Contents

<u>Document Version Control</u>

Date Issued	Version	Description	Author
08/06/2021	1	Initial HLD – V1.0.1	Pankaj Jaiswal Siddhiraj Kolwankar

Abstract

There are many ways through which we travel from one place to another but now-days traveling through flight is more preferred way by peoples due to its less traveling time. Due to such a high demand in traveling, flight tickets price changing continuously. Though there are many other factor on which flight fare depends on like flight timing, source and destination of the flights, it depends upon festive or holidays season and more like this. Airlines company have different kind of computational techniques through which they maintain there revenues but when it comes to customer, they don't have any technique by which they estimate the flight fare so they can plan their journey in advance. Because of this I'm building some prediction model using different different regression algorithms through which travelers will get an idea of the flight fare, so that they can plan their journey in prior.

1. Introduction

1.1 Why this High-Level Design Document ?

The purpose of this High-Level Design (HLD) Document is to add the important details about this project. Through this HLD Document, I’m going to describe every small and big things about this project.

2. General Description

2.1 Product Perspective

The Flight Fare Prediction predict the flight price using regression based Machine Learning algorithms.

2.2 Problem statement

Traveling through flights has become an integral part of today's lifestyle as more and more people are opting for faster traveling options. The flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, and duration of flights various occasions such as vacations or festive season. Therefore, having some basic idea of the flight fares before planning the trip will surely help many people save money and time. The main goal is to predict the fares of the flights based on different factors available in the provided dataset.

2.3 Proposed Solution

The solution here is an Regression based Machine Learning model. It can be implemented by different regression algorithms (like Linear Regression, RandomForest Regression, Decision Tree Regression, SVR and so on.).Here First we are performing Data preprocessing step, in which feature engineering, feature selection, feature scaling steps are performed and then we are going to build model.

2.4 Technical Requirements

In this Project the requirements to get flight fare through various platform. For that, in this project we are going to use different technologies. Here is some requirements for this project.

- Model should be exposed through API or User Interface, so that anyone can test model.
- Model should be deployed on cloud (Azure, AWS, GCP).
- Cassandra database should be integrated in this project for any kind of user input.

2.5 Data Requirements

Data Requirement completely depend on our problem.

- For training and testing the model, we are using flight fare prediction dataset from kaggle.
- From user we are taking following input :

  → Airlines Service – IndiGo, Air India', Jet Airways, SpiceJet, Multiple carriers, GoAir, Vistara, Air Asia, Vistara Premium economy, Jet Airways Business, Multiple carriers Premium economy, Trujet.
  → Source - Banglore, Kolkata, Delhi, Chennai, Mumbai.
  → Destination - Delhi, Banglore, Cochin, Kolkata, Hyderabad.
  → Total Stops - non-stop, 1 stop, 2 stops ,3 stops, 4 stops.
  → Full date of journey - mm/dd/yyyy format.
  → Total Duration – from 1 to 10 hr.

