

Architecture

PHISHING DOMAIN DETECTION

Revision Number – 1.3

Last Date of Revision – 10-FEB-2023

Abhishek

Document Version Control:

Date	Version	Author	Description
31-01-2023	1.0	Abhishek	Abstract, Introduction
02-02-2023	1.1	Abhishek	Architecture
10-02-2023	1.2	Abhishek	Architecture Design

Contents

Abstract

1 Introduction ----- 3

1.1 What is Architecture -Level design document? ----- 3

1.2 Scope ----- 3

2 Architecture ----- 4

3 Architecture Descriptions ----- 4

3.1 Data Gathering ----- 4

3.2 Data Description ----- 5

3.3 Tool Used ----- 5

3.4 Export data from database-----5

3.5 Data Pre-processing ----- 6

3.6 Model Building ----- 6

3.7 Data from User ----- 6

3.8 Data Validation ----- 6

3.9 Rendering Result ----- 6

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 customer’s 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

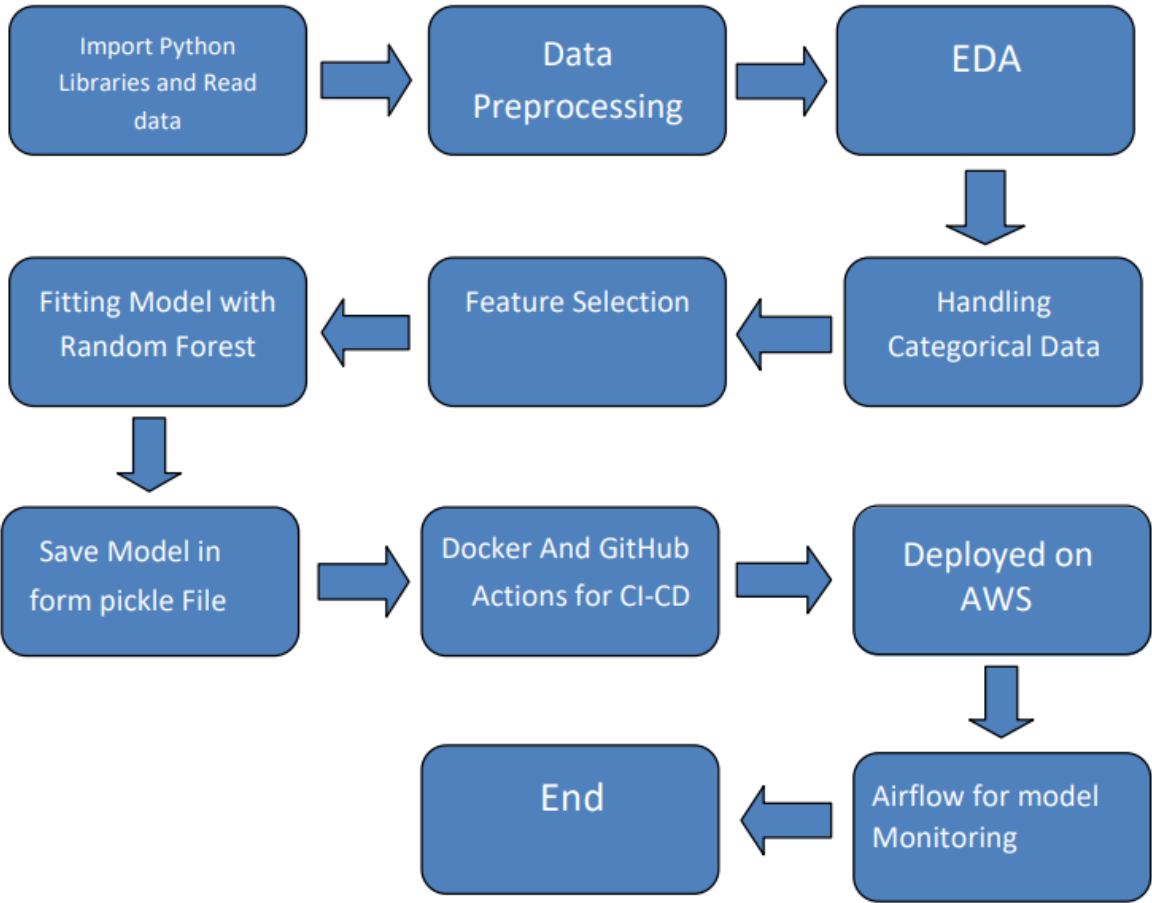
1.1. What is Architecture design document?

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-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



2. Architecture Description

This project is to make associate interface for the user to grasp their approximate flight price ticket worth, additionally to the present, in would like of obtaining the important time project expertise we have a tendency to square measure mercantilism the gathered information into our own information then begin the project from the scratch.

3.1. Data Gathering

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

<https://data.mendeley.com/datasets/72ptz43s9v/1>

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.

1	qty_dot_u	qty_hyph	qty_under	qty_slash	qty_quest	qty_equal	qty_at	qty_and	u_qty_excl	qty_space	qty_tilde	i_qty_comm	qty_plus	s_qty_aster	qty_hash	qty_dollar	qty_perce	qty_tld	ur_length	url_qty_dot	d_qty_hyph	qty_under	qty_slash	qt
2	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	25	2	0	0	0	0
3	5	0	1	3	0	3	0	2	0	0	0	0	0	0	0	0	0	3	223	2	0	0	0	0
4	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	2	0	0	0	0
5	4	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	81	2	0	0	0	0
6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	19	2	0	0	0	0
7	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	22	1	0	0	0	0
8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	27	2	0	0	0	0
9	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	46	2	0	0	0	0
10	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	16	2	0	0	0	0
11	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	1	0	0	0	0
12	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	19	2	1	0	0	0
13	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	58	1	0	0	0	0
14	2	2	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	45	1	1	0	0	0
15	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	21	2	0	0	0	0
16	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	33	3	0	0	0	0
17	3	0	1	5	0	3	0	2	0	0	0	0	0	0	0	0	0	1	213	2	0	0	0	0
18	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13	2	1	0	0	0
19	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	30	3	0	0	0	0
20	4	0	0	2	0	1	1	0	0	0	0	0	0	0	0	0	0	2	57	1	0	0	0	0
21	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	17	3	0	0	0	0
22	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	21	4	0	0	0	0
23	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	20	2	1	0	0	0
24	4	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	81	2	0	0	0	0
25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13	2	0	0	0	0
26	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	18	2	0	0	0	0
27	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	21	3	0	0	0	0
28	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	17	2	0	0	0	0
29	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	23	2	0	0	0	0

3.3. Tool Used

- Python 3.9 is employed because the programming language and frameworks like
- Numpy , Pandas , sklearn and alternative modules for model building
- VsCode is employed as IDE
- For visualization seaborn and matplotlib is used
- Github is Employed for version management
- For deployment AWS is Used
- Dockers and Github Actions is used for CI-CD pipelining
- For model monitoring and trigger , Apache Airflow is used

3.4. Export data from database

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

3.5. 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

3.4. Model Building

The pre-processed information is then envisioned and every one the specified insights are being drawn. Though from the drawn insights, the info is at random unfold however still modeling is performed with completely different machine learning algorithms to form positive we tend to cowl all the chances. And eventually, for sure random forest regression performed well and any hyper parameter calibration is finished to extend the model's accuracy.

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.