FLIGHT PRICE PREDICTION

Submitted by:

Lokesh Bisen

ACKNOWLEDGMENT

INTRODUCTION

- **Business and Conceptual Background of the Domain Problem**

Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. In this project we majorly targeted to uncover underlying trends of flight prices in India using historical data and also to suggest the best time to buy a flight ticket.

For this project, we have collected data from major routes across India while the data of 4 routes were extensively used for the analysis due to the sheer volume of data collected over 4 months resulting in 1500 data points each across the Mumbai-Delhi and Delhi-Mumbai route and other major routes.The project implements the validations or contradictions towards myths regarding the airline industry, a comparison study among various models in predicting the optimal time to buy the flight ticket and the amount that can be saved if done so. These models have led to significant savings and produced average positive savings on each transaction.

Remarkably, the trends of the prices are highly sensitive to the route,time of departure, Highly competitive routes like most business routes (tier 1 to tier 1 cities like Mumbai-Delhi) had a non-decreasing trend where prices increased as days to departure decreased, however other routes (tier 1 to tier 2 cities like Delhi - Chandigrah) had a specific time frame where the prices are minimum. The data also validated the fact that, there are certain time-periods of the day where the prices are expected to be maximum.

- ## Review of Literature

  In order to create this model I have collected data from various websites which are offers the flights tickets online. Our data set contains 1562 data points of different cities. There are no null values in the data set for in the arrival time columns some of the data points has time in nanoseconds which I have to remove and replace them.

- ## Motivation for the Problem Undertaken

  Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, it will be a different story. We might have often heard travelers saying that flight ticket prices are so unpredictable. As data scientists, we are gonna prove that given the right data anything can be predicted. Here you will be provided with prices of flight tickets for various airlines between the months of October and November of 2021 and between various cities.

  With a high probability (about 20-25%) that a person has to wait to buy a ticket, the scope of the project can be extensively extended across the various routes to make significant savings on the purchase of flight prices across the Indian Domestic Airline market.

Analytical Problem Framing

● Data Preprocessing Done

Data exploration and preprocessing is the first step in data analysis and typically involves summarizing the main characteristics of a data set, including its size, accuracy, initial patterns in the data and other attributes. It is commonly conducted by data analysts using visual analytics tools, but it can also be done in more advanced statistical software, Python. Before it can conduct analysis on data collected by multiple data sources and stored in data warehouses, an organization must know how many cases are in a data set, what variables are included, how many missing values there are and what general hypotheses the data is likely to support. An initial exploration of the data set can help answer these questions by familiarizing analysts with the data with which they are working. In addition In the Data preprocessing we have checked null values using pandas library and in order to get better insights from data set used visualization technique using matplotlib and seaborne library.

● Hardware and Software Requirements and Tools Used

In order to build a good model it is necessary to have good computational power hardware and software's. The used hardware for this project 8GB RAM laptop with I5 processor.And we have used python and its respective libraries.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods) and Testing , Evaluations of model scores

After exploratory data analysis and preprocessing the main steps comes up that is model building and evaluation is the technique where we make assumptions that which model giving the best accuracy. Moreover , we have used hyper parameter technique to check the chances in increment of accuracy.

CONCLUSION

- Key Findings and Conclusions of the Study and Learning Outcomes of the Study in

Data analysis is a proven way for any organizations and enterprises to gain the information they need to make better decisions, serve their customers, and increase productivity and revenue. The benefits of data analysis are almost too numerous to count, and some of the most rewarding benefits include getting the right information for business, getting more value out of it. And it creates more effective marketing campaigns, gaining a better understanding of customers, and so on.To summarize, for this project first we have collected data. Since the APIs by Indian companies like yatra.com returned data in a complex format resulting in a lot of time to clean the data before analyzing, therefore we decided to build a web spider that extracts the required values from a website and stores it as a CSV file. In addition, The data was further processed based on the parameters like departure time , arrival time , duration and all. Furthermore, after visualization to get better insights from data points. There were few columns contains categorical values which converted into numeric using pandas get_dummies function. And used various advanced model to build ML project. And as per cross validation we have found that decision tree model giving the good accuracy in order to predict out flight price which has been saved as base model using pickle library.