2.6 Tools Used

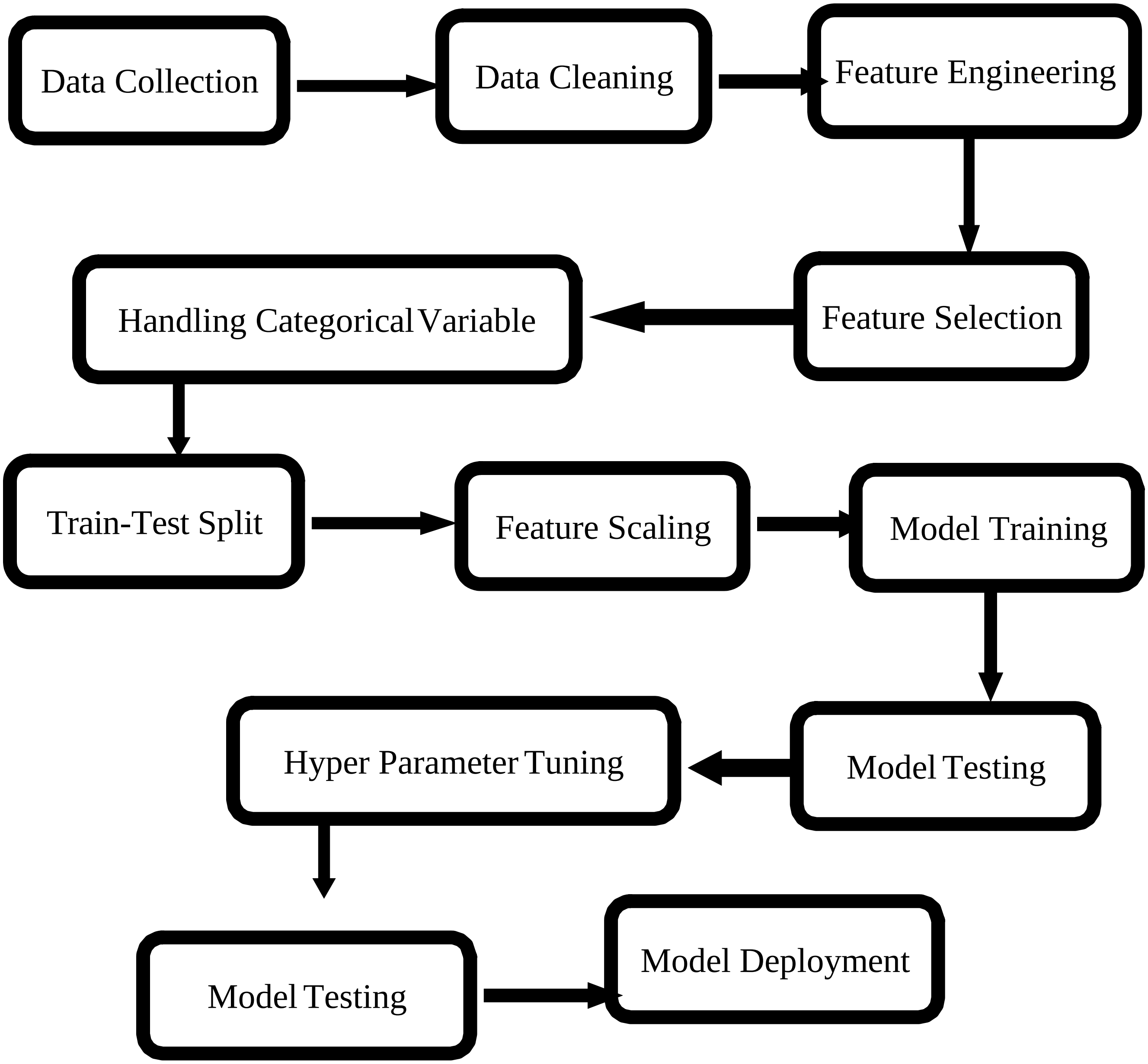
- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Azure is used for deployment of the model.
- Cassandra is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML/CSS, Flask is used for backend development and for API development.
- GitHub is used as version control system.

