

# WEB PHISHING DETECTION

## A PROJECT REPORT

*Submitted by*

ABULHASAN A

ASIKAHAMED Y

AJITHKUMAR B

GOKULRAM D

*of*

ELECTRONICS AND COMMUNICATION ENGINEERING

V.S.B ENGINEERING COLLEGE

KARUR – 639111

| CHAPTER NO | TITLE                                    | PAGE NO |
|------------|--|---------|
| <b>1.</b>  | <b>INTRODUCTION</b>                      |         |
|            | 1.1 Project Overview                     | 4       |
|            | 1.2 Purpose                              | 4       |
| <b>2.</b>  | <b>LITERATURE SURVEY</b>                 |         |
|            | 2.1 Existing problem                     | 5       |
|            | 2.2 References                           | 5       |
|            | 2.3 Problem Statement Definition         | 6       |
| <b>3.</b>  | <b>IDEATION &amp; PROPOSED SOLUTION</b>  |         |
|            | 3.1 Empathy Map Canvas                   | 7       |
|            | 3.2 Ideation & Brainstorming             | 7       |
|            | 3.3 Proposed Solution                    | 8       |
|            | 3.4 Problem Solution fit                 | 9       |
| <b>4.</b>  | <b>REQUIREMENT ANALYSIS</b>              |         |
|            | 4.1 Functional requirement               | 10      |
|            | 4.2 Non-Functional requirements          | 10      |
| <b>5.</b>  | <b>PROJECT DESIGN</b>                    |         |
|            | 5.1 Data Flow Diagrams                   | 11      |
|            | 5.2 Solution & Technical Architecture    | 11      |
|            | 5.3 User Stories                         | 13      |
| <b>6.</b>  | <b>PROJECT PLANNING &amp; SCHEDULING</b> |         |
|            | 6.1 Sprint Planning & Estimation         | 14      |
|            | 6.2 Sprint Delivery Schedule             | 14      |
|            | 6.3 Reports from JIRA                    | 15      |
| <b>7.</b>  | <b>CODING &amp; SOLUTIONING</b>          |         |
|            | 7.1 Feature 1                            | 17      |
|            | 7.2 Feature 2                            | 33      |
|            | 7.3 Database Schema                      | 34      |
| <b>8.</b>  | <b>TESTING</b>                           |         |
|            | 8.1 Test Cases                           | 36      |
|            | 8.2 User Acceptance Testing              | 37      |

|            |                                       |    |
|------------|---------------------------------------|----|
| <b>9.</b>  | <b>RESULTS</b>                        |    |
|            | 9.1 Performance Metrics               | 38 |
| <b>10.</b> | <b>ADVANTAGES &amp; DISADVANTAGES</b> | 39 |
| <b>11.</b> | <b>CONCLUSION</b>                     | 40 |
| <b>12.</b> | <b>FUTURE SCOPE</b>                   | 40 |
| <b>13.</b> | <b>APPENDIX</b>                       |    |
|            | 13.1 Source Code                      | 41 |
|            | 13.4 GitHub & Project Demo Link       | 76 |

## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1 Project Overview:**

PhisherFind is a website which is used to detect phishing sites to improve the customer's sense of safety whenever he/she attempts to provide any sensitive information to a site. Also, by which people won't access them which will reduce the revenue of malicious site owners. This application can be accessed online without paying instead, can be accessed via any browser of the customer's choice to detect any site with high accuracy. This system uses machine learning algorithm which implements classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy.

The design and implementation of a comprehensive web phishing detection system instils a cyber security culture which prevents the need for the deployment of targeted anti-phishing solutions in a corporate to meet industry's compliance obligations.

#### **1.2 Purpose:**

Web phishing is a threat in various aspects of security on the internet, which might involve scams and private information disclosure. Some of the common threats of web phishing are:

- Attempt to fraudulently solicit personal information from an individual or organization.
- Attempt to deliver malicious software by posing as a trustworthy organization or entity.
- Installing those malwares infects the data that cause a data breach or even nature's forces that takes down your company's data headquarters, disrupting access.

