**High Level Design (HLD)**

**FLIGHT**

**FARE**

**PREDICTION**

**Submitted By –**

**Abhishek**

**Monarch**

**Emmanual**

**Table of Contents**

**Abstract**

**INTRODUCTION**

**Why this HLD documentation?**

[***Key points:***](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.3znysh7)

[**1 Description**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.2et92p0)

[**1.1 Problem Perspective**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.tyjcwt)

**1.2 Problem Statement**

**1.3 Proposed Solution**

[**1.4 Solution Improvements**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.4d34og8)

[**1.5 Technical Requirements**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.2s8eyo1)

[**1.6 Data Requirements**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.17dp8vu)

[**1.7 Tools Used**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.3rdcrjn)

[**1.8 Constraints**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.26in1rg)

[**1.9 Assumptions**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.lnxbz9)

[**2 Design Flow**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.35nkun2)

[**2.1 Modelling Process**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.1ksv4uv)

[**2.2 Deployment Process**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.44sinio)

[**2.3 Logging**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.3j2qqm3)

[**2.4 Error Handling**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.1y810tw)

[**3 Performance Evaluation**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.4i7ojhp)

[**3.1 Reusability**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.2xcytpi)

[**3.2 Application Compatibility**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.1ci93xb)

[**3.3 Resource Utilization**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.3whwml4)

[**3.2 Deployment**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.2bn6wsx)

[**Conclusion**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit?pli=1#heading=h.qsh70q)

**Abstract**

**The recent global situations had a huge impact on the aviation sector due to many reasons. This impact has two category people, the first is business perspective and the second is the customers perspective. As safety is the major reason for such impact on the aviation sector, the governments around the world amended different rules to their respective airlines companies. These restrictions had made the availability of the flights and their attendee capacity less. Taking all these factors in consideration the cost of the flight tickets has increased and vary from one place to the other. Booking a flight ticket has split into two, one is the online and the other is the offline bookings. Both these have their respective criteria for cost of the ticket, one such example is the server load and the number of booking requests. In this machine learning implementation, we will see various factors that impact the cost of the flight ticket and predict the appropriate price of the ticket.**

**INTRODUCTION**

**Why this HLD documentation?**

The main purpose of this HLD documentation is to add the necessary details of the project and provide the description of the machine learning model and the written code. This also provides the detailed description on how the entire project has been designed end-to-end.

***Key points:***

* Describes the design flow
* Implementations
* Software requirements
* Architecture of the project
* Non-functional attributes like:
* Reusability
* Portability
* Resource utilization

**1 Description**

1. **Problem Perspective**

The flight fare prediction is a machine learning model which helps us to predict the cost of the flight ticket and helps the users to know the cost of their journey.

**1.2 Problem Statement**

**The main goal of the project is to create a user interface which predicts the cost of the flight ticket by taking certain input from the user like date of journey, onboard location and destination etc.**

**1.3 Proposed Solution**

**The solution proposed to take the required input of user from the created interface and process all the provided data to meet the requirements of the machine learning model and finally display the output saying so and so amount is the predicted cost.**

**1.4 Solution Improvements**

We can even predict the cost of ticket considering whether is it a weekday, holiday season or other social reasons. But considering from the perspective of business, if we process such data and predict the cost of the discounted ticket it will bring some loss to the airlines company. Hence this method is not considered.

**1.5 Technical Requirements**

There are no hardware requirements required for using this application, the user must have an interactive device which has access to the internet and must have the basic understanding of providing the input. And for the backend part the server must run all the software that is required for the processing the provided data and to display the results.

**1.6 Data Requirements**

The data requirement is completely based on the problem statement. And the data set is available on the Kaggle in the form of excel sheet(.xlsx). As the main theme of the project is to get the experience of real time problems, we are again importing the data into the Cassandra data base and exporting it into csv format.

**1.7 Tools Used**

* Python 3.9 is used as the programming language and frame works like numpy, pandas, sklearn and other modules for building the model.
* PyCharm is used as IDE.
* For visualizations seaborn and parts of matplotlib are being used.
* For data collection Cassandra database is being used.
* Front end development is done using streamlit.
* streamlit is used for both data and backend deployment.
* GitHub is used for version control.
* Streamlit Cloud is used for deployment.