2.7 Constraints

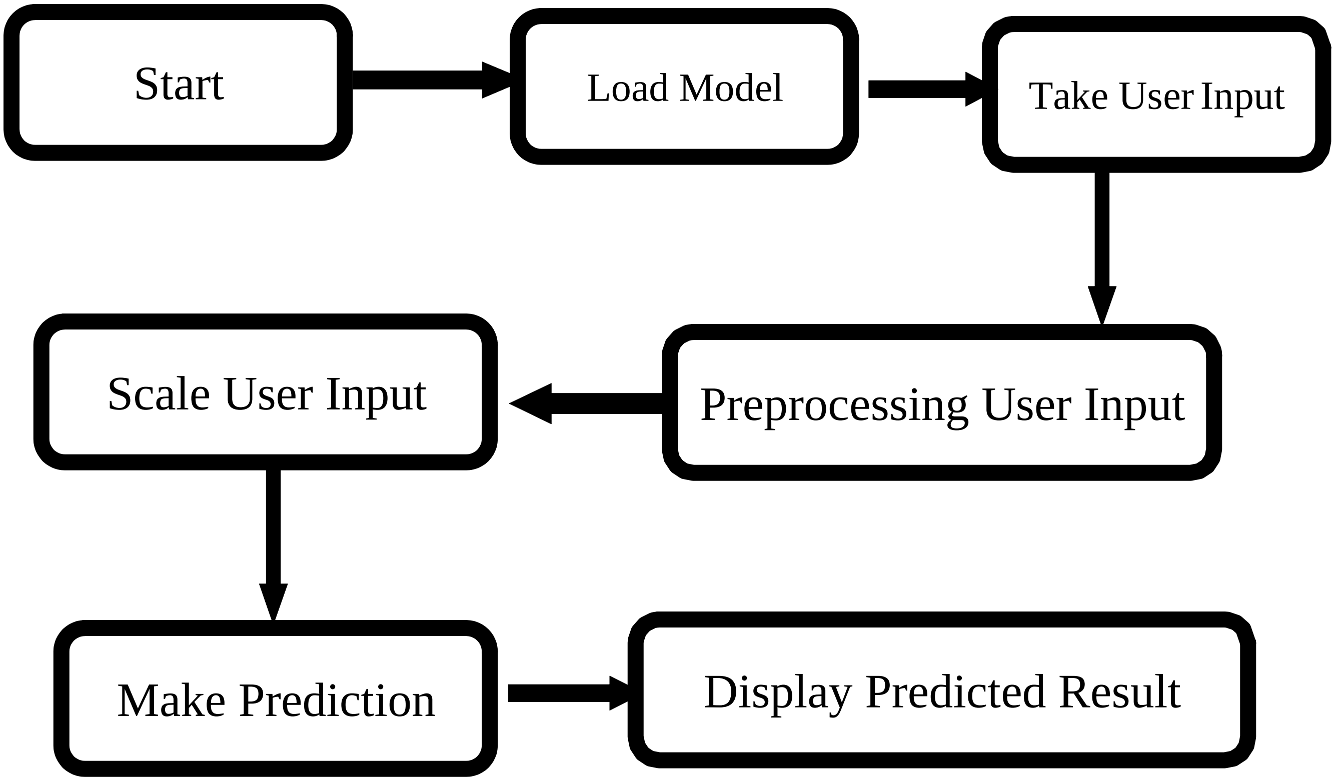
The Flight fare prediction system must be user friendly, errors free and users should not be required to know any of the back-end working.

3 Design Details

3.1 Process Flow



3.2 Deployment Process



3.3 Event Log

In this Project we are logging every process so that the user will know what process is running internally.

Step-By-Step Description:

- In this Project we defined logging for every function, class.
- By logging we can monitor every insertion , every flow of data in database.
- By logging we are monitor every step which may create problem or every step which is important in file system.
- We have designed logging in such a way that system should not hang even after so many logging's, so that we can easily debug issues which may arises during process flow.

3.3 Error Handling

We have designed this project in such a way that, at any step if error occur then our application should not terminate rather it should catch that error and display that error with proper explanation as to what went wrong during process flow.

4. Performance

Solution of Flight fare prediction is used to predict the flight fare in advance, so it should be as accurate as possible so that it should give as much as possible accurate price prediction.

That’s why before building this model we followed complete process of Machine Learning . Here are summary of complete process:

1. First we cleaned our dataset properly by removing all null value and duplicate value present in dataset.

2. Then we performed feature extraction, in which I extracted journey date, month and departure and arrival hour , minutes in new separate column.

3. After that I performed feature engineering step in which I created one new feature “Total_Duration”. In this feature what I have done is , I converted total time in minutes.

4. Then I performed feature selection step in which I dropped some feature like(Route, Date_of_journey, Departure_time, arrival_time and Additional_Info).

5. Then I handled categorical variable by performing One-Hot encoding.

6. Then I split the hole data set train-test split. After that I performed scaling on X_train and X_test.

7. After performing above step I was ready for model training. In this step, I trained my dataset on different Regression Learning algorithm(Linear, Random-Forest, K-NN, DecisionTreeRegression, SVR, Ridge, Lasso and Elastic net). After training the dataset on different algorithms I got highest accuracy of 80% on RandomForrestRegression.

8. After that I applied hyper-parameter tuning on all model which I have described above. Here also I got highest accuracy of 85% on test dataset by same RandomForrestRegression.

9. After that I saved my model in pickle file format for model deployment.

10. After that my model was ready to deploy. I deployed this model on various cloud storage(Azure, AWS and heroku) and also dockerize this model.

4.1 Re-usability

We have done programming of this project in such a way that it should be reusable. So that anyone can add and contribute without facing any problems.

4.2 Application Compatibility

The different module of this project is using Python as an interface between them. Each modules have it’s own job to perform and it is the job of the Python to ensure the proper transfer of information.

