

High Level Design (HLD)

Flight Fare Prediction System



Document Version Control

Date	Version	Description	Author
08/1/2022	1.0	Abstract, General Description	Abhishek Saha
10/1/2022	1.1	Design Details, Performance	Abhishek Saha
13/1/2022	1.2	Deployment, Further Improvement, Conclusion	Abhishek Saha

Contents

Document Control.....	1
Abstract.....	3
1. Introduction.....	4
1.1 Why this High-Level Design-Document.....	4
1.2 Scope.....	4
1.3 Definition.....	4
2. General Description.....	5
2.1 Product Perspective.....	5
2.2 Problem Statement.....	5
2.3 Proposed Solution.....	5
2.4 Data Requirement.....	5
2.5 Tool Used.....	6
2.6 Constraint.....	7
2.7 Assumption.....	7
3. Design Details.....	8
3.1 Process Flow.....	8
3.1.1 Model training and evaluation.....	8
3.1.2 Deployment Process.....	9
3.2 Optimization.....	9
3.2.1 Data Strategy drive performance.....	9
3.2.2 Reduce the data points.....	9
3.3 Event.....	9
3.4 Error Handling.....	9
4. Performance.....	10
4.1 Reusability.....	10
4.2 Application Compatibility.....	10
4.3 Resource Utilization.....	10
5. Deployment.....	10
6. Further Improvement.....	11
7. Conclusion.....	11

Abstract:

Recent trends in travel through Flight has become an integral part of today's lifestyle. As more and more people are opting Flight for faster travelling options. Flight ticket Fare can be something hard to guess, today we might see a Fare, check out the Fare of the same Flight tomorrow, it will be a different. We might have often heard travelers saying that Flight ticket Fare are so unpredictable. The Flight ticket Fare increase or decrease every moment and also Fare depending on various factors like timing of the Flights, destination, and duration of Flights various occasions such as vacations or festive season etc. Here we are determining Fare of Flight tickets for various airlines depend on various factors. Therefore, having some basic idea of the Flight fares before planning the trip will surely help many people save money and time.

1. Introduction:

1.1 Why this High-Level-Design Document?

The purpose of High-Level-Design (HLD) Document is to add the necessary details to the current object description to represent a suitable Model for coding. The document also intended to help detect contradiction prior to coding and can be used as a reference manual for how the modules interact at 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 interface.
- Describe the performance requirements.
- Include design features and the Architecture of the project.
- List and describe the non-functional attribute like:
 - Security
 - Reliability
 - Maintainability
 - Portability
 - Reusability
 - Application compatibility
 - Resource Utilization

1.2 Scope:

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

1.3 Definitions:

<i>Term</i>	<i>Description</i>
Database	Collection of Information (Dept time, Duration, Route etc.)
IDE	Integrated Development Environment

2. General Description:

2.1 Product Perspective:

Flight Fare

2.2 Problem Statement:

Now a days Flight ticket Fare can be something hard to guess, today we might see a Fare, check out the Fare of the same Flight tomorrow, it will be a different. We might have often heard travelers saying that Flight ticket Fare are so unpredictable.

Here I have built an AI solution to predict Flight Fare for the following use case.

- Detect the Flight Fare on various parameter so that traveler can make plan for their trip.

2.3 Proposed Solution:

The Solution proposed here is a Flight Fare Prediction based system that can be implemented to perform above mentioned use case. In this case we have dataset which contain several information like Date-of-journey, Source, Destination, Route, Number-of-stoppage, Travel duration, Additional information, Fare depend on this various factor we trying to build a Machine Learning Model. After training the Model we are saving it for future prediction, so depend on prediction value Passenger make their tour plan and also save money and time.

2.4 Data Requirements:

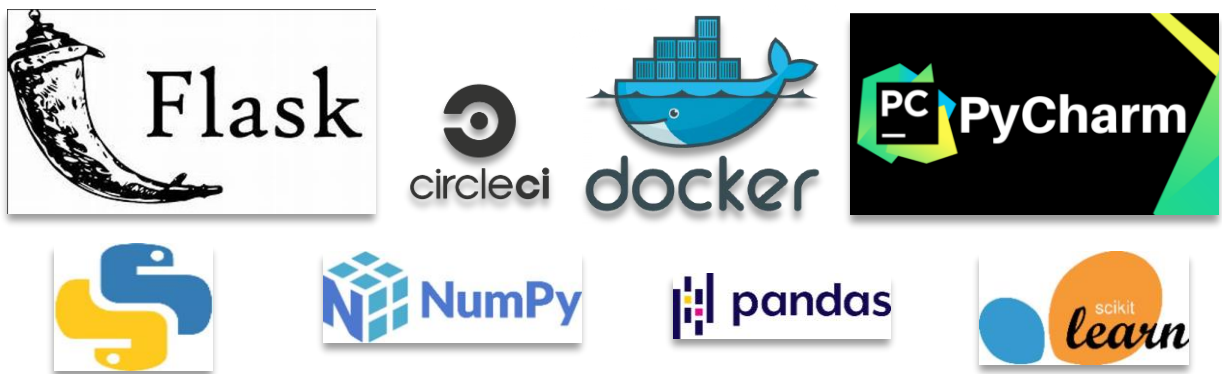
Data requirements completely depend on the problem statement.

