**Flight Fare Prediction**

**Detail Project Report (DPR)**

FLIGHT FARE PREDICTION

**Revision Number – 1.3**

**Last Revised Date: 10-MAR-2023**

Kritika

Document Version Control:

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 08-03-2023 | 1.0 | Abstract,  Introduction | Kritika |
| 09-03-2023 | 1.2 | Process | Kritika |
| 10-03-2023 | 1.3 | Q and A | Kritika |

P a g e | **1**

**Flight Fare Prediction**

**Contents**

**Abstract----------------------------------------------------------------------------3**

**1 Introduction ----------------------------------------------------------------------------- 3**

## 1.1 What is High-Level design document? ----------------------------------- 3

**1.2 Scope ------------------------------------------------------------------------------- 3**

1. **Description ------------------------------------------------------------------------------- 3**

* 1. **Problem Perspective ----------------------------------------------------------- 3**

* 1. **Problem Statement ------------------------------------------------------------ 3**

* 1. **Purposed Solution -------------------------------------------------------------- 3**

* 1. **Solution Improvements ------------------------------------------------------- 4**

* 1. **Technical Requirements ------------------------------------------------------ 4**

* 1. **Data Requirements ------------------------------------------------------------ 4**

* 1. **Tool Used ------------------------------------------------------------------------- 4**

* 1. **Data Gathering ------------------------------------------------------------------ 4**

* 1. **Data Description ---------------------------------------------------------------- 4**

1. **Data Pre-Processing ------------------------------------------------------------------- 5**

1. **Design Flow ----------------------------------------------------------------------------- 5**

* 1. **Modeling ------------------------------------------------------------------------- 5**

* 1. **Modeling Process -------------------------------------------------------------- 6**

1. **Data from User ------------------------------------------------------------------------ 6**

1. **Data Validation ------------------------------------------------------------------------ 6**

1. **Rendering Result ---------------------------------------------------------------------- 6**

1. **Conclusions ------------------------------------------------------------------------------ -7**

1. **Q & A ------------------------------------------------------------------------------------- 7**

**P a g e | 2**

**Flight Fare Prediction**

**Abstract**

The recent global situations had a huge impact on the aviation sector due to many reasons. This impact has two category people, the first is business perspective and the second is the customers perspective. As safety is the major reason for such impact on the aviation sector, the governments around the world amended different rules to their respective airlines companies. These restrictions had made the availability of the flights and their attendee capacity less. Taking all these factors in consideration the cost of the flight tickets has increased and vary from one place to the other. Booking a flight ticket has split into two, one is the online and the other is the offline bookings. Both these have their respective criteria for cost of the ticket, one such example is the server load and the number of booking requests. In this machine learning implementation, we will see various factors that impact the cost of the flight ticket and predict the appropriate price of the ticket.

# 1. INTRODUCTION

**1.1. What is High-Level design document?**

The main purpose of this HLD 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**

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

# 2. Description

**2.1. Problem Perspective**

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

**2.2. Problem Statement**

The most goal of the project is to form a programme that predicts the price of the flight price tag by taking bound input from the user like date of journey, aboard location and destination etc.

**2.3. Purposed Solution**

Projected to require the desired input of user from the created interface and method all the provided information to satisfy the wants of the machine learning model and at last show the output oral communication so and then quantity is that the expected value.

P a g e | **3**

**Flight Fare Prediction**

**2.4. Solution Improvements**

We will even predict the price of price tag considering whether or not is it a weekday, season or alternative social reasons. However, considering from the angle of business, if we have a tendency to method such information and predict the price of the discounted price tag it'll bring some loss to the airlines company. Therefore, this technique isn't thought-about.

**2.5. Technical Requirements**

There are not any hardware needs needed for victimization this application, the user should have AN interactive device that has access to the web and should have the fundamental understanding of providing the input. And for the backend half the server should run all the package that's needed for the process the provided information and to show the results.

