

# Architecture

## PHISHING DOMAIN DETECTION

Document Version: 0.1

Last Revised Date: 14-May-2023

Parvej alam Ansari

Document Version Control:

Version	Date	Author	Description
0.1	14-04-2023	Parvej alam	Introduction Architecure
0.1	14-04-2023	Parvej alam	Architecture Design

## Contents

<b>1</b>	<b>Introduction-----</b>	<b>3</b>
1.1	About Architecture Design-----	3
1.2	Scope-----	3
<b>2</b>	<b>Architecture Design-----</b>	<b>3</b>
<b>3</b>	<b>Architecture Description-----</b>	<b>3</b>
3.1	Data Ingestion from MongoDB-----	3
3.2	Data Validation-----	4
3.3	Data Transformation-----	4
3.4	Model Building using Model Trainer-----	4
3.5	Model Evaluation-----	4
3.6	Model Training using FAST API-----	4
3.7	Saving Model and Preprocessing File-----	4
3.8	Cloud Setup and Model Pushing-----	4
<b>4</b>	<b>User Input-Output-----</b>	<b>5</b>

## 1. Introduction

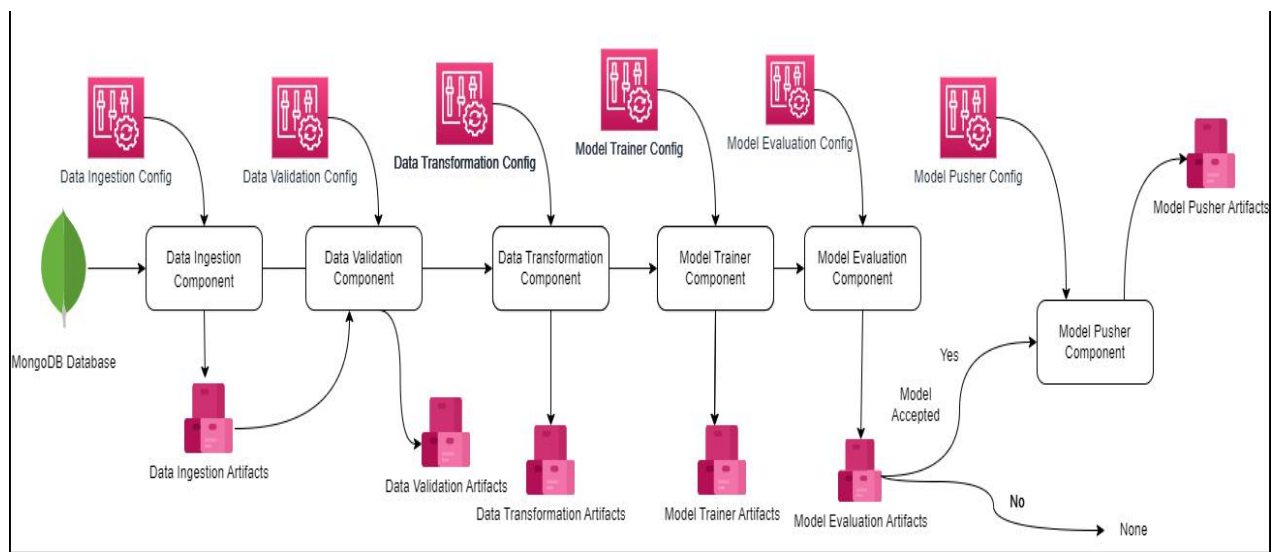
### 1.1 About Architecture Design

Architecture Design (AD) aims to give the internal design of the actual program code for the 'Heating Load and Cooling Prediction'. AD describes the class diagrams with the methods and relation between classes and program specifications. It describes the modules so that the programmer can directly code the program from the document.

### 1.2 Scope

Architecture Design (AD) 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. And the complete workflow.

## 2. Architecture Design



## 3. Architecture Description

### 3.1. Data Ingestion from MongoDB

Data Ingestion components takes data ingestion config as input and returns data ingestion artifact as an output which works as input for the data validation component.

#### Data Insertion into Database

- a. Database creation and connection - create a database with name passed. If the

- database is already created, open the connection to the database.
- b. Table creation in the database.
- c. Insertion of files in the table

### **Export Data from Database**

Data export from database - The data in a stored database is exported as a .CSV file to be used for data pre-processing and model Training.

### **3.2. Data Validation**

Data validation takes data ingestion artifact and data validation config as inputs and produces data validation artifact as output.

### **3.3. Data Transformation**

Data Transformation component uses data validation artifact and data transformation config as inputs and gives data transformation artifact as an output for the next stage.

As soon as data is successfully validated from data validation stage, The validated data is then sent to the data transformation stage where data is wrangled, cleaned by using some data imputation techniques. Furthermore if outliers are present then apply some outliers handling techniques. Next, if dataset is not having proper scaling then will apply proper scaling techniques to transform the data. Now data is ready to next stage i.e. Model Trainer.

### **3.4. Model Building using Model Trainer**

Model trainer component takes Data transformation artifact and model trainer config as inputs and returns model trainer artifact.

In this stage, Number of classification type supervised machine learning models should be applied and will try to find out the best suitable model as per our available dataset by using some metric evaluation techniques. After getting the best suitable classification model, its training and testing accuracy will be determined again.

### **3.5. Model Evaluation**

Model evaluation component takes model trainer artifact and model evaluation config as inputs and gives model evaluation artifact as output.

In this stage, selected and hypertuned model is evaluated using some criteria like training and testing accuracy with the help of proper metric evaluation techniques, overfitting and underfitting etc. once this selected model is properly evaluated, it will be passed to the pipeline stage.

### **3.6. Model Training using FAST API**

After successful run of individual component, a pipeline will be created where each and every component discussed above will merge and that pipeline will be triggered using FAST API where the finalised model shall be trained with available training dataset.

### 3.7. Saving Model and Preprocessing File

In this stage, the trained model file plus the preprocessing file will be stored in pickeled format.

### 3.8. Cloud Setup and Model Pushing

Model pushing component takes model evaluation artifact and model pusher config takes as an input and response back model pusher artifact as an output.

After the model/product/app is ready, the app is deplyed to AWS by using some AWS services Elastic Beanstalk (EB) and Docker hub. This cloud deplymnet will help user to access the application through any internet devices.

## 4. User Input-Output