**1.8 Constraints**

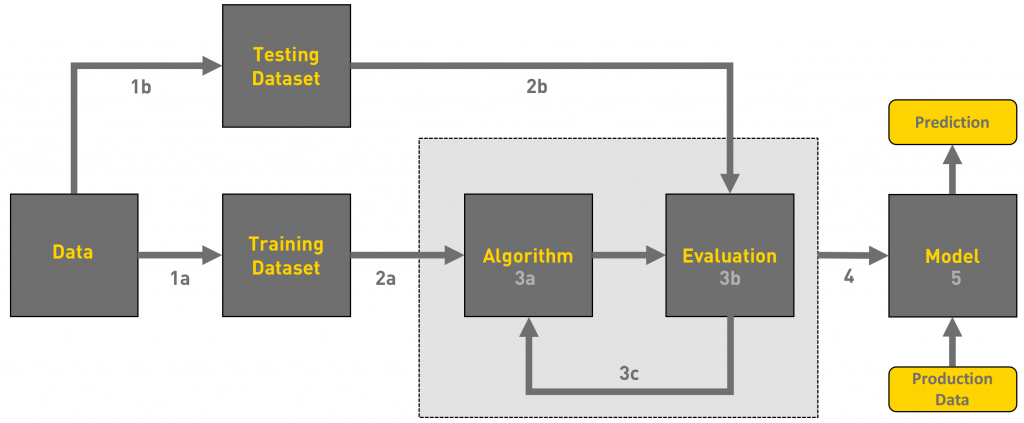
The flight fare prediction solution must be user friendly, as automated as possible and the user must not be required to know any of the working.