4.3 Resource Utilization

In this project, when any task is performed, it will likely that the task will use all the processing power available in that particular system until it’s job finished. By keeping this in mind, In this project we have used the concept of multi-threading.

4.4 Deployment

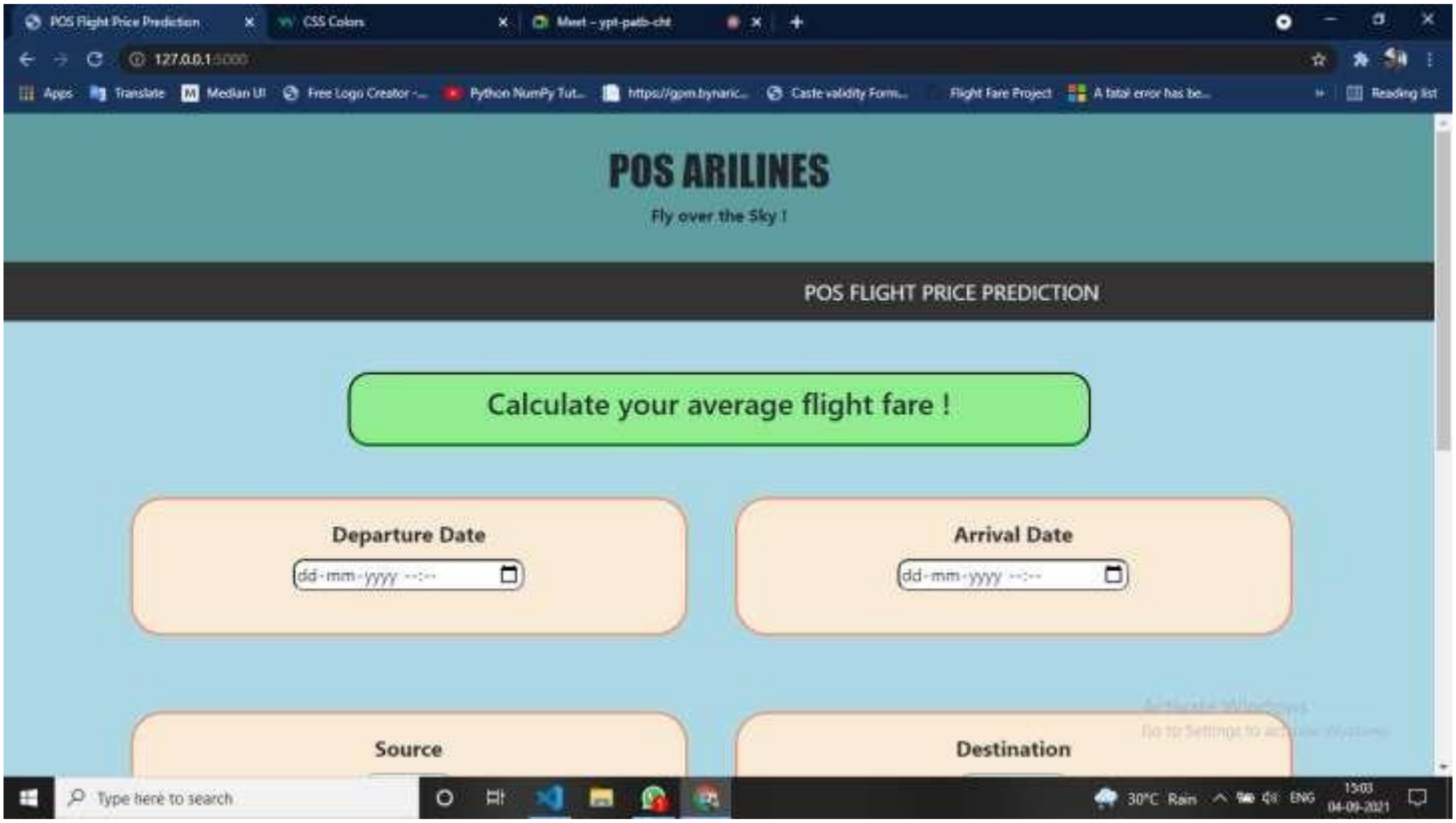
we have deployed on AWS (Amazon Web Services)

