High Level Design (HLD)

Flight Fare Prediction

Gautam Sharma

Document Version Control

Date Issued	Version	Description	Author
13 September 2021	1.1	Initial HLD v1.0	Gautam Sharma

Contents

Contents

Abstract

Travelling through flights has become an integral part of today's lifestyle as more and more people are opting for faster travelling options. The flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, 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. In the proposed system a predictive model will be created by applying machine learning algorithms to the collected historical data of flights. This system will give people the idea about the trends that prices follow and also provide a predicted price value which they can refer to before booking their flight tickets to save money. This kind of system or service can be provided to the customers by flight booking companies which will help the customers to book their tickets accordingly.

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 manual for 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
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HI-D documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HI-D uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

.

1.3 Definations

Terms	Description
FFD	Flight Fare Prediction
IDE	Integrated development system
Heroku	Deployment server

2 General Description

2.1 Product Perspective

The flight dare prediction based system is a machine learning-based project by this we determine the flight fare on different occasions and find which flight is less costly.

2.2 PROPOSED SOLUTION

In this we developed a novel method for fraud detection, where customers are grouped based on their transactions and extract behavioural patterns to develop a profile for every cardholder. Then different classifiers are applied on three different groups later rating scores are generated for every type of classifier. This dynamic changes in parameters lead the system to adapt to new cardholder's transaction behaviours timely. Followed by a feedback mechanism to solve the problem of concept drift. We observed that the Matthews Correlation Coefficient was the better parameter to deal with imbalance dataset. MCC was not the only solution. By applying the SMOTE, we tried balancing the dataset, where we found that the classifiers were performing better than before. The other way of handling imbalance dataset is to use one-class classifiers like one-class SVM. We finally observed that Logistic regression, decision tree and random forest are the algorithms that gave better results.

2.3 Technical Requirements

We should be able to log every activity done by the user.

- The System identifies the source and destination

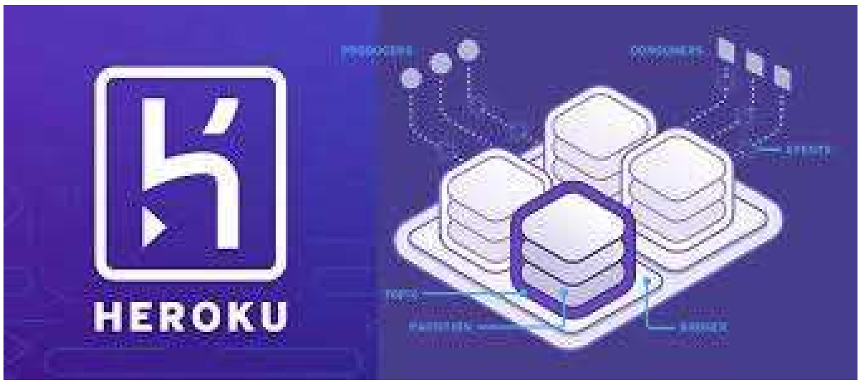
2.4 Data Requirements

In the Flight Fare Predictions system we use the kaggle dataset.

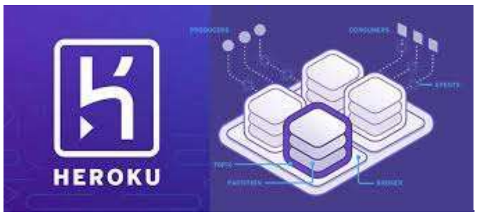
.

