High Level Design (HLD)

Flight Fare Prediction

Document Version Control

Date Issued	Version	Description	Author
12/10/2022	1	HLD – V1.0	Diana Laveena DSouza

Contents

Document Version Control ……2

Abstract ……..4

1 Introduction …….....5

1.1 Why this High-Level Design Document? …….....5

1.2 Scope …….....5

1.3 Definitions ……..……..5

2 General Description ……...……6

2.1 Productive Perspective …….....6

2.2 Problem Statement ……6

2.3 Proposed Solution …….....6

2.4 Data Requirements ……...6

2.5 Tools Used …….....6

3 Design Details ……8

3.1 Process Flow ……8

3.1.1 Model Training and Evaluation ……9

3.1.2 Deployment Process ……10

3.2 Event Log ……...10

3.3 Error Handling ……11

4 Performance ……..11

4.1 Reusability ……..11

4.2 Application Compatibility ……..11

4.3 Resource Utilization ……...11

4.4 Deployment ……11

5 Dashboards……12

5.1 KPIs (Key Performance Indicators)……13

6 Conclusion……13

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, and duration of flights on various occasions such as vacations or the 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.

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 before 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 is 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 HLD documentation presents the system's structure, including the database architecture, application architecture, application flow and technology architecture. The HLD uses non-technical to mildly technical terms, which should be understandable to the system's administrators.

1.3 Definitions

Term	***Description***
Database	Collection of all the information monitored by this system
IDE	Integrated Development Environment

GCP	Google Cloud Platform

2 General Description

2.1 Product Perspective

Flight Fare Prediction is a machine learning model to predict the fares of flights.

2.2 Problem Statement

To create a regression model to predict the fares of the flights.

- To detect the flight ticket prices increase or decrease every now and then.

2.3 Proposed Solution

The solution proposed here is to build a regression model to detect whether flight ticket prices increase or decrease every now and then.

2.4 Data Requirements

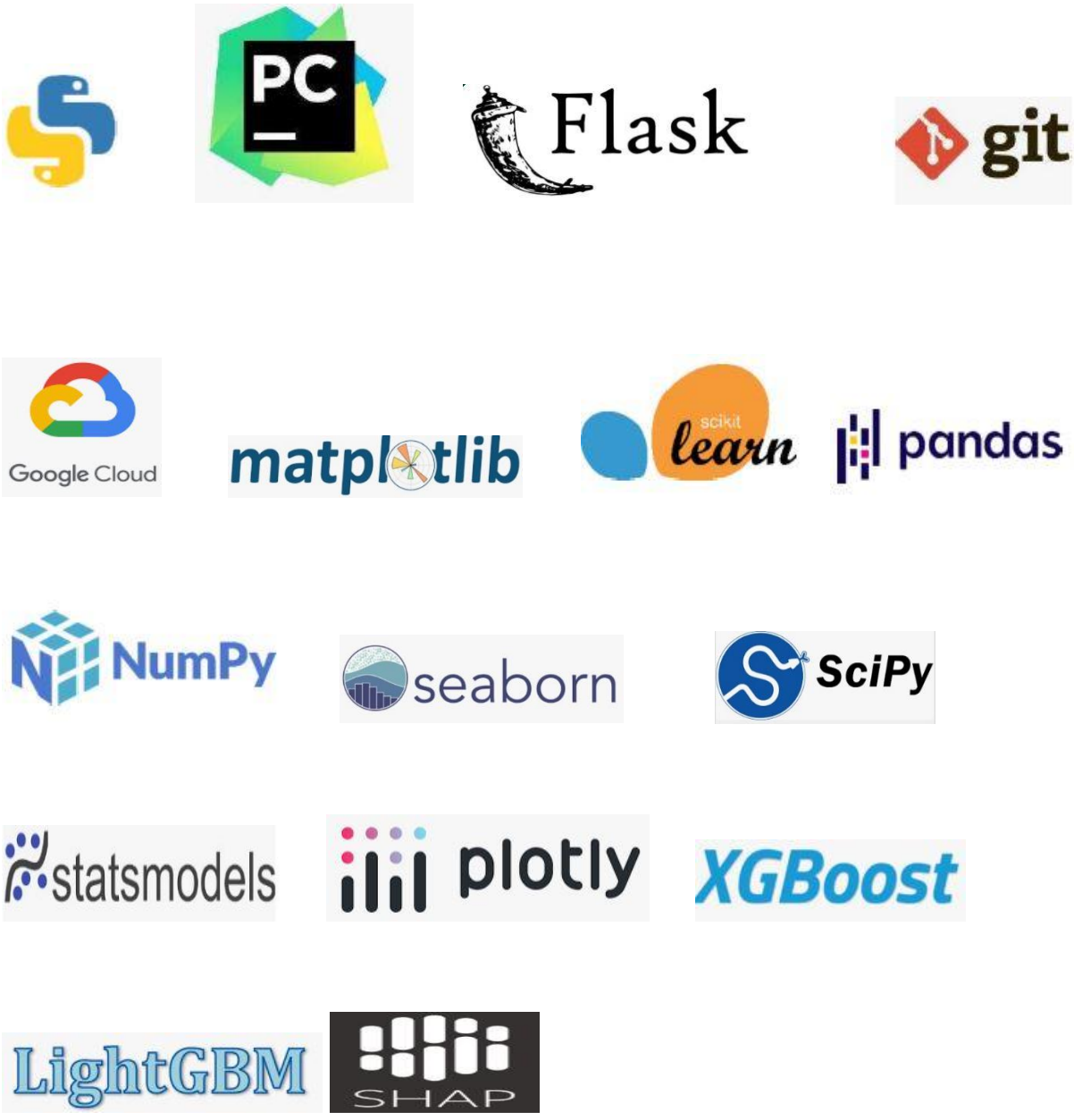
Data requirements depend entirely on our problem statement.