**2.6. Data Requirements**

The info demand is totally supported the matter statement. And also, the information set is accessible on the Kaggle within the type of standout sheet(.xlsx). Because the main theme of the project is to induce the expertise of real time issues, we have a tendency to once more mercantilism {the information into the prophetess data base and commerce it into csv format}.

**2.7. Tool Used**

* + Python 3.9 is employed because the programming language and frame works like numpy, pandas, sklearn and alternative modules for building the model.
  + PyCharm is employed as IDE.
  + For visualizations seaborn and components of matplotlib are getting used.
  + For information assortment prophetess info is getting used.
  + Front end development is completed by Flask.  Github is employed for version management.

**2.8. Data Gathering**

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

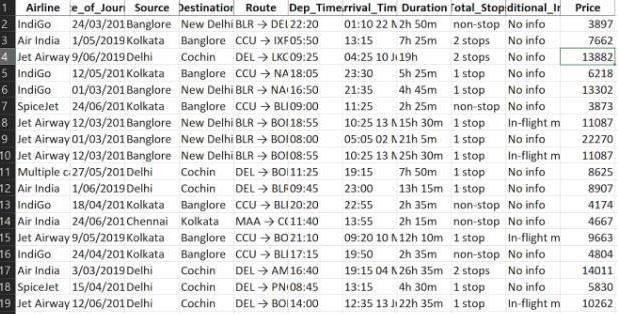
https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh

**2.9. 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.

### P a g e | 4

**Flight Fare Prediction**



**3. Data Pre-processing**

Steps performed in pre-processing are:

* First the info sort’s 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

# 4. Design Flow

**4.1. Modeling**

The pre-processed data is then visualized and all the required insights are being drawn. Although from the drawn insights, the data is randomly spread but still modeling is performed with different machine learning algorithms to make sure we cover all the possibilities. And finally, as expected random forest regression performed well and further hyper parameter tuning is done to increase the model’s accuracy.

P a g e | **5**

**Flight Fare Prediction**

**4.2. Modeling Process**

EDA

Create a webpage

using HTML

Input Value and

Predict Final Result

End

Import python

libraries and read

data

Data

Processing

Saving Data in

Pickle File

Hyper Parameter

Tuning

Model Building

Feature Selection

Handling

categorical data

EDA

**1.1. Data from User**

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

**1.2. 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.

**1.3. Rendering Result**

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

### P a g e | 6

**Flight Fare Prediction**

1. **Conclusion**

The flight fare prediction will predict the worth supported the trained knowledge set within the rule. Therefore, the user will recognize the approximate value for his or her journey.

1. **Q & A**

Q1) what’s the source of data?

The data for training is provided by the client in multiple batches and each batch contains multiple files.

Q 2) what was the type of data?

The data was the combination of numerical and Categorical values.

Q 3) What’s the complete flow you followed in this Project?

Refer Page no 6 for better Understanding.

Q 4) After the File validation what you do with incompatible file or files which didn’t pass the validation?

Files like these are moved to the Achieve Folder and a list of these files has been shared with the client and we

Removed the bad data folder.

Q 5) How logs are managed?

We are using different logs as per the steps that we follow in validation and modeling like File validation log, Data

Insertion, Model Training log, prediction log etc.

Q 6) What techniques were you using for data pre-processing?

* Removing outliers

* Cleaning data and imputing if null values are present.

* Converting categorical data into numeric values.

Q 7) How training was done or what models were used?

* Before dividing the data in training and validation set, we performed pre-processing over the data set and made the final dataset.

* As per the dataset training and validation data were divided.

* Algorithms like Linear regression, SVM, Decision Tree, Random Forest, and XGBoost were used based on the recall, final model was used on the dataset and we saved that model.

Q 8) How Prediction was done?

The testing files are shared by the client. We performed the same life cycle on the provided dataset. Then, on the basis of dataset, model is loaded and prediction is performed. In the end we get the accumulated data of predictions. P a g e | **7**