



FLIGHT PRICE PREDICTION

Submitted by:

PUROBI DEBNATH

ACKNOWLEDGMENT

I would like to thank Flip Robo Technologies for giving projects like this and thanks to Keshav bansal who thought that I would be able to do this project .

This project helped me understand more about data analysis and I got to improve my skills as well. I have taken a lot of efforts to complete this project, but without the help of internet and our team leader could not have been able to complete this project.

INTRODUCTION

- Business Problem Framing

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 does not have limit, it's up to you and your creativity. Generally these columns are airline names, arrival time, duration, total stops and the target variable price. You can make changes to it, you can add or remove some columns. It completely depends on the website from which you are fetching the data.

- Conceptual Background of the Domain Problem

First I scrapped the data from yatra.com,skyscanner.com and all the official websites of airline and then built the machine learning model that understands us which flight is costlier or cheaper. I have performed different visualisation analysis to get graphically understand the data. I have also done some statistical techniques of the data.

- Review of Literature

The problem I sought to solve was predicting the price of the flight, which gives proper understanding to the client where these stands.

Our goal is to build a prototype of predicting flight price that can control the loss and the profit in the future.

- Motivation for the Problem Undertaken

Every project begin with the idea that are further developed and inspired variety of situation and circumstances..

The client want some predictions of the prices that could help them in further investment and improvement in selection of tickets. So to help them I made this project. My motivation behind this project is to upgrade my skills and to learn new things from internet while doing this project. So after doing all the necessary data analysis and programming to give perfect solution to the client.

Analytical Problem Framing

- **Mathematical/ Analytical Modeling of the Problem**

First I checked the information of the data set that is, the shape, info, columns and by applying pandas it gives all the information about the data types. After that I checked null values, their were null values present in the data set and I filled the null values with mean and median.

- **Data Sources and their formats**

The data which I scrapped all in excel csv format. The data set contains the training set which has 10683 rows and 11 columns and in the test data set it has 2671 rows and 10 columns.

Training simple data contains 11 features which includes Airline, Date of journey, source, destinationroute, de_time, arrival_time, duration, total_stops, additional_info, price.

- **Data Preprocessing Done**

First I checked the information of the data set and the shape of the data set, it gives the information about the type of data present in the data set and data type of columns which is integer, object or float. After that I have created new columns for feature selection and data mining.

- Hardware and Software Requirements and Tools Used

*Processor- Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz 2.00
GHz*

Installed ram- 8.00 GB

System type- 64-bit operating system, x64-based processor