- We need a large amount of data, nearly 10462 records.

We use the Cassandra database by inserting records into it and then exporting them to a CSV file for further use.

2.5 Tools used

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



- PyCharm is used as IDE.
- Matplotlib, Seaborn, Plotly, statsmodels are used for Data Vizualization.
- Scipy is used for statistical tests. i.e. Hypothesis Testing.
- Sci-kit learn, Xgboost, and LightGBM are used for model building.
- Shap for model interpretation.
- GCP is used for the deployment of the model.
- Cassandra is used to retrieve, insert, delete and update the database.
- Front-end development is done using HTML.

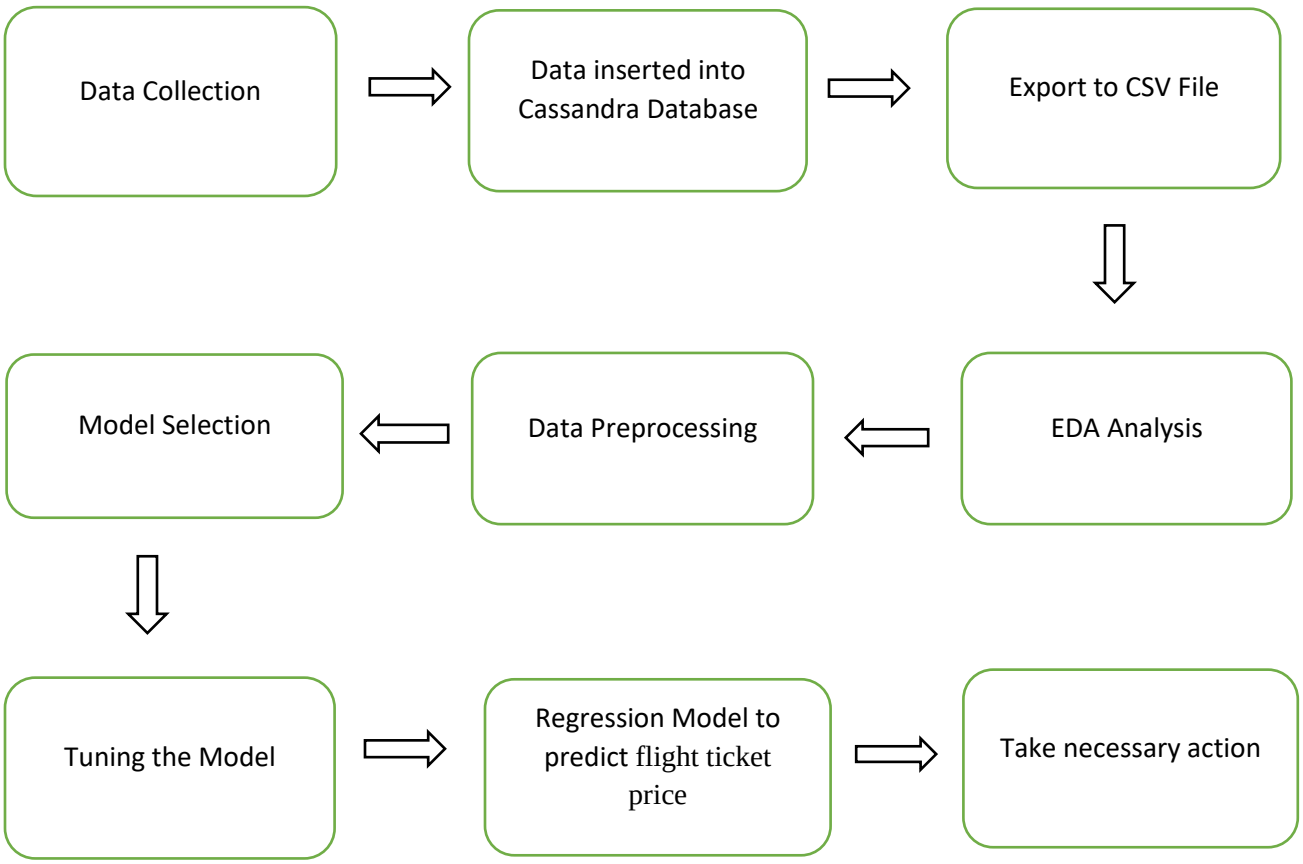
- Python Flask is used for backend development.
- GitHub is used as a version control system.