- Training and Prediction dataset should get from User.
- We need as much as data so that Model training done in an efficient way.
- Data does not contain too many duplicate and null values.
- In Dataset should have all those columns mentioned in Schema file.
- For training Dataset have 10 independent feature and 1 dependent feature. For Prediction dataset have only 10 independent features as dependent column predict by Model.
- All necessary details of a Flight should be present like Source, Destination, Route, Date of journey, Departure time, arrival time, duration etc.
- Before pass the dataset into Model we are perform different validation operation. File name like “FlightPrice_01012002_090923”.
 - FlightPrice is case sensitive it anything is not match file move into bad_Data Directory.
 - In file name second field indicate datestamp if datestamp length is not match with schema file defined datestamp length, file move into Bad_Data Directory else store into Good_Data directory.
 - In file name third field indicate timestamp if timestamp length is not match with schema file defined timestamp length, file move into Bad_Data Directory else store into Good_Data directory.
- If in the dataset may have any column entirely contain Null value, that file should be move into Bad_Data Directory.

2.5 Tool Used:

Python programming language and its package such as NumPy, Seaborn, Pandas, Scikit-learn are used to read CSV file, EDA, visualization. Flask is used to build the web application.

We are using docker to dockerize entire application and with the help of Ci-cd pipeline deploy the entire application on Heroku Cloud Platform



- PyCharm is used as IDE
- SQLite is to retrieve, insert, delete and update the database.
- For visualization Matplotlib, Seaborn are used.
- Flask framework is used to create the web application.
- Github is used as version control.
- Frontend development done by HTML and CSS.
- Docker is used to hold the entire application.
- Ci-cd pipeline is used to host entire docker container in Cloud platform.

2.6 Constraint:

Flight Fare prediction solution system must be user friendly. After prediction done it share the final outcome into user specified location in CSV format.

2.7 Assumptions:

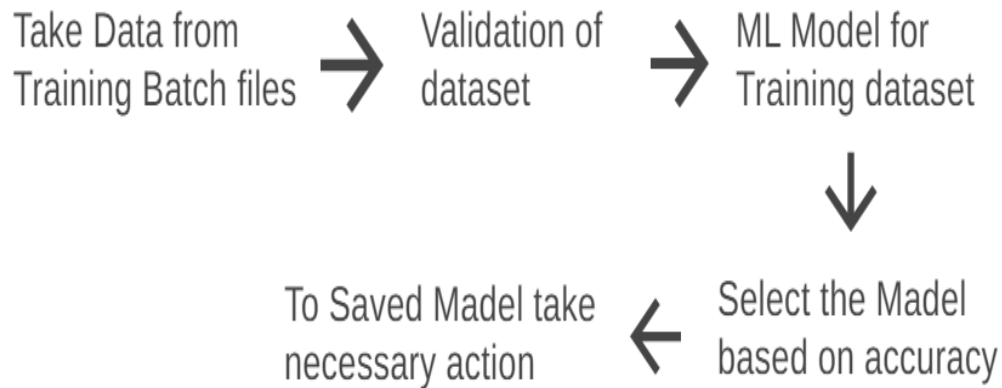
The main objective of this project is to implement the use cases as previously mentioned (2.2 Problem Statement) for dataset shared by the Client. After train the Model by user shared Training batch files, Model will be saved for prediction. Model predict the result based on predict batch files depend on various feature and then result store in client specified location. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

