



Name of the project

Flight price prediction

Submitted by:

Pooja Rajpal

Internship batch 33

Acknowledgement:

I would like to express my appreciation to my company

(Flip Robo technologies)

For the sample documentation that helped us in making the project report.

And also thankful to the mentors who gave very good training for the web scrapping.

This project has been source of learning and bring our theoretical knowledge into real time projects .

I would really acknowledge my teacher's help and guidance.

Also the recordings provided by my institute that helped me for completing my project.

Introduction

- **Business problem statement:**

This problem statement is about predicting the price of flights on a particular date. In this project first we have to scrap the details such as airline name ,date of journey,departure,source,route duration from various sites namely yatra.com,skyscanner.com,kayak .com and many more.

- **Conceptual Background of the Domain Problem**

For better understanding of the project it is necessary to understand the problem .This is a flight price prediction dataset that tells about the price of the flights of particular airline on a particular date.

- **Review of Literature**

This is a FLIGHT PRICE PREDICTION project in which data is scrapped through web selenium scrapping .Details which are scrapped includes airline name, date of journey ,source,destination ,duration of journey and Price (label).

The details are scrapped through varios sites .After collecting the data ,the data is collected in a dataframe and the dataframes are concatenated .

- **Motivation for the Problem Undertaken**

Building projects improves the skill and it also gives immense pride when the model is completed with good score.

Though completing the project requires lot of steps.but it is necessary to improve our vision.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

The first step here is to import necessary libraries.And load the dataset in jupyter notebook to analyse it.

Statistical modelling is necessary as it tells about the relationship between variables.

Statistical techniques are used for the EDA .The goal is to collect and clean the data and make predictions .

- Data Sources and their formats

The data is collected through web scrapping through various sites and collected in a dataframe .Data can be collected through structured or unstructured format .Web selenium is used to scrap the data from various sites.Data collection through webscrapping is primary data .

- Data Preprocessing Done

Step1:collection of data that includes scrapping of data and then load the dataset in jupyter notebook using pandas and for that we need to import pandas library .

Step2: converting the collected data into dataframes .data cleaning that include checking for nulls and treat them using imputers.Checking for white spaces ,replacing them with nulls and then treat them.

Step3:checking for datatypes and if object datatype convert them into integer as computers only understand numeric data

Step 4:data visualisation

Step5:checking for skewness if any and treat them using Power transformer

and finally split the dataset

- Data Inputs- Logic- Output Relationships

All the other variables except price are variables except price are features .The features are dependent on the price of flight such as flight name,date are directly dependent on label price. All the input variables have more or less relationship with output variable.

- Hardware and Software Requirements and Tools Used
- Laptop and

- Anaconda navigator is desktop graphical user interface that allows to launch applications
- Jupyter notebook is open source software.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

Regularization techniques, hyper parameter tuning to increase the accuracy.

- Testing of Identified Approaches (Algorithms)
As it is a regression problem Linear Regression and RandomForestRegression and kneighborsregressor algorithms are used.

- Run and Evaluate selected models
To do this you use the model to predict the evaluation dataset and comparing the predicted target to the actual answer. Number of metrics are used in ML to measure its accuracy.

- Key Metrics for success in solving problem under consideration

Metrics are used to measure the performance of the model. I used the accuracy score to find the accuracy of the model because according to me it gives the most reliable results and its easy to

determine what is the level of accuracy of the model built.

- Visualizations

it is a simple technique that allows the business users to recognize relationship and pattern between the data .it helps the people to see,interact and better understand data.

- Interpretation of the Results

Regression model works with this type of project

CONCLUSION

The data is collected through web scrapping from many sites such as yatra.com,kayak.com etc and then these details are converted into dataframe

Then we will merge the dataframes .Checking for nulls,whitespaces,skewness and preprocessing steps are to be followed to clean the data.After cleaning the data is divided into features and label

Building model and checking the accuracy is another step.

Finally hyperparameter tuning is to be done to increase the accuracy.

- Limitations of this work and Scope for Future Work
- I think I should have worked harder to increase the accuracy .

