**Architecture Design**

[Flight Fare Prediction](https://github.com/ganeshss0/flight-fare-prediction)

Revision Number: 1.0

Last Date of Revision: 31/01/2024

[**Ganesh Swami**](https://linkedin.com/in/ganesh0)

**Libraries Requirement**

* [Pandas](https://pandas.pydata.org/)
* [Numpy](https://numpy.org/)
* [Scikit-Learn](https://scikit-learn.org/stable/)
* [Openpyxl](https://openpyxl.readthedocs.io/en/stable/tutorial.html)
* [Matplotlib](https://matplotlib.org/stable/)
* [Seaborn](https://seaborn.pydata.org/)
* [LightGBM](https://lightgbm.readthedocs.io/en/latest/index.html)
* [Optuna](https://optuna.org/)
* [Tqdm](https://tqdm.github.io/)
* [XGBoost](https://xgboost.readthedocs.io/en/stable/)
* [Streamlit](https://docs.streamlit.io/get-started/installation)
* [Joblib](https://joblib.readthedocs.io/)

Contents

|  |  |
| --- | --- |
| * Abstract | 4 |
| 1. Introduction  * Why this Architecture design documentation? | 5  5  5 |
| 1. Architecture  * Architecture Design * Data Gathering * Data Description * Data pre-processing * Modelling * UI Integration * Data from User * Data Validation * Rendering the results * Deployment | 6  7  7  7  7  7  7  7  7  8  8  8 |
| 1. Application Screenshot | 9 |

* **Abstract**

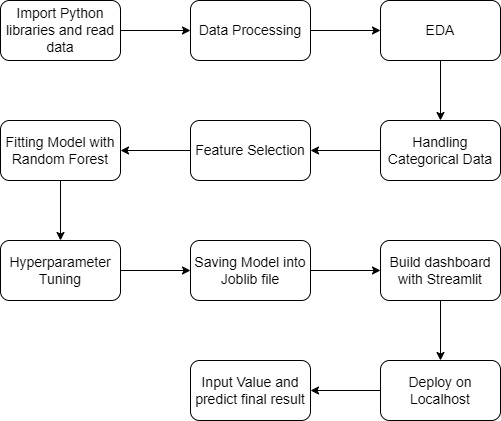
The recent global situations had a huge impact on the aviation sector due to many reasons. This impact has two category people, the first is business perspective and the second is the customers perspective. As safety is the major reason for such impact on the aviation sector, the governments around the 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 fare of the ticket.

1. Introduction

* Why this Architecture Design documentation?

The main objective of the Architecture design documentation is to provide the internal logic understanding of the flight fare prediction code. The Architecture design documentation is designed in such a way that the programmer can directly code after reading each module description in the documentation.

1. Architecture



**2.1. Architecture design**

This project is to create an interface for the user to know their approximate flight ticket price, in addition to this, in need of getting the real time project experience we are importing the gathered data into our own database and then start the project from the scratch.

**2.2. Data gathering from main source**

The data for the current project is being gathered from Kaggle dataset, the link to the data is: [Indian Flight Fare Dataset (kaggle.com)](https://www.kaggle.com/datasets/dopodix/indian-flight-fare-dataset)

**2.3. Data description**

There are about 10k+ records of flight information such as “Airlines”, “Date of Journey”, “Source”, “Destination”, “Departure Time”, “Arrival Time”, “Duration”, “Total Stops”, “Additional Information” and “Price”.

**2.4. Data Pre-Processing**

Steps performed in pre-processing are:

* First the data types are being checked and found only the price column is of type integer.
* Checked for null values as there are few null values, those rows are dropped.
* Converted all the required column into the date time format.
* Performed one-hot encoding for the required columns.
* Scaling is performed for required data.
* The data is ready for passing to the machine learning algorithm.

**2.5. Modelling**

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 modelling 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 hyperparameter tuning is done to increase the model’s accuracy.

**2.6. UI integration**

Streamlit library is used to create the webpage of project. The webapp is created inside `app.py` file.

**2.7. Data from user**

The data from the user is retrieved from the created Streamlit webpage.

**2.8. 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.

**2.9. Rendering the results**

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

**2.11. Deployment**

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

1. Application Screenshot