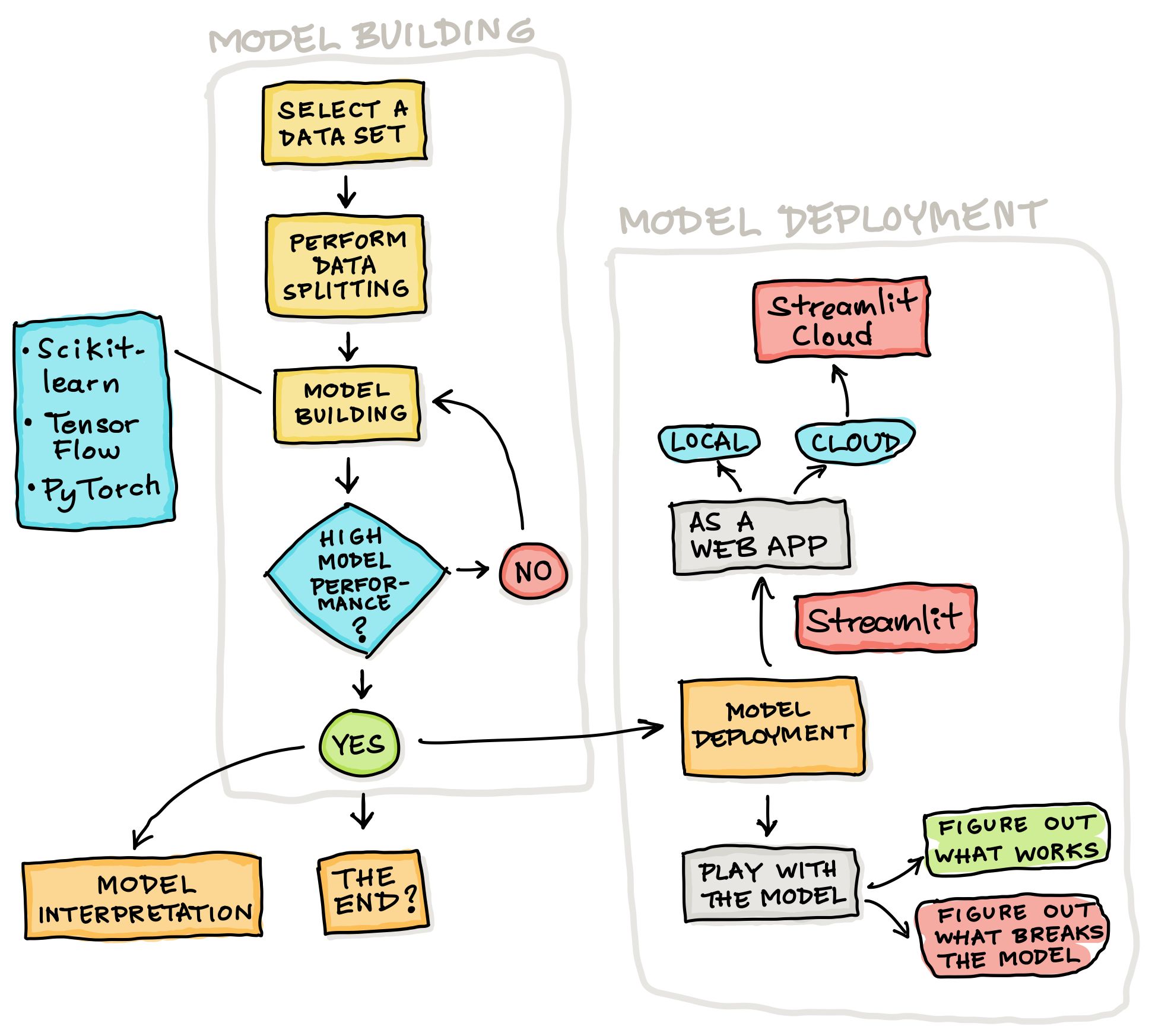
**1.9 Assumptions**

The main objective of the project is to implement the use cases as previously mentioned (1.2 Problem Statement) for the new dataset that user provides through the user interface. Machine learning model is used for processing the above input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

**2 Design Flow**

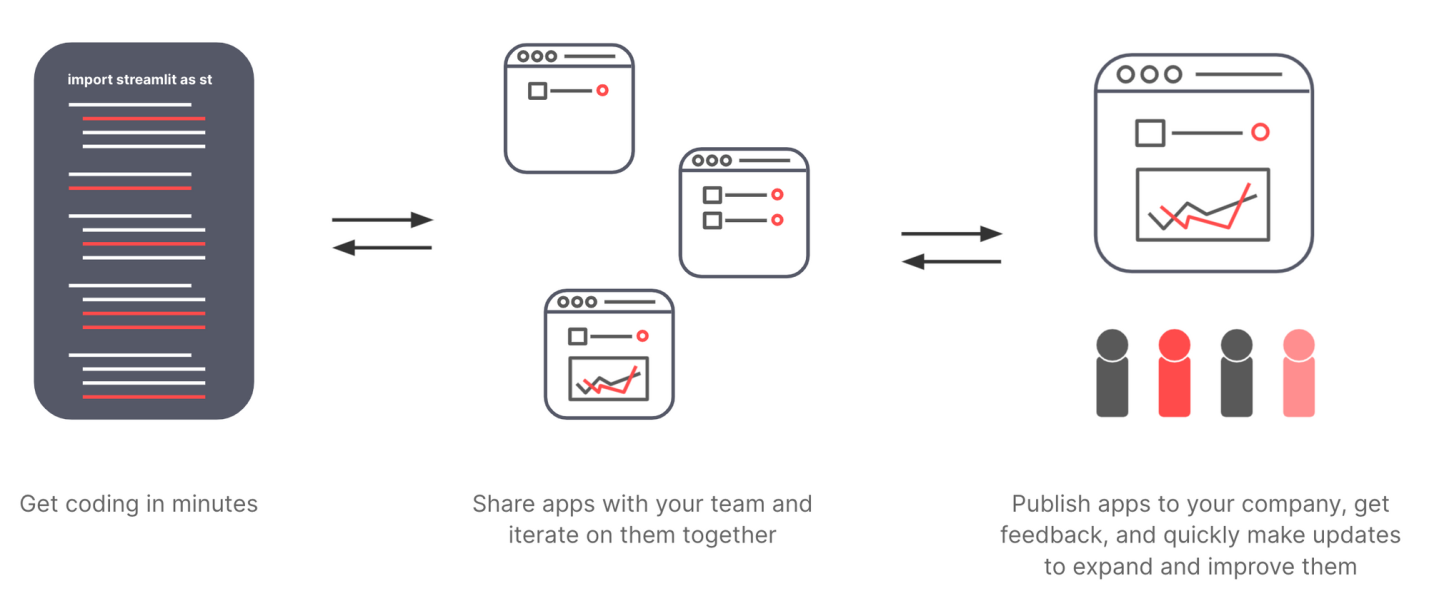
**2.1 Modelling Process**

****

****

**2.2 Deployment Process**

Deployement is on Streamlit and Render.

****

**2.3 Logging**

Every step is being logged in the system that runs internally, which shows the date time and the processed that has been performed, logging is done in different layers as INFO, DEBUG, ERROR, WARNINGS. This gives us the understand of the logged information.

**2.4 Error Handling**

When ever an error is occurred, the explanation will be logged in its respective log file, so that the developer can rectify the error.

**3 Performance Evaluation**

**3.1 Reusability**

Parts of the code written can be used to other applications and the rest can be modified and be reused.

**3.2 Application Compatibility**

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

**3.3 Resource Utilization**

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

**3.2 Deployment**

The model is being deployed on render https://internship-flight-fare-prediction.onrender.com/.



**Conclusion**

The flight fare prediction can predict the price based on the trained data set in the algorithm. Hence the user can know the approximate cost for their journey.