Low-Level Design (LLD)
Flight Fare Estimator

Revision Number – 1.1
Last Date of Revision: 12–01–2023

Shubham Shaah

Document Version Control

Date	Version	Description	Author
10 – 11 - 2022	1.0	Abstract Introduction Architecture	Shubham
12 – 01 – 2023	1.1	Architectural Design	Shubham

Contents

Abstract

The aviation market has seen significant changes and challenges in recent years. The COVID-19 pandemic has had a major impact on the industry, causing widespread disruptions and reductions in demand for air travel. Despite this, the industry is beginning to recover as vaccine rollouts and declining case numbers lead to increased consumer confidence. However, the industry is still facing challenges, such as shifting consumer preferences for travel and a rise in low-cost carriers. In response, many airlines and travel companies are investing in new technologies, such as flight fare prediction systems, to remain competitive and meet the evolving needs of travellers.

The prediction of flight fares is important because it helps travellers to make informed decisions when booking their travels. By having access to accurate and up-to-date predictions of flight fares, travellers can save money and avoid overpaying for their flights. Additionally, a prediction system for flight fares can also help airlines and travel companies to optimize pricing strategies and make data-driven decisions. In a highly competitive market, being able to predict flight fares can provide a significant advantage to both travellers and companies.

1. Introduction

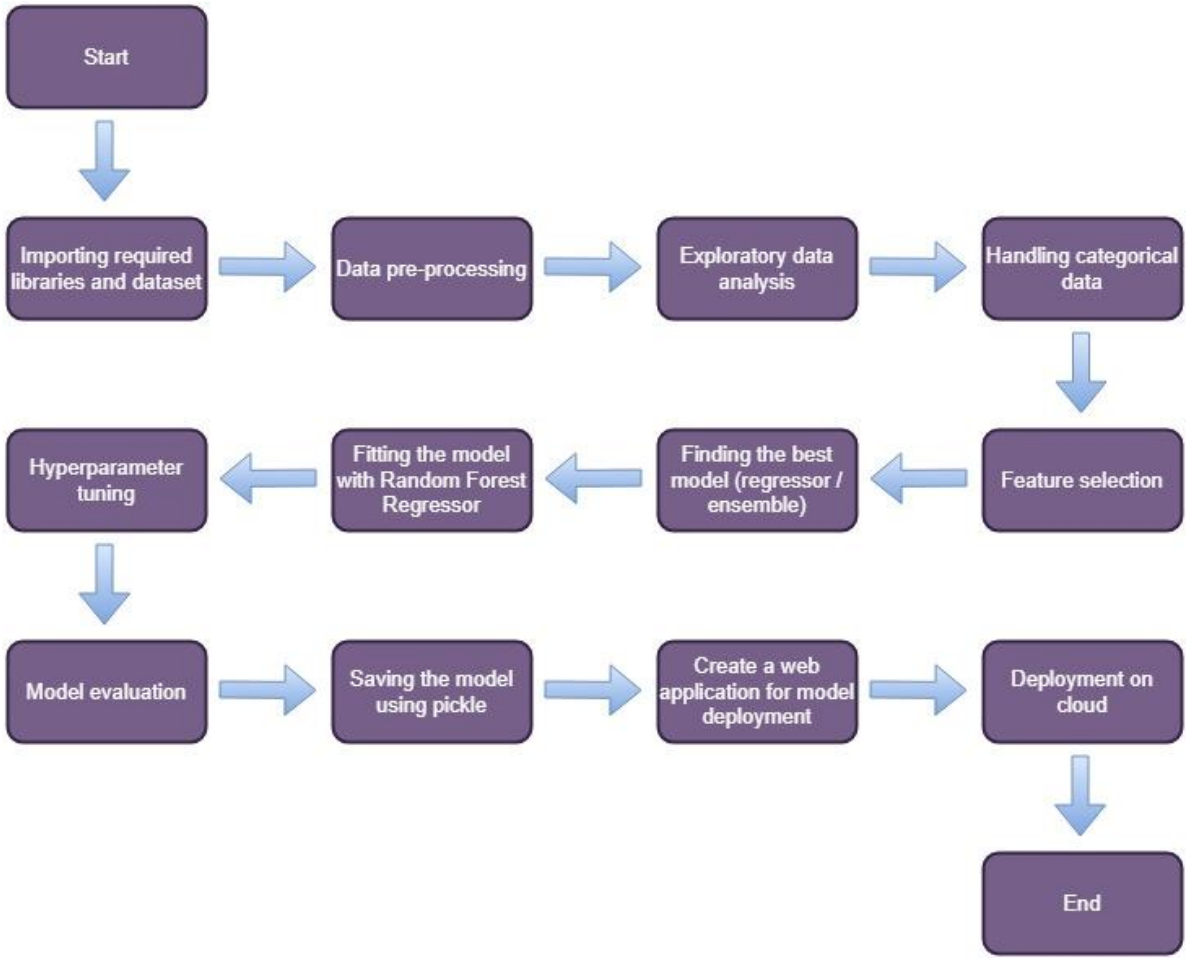
1.1 Why is the LLD Document?