2.5 Deployment

1. Heroku



2.6 Tools Used

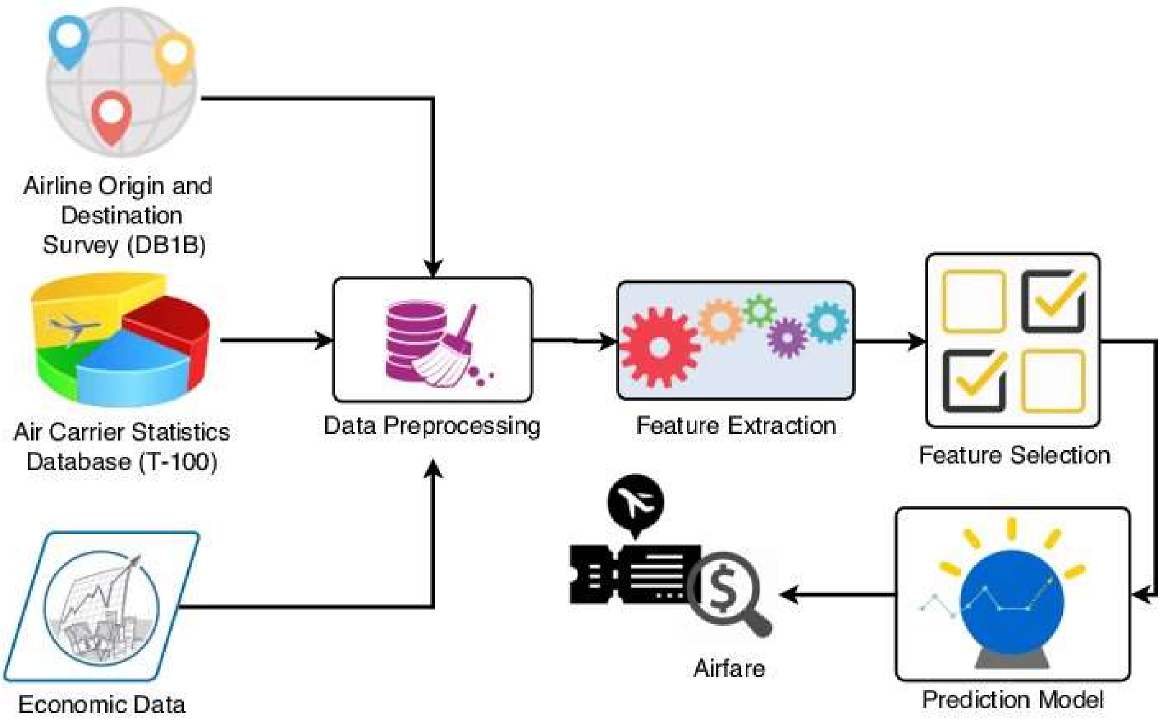


Python programming language and frameworks such as NumPy, Pandas, Scikit-learn are used to build the whole model.

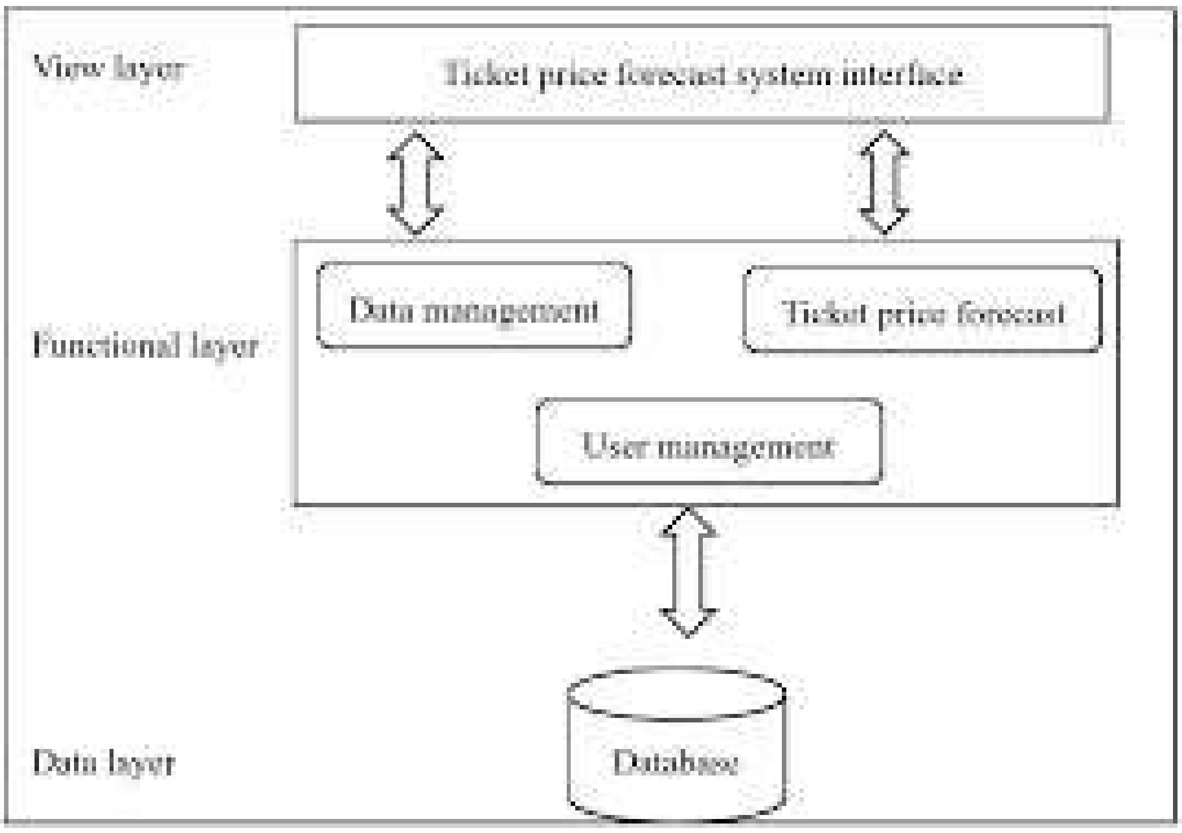
- Spyder is used as IDE
- Heroku is used as model deployment
- For visualization of the plots, Matplotlib, and Plotly are used.

