

# Architecture

## FLIGHTFARE PREDICTION

Document Version: 0.3

Last Revised Date: 14-06-2023

Rajesh Singh Gurjar

### Document Version Control:

Version	Date	Author	Description
0.1	10-06-2023	Rajesh	Introduction
0.2	11-06-2023	Rajesh	Architecture
0.3	14-06-2023	Rajesh	Architecture Design

## Contents

<b>1</b>	<b>Introduction -----</b>	<b>3</b>
	<b>1.1 What is Architecture -Level design document ? -----</b>	<b>3</b>
	<b>1.2 Scope-----</b>	<b>3</b>
<b>2</b>	<b>Architecture -----</b>	<b>3</b>
<b>3</b>	<b>Architecture Description -----</b>	<b>4</b>
	<b>3.1 Data Gathering -----</b>	<b>4</b>
	<b>3.2 Data Description -----</b>	<b>4</b>
	<b>3.3 Tool Used -----</b>	<b>5</b>
	<b>3.4 Data Pre-processing-----</b>	<b>5</b>
	<b>3.5 Model Building -----</b>	<b>5</b>
	<b>3.6 Data from User -----</b>	<b>6</b>
	<b>3.7 Data Validation-----</b>	<b>6</b>
	<b>3.8 Rendering Result-----</b>	<b>6</b>
	<b>3.9 Deployment -----</b>	<b>6</b>

## 1. Introduction

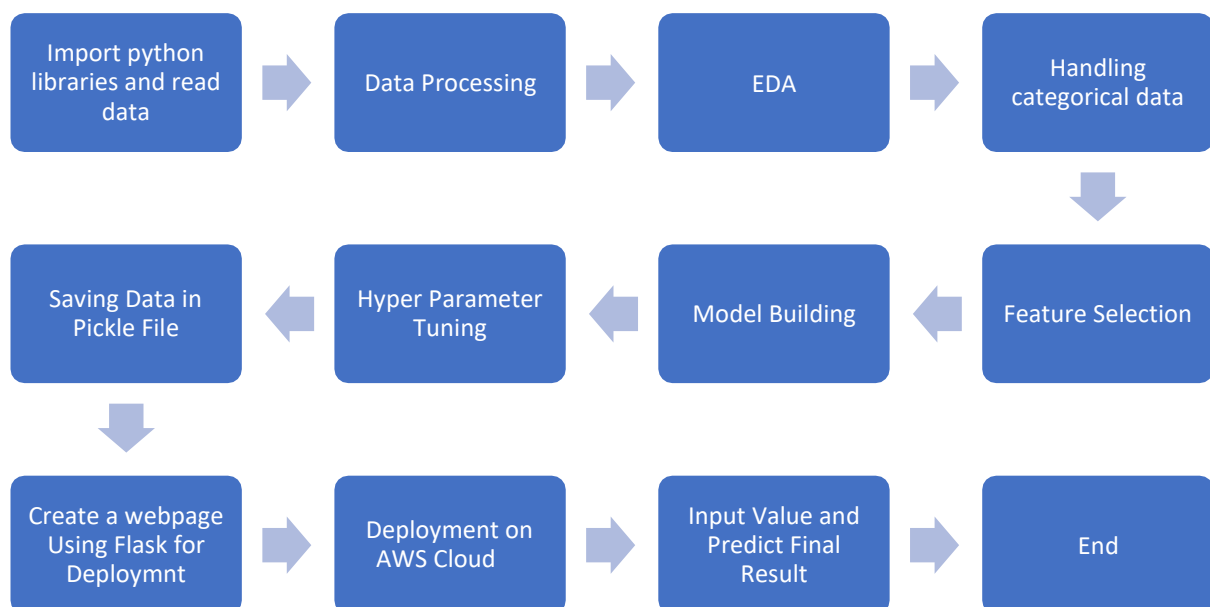
### 1.1. What is Architecture design document?

The main purpose of this LLD documentation is to feature the required details of the project and supply the outline of the machine learning model and also the written code. This additionally provides the careful description on however the complete project has been designed end-to-end.

### 1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-bystep refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work

## 2. Architecture



### 3. Architecture Description

This project aims to create a user interface that provides users with approximate flight ticket prices. Additionally, to gain real-time project experience, we have extracted the data, processed it, build model by inferring the data, did hyperparameter tuning to optimize the model, built batch prediction pipeline, instance prediction API and then deployed it on AWS EC2 instance.

#### 3.1. Data Gathering

The data for the current project is being gathered from Kaggle dataset, the link to the data is:

<https://www.kaggle.com/nikhilmittal/flight-fare-prediction-mh>

#### 3.2. Data Description

There are about 10k+ records of flight information such as airlines, data of journey, source, destination, departure time, arrival time, duration, total stops, additional information, and price. A glance of the dataset is shown below.

