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.

## Data Used

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

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

## 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.

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

Data Preparation

Before starting data preparation let’s have a glimpse of data first.

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.

Conclusion

In this article, we saw how to apply Laze 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 LazyPredict library’s functionalities.

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

Please reach out in the comment section in case of any query.