3 Design Details

Front End	HTML/CSS/
Backend	Python Flask
Database	Kaggle
Deployment	Heroku



3.1 Process Flow



3.2 Event log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

1. we have implemented the machine learning life cycle to create a basic web application which will predict the flight prices by applying machine learning algorithm to historical flight data using python libraries like Pandas, NumPy, Matplotlib, seaborn and sklearn.
2. In the exploratory data analysis step, we cleaned the dataset by removing the duplicate values and null values. If these values are not removed it would affect the accuracy of the model. We gained further information such as distribution of data.
3. Next step is data pre-processing where we observed that most of the data was present in string format. Data from each feature is extracted such as day and month is extracted from date of journey

in integer format, hours and minutes is extracted from departure time

4 Formula

MAE (Mean Absolute Error)

$$MAE = 1/n[\sum(y-\acute{y})]$$

MSE (Mean Square Error)

$$MSE = 1/n[\sum(y-\acute{y})^2]$$

RMSE (Root Mean Square Error)

$$RMSE = \surd 1/n[\sum(y-\acute{y})^2]$$

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

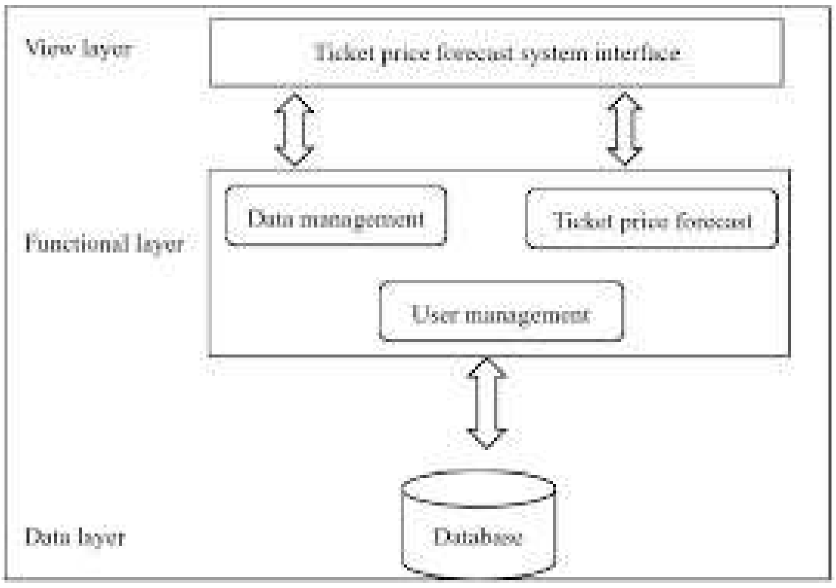
4.2 Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

