

# Architecture Design

## Phishing Domain Detection



Project by – Tejas Muley

# Content

1. Introduction.....	3
1.1 What is Architecture Design Document.....	3
1.2 Scope.....	3
2. Architecture.....	4
2.1 Machine Learning Model.....	4
2.2 Input – Output Flow of Project.....	5
2.3 Architecture Description.....	6

# 1. Introduction

## 1.1 Why this Architecture Design Document?

Any software needs the architectural design to represent the design of the software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures. Each style will describe a system category that consists of:

- A set of components (eg: a database, computational modules) that will perform a function required by the system.
- The set of connectors will help in coordination, communication, and cooperation between the components.
- Conditions that how components can be integrated to form the system.
- Semantic models help the designer to understand the overall properties of the system.

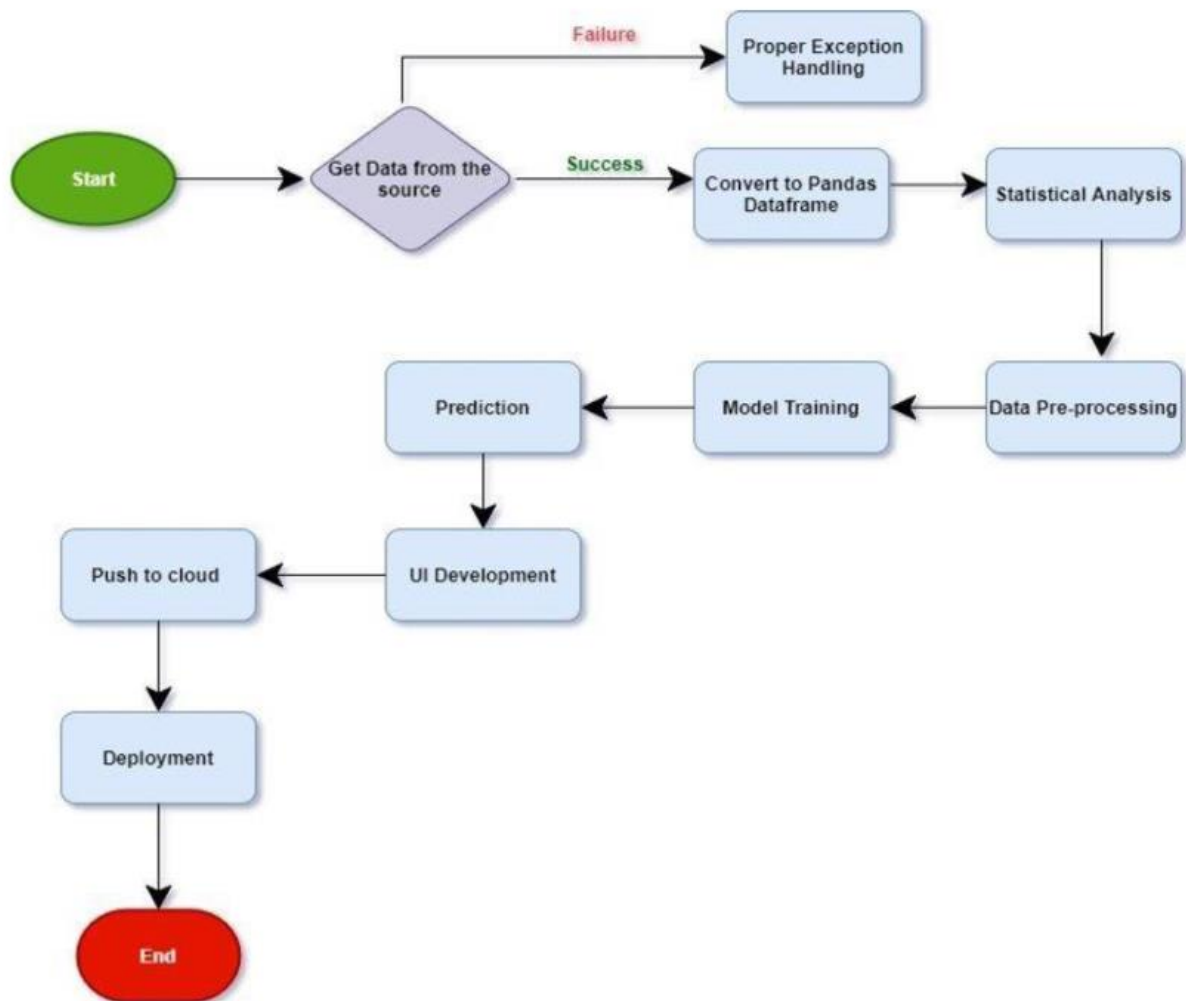
## 1.2 Scope

Architecture Design Document (ADD) is an architectural design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.

