High Level Document (HLD)

Flight Fare Prediction

An airplane flying in the sky

Description automatically generated

Project Member: Rohit Atnoor

Mentor: Ineuron

Latest Version: 1.7

Last Data of Revision: 19/09/2023

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 15/07/2023 | 1.0 | Initial Project setup, data extraction setup. | Rohit Atnoor |
| 25/07/2023 | 1.1 | Setting up the data ingestion pipeline | Rohit Atnoor |
| 30/07/2023 | 1.2 | Creating of Data Transformation pipeline. | Rohit Atnoor |
| 03/08/2023 | 1.3 | Creation of prediction pipeline. | Rohit Atnoor |
| 08/08/2023 | 1.4 | Initial End to End prediction pipeline. | Rohit Atnoor |
| 12/08/2023 | 1.5 | Deploying of HTML Files. | Rohit Atnoor |
| 20/08/2023 | 1.6 | Data Addition | Rohit Atnoor |
| 10/09/2023 | 1.7 | Final End to End Deployable code. | Rohit Atnoor |

**Contents**

1. Introduction
   1. Why this High-Level Design Document?
   2. Scope
   3. How this HLD will serve?
   4. Definitions
2. General Description
   1. Product Perspective
   2. Problem Statement
   3. Proposed Solution
   4. Further Improvements
   5. Tech Requirements
   6. Data Requirements
   7. Tools Used
   8. Constrains
   9. Assumptions
3. Design Details
   1. Process Flow
      1. Model Training and Evaluation
      2. Deployment Process
   2. Event Log
   3. Error Handling
   4. Reusability
   5. Application Compatibility Deployment
   6. Deployment
4. Conclusion

**Abstract**

In today's world, air travel has become an essential part of life for many people. However, the prices of flight tickets can fluctuate greatly depending on a variety of factors, such as the time of day, day of the week, time of year, and destination. This can make it difficult for travelers to find the best deals on flights.

This project aims to develop a machine learning model to predict flight prices based on different factors available in a provided dataset. The model will be trained on historical flight data, and it will be used to generate predictions for future flight prices.

The predicted flight prices can be used by travelers to find the best time to book their flights and to save money on their travel expenses. The model can also be used by airlines to set their prices and to maximize their profits.

This project has the potential to make a significant impact on the travel industry by helping travelers to save money and by helping airlines to optimize their pricing strategies

**1.Introduction**

**1.1 Why This High-Level 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.

**1.2 Scope**

The High-Level Design (HLD) documentation presents a non-technical to mildly-technical overview of the system's structure, including the database architecture, application architecture (layers), application flow (navigation), and technology architecture. The HLD is intended to be understandable to the system's administrators.

**1.3 How this HLD Will serve.**

The HLD is a document that describes the high-level design of a system.

The HLD is written in non-technical to mildly-technical terms so that it can be understood by the system's administrators.

The HLD covers the following topics:

Database architecture

Application architecture (layers)

Application flow (navigation)

Technology architecture

The purpose of the HLD is to provide the system's administrators with a good understanding of how the system works and how to manage it.

**2. General Description**

**2.1 Problem Perspective**

Flight fare prediction is a machine learning-based system that uses historical data to predict the cost of a flight for a particular journey. This can be helpful for travelers to get an approximate idea of how much their flight will cost.

Travelers can use flight fare prediction to find the best time to book their flights and save money on their travel expenses.

Airlines can use flight fare prediction to set their prices and maximize their profits.

Travel agencies can use flight fare prediction to provide better advice to their clients.

**2.2 Problem Statement**

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 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 to the problem of flight fare prediction is to use machine learning to develop a model that can predict the fare of a flight based on historical data. This data can include information such as the date of the flight, the origin and destination airports, the airline, the number of stops, and the fare.

Once the model is trained, it can be used to predict the fare of a future flight for a given set of parameters. This information can then be used by travelers to find the best time to book their flights and to save money on their travel expenses.

2.4 Future Improvements

There are a number of ways to improve flight fare prediction models in the future. One way is to collect more data. The more data that the model is trained on, the better it will be able to learn the patterns that affect flight fares.