For this purpose, the objective of our project involves building an efficient and intelligent system to detect such websites by applying a machine-learning algorithm which implements classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy and as a result of which whenever a user makes a transaction online and makes payment through an e-banking website our system will use a data mining algorithm to detect whether the e-banking website is a phishing website or not.

## CHAPTER 2

### LITERATURE SURVEY

#### 2.1 Existing problem:

There are phishing detection sites out in the web. But they charge users after a limit of usage. Most of them are built on a clean set of features. We have carefully analysed and identified several factors that could be used to detect a phishing site. These factors fall under the categories of address bar-based features, domain-based features, HTML & JavaScript based features. Using these features, we build an intelligent system which can identify a phishing site with high accuracy and efficiency. It is also an open-source website which will be easily accessible to all users.

#### 2.2 References:

- [https://en.wikipedia.org/wiki/Web\\_service](https://en.wikipedia.org/wiki/Web_service).

[1] Farashazillah Yahya, Ryan Isaac W Mahibol, Chong Kim Ying, Magnus Bin Anai, Sidney Allister Frankie, Eric Ling Nin Wei and Rio Guntur Utomo, "Detection of Phishing Websites using Machine Learning Approaches", 2021 International Conference on Data Science and Its Applications (ICoDSA).

[2] Prajakta Patil, Rashmi Rane and Madhuri Bhalekar, "Detecting spam and phishing mails using SVM and obfuscation URL detection algorithm", 2017 International Conference on Inventive Systems and Control (ICISC).

[3] Gaurav Varshney, Manoj Mishra and Pradeep K. Atrey, "A phish detector using lightweight search features", Computers & Security, 2016.

[4] Antonio Hernández Domínguez and Walter Baluja García, "Updated Analysis of Detection Methods for Phishing Attacks", Futuristic Trends in Network and Communication Technologies, vol.1395, pp.56, 2021.

[5] Anggit Ferdita Nugraha and Luthfia Rahman, "Meta-Algorithms for Improving Classification Performance in the Web-phishing Detection Process", 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp.271-275, 2019.

[6] Yoga Pristyanto and Akhmad Dahlan, "Hybrid Resampling for Imbalanced Class Handling on Web Phishing Classification Dataset", 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp.401-406, 2019.

[7] Athulya A.A and Praveen K, "Towards the Detection of Phishing Attacks", 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)

[8] Miyamoto D, Hazeyama H and Kadobayashi Y, "An evaluation of machine learning-based methods for detection of phishing sites", International Conference on Neural Information Processing pp. 539-546. Springer, Berlin, Heidelberg. (2008)

[9] K S Swarnalatha, K C Ramchandra, Kaushar Ansari, Love Ojha and Sanjok Subedi Sharma, "Real-Time Threat Intelligence-Block Phishing Attacks", 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)

[10] Salvi Siddhi Ravindra, Shah Juhi Sanjay, Shaikh Nausheenbanu Ahmed Gulzar and Khodke Pallavi, "Phishing Website Detection Based on URL", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp.589, 2021.

### **2.3 Problem statement definition:**

Web Phishing is a form of cyber fraud, which implies that fraudsters use various means to impersonate the URL address and page content of a real website or use vulnerabilities in the server program of a real website to insert dangerous HTML code in certain pages of the site.

It is a threat in various aspects of security on the internet, which might involve scams and private information disclosure. Some of the common threats of web phishing are:

- Obtaining personal information from an individual or organization.
- Impersonating as a trustworthy organization to deliver malicious websites.

To avoid these threats, we build an efficient and intelligent system to detect such websites using machine-learning algorithms which implements classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy.

This project can also be further extended by creating a browser extension or developing a GUI which takes the URL and analyses its nature to determine if it is a legitimate or a phishing website.

## CHAPTER 3

### IDEATION & PROPOSED SOLUTION

#### 3.1 Empathy Map Canvas:

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. Empathy maps should be used throughout any UX process to establish common ground among team members and to understand and prioritize user needs. In user-centered design, empathy maps are best used from the very beginning of the design process.