3 Design Details

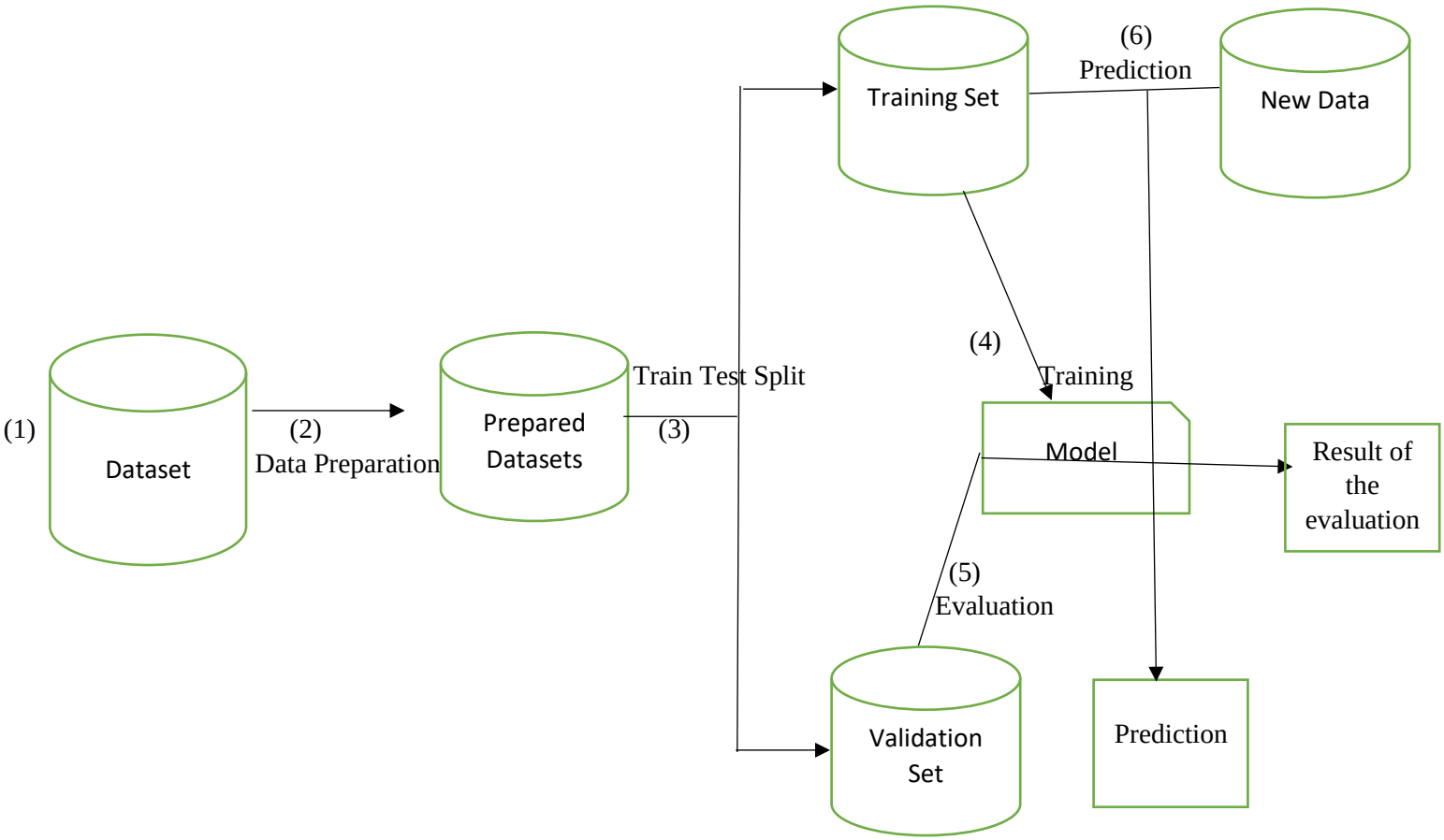
3.1 Process Flow

To predict the fares of the flights, we will use the regression model. Below is the process flow diagram is as shown below.