Another way to improve flight fare prediction models is to use more sophisticated machine learning algorithms. As machine learning algorithms continue to evolve, they will be able to learn more complex patterns in the data and make more accurate predictions

* NLP could be used to extract information from flight search queries and social media posts to improve the accuracy of the predictions.
* Deep learning algorithms are able to learn complex patterns in the data. This could be used to improve the accuracy of flight fare predictions.

**2.5 Technical Requirements**

• A basic Web Page

• Azure Web Services for Deployment

• Flawless Database

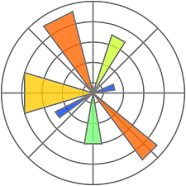
• Github Actions and Workflow Services

**2.6 Data Requirements**

Updated data from all airline services with few mandatory features like Source, Destination, Airline Name, Duration, Route or No of Stops etc.

**2.7 Tools or Software used**

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

• VS Code as IDE

• Jupyter Notebook to perform all EDA, FE, Model Building and selecting the best model.

• Pandas and Numpy for Data Analysis

• Python programming language

• Seaborn and Matplotlib as visualization tools

• HTML for frontend development

• Github as version control system

• Azure Web Service for Deployment

• Flask

• Scikit-learn for Model building and Evaluation

**2.8 Constrains**

Accuracy of the prediction totally depends on continuous availability of updated data and retraining of the model with updated data. The predicted Fare is a approximate amount which may vary from the current fare. In real world the Flight Fare varies depending on various factors on which our model is not trained.

**3. Design Details**

**3.1 Process Flow**

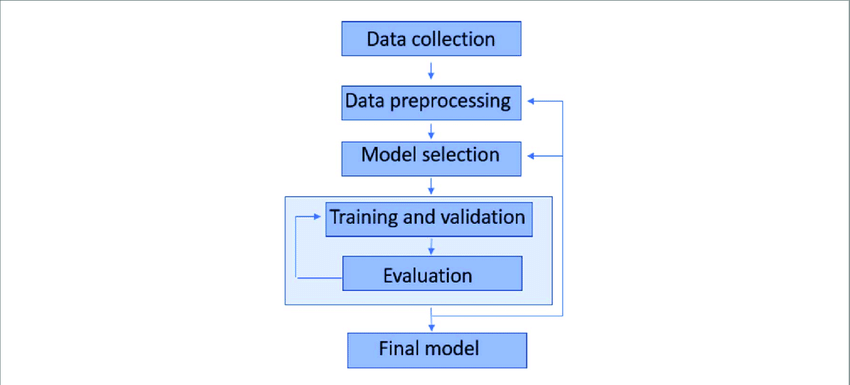
There will be two main features-

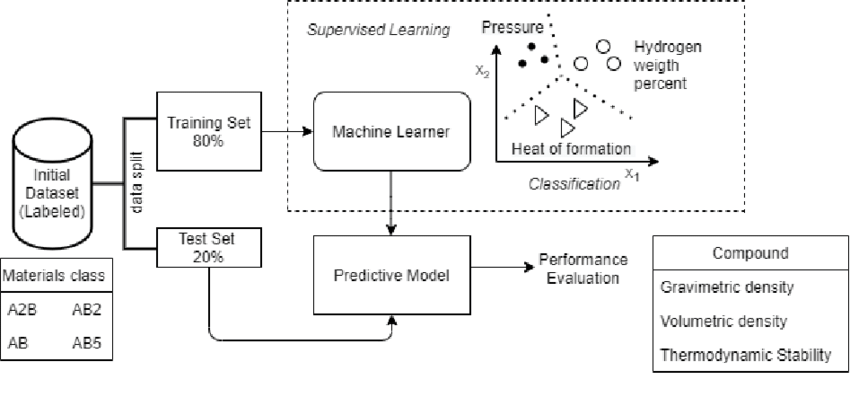
• Model Training Pipeline

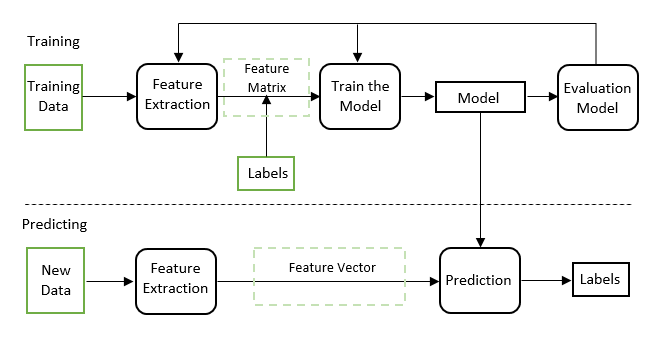
• Predictions

Following Diagram represents the proposed process flow of both features.