# Empathy Map

Dive into the mind of the user for focused product development

● Web-Phishing



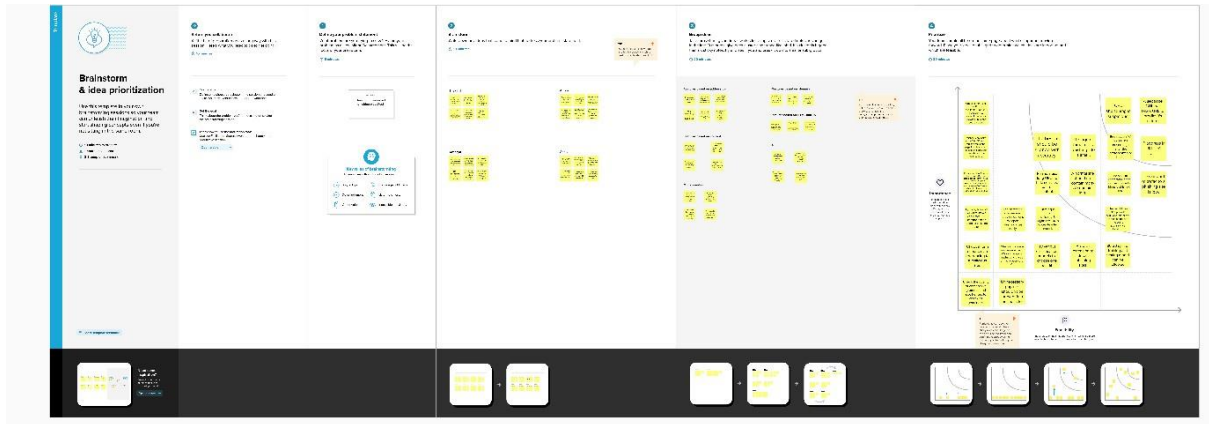
Share your feedback



### **3.2 Ideation & Brainstorming:**

Ideation essentially refers to the whole creative process of coming up with and communicating new ideas. Ideation is innovative thinking, typically aimed at solving a problem or providing a more efficient means of doing or accomplishing something.

Ideation is often closely related to the practice of brainstorming, a specific technique that is utilized to generate new ideas. A principal difference between ideation and brainstorming is that ideation is commonly more thought of as being an individual pursuit, while brainstorming is almost always a group activity.



### 3.3 Proposed Solution:

| S.No. | Parameter                                | Description  |
|-------|--|--|
| 1.    | Problem Statement (Problem to be solved) | Novel phishing approaches suffer low detection accuracy. The most common technique used is the blacklist-based method. It has become inefficient since registering a new domain has become easier. No comprehensive blacklist can ensure a perfect up-to-date database.  |
| 2.    | Idea / Solution description              | Our solution is to build an efficient and intelligent system to detect phishing sites by applying a machine learning algorithm which implements classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy.   |
| 3.    | Novelty / Uniqueness                     | We have carefully analysed and identified various factors that could be used to detect a phishing site. These factors fall under the categories of address bar based features, domain based features, HTML & Javascript based features. Using these features we can identify a phishing site with high accuracy. |
| 4.    | Social Impact / Customer Satisfaction    | By using this application the customer has the sense of safety whenever he attempts to provide sensitive information to a site.  |
| 5.    | Business Model (Revenue Model)           | By generating leads we can improve our business model. By detecting the phishing sites, people won't access them which will reduce the revenue of malicious site owners.   |

|    |                             |   |
|----|-----------------------------|---|
| 6. | Scalability of the Solution | This application can be accessed online without paying. It can be accessed via any browser of your choice. It can detect any site with high accuracy. |
|----|-----------------------------|---|

### 3.4 Problem Solution fit

The Problem-Solution Fit simply means that you have found a problem with your customer and that the solution you have realized for it solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioural patterns and recognize what would work and why.