	Airline	Date of Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
1	IndiGo	24/03/2019	Bangalore	New Delhi	BLR → DEL	22:20	01:10 22	12h 50m	non-stop	No info	3897
2	Air India	1/05/2019	Kolkata	Bangalore	CCU → IXF	05:50	13:15	7h 25m	2 stops	No info	7662
3	Jet Airway	9/06/2019	Delhi	Cochin	DEL → LKO	09:25	04:25 10	19h	2 stops	No info	13882
4	IndiGo	12/05/2019	Kolkata	Bangalore	CCU → NA	18:05	23:30	5h 25m	1 stop	No info	6218
5	IndiGo	01/03/2019	Bangalore	New Delhi	BLR → NA	16:50	21:35	4h 45m	1 stop	No info	13302
6	SpiceJet	24/06/2019	Kolkata	Bangalore	CCU → BLI	09:00	11:25	2h 25m	non-stop	No info	3873
7	Jet Airway	12/03/2019	Bangalore	New Delhi	BLR → BOI	18:55	10:25 13	15h 30m	1 stop	In-flight m	11087
8	Jet Airway	01/03/2019	Bangalore	New Delhi	BLR → BOI	08:00	05:05 02	12h 5m	1 stop	No info	22270
9	Jet Airway	12/03/2019	Bangalore	New Delhi	BLR → BOI	08:55	10:25 13	12h 30m	1 stop	In-flight m	11087
10	Multiple c	27/05/2019	Delhi	Cochin	DEL → BOI	11:25	19:15	7h 50m	1 stop	No info	8625
11	Air India	1/06/2019	Delhi	Cochin	DEL → BLF	09:45	23:00	13h 15m	1 stop	No info	8907
12	IndiGo	18/04/2019	Kolkata	Bangalore	CCU → BLI	20:20	22:55	2h 35m	non-stop	No info	4174
13	Air India	24/06/2019	Chennai	Kolkata	MAA → CC	11:40	13:55	2h 15m	non-stop	No info	4667
14	Jet Airway	9/05/2019	Kolkata	Bangalore	CCU → BO	21:10	09:20 10	12h 10m	1 stop	In-flight m	9663
15	IndiGo	24/04/2019	Kolkata	Bangalore	CCU → BLI	17:15	19:50	2h 35m	non-stop	No info	4804
16	Air India	3/03/2019	Delhi	Cochin	DEL → AM	16:40	19:15 04	12h 35m	2 stops	No info	14011
17	SpiceJet	15/04/2019	Delhi	Cochin	DEL → PN	08:45	13:15	4h 30m	1 stop	No info	5830
18	Jet Airway	12/06/2019	Delhi	Cochin	DEL → BOI	14:00	12:35 13	12h 35m	1 stop	In-flight m	10262

### 3.3. Tool Used

- Python 3.8 is used as the programming language, along with frameworks like numpy, pandas, scikit-learn, and other modules for building the model.
- VS Code is employed as IDE.
- For visualization, seaborn and components of matplotlib are used.
- GitHub is used as a repository for the source code. It is used for the version management. GitHub Actions is used, which is a powerful workflow automation and CI/CD (Continuous Integration/Continuous Deployment) platform provided by GitHub. It allows you to automate tasks and workflows directly within your GitHub repository.
- AWS EC2 is used for deployment. At AWS, ECR is used to maintain the Docker image of the project.

### 3.4. Data Pre-processing

Steps performed in pre-processing are:

- First the info sorts square measure being checked and located solely the value column is of sort number.
- Checked for null values as there square measure few null values, those rows square measure born.
- Converted all the desired column into the date time format.
- Performed one-hot cryptography for the desired columns.
- Scaling is performed for needed information.
- And, the info is prepared for passing to the machine learning formula

### 3.4. Model Building

The pre-processed data is visualized, and relevant insights are extracted. Although the insights are randomly distributed, we perform modeling using various machine learning algorithms to cover all possibilities. Finally, random forest regression proves to be the best-performing algorithm, and hyperparameter tuning is conducted to enhance the model's accuracy.

### **3.5. Data from User**

The data from the user is retrieved from the created HTML web page.

### **3.6. Data Validation**

The data provided by the user is then being processed by app.py file and validated. The validated data is then sent for the prediction.

### **3.7. Rendering Result**

The data sent for the prediction is then rendered to the web page.

### **3.8. Deployment**

The tested model is then deployed to AWS EC2 machine. So, users can access the project from any internet devices.