Low Level Design (LLD)

PHISHING DOMAIN DETECTION

Revision Number – 1.3

Last Date of Revision – 06-feb-2023

Abhishek

Document Version Control:

Date	Version	Description	Author
05-02-2023	1.0	Abstract, Introduction Architecture	Abhishek
06-02-2023	1.1	Data Preprocessing	Abhishek

Contents

Abstract

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

1.1 What is Low-Level design document ? ---------- 4

1.2 Scope ---------- 4

2 Architecture ---------- 5

3 Architecture Description ---------- 5

3.1 Data Gathering ---------- 6

3.2 Data Description ---------- 6

3.3 Tool Used ---------- 7

3.4 Data Pre-processing ---------- 7

3.5 Model Building ---------- 8

3.6 Data from User ---------- 8

3.7 Data Validation ---------- 8

3.8 Rendering Result ---------- 8

Abstract

The recent international things had a large impact on the aviation sector because of several reasons. This impact has 2 class folks, the primary is business perspective and therefore the second is that the customers perspective. As safety is that the major reason for such impact on the aviation sector, the governments round the world amended totally different rules to their various airlines firms. These restrictions had created the supply of the flights and their attendant capability less. Taking of these factors in thought the value of the flight tickets has accrued and vary from one place to the opposite. Booking a flight price tag has split into 2, one is that the on-line and therefore the alternative is that the offline bookings. Each these have their various criteria for value of the price tag, one such example is that the server load and therefore the range of booking requests. during this machine learning implementation, we are going to see numerous factors that impact worth of the flight ticket price and predict the acceptable price of the ticket.

1. Introduction

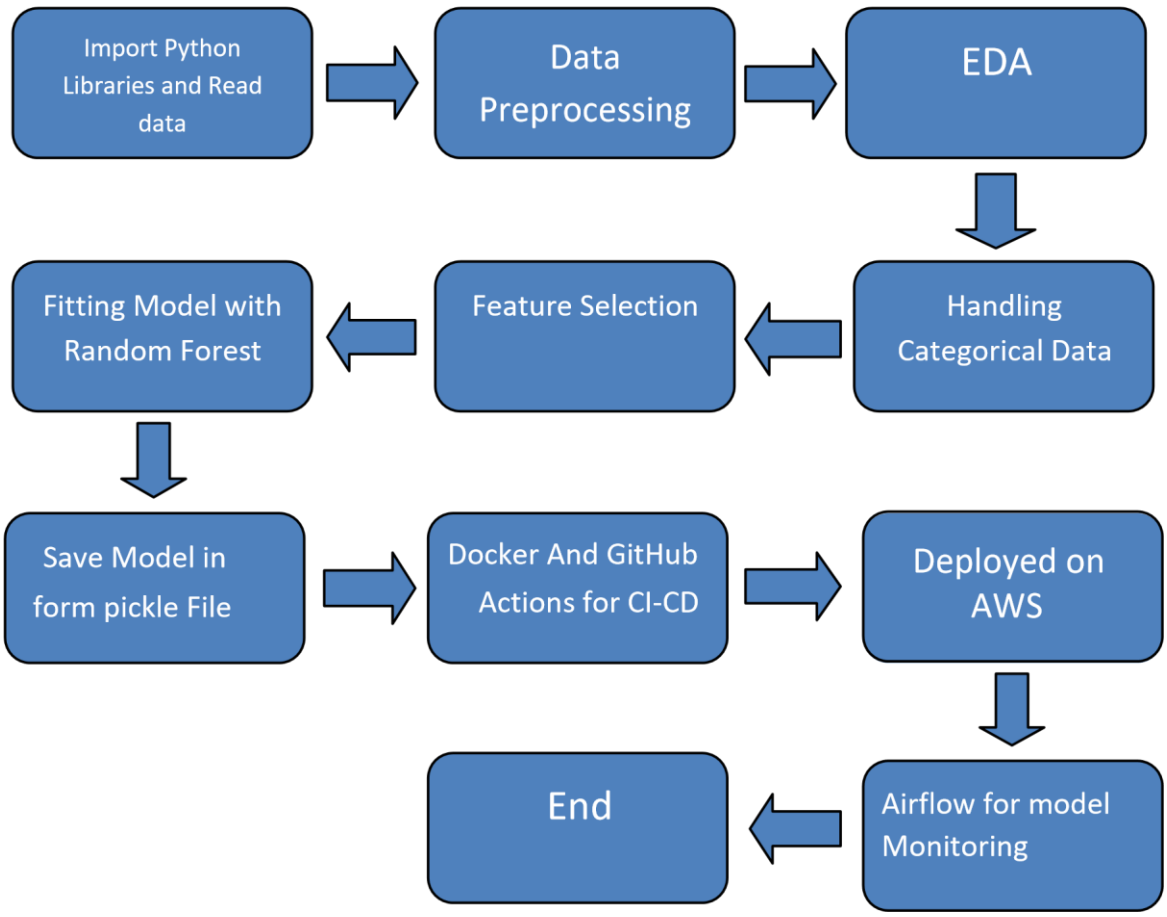
1.1. What is Low-Level design document?

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



3. 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

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

3.2. Data Description

Total number of instances: 88,647 Number of legitimate website instances (labeled as 0): 58,000 Number of phishing website instances (labeled as 1): 30,647 Total number of features: 111

1	qty_dot_u	qty_hyphe	qty_under	qty_slash_	qty_questi	qty_equal	qty_at_url	qty_and_u	qty_exclar	qty_space	qty_tilde_	qty_comm	qty_plus_	qty_asteri	qty_hasht	qty_dollar	qty_perce	qty_tld_ur	length_url	qty_dot_d	qty_hyphe	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	
3	5	0	1	3	0	3	0	2	0	0	0	0	0	0	0	0	0	3	223	2	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	
5	4	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	81	2	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	
7	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	22	1	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	
9	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	46	2	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	
11	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	1	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	
13	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	58	1	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	
15	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	21	2	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	
17	3	0	1	5	0	3	0	2	0	0	0	0	0	0	0	0	0	1	213	2	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	
19	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	30	3	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	
21	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	17	3	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	
23	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	20	2	1	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	
25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13	2	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	
27	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	21	3	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	
29	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	23	2	0	0	0	

3.3. 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.
- VsCode is employed as IDE.
- For visualizations seaborn and components of matplotlib and seaborn are getting used.
- GitHub is employed for version management.
- For Deployment AWS is used
- Dockers and Github Actions is used for CI CD piple lining

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

3.5. 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 from Industry

The data from the is is retrieved Apache Airflow for production usecases

3.7. 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.