Purpose:

- ☐ Solve complex problems in a way that fits the state of your customers.
- ☐ Succeed faster and increase your solution adoption by tapping into existing mediums and channels of behaviour.
- ☐ Sharpen your communication and marketing strategy with the right triggers and messaging.
- ☐ Increase touchpoints with your company by finding the right problem-behaviour fit and building trust by solving frequent annoyances, or urgent or costly problems.
- ☐ Understand the existing situation in order to improve it for your target group.

**Problem-Solution fit canvas 2.0**

Purpose / Vision To detect phishing sites.

|   |   |   |
|---|---|---|
| <b>1. CUSTOMER SEGMENT(S)</b><br><small>Who is your customer?<br/>i.e. working parents of 0-5 y.o. kids</small><br><b>CS</b>  | <b>6. CUSTOMER CONSTRAINTS</b><br><small>What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices</small><br><b>CC</b>   | <b>5. AVAILABLE SOLUTIONS</b><br><small>Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros &amp; cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking</small><br><b>AS</b>        |
| <b>2. JOBS-TO-BE-DONE / PROBLEMS</b><br><small>Which jobs-to-be-done (or problems) do you address for your customers?<br/>There could be more than one; explore different sides.</small><br><b>J&amp;P</b>                                      | <b>9. PROBLEM ROOT CAUSE</b><br><small>What is the real reason that this problem exists?<br/>What is the back story behind the need to do this job?<br/>i.e. customers have to do it because of the change in regulations.</small><br><b>RC</b>   | <b>7. BEHAVIOUR</b><br><small>What does your customer do to address the problem and get the job done?<br/>i.e. directly related: find the right solar panel installer, calculate usage and benefits;<br/>indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)</small><br><b>BE</b> |
| <b>3. TRIGGERS</b><br><small>What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.</small><br><b>TR</b>   | <b>10. YOUR SOLUTION</b><br><small>If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.<br/>If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.</small><br><b>SL</b> | <b>8. CHANNELS OF BEHAVIOUR</b><br><b>8.1 ONLINE</b><br><small>What kind of actions do customers take online? Extract online channels from #7</small><br><b>CH</b>  |
| <b>4. EMOTIONS: BEFORE / AFTER</b><br><small>How do customers feel when they face a problem or a job and afterwards?<br/>i.e. lost, insecure = confident, in control - use it in your communication strategy &amp; design.</small><br><b>EM</b> |   | <b>8.2 OFFLINE</b><br><small>What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.</small>  |

Problem-Solution fit canvas is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 license  
Created by Daria Neprukhina / Amaltama.com

**AMALTAMA**

**CHAPTER 4**  
**REQUIREMENT ANALYSIS**

**4.1 Functional requirements:**

| FR No. | Functional Requirement (Epic) | Description  |
|--------|-------------------------------|--|
| FR-1   | User Input                    | User inputs an URL in the form to check whether it is a malicious website.                         |
| FR-2   | Website comparison            | The model compares the given URL with the list of phishing URLs present in the database.           |
| FR-3   | Feature Extraction            | If it is found none on the comparison it extracts the HTML and domain-based features from the URL. |
| FR-4   | Prediction                    | The model predicts the URL using machine Learning algorithms such as Random Forest technique.      |
| FR-5   | Classifier                    | Model then sends the output to the classifier and produces the result.                             |
| FR-6   | Announcement                  | The model finally displays whether the given URL is phishing or not.                               |

**4.2 Non-functional requirements:**

| FR No. | Non-Functional Requirement | Description   |
|--------|----------------------------|---|
| NFR-1  | Usability                  | It is an easy to use and access interface which results in greater efficiency.                              |
| NFR-2  | Security                   | It is a secure website which protects the sensitive information of the user and prevents malicious attacks. |
| NFR-3  | Reliability                | The system can detect phishing websites with greater accuracy using ML algorithms.                          |
| NFR-4  | Performance                | The system produces responses within seconds and execution is faster.                                       |
| NFR-5  | Availability               | Users can access the website via any browser from anywhere at any time.                                     |
| NFR-6  | Scalability                | This application can be accessed online without paying. It can detect any web site with high accuracy.      |

