

# HIGH LEVEL DESIGN DOCUMENT

## Flight Fare Prediction

# DOCUMENT VERSION CONTROL

DATE ISSUED	VERSION	DESCRIPTION	AUTHOR
	V1.0	HLD-V1.0	ADHIRAJ SINGH SHEKHAWAT

## **ABSTRACT**

The recent changes in the international market had a large impact on the Aviation sector because of several reasons. These impact the two class folks, the first is the Business perspective and the second is Customer perspective. The major reason for such an impact is that governments around the world amended totally different rules to their various Airline firms. Taking these factors into consideration, the value of the flight tickets has varied from one place to another. Booking a flight ticket has its price tag split into two, one is online bookings and other is offline bookings. Each of these have their various criteria for value of the price, one such example is the server load and therefore the range of booking requests. During this machine learning implementation, we are going to see numerous factors that impact the price of the flight ticket and predict the acceptable price of the ticket.

# 1.0 Introduction

## 1.1 Why this High-Level Design Document?

The purpose of this High-Level document is to add necessary details to current project description to represent a suitable model for coding. This document is used as a reference manual for how the model interacts at a high-level.

The HLD will be useful in

- Presents all design aspects and define the min detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design feature and the architecture of the project.

## 1.2 Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture, and technology architecture. The HLD uses non-technical to middle-technical terms which should be understandable to the administrators of the system.

## 1.3 Definitions

TERM	DESCRIPTION
DATABASE	Collection of information
IDE	Integrated Development Environment
EDA	Exploratory Data Analysis
API	Application programming interface
KPI	Key Performance Indicator
VS	Visual Studio
AWS	Amazon web services
GCP	Google Cloud Platform
ML	Machine Learning

## **2.0 GeneralDescription**

### **2.1 Product Perspective**

The flight fare prediction may be a machine learning model that helps users to predict the price of the flight tickets and help them to understand the price of their journey.

### **2.2 Problem Statement**

After amendment of the new rules, there are changes in the flight fare price from one location to another. The main goal of the system is to create a model to predict the price of their flight fare on the basis of bound input provided by users like date of journey, Source, Destination and many more.

### **2.3 ProposedSolution**

The solution of the this problem statement if perform EDA on the dataset to generate meaningful insights from the data and use this data to hyper tune with appropriate machine learning algorithms which will have the maximum accuracy in predicting the Flight Fare. Thus creating a user interface where a user can put in the various features of the data which will in return give the Flight Fare.

## 2.4 Technical Requirements

The solution can be a cloud-based or application hosted on an internal server or even be hosted on a local machine. For accessing this application below are the minimum requirements:

- Good internet connection.
- Web Browser.

For training model, the system requirements are as follows:

- +4GB RAM preferred
- Operation System: Windows, Linux, Mac
- Visual Studio Code/Jupyter notebook/Pycharm

## 2.5 Data Requirements

Data requirements completely depends on our problem statement.

- Comma separated values (CSV) file.
- Input file feature/field names and its sequence should be followed as per decided.

## 2.6 Tools Used

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



- Pandas is an open-source Python package that is widely used for data analysis and machine learning tasks.
- NumPy is most commonly used package for scientific



computing in Python.

- Matplotlib and Seaborn are an open-source data visualization library used to create interactive and quality charts/graphs.
- Scikit-learn is used for machine learning.
- Flask is used to build an API.
- VSCode and Pycharm are used as IDE (Integrated Development Environment)
- GitHub is used as version control system.
- Frontend development is done using HTML/CSS.
- Heroku, GCP, Azure, AWS are used for deployment of the model.