The main goal of the LLD document is to give the internal logic design of actual code implementation and supply the outline of the machine learning model and its implementation. Additionally, it provides the description of how our project will be designed end-to-end.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step 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 Design

This project is designed to make an interface for the User to predict its approximate flight ticket price.

3.1 Data Collection

The data for this project is collected from the Kaggle Dataset, the URL for the dataset is kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh

3.2 Data Description

Flight Fare Prediction is a 10K+ dataset publicly available on the Kaggle. The information in the dataset is present in two different excel files named as train.xlsx and test.xlsx. The dataset contains 10683 rows which shows the information such as Date of Journey, Source, Destination, Arrival Time, Departure Time, Total stops, Airlines, Additional Info, and Price. A glance of the Dataset is:

Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302
SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BLR	09:00	11:25	2h 25m	non-stop	No info	3873
Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	18:55	10:25 13 Mar	15h 30m	1 stop	In-flight meal not included	11087
Jet Airways	01/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:00	05:05 02 Mar	21h 5m	1 stop	No info	22270
Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:55	10:25 13 Mar	25h 30m	1 stop	In-flight meal not included	11087
Multiple carriers	27/05/2019	Delhi	Cochin	DEL → BOM → COK	11:25	19:15	7h 50m	1 stop	No info	8625
Air India	1/06/2019	Delhi	Cochin	DEL → BLR → COK	09:45	23:00	13h 15m	1 stop	No info	8907
IndiGo	18/04/2019	Kolkata	Banglore	CCU → BLR	20:20	22:55	2h 35m	non-stop	No info	4174
Air India	24/06/2019	Chennai	Kolkata	MAA → CCU	11:40	13:55	2h 15m	non-stop	No info	4667
Jet Airways	9/05/2019	Kolkata	Banglore	CCU → BOM → BLR	21:10	09:20 10 May	12h 10m	1 stop	In-flight meal not included	9663
IndiGo	24/04/2019	Kolkata	Banglore	CCU → BLR	17:15	19:50	2h 35m	non-stop	No info	4804
Air India	3/03/2019	Delhi	Cochin	DEL → AMD → BOM → COK	16:40	19:15 04 Mar	26h 35m	2 stops	No info	14011
SpiceJet	15/04/2019	Delhi	Cochin	DEL → PNQ → COK	08:45	13:15	4h 30m	1 stop	No info	5830
Jet Airways	12/06/2019	Delhi	Cochin	DEL → BOM → COK	14:00	12:35 13 Jun	22h 35m	1 stop	In-flight meal not included	10262
Air India	12/06/2019	Delhi	Cochin	DEL → CCU → BOM → COK	20:15	19:15 13 Jun	23h	2 stops	No info	13381
Jet Airways	27/05/2019	Delhi	Cochin	DEL → BOM → COK	16:00	12:35 28 May	20h 35m	1 stop	In-flight meal not included	12898
GoAir	6/03/2019	Delhi	Cochin	DEL → BOM → COK	14:10	19:20	5h 10m	1 stop	No info	19495
Air India	21/03/2019	Banglore	New Delhi	BLR → COK → DEL	22:00	13:20 19 Mar	15h 20m	1 stop	No info	6955

3.3 Exporting Data from Database

In the above-created API, the download URL is also being created, which downloads the data into a CSV file format.

3.4 Data Pre-processing

- Checked for info on the Dataset, to verify the correct datatype of the Columns.
- Checked for Null values, because the null values can affect the accuracy of the model.
- Converted all the desired columns into Date time format.
- Performed One – Hot encoding on the desired columns.
- Checking the distribution of the columns to interpret their importance.
- Now, the info is prepared to train a Machine Learning Model.

3.5 Model Creation

The Pre-processed info is now envisioned and drawn insights help us to select the feature that improves the accuracy of the model. The info is randomly used for modelling with different machine learning algorithms to create a model to predict the Flight ticket price. After performing on different algorithms, we use Random Forest Regression to create a model and then also perform Hyperparameter Tuning to improve the accuracy of the model.

3.6 Data from User

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

3.7 Data Validation

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

3.8 Rendering the Results

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

4. Deployment

The tested model is then deployed to Streamlit. So, users can access the project from any internet device.

4.1 Unit Test Cases

Test Case Description	Pre - Requisites	Expected Results
Verify whether the webpage is accessible to the user or not.	Webpage URL should bedefined.	Webpage should be accessible to the User.
Verify whether the webpage completely loads for the user or not.	Webpage URL is accessible. Webpage is deployed.	The webpage should be completely loaded for the user when it is accessed.
Verify whether the user can enter data in inputfields or not.	Webpage URL isaccessible. Web page is deployed. Webpage input fields are editable.	The User can enter the data in input fields.
Verify whether the user can submit details or not.	Webpage URL is accessible. Webpage is deployed. Webpage input fields are editable.	The user can submit details in the process.
Verify whether the user gets recommended results on submitting thedetails or not.	Webpage URL isaccessible. Webpage is deployed. Webpage input fields are editable.	The user gets recommended results on submitting the details.