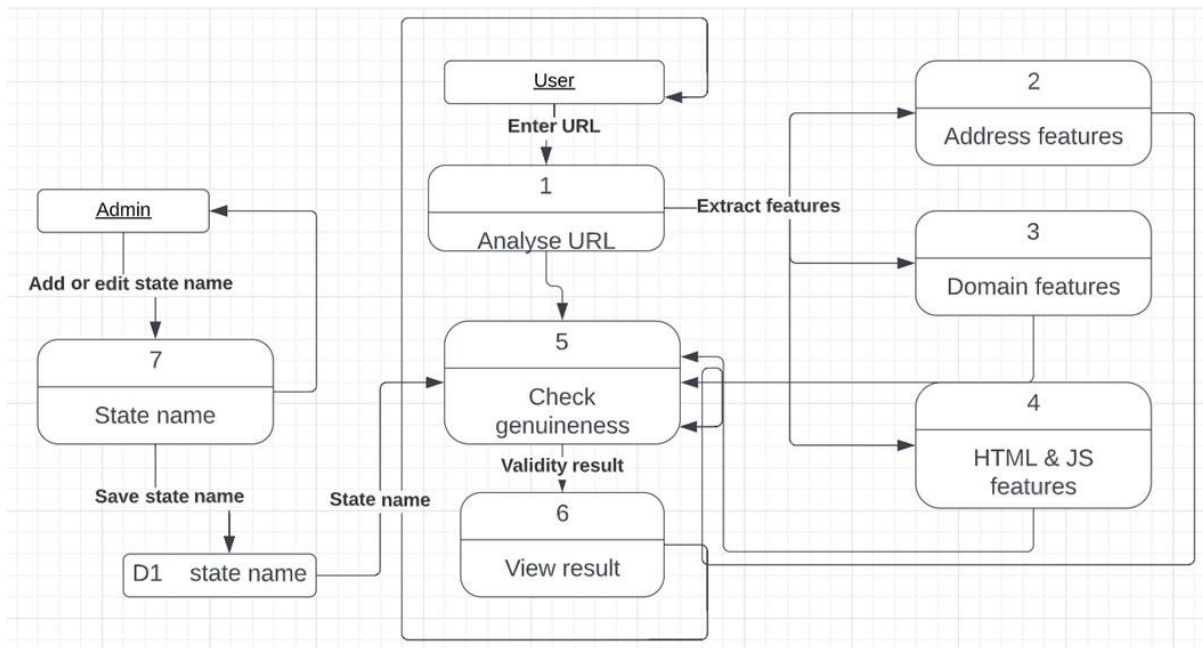
## CHAPTER 5

### PROJECT DESIGN

#### 5.1 Data Flow diagram:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

DFD level 0:



