predict(X_test))))
```

```
R2 score 0.65587020820137
MAE: 62.33693311888358
MSE: 31719.42027579546
RMSE: 178.09946736527726
```

The final model is able to predict flight ticket prices within around = $62.33