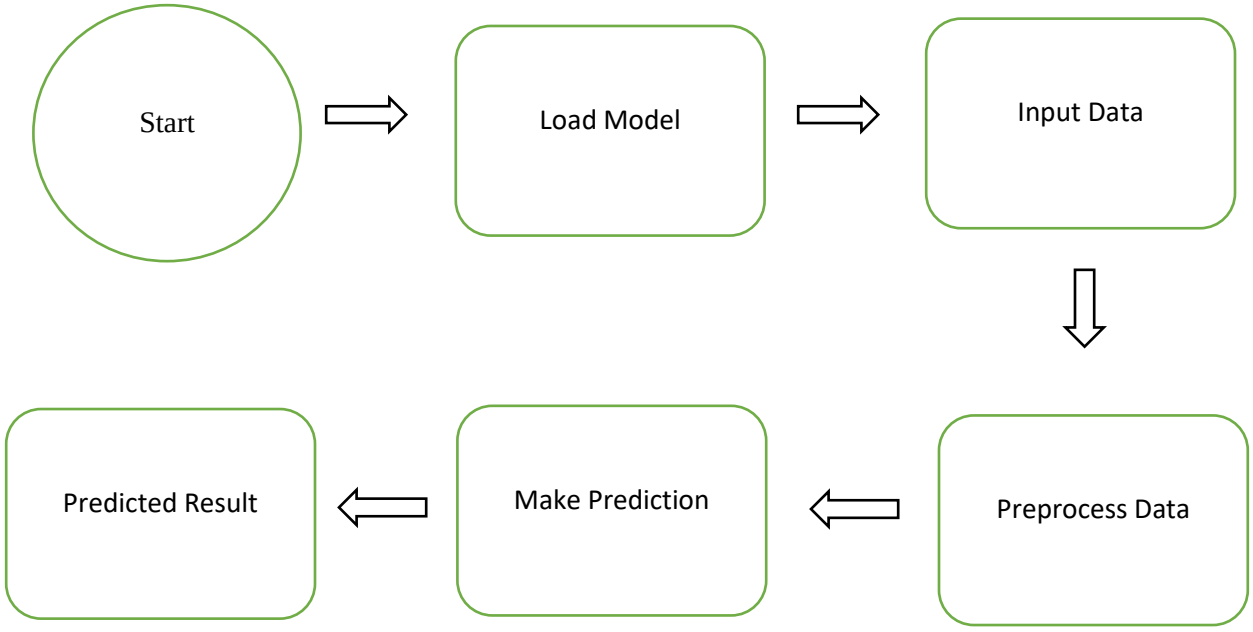
Proposed Methodology



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



3.2 Event Log

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

Initial Step-By-Step Description:

1. The System identifies at what step logging is required.
2. The System should be able to log each system flow.
3. Developer can choose the logging method. You can choose database logging/File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues, so logging is mandatory too.

3.3 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 Regression Model (XGBOOST) is used to detect the flight ticket prices increase or decrease now and then. So, people can plan to trip will surely help them to save money and time.

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 use Python as an interface between them. Each component will have its task to perform, and it is the job of Python to ensure the proper transfer of information.

4.3 Resource Utilization

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

4.4 Deployment

5 Dashboards

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the unveiled problems that if not addressed in time could cause catastrophes of unimaginable impact.

5.1 KPIs (Key Performance Indicators)

1. Key indicators display a summary of the Air Flights.
2. Chart showing total flights travelled through different routes for a particular date.
3. Summary statistics for each Flight.
4. Total flights travelled from source to destination based on the number of stops.
5. The total flights and average duration taken by different flights to reach the destination in minutes over a period.
6. Average flight price travelled from source to destination over a period.
7. Relationships between average duration and the average price of each flight travelled from source to destination.
8. Summary statistics for each stop.
9. Flight fare distribution travelled from source to destination.

6 Conclusion

The Regression Model (XGBOOST) is used to detect the flight ticket prices increase or decrease now and then. So, people can plan trips that will surely help them to save money and time.

7 References

1. https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh
2. https://www.ijraset.com/research-paper/flight-fare-prediction-system-using-ml