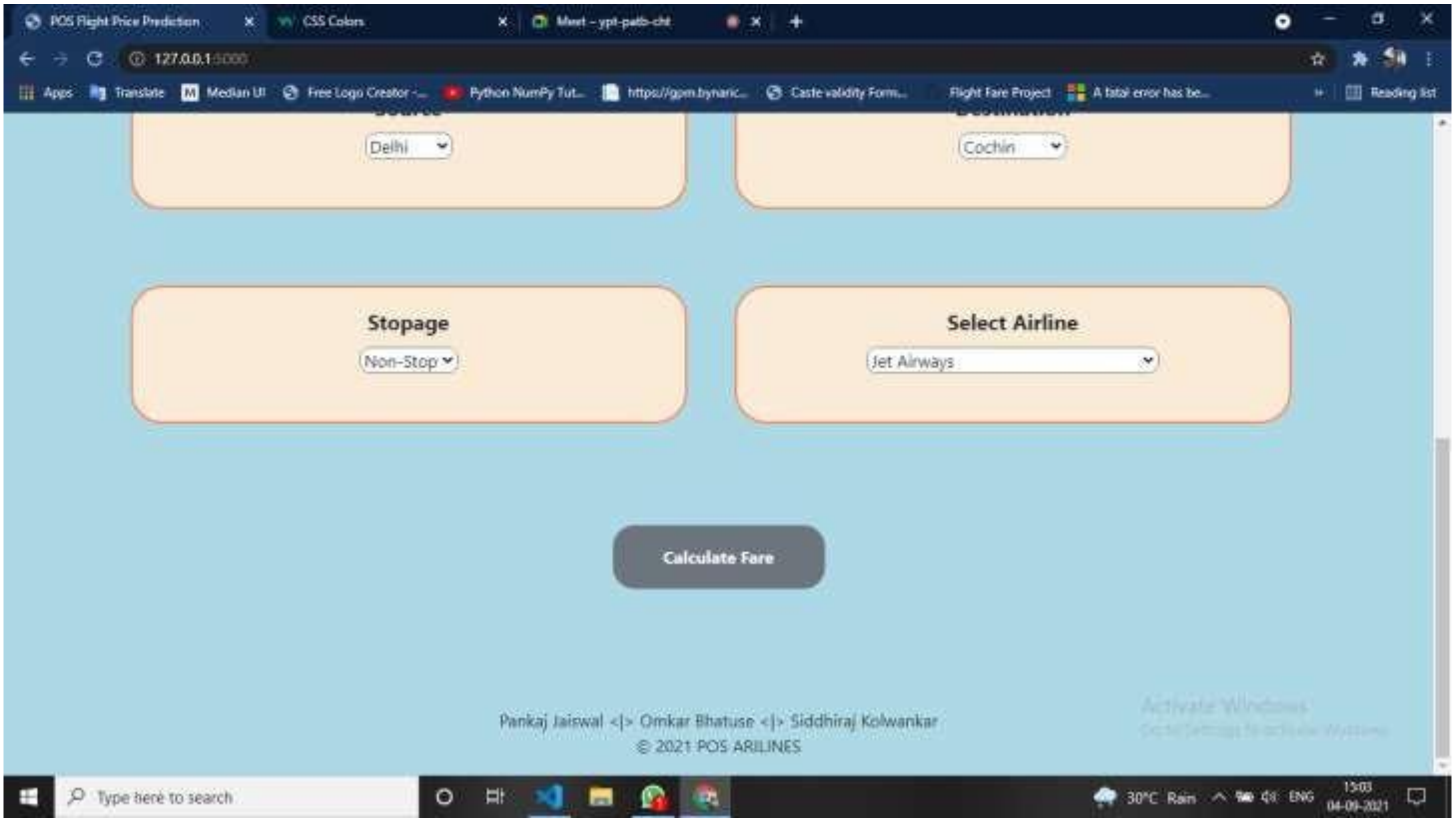
4.5 User Interface

We have created an UI for user by using HTML and CSS.

5. Conclusion

The Flight Fare prediction model will predict the fare of flight in prior so that costumer can get the idea of how much money they are going to spend on traveling.





Source: Delhi

Destination: Cochin

Stopage: Non-Stop

Select Airline: Jet Airways

Pankaj Jaiswal <|> Omkar Bhatuse <|> Siddhiraj Kolwankar

© 2021 POS ARILINES