

## FLIGHT PRICE PREDICTION

**STEP 1. Data Collection** You have to scrape at least 1500 rows of data. You can scrape more data as well, it's up to you, More the data better the model In this section you have to scrape the data of flights from different websites (yatra.com, skyscanner.com, official websites of airlines, etc). The number of columns for data doesn't have limit, it's up to you and your creativity. Generally, these columns are airline name, date of journey, source, destination, route, departure time, arrival time, duration, total stops and the target variable price. You can make changes to it, you can add or you can remove some columns, it completely depends on the website from which you are fetching the data.

**2. Model Building** After collecting the data, you need to build a machine learning model. Before model building do all data pre-processing steps. Try different models with different hyper parameters and select the best model. Follow the complete life cycle of data science. Include all the steps like 1. Data Cleaning 2. Exploratory Data Analysis 3. Data Pre-processing 4. Model Building 5. Model Evaluation 6. Selecting the best model

1)Data cleaning: we first clean the data i.e. checking for the null values if any

If the null values are found drop the rows with the null values.

So the null values are removed

Date of journey column is converted into Datetime datatype.

So after converting we drop the Date of journey column

Next we convert ArrivalTime,Dept time columns and drop them.

2)EDA:

Outlier detection is performed .Outliers are removed.

Different plottings i.e. distplot,boxplot,pairplot..etc

3)Data Pre-Processing: we convert all the categorical into continuous columns

Then we use the standard scaler method

4)Model building :we convert the dataframe into x and Y

Then we split the data into train and test split

Then we use the mutual info classify method for feature selection.

Then we train the model using different ml models

i.e.LogisticRegression,svr,adaboostClassifier,GradientBoostingClassifier,RandomForestClassifier.etc

We select the one with the best accuracy i.e.RandomForestClassifier in our case.

5)Evaluate the data using the best model.

6)HyperparameterTuning is used with RandomForestRegressor with optimum parameters to get the accuracy high with the best model i.e.RandomForestRegressor.

Then we save the model