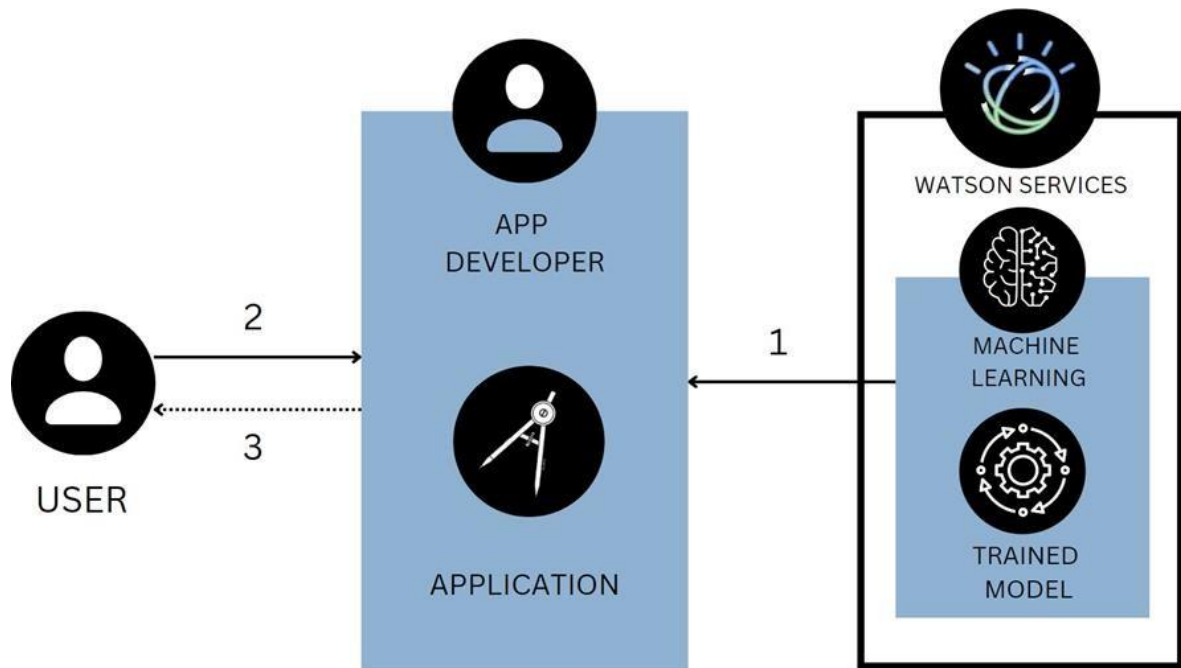
#### 5.2 Solution & Technical Architecture:

##### SOLUTION:

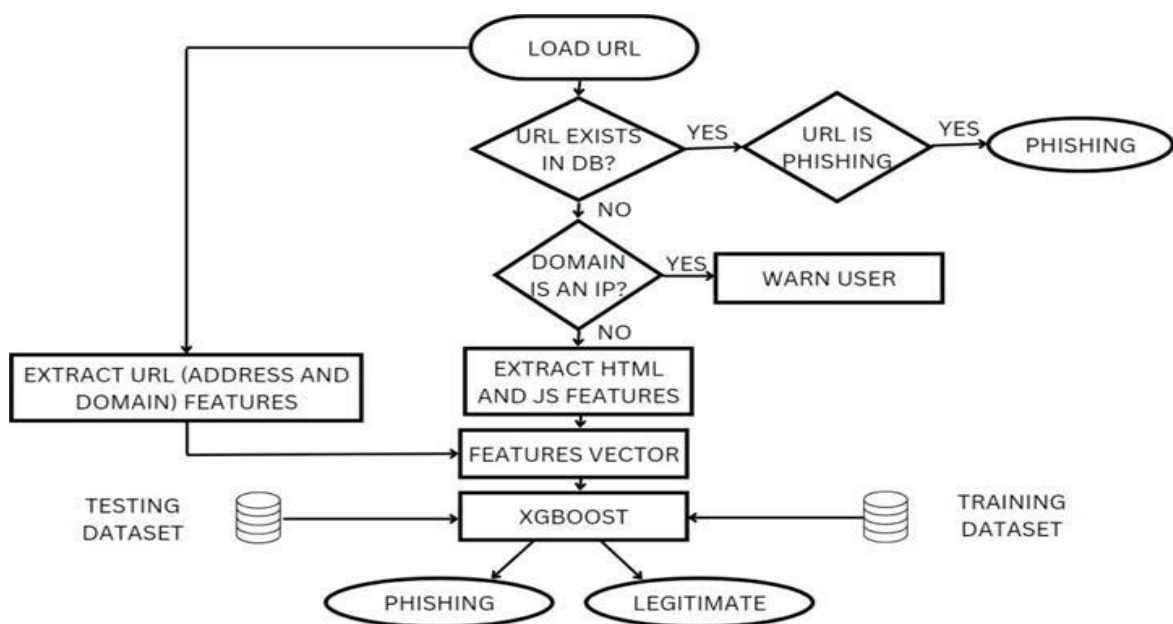
Our solution is to build an efficient and intelligent system to detect phishing sites by applying a machine learning algorithm which implements classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy by carefully analysing and identifying various factors that could be used to detect a phishing site. These factors fall under the categories of address bar-based features, domain-based features, HTML & JavaScript based features. Using these features, we can identify a phishing site with high accuracy.

