iNeuron

High Level Design
(HLD)

Flight Price Prediction

Document Version Control

Date Issued	Version	Description	Author
05-12-2021	1.0	Final Update	Pavan Dasari

Contents

Abstract

The travel history of all the travelers had increased worldwide.Way back, there is very minimal travel history,With increase in technology ,the comfort level of travel increased along with safety,This made every one choose modern technological way of transport.Every mean of transport has rapidly increased with time.Among them airways tops the list.Airways are the fastest means of transport among all.The comfort level depends on the price and airlines.The number of airlines had grew in a very vast number when compared to past.The safety levels of aeroplanes has also increased with technological advancements.

The aim of our project is to predict the fare of airline,between source and destination based on number of stops and class of airline.Using machine learning we have predicted the estimated price of the ticket of airline.

dssf

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manualfor how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project.

2 General Description

2.1 Product Perspective

Flight fare estimator is a prediction web app for predicting prices of airlines between source and destination.It is regression based machine learning algorithm.

2.2 Problem statement

Travelling had become passion for some people while It has been the integral part of life for some people for faster travelling options.The airlines charges the ticket price based on the source and destination and the level of comfort they are providing to the customer.

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 proposed here is the machine learning algorithm which predicts the fare of the ticket based on various factors such as source,destination,stops etc..It can be developed using various algorithms.The best accuracy score is the one we choose to implement.The model is implemented by performing various data pre processing and post processing steps and then deployed in deployment platform.

2.3 Technical Requirements

The model created is exposed to user through GUI.and the data is stored in database.The model is deployed in any one of the cloud platforms like GCP.AWS,AZURE.

Having knowledge in database connections and cloud deployment is the main technical requirement of the project.

2.4 Data Requirements

Data requirement is based on the problem statement.

For training and testing the model, we are using flight fare prediction

dataset from kaggle.
From user we are taking following input :

→ **Airlines Servic**e – 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.

2.4Tools used

- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- Heroku 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.
- GitHub is used as version control system.

2.4 Constraints

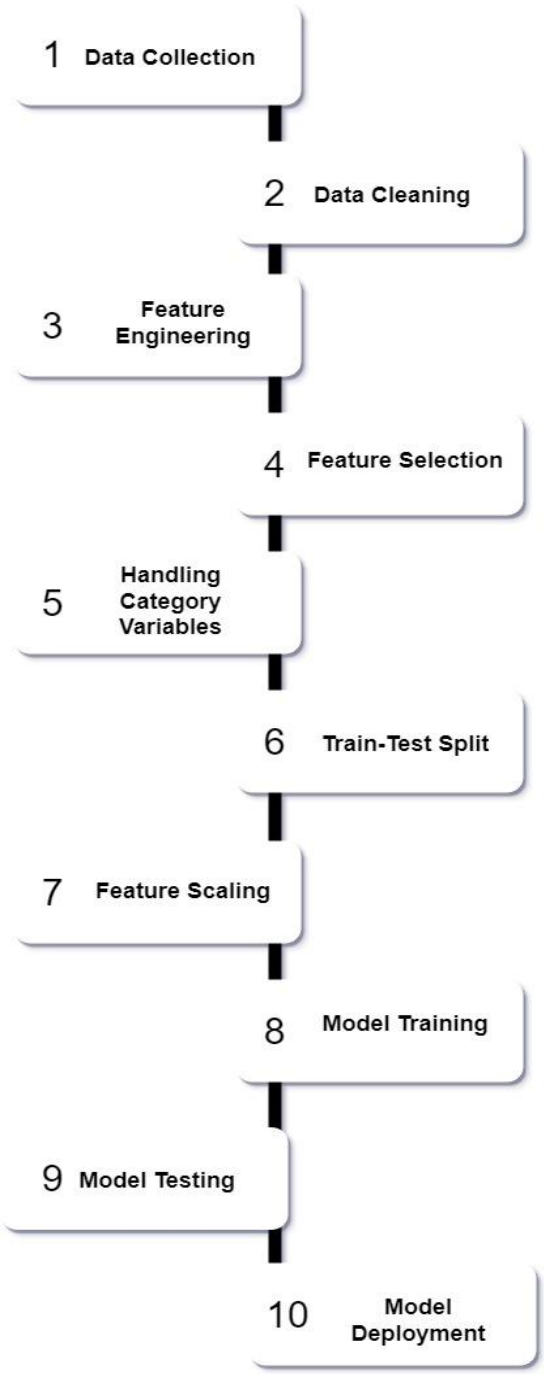
The machine learning model should predict well with less error rate.The chance of false predictions should be less.

