

Flight Price Prediction

Submitted by:

Suganya

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

**INTRODUCTION**

* Business Problem Framing

Predicting the flight Price is the problem statement.There are two predominant use instances of flight rate prediction in the tour industry. OTAs and different tour structures combine this function to entice extra traffic searching for the excellent rates. Airlines appoint the science to forecast costs of rivals and regulate their pricing techniques accordingly.

* Conceptual Background of the Domain Problem

Pricing in the airline enterprise is frequently in contrast to a Genius recreation between carriers and passengers the place every birthday party pursues the exceptional rates. Carriers love promoting tickets at the easiest fee viable — whilst nevertheless now not dropping buyers to competitors. Passengers are loopy about shopping for flights at the lowest price accessible — whilst now not lacking the threat to get on board. All this makes flight expenditures fluctuant and challenging to predict. But nothing is not possible for human beings armed with mind and algorithms..

* Review of Literature

Flight Fare Prediction Using Machine Learning -Aman Preet Gulati — January 18, 2022: By comparing all the models (Ridge Regression, Lasso Regression, Decision Tree Regressor, Random Forest Regressor), we can conclude that Decision Tree Regressor and Random Forest Regressor performs the best.

* Motivation for the Problem Undertaken

EDA: Learn the complete process of EDA

Data analysis: Learn to withdraw some insights from the dataset both mathematically and visualize it.

Data visualization: Visualising the data to get better insight from it.

Feature engineering: We will also see what kind of stuff we can do in the feature engineering part.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

There are null values and dropped unnecessary rows.Used numerical code for categorical data.

* Data Sources and their formats

Data are collected from Yatra.com, makemytrip, Paytm.

* Data Preprocessing Done

Dropped No values in rows, unnecessary data.

Label Encode and Hot Encode for Categorical Columns.

Changed Object and Float into int.

* Data Inputs- Logic- Output Relationships

Departure place and Duration decides the Price.

* State the set of assumptions (if any) related to the problem under consideration

As booking the tickets in (advance)reservation makes the price less.

* Hardware and Software Requirements and Tools Used

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import mean\_squared\_error as mse

from sklearn.metrics import r2\_score

from math import sqrt

from sklearn.linear\_model import Ridge

from sklearn.linear\_model import Lasso

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

from sklearn.preprocessing import LabelEncoder

from sklearn.model\_selection import KFold

from sklearn.model\_selection import train\_test\_split

from sklearn.model\_selection import GridSearchCV

from sklearn.model\_selection import RandomizedSearchCV

from sklearn.linear\_model import LinearRegression

import pickle.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Finding the Correlation between the feature and output, which will decide the price.

* Testing of Identified Approaches (Algorithms)

Linear Regression, Decision Tree Regressor, Forest Regressor.

* Run and Evaluate selected models

Linear Regression, Decision Tree Regressor, Forest Regressor.

* Key Metrics for success in solving problem under consideration

Train-test split, Kfold, GridsearchCV, RandomizedSearchCV.

* Visualizations

Catplot, distplot, heatmap

* Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Describe the key findings, inferences, observations from the whole problem.

* Learning Outcomes of the Study in respect of Data Science

List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

* Limitations of this work and Scope for Future Work

Data collected with very few websites, other than this features many other affects the prediction of flight price, for example, environmental situation, etc.