##### TECHNICAL ARCHITECTURE:

Technical architecture which is also often referred to as application architecture includes the major components of the system, their relationships, and the contracts that define the interactions between the components. The goal of technical architects is to achieve all the business needs with an application that is optimized for both performance and security.



1. The application developer builds a Python-based app and deploys it.
2. The user enters the URL of a website in the application to check for its genuineness.
3. The user submits the URL through the web-based application and gets back the result.
4. The user makes a decision whether to proceed surfing in that website or move to another one.



### 5.3 User Stories:

| User type           | Functional requirement (Epic) | User Story number | User story/task   | Acceptance criteria                 |
|---------------------|-------------------------------|-------------------|---|-------------------------------------|
| Customer (web user) | Login                         | USN-1             | As a user, I can navigate into the website.   | I can access the page.              |
|                     | Dashboard                     | USN-2             | As a user, I will paste the URL that needs to be checked if it's a phishing website or not. | I can paste the URL in the textbox. |
|                     |                               | USN-3             | As a user, I can see the output.  | I can see if it's a safe site.      |
| Administrator       |                               | USN-4             | If the new URL is found, I can add the new state into the database.                         | I can add the new URL.              |

## CHAPTER 6

### PROJECT PLANNING & SCHEDULING

#### 6.1 Sprint Planning & Estimation:

| Sprint   | Functional Requirement (Epic) | User Story Number | User Story / Task  | Story Points | Priority | Team Members |
|----------|-------------------------------|-------------------|--|--------------|----------|--------------|
| Sprint-1 | Login                         | USN-1             | As a user, I can navigate into the website.                                    | 1            | High     | Asikahamed   |
| Sprint-1 | Dashboard                     | USN-2             | As a user, I will input any site's URL in the form to check its genuineness.   | 1            | High     | Asikahamed   |
| Sprint-1 |                               | USN-3             | As a user, I can see the output.   | 2            | High     | Abulhasan    |
| Sprint-2 | Backend                       | USN-4             | As an admin, if a new URL is found, I can add the new state into the database. | 3            | Medium   | Asikahamed   |
| Sprint-3 | Report                        | USN-5             | As a user, I can ask my queries and report suspicious sites in the report box. | 1            | Low      | Abulhasan    |
| Sprint-4 |                               | USN-6             | As an admin, I can take actions to the queries asked by the user.              | 2            | Low      | Asikahamed   |



## 6.2 Sprint Delivery Schedule:

| Sprint   | Total Story Points | Duration | Sprint Start Date | Sprint End Date (Planned) | Story Points Completed (as on Planned End Date) | Sprint Release Date (Actual) |
|----------|--------------------|----------|-------------------|---------------------------|---|------------------------------|
| Sprint-1 | 20                 | 6 Days   | 24 Oct 2022       | 29 Oct 2022               | 20  | 29 Oct 2022                  |
| Sprint-2 | 20                 | 6 Days   | 31 Oct 2022       | 05 Nov 2022               | 20  | 05 Nov 2022                  |
| Sprint-3 | 20                 | 6 Days   | 07 Nov 2022       | 12 Nov 2022               | 20  | 12 Nov 2022                  |
| Sprint-4 | 20                 | 6 Days   | 14 Nov 2022       | 19 Nov 2022               | 20  | 19 Nov 2022                  |