2.5 Assumptions

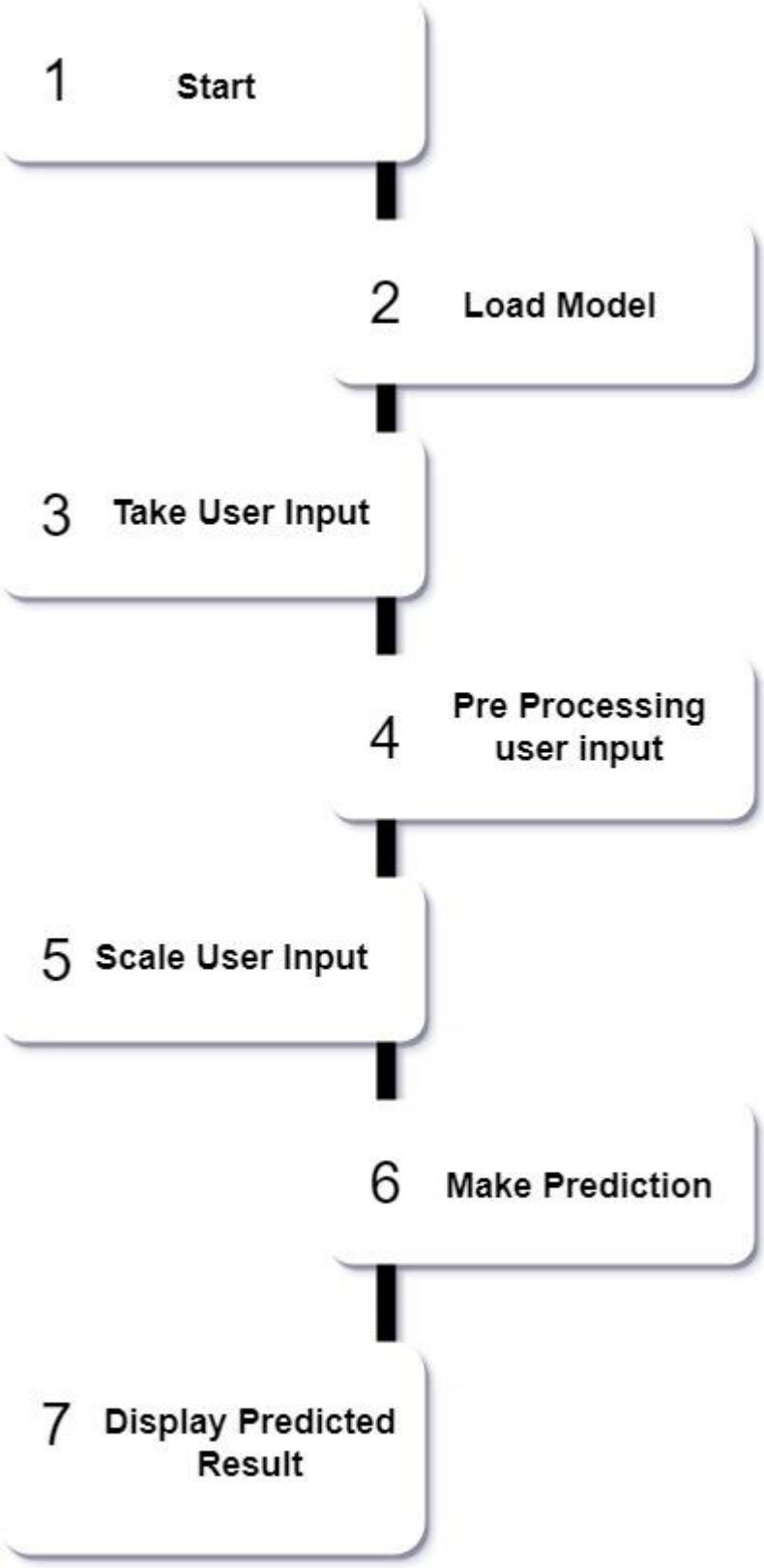
The data is assumed to be clean and neat.Data extracted or data given to us is assumed to be uniform and without any noise.

