# Step for making Model on Flight\_Ticket\_Predtion

By-Abhishek Ranjan

Step1: Importing necessary libraries.

Step2:Initializing the dataFrame and seeing the dataset.

Conclusion:

.There are 10683 rows and 11 columns.

.Except the label column ‘Price’ , all the columns

Are having ‘Object’ datatype.

Step3:Checking the count of non-null and datatype of the dataset.

Conclusion:

.Null values are present in two columns.

.There are 11 column of ‘Object’ as model doesn’t accept Categorical values. So it must be Encode into int or float datatype.

Step 4: We Can drop ‘Dep\_Time’ and ‘Arrivial\_Time’ and ‘Date\_of\_Journey’ column as it has no important value in creating dataset.

Step 5: Now, We have to fill the null-value of all the remaining

Column Using mean and mode method.

Conclusion:

.Now all the null- values gets filled. Now, we

Are ready to move forward.

Step 6: We have to drop duplicates data .if found by using

.drop\_duplicates().

Conclusion: Now there are total 10463 rows left.

Step 7: We are going to Encode those column, having

‘Object’ datatype Using OrdinalEncoder.

Conclusion:

.All the column having ‘Object’ datatype

are now encoded into ‘int’ or ‘float’

datatype.

Step 8: OverRiding all the encoded column into dataframe.

Conclusion:

.Now model is ready to use further.

Step 9: Getting the summary of the dataset using describe ()

Method.

Conclusion:

.No Null value found.

. There may be outliers Present in the dataset

As the difference between mean and std is

Less it must be as large as possible.

Step 10: Finding and Removing Outliers by using boxplot.

Conclusion: Before outliers removal ,these are 10463 rows

Present but after removal of outliers only 7069

rows left in the dataset.

Step 11: Checking correlation between the features, to check

Multicolinearity .

Conclusion: We were checking correlation using vif.

All the vif values are less then 5 and very low.

That means no multicollinearity in dataset.

Step 12: Dividing the dataset into feature and label.

Step 13: Doing Data Scaling before training data.

Step14:Splitting the Dataset into train and test data.

Step 15: Importing Linear Regression algorithm for model

training and prediction.

Step 16: After fitting the model,Prediction done by using

.Predict() .

Step 17: Doing model Evaluation.

Conclusion: Model fits 31% of dataset.

Step 18: Doing HyperParameter Tunning Using

GridSearchCV Using RandomForestRegressor.

Conclusion: Model fits 70% of data.

Step 19: We use LassoCV and RidgeCV to check wheather the

Model is overfitting or not

Conclusion: we get same r2 score in both LassoCV and RidgeCV

.so,we can say that model is no overfitting .

Step20: Saving the model using Pickle.

**THANK YOU.**