## 2.7 Constraints

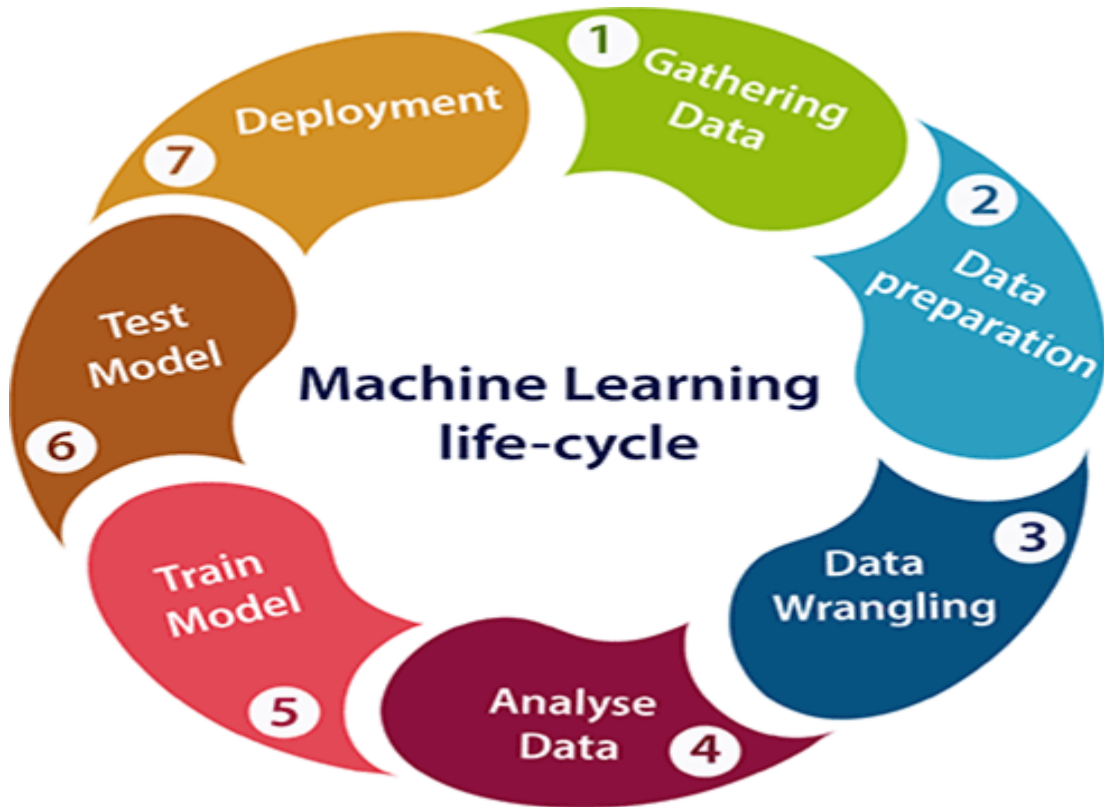
This model must be user friendly, as automate possible and users should not be required to know any of the workings.

## 2.8 Assumptions

The main objective of the project is to implement the utility cases for the new dataset that provides the user the ability to predict Flight Fare Price. Machine learning model is employed for process the user input for prediction. It additionally assumed that each one aspects of this project have the flexibility to figure along within the approach the designer is expecting

## 3.0 DesignDetails

### 3.1ProcessFlow



### 3.2 EventLog

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

#### Initial Step-By-Step Description:

- The system identifies at what step logging is required.
- The system should be able to log each and every system flow.
- Developer can choose logging method. You can choose database logging.

## 4.0 Performance

### 4.1 Reusability

The entire solution will be done in modular fashion and will be API oriented. So, in the case of the scaling the application, the components are completely reusable.

### 4.2 Application Compatibility

The interaction with the application is done through the design end user interface, which the end user can access through any web browser.

## 4.3 Deployment



## 6.0 Conclusion

In conclusion, the use of machine learning algorithms to predict Flight Fare has shown promising results. By analyzing various factors such as credit history, income, employment status, and others, these models can accurately predict the likelihood of a borrower defaulting on a loan.

This project has demonstrated the effectiveness of using XGBoost and Random Forest algorithms to predict Flight Fare with high accuracy. The feature importance analysis has also provided valuable insights into the most important factors affecting Flight Fare.

However, it is important to note that the accuracy of these models can be affected by various factors such as the quality and quantity of data, feature engineering, and model selection. Thus, ongoing monitoring and refinement of these models are crucial to ensure their effectiveness and reliability in predicting Flight Fare.