3 Design Details

3.1 Process Flow



3.1.2 Deployment Process



3.2 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 Reusability

The code is written such that anyone can use the program further.

4.2 Application Compatibility

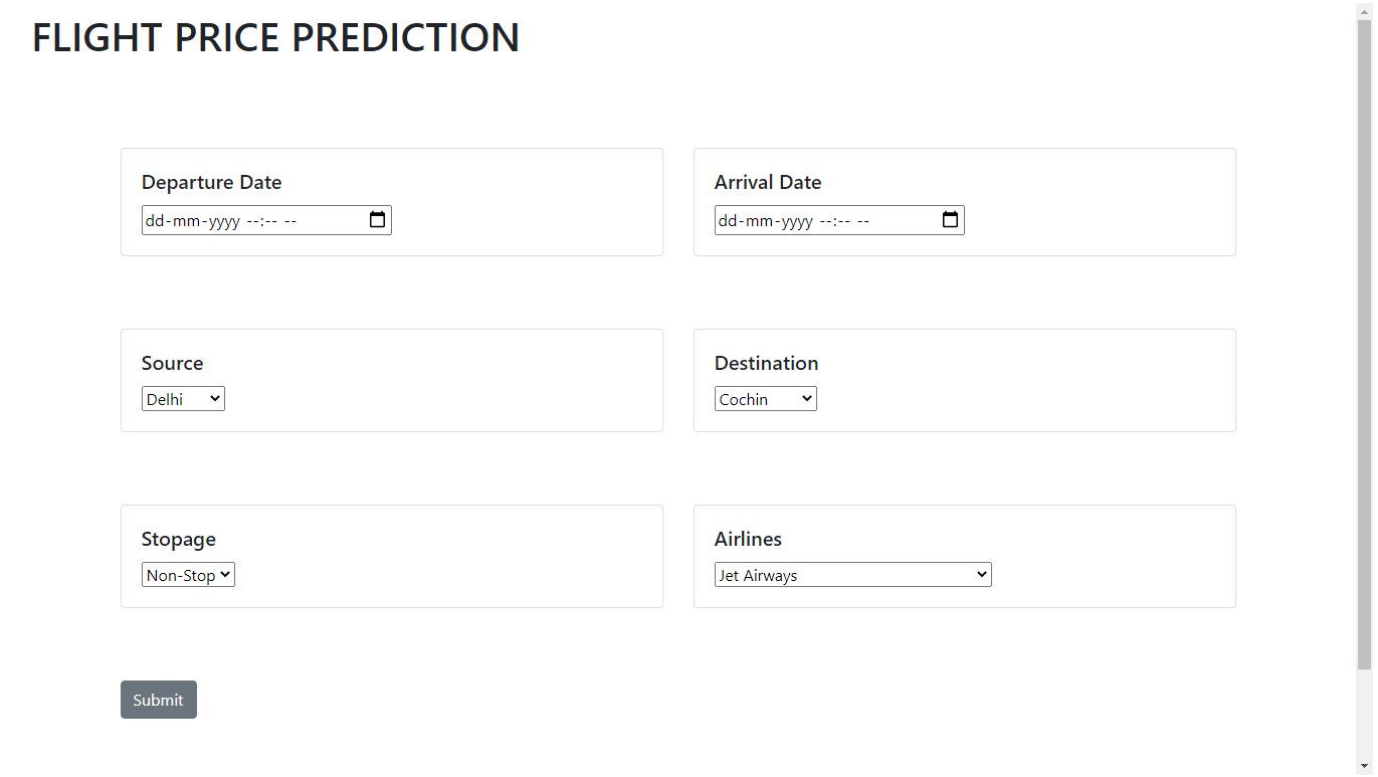
The different module of this project is using Python as an interface between them.Each modules have it's own job to perform and it is the job of the Python to ensure the proper transfer of information..

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

User Interface



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.