## 2. Architecture

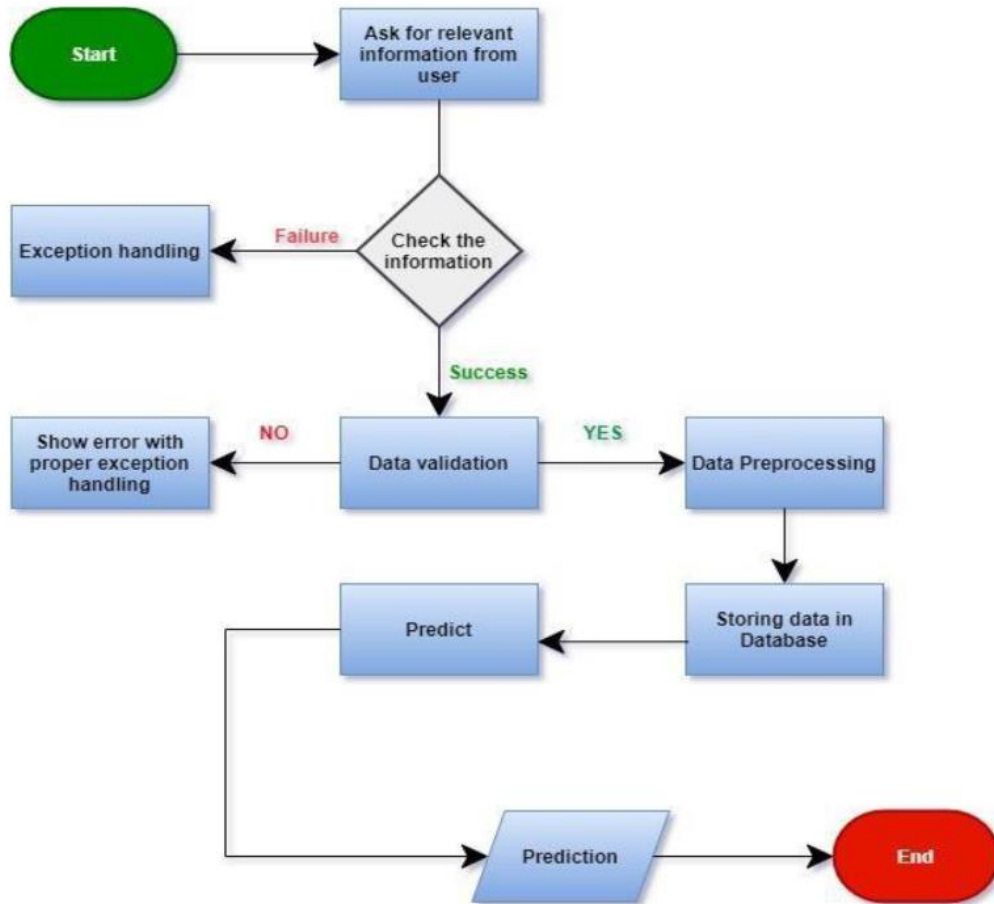
### 2.1 Machine Learning Model

This machine learning use case architecture is defining how the machine learning model is working from starting to end and what the success and failure scenarios are.

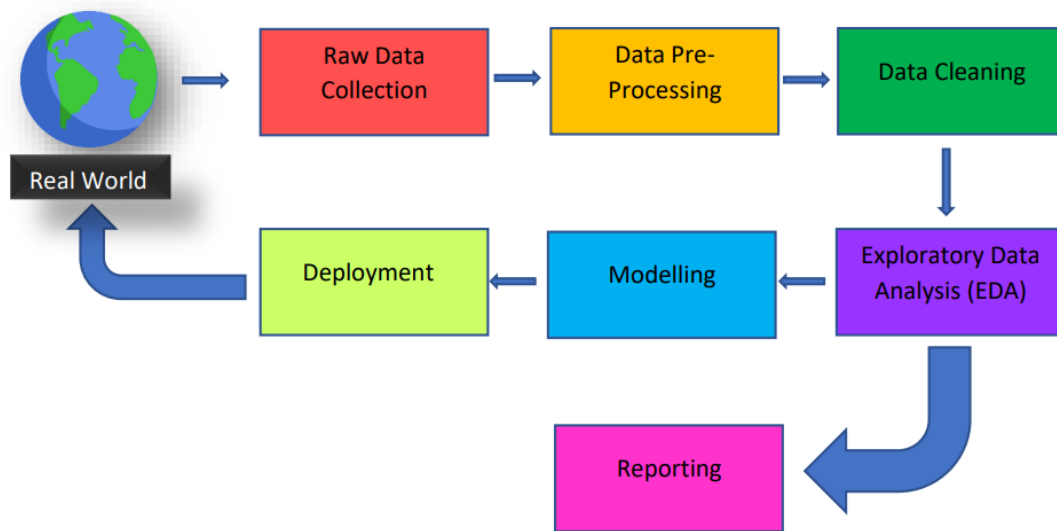


## 2.2 Input – Output Flow of the Project

This input output flow will define the project's end to end flow with all success and failure cases.



## 2.3 Architecture Description



### 1. Raw Data Collection

The Dataset was taken from iNeuron's Provided Project Description Document.

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

### 2. Data Pre-Processing

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data needed to the model to train. This Process includes a) Handling Null/Missing Values b) Handling Skewed Data c) Outliers Detection and Removal.

### 3. Data Cleaning and Exploratory Data Analysis

Data Cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. a) Remove duplicate or irrelevant observations b) Filter unwanted Outliers c) Renaming required attributes.

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

#### 4. Model Building

Model Building is the process of developing a probabilistic model that best describes the relationship between the dependent and independent variables.

#### 5. Reporting

Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy and self-explanatory report because your model will be used by many stakeholders who are not from technical background. a) High Level Design Document b) Low Level Design Document (LLD) c) Architecture d) Wireframe e) Detailed Project Report f) Power Point Presentation

#### 6. Modelling

Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

#### 7. Deployment using Flask on Azure Cloud.