Flight Price Prediction

Submitted by:

Gourav kumar

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

Airline companies use complex algorithms to calculate flight prices given various conditions present at that particular time. These methods take financial, marketing, and various social factors into account to predict flight prices.

Nowadays, the number of people using flights has increased significantly. It is difficult for airlines to maintain prices since prices change dynamically due to different conditions. That's why we will try to use machine learning to solve this problem. This can help airlines by predicting what prices they can maintain. It can also help customers to predict future flight prices and plan their journey accordingly.

Airline companies use complex algorithms to calculate flight prices given various conditions present at that particular time. These methods take financial, marketing, and various social factors into account to predict flight prices.

Nowadays, the number of people using flights has increased significantly. It is difficult for airlines to maintain prices since prices change dynamically due to different conditions. That's why we will try to use machine learning to solve this problem. This can help airlines by predicting what prices they can

maintain. It can also help customers to predict future flight prices and plan their journey accordingly.

Analytical Problem Framing

This article was published as a part of the Data Science Blogathon

This Blog deals with the problem of flight price prediction.

1. Objective
The objective of this article is to predict flight prices given the various parameters. Data used in this article is publicly available at Kaggle. This will be a regression problem since the target or dependent variable is the price (continuous numeric value).

2. Introduction
Airline companies use complex algorithms to calculate flight prices given various conditions present at that particular time. These methods take financial, marketing, and various social factors into account to predict flight prices.

Nowadays, the number of people using flights has increased significantly. It is difficult for airlines to maintain prices since prices change dynamically due to different conditions. That's why we will try to use machine learning to solve this problem. This can help airlines by predicting what prices they can maintain. It can also help customers to predict future flight prices and plan their journey accordingly.

3. Data Used
Data was used from Kaggle which is a freely available platform for data scientists and machine learning enthusiasts.

Loading Image

Source: https://www.kaggle.com/nikhilmittal/flight-fare-prediction-mh

We are using jupyter-notebook to run Flight Price Prediction task.

4. Data Analysis
The procedure of extracting information from given raw data is called data analysis. Here we will use eda module of data-prep library to do this step.

```
from dataprep.eda import create_report
```

```
import pandas as pd
dataframe = pd.read_excel("../output/Data_Train.xlsx")
create_report(dataframe)
```

Lazy Prediction overview

After running the above code you will get a report as shown in the above figure. This report contains various sections or tabs. 'Overview' section of this report provides us with all the basic information of the data we are using. For the current data we are using we got the following information:

Number of variables = 11
Number of rows = 10683
Number of categorical type of feature = 10
Number of numerical type of feature = 1
Nuplicate rows = 220 e.t.c

Let's explore other sections of the report one by one.

4.1 Variables
After you select the variable section you will get information as shown in the below figures.

Lazy Prediction variables

Lazy Prediction variables 2

Lazy Prediction variables 2

This section provides the type of each variable along with a detailed description of the variable.

4.2 Missing Values
This section has multiple ways using which we can analyze missing values in variables. We will discuss three mostly used methods, bar-chart, spectrum, and Heat Map. Let's explore each one by one.

4.2.1 Bar Chart
Lazy Prediction bar chart

The bar chart method shows the 'number of missing and present values' in each variable in a different color.

4.2.2 Spectrum
Lazy Prediction spectrum

The spectrum method shows the percentage of missing values in each variable.

4.2.3 Heat Map
Heat map Lazy Prediction

The heat Map method shows variables having missing values in terms of correlation. Since 'Route' and 'Total_Stops' both are highly correlated, they both have missing values.

As we can observe 'Route' and 'Total_Stops' variables have missing values. Since we did not find any missing values information from Bar-Chart and Spectrum method but we found missing value variables using the Heat Map method. Combining both of these information, we can say that the 'Route' and 'Total_Stops' variables have missing values but are very low.

5. Data Preparation
Before starting data preparation let's have a glimpse of data first.

dataframe.head()
Lazy Prediction head

As we saw in Data Analysis there are 11 variables in the given data. Below is the description of each variable.

Airline: Name of the airline used for traveling

Date_of_Journey: Date at which a person traveled

Source: Starting location of flight

Destination: Ending location of flight

Route: This contains information on starting and ending location of the journey in the standard format used by airlines.

Dep_Time: Departure time of flight from starting location

Arrival_Time: Arrival time of flight at destination

Duration: Duration of flight in hours/minutes

Total_Stops: Number of total stops flight took before landing at the destination.

Additional_Info: Shown any additional information about a flight

Price: Price of the flight

Few observations about some of the variables:

1. 'Price' will be our dependent variable and all remaining variables can be used as independent variables.

2. 'Total_Stops' can be used to determine if the flight was direct or connecting.

Model/s Development and Evaluation

- Applying KNeighborsRegressor Prediction
- One of the problems of the model-building exercise is 'How to decide which machine learning algorithm to apply ?'

- This is where KNeighborsRegressor Prediction comes into the picture. KNeighborsRegressor Prediction is a machine learning library available in python that can quickly provide us with performances of multiple standard classifications or regression models on multiple performance matrices.

CONCLUSION

- In this article, we saw how to apply KNeighborsRegressor Prediction library to choose the best machine learning algorithm for the task at hand.

- Lazy Prediction saves time and efforts to build a machine learning model by providing model performance and training time. One can choose either based on the situation at hand.

- It can also be used to build an ensemble of machine learning models. There are so many ways one can use the KNeighborsRegressor library's functionalities.

- I hope this article helped you to understand Data Analysis, Data Preparation, and Model building approaches in a much simpler way