3. Design Details

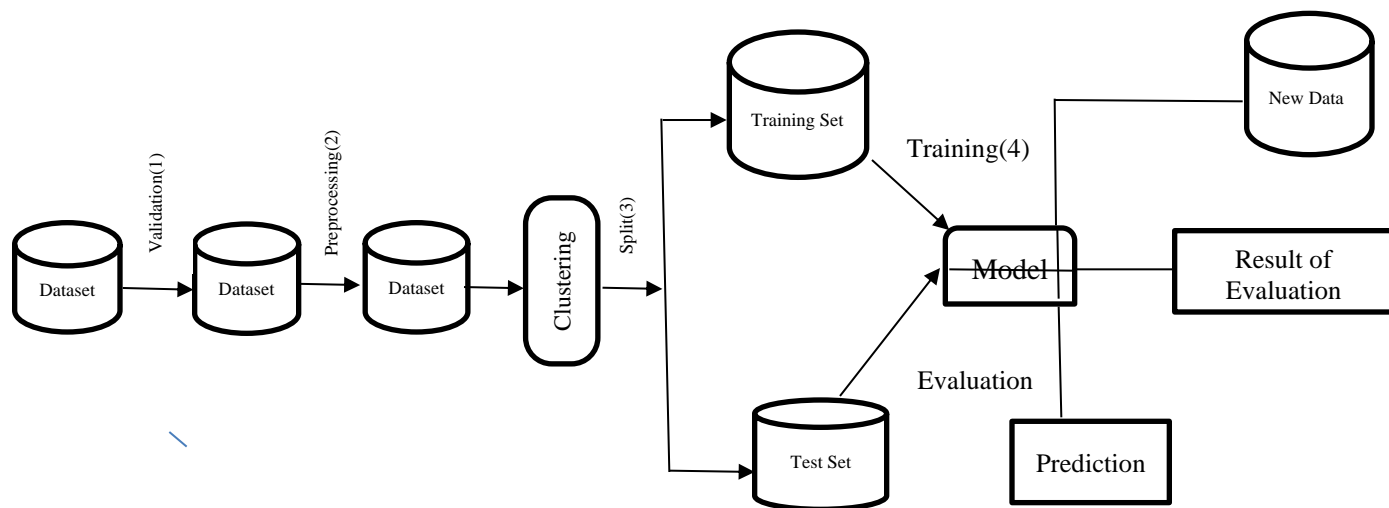
3.1 Process Flow:

To Predict Flight Fare, we are using Machine Learning based Model. Below is the process flow diagram as shown below:

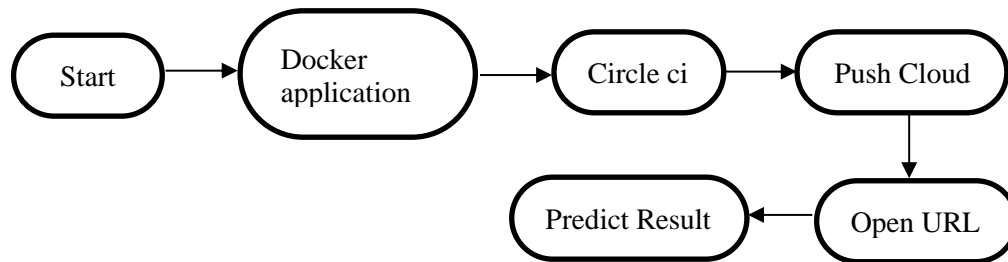
Proposed Methodology



3.1.1 Model Training and Evaluation:



3.1.2 Deployment Process:



3.2 Optimization:

3.2.1 Data Strategy drive performance

- Minimize the number of fields.
- Minimize the number of records.
- Optimize extract to speed up future queries by materializing calculations, removing columns.

3.2.2 Reduce the data points

- Remove unneeded dimensions for detail itself.
- Explore. Try displaying data in different views.

3.3 Event:

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

Initial Step-By-Step Description:

1. The system should be able to maintain log each and every flow.
2. System should not hang even after using so many loggings. Logging is used just because we can easily debug if any issue came up as well as maintain system working flow.

3.4 Error Handling:

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

4. Performance:

The Flight Fare Prediction system is used for detect the Flight Fare based on various factors of the dataset. It will helpful for those people most of the time who are opting Flight to travel from one place to another place. Therefore, they have some basic idea of the Flight fares before planning the trip will surely help many people save money and time.

4.1 Reusability:

The Code written and the components used should have the ability to be reused without any problem.

4.2 Application Compatibility:

The different component 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. Deployment:

Prioritizing data and analytics couldn't come at a better time. Any Organization, no matter what size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying.

Deploy application in a Cloud platform is another essential activity. So that we can Provide the URL to Client and client can easily access application for prediction operation.

6. Further Improvement:

This project can be used for many use cases like Ticket Fare for train, bus, ship etc. so that traveller depend on Fare can make their plan.

After data pre-processing, we have near about 16k rows. It will be better if the Model can be trained by more data.

Hence, I have done this project in Machine Learning, if this project can implement using Deep Learning, I think Model accuracy may be increase and get more better outcome.

7. Conclusion:

The Designed Flight Fare Prediction System predict the Fare based on Dataset shared by the user. So that we developed a Model train the Model based on training data after that Model make the prediction based on various parameters of the prediction dataset, so that traveler having some basic idea of the Flight fares before planning the trip. It will surely help many people by saving money as well as time.