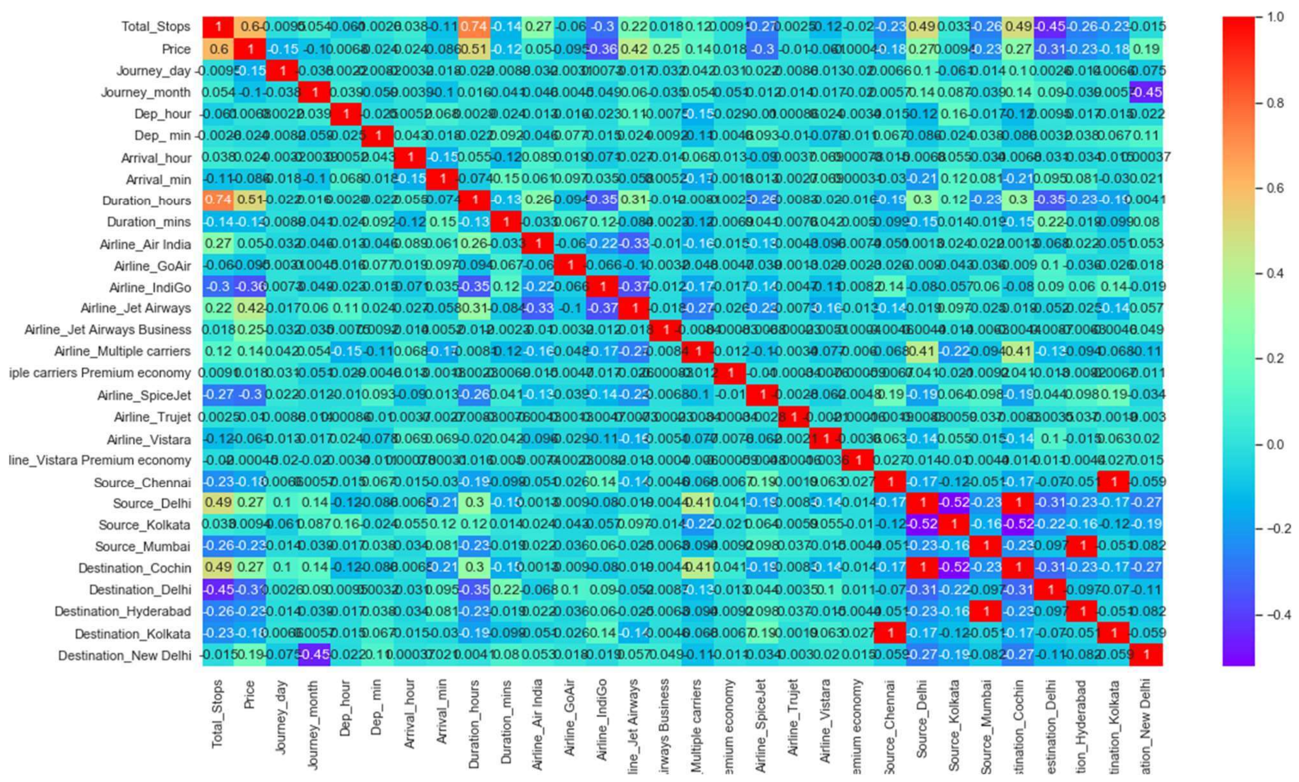
4.3 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

5 Model training/validation workflow



6 Correlations :



7 Conclusion

A proper implementation of this project can result in saving money of inexperienced people by providing them the information related to trends that flight prices follow and also give them a predicted value of the price which they use to decide whether to book ticket now or later. In conclusion this type of service can be implemented with good accuracy of prediction. As the predicted value is not fully accurate there is huge scope for improvement of these kind of service.