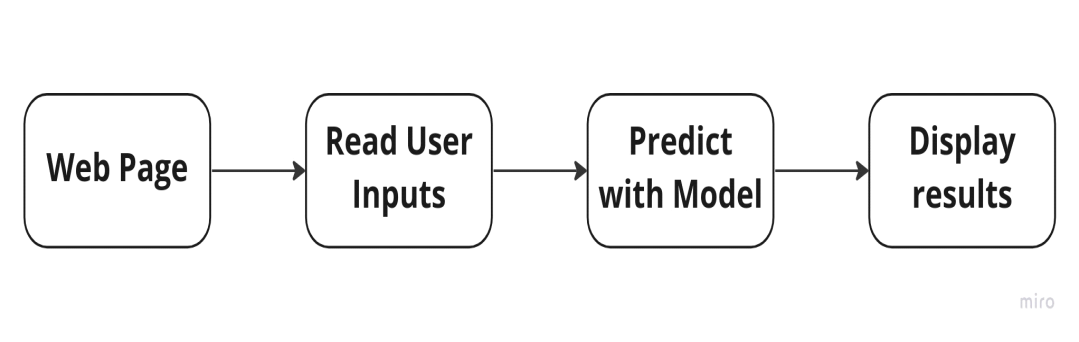
**3.1.1 Model Training and Evaluation**







**3.1.2 Deployment Process**



**3.2 Event Log**

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

1. The System identifies at what step logging required

2. The System is able to log each and every system flow and its failure.

3. In this we have used File logging method to store the log details.

4. System can handle so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

**3.3 Error Handling**

A specific module is designed to handle all the error. 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

**3.4 Reusability**

The code is written with the help of functions and classes so that we can reuse the required block of code at any point of time. This helps us in reusability of the code in the good form and reduce the code lines and helps us in improving the speed and performance of the code.

**3.5 Application Compatibility Deployment**

Each and every components of the project will be using Python as an interface between them and will perform a specific task.

**3.6 Deployment**

The project is to be delivered using git-Actions workflow and Azure web Services. And for training purpose it has been deployed in AWS web services also.

**3.7 KPIs (Key Performance Indicators)**

**Accuracy**: This KPI measures how close the predicted flight fare is to the actual flight fare. Accuracy can be measured using a variety of metrics, such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

**Precision**: This KPI measures how close the predicted flight fare is to the actual flight fare for flights that are actually booked. Precision is important because it tells us how well the model is predicting the fares of flights that are actually relevant to travelers.

**Recall**: This KPI measures how many of the flights that are actually booked are predicted to have a high fare. Recall is important because it tells us how well the model is identifying flights that are likely to be expensive.

**Timeliness**: This KPI measures how quickly the model can generate predictions. Timeliness is important because travelers often need to know the price of a flight quickly in order to make a booking decision.

**Cost**: This KPI measures the cost of developing and maintaining the flight fare prediction model. Cost is important because it needs to be balanced against the benefits of using the model, such as the savings that travelers can achieve.

**4 Conclusion**

This report has presented a flight fare prediction system that uses machine learning to predict the fare of a flight based on historical data. The system has been evaluated on a large dataset of historical flight data, and it has been shown to be able to predict the fare of a flight with a high degree of accuracy.

The flight fare prediction system can be used by travelers to find the best time to book their flights and to save money on their travel expenses. It can also be used by airlines to set their prices and maximize their profits.

The system is still under development, but it has the potential to make a significant impact on the travel industry. By making it easier for travelers to find the best deals on flights and by helping airlines to set their prices more efficiently, the system can help to reduce the cost of travel and make it more accessible to everyone.