## 6.3 Reports from JIRA:

### Backlog:

▼ Backlog (6 issues) 10 0 0 [Create sprint](#)

|  |   |       |   |
|--|---|-------|---|
| WPD-1 As a user, I can navigate into the website.                                    | 1 | TO DO | A |
| WPD-2 As a user, I will input any site's URL in the form to check its genuineness.   | 1 | TO DO | A |
| WPD-3 As a user, I can see the output.   | 2 | TO DO | S |
| WPD-4 As an admin, if a new URL is found, I can add the new state into the database. | 3 | TO DO | A |
| WPD-5 As a user, I can ask my queries and report suspicious sites in the report box. | 1 | TO DO | S |
| WPD-6 As an admin, I can take actions to the queries asked by the user.              | 2 | TO DO | A |

[+ Create issue](#)

### Sprint 1:

▼ Sprint-1 24 Oct – 31 Oct (3 issues) 4 0 0 [Start sprint](#) [...](#)

Create the web phishing detection site with basic forms to get input and display output.

|  |   |       |   |
|--|---|-------|---|
| WPD-1 As a user, I can navigate into the website. <a href="#">LOGIN</a>                                      | 1 | TO DO | A |
| WPD-2 As a user, I will input any site's URL in the form to check its genuineness. <a href="#">DASHBOARD</a> | 1 | TO DO | A |
| WPD-3 As a user, I can see the output. <a href="#">DASHBOARD</a>   | 2 | TO DO | S |

[+ Create issue](#)

Projects / Web Phishing Detection

### Sprint-1

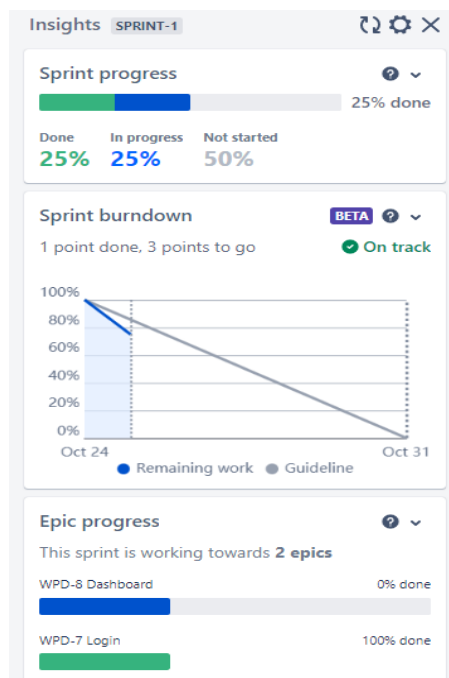
Create the web phishing detection site with basic forms to get input and display output.

[4 days remaining](#) [Complete sprint](#) [...](#)

GROUP BY: [None](#) [Insights](#)

| TO DO 1 ISSUE  | IN PROGRESS 1 ISSUE  | DONE 1 ISSUE  |
|--|--|---|
| <p>As a user, I can see the output. <a href="#">DASHBOARD</a></p> <p>WPD-3 2 S</p> | <p>As a user, I will input any site's URL in the form to check its genuineness. <a href="#">DASHBOARD</a></p> <p>WPD-2 1 A</p> | <p>As a user, I can navigate into the website. <a href="#">LOGIN</a></p> <p>WPD-1 1 A</p> |

## Insights:



## Sprint 2:

▼ **Sprint-2** 31 Oct – 7 Nov (1 issue) 3 0 0 Start sprint ...

Create admin privileges.

WPD-4 As an admin, if a new URL is found, I can add the new state into the database. BACKEND 3 TO DO ...

+ Create issue

Quickstart

## Sprint 3:

▼ **Sprint-3** 7 Nov – 14 Nov (1 issue) 1 0 0 Start sprint ...

Create report functionality in the site.

WPD-5 As a user, I can ask my queries and report suspicious sites in the report box. REPORT 1 TO DO ...

+ Create issue

## Sprint 4:

▼ **Sprint-4** 14 Nov – 21 Nov (1 issue) 2 0 0 Start sprint ...

Update admin privileges.

WPD-6 As an admin, I can take actions to the queries asked by the user. REPORT 2 TO DO ...

+ Create issue

## CHAPTER 7

### CODING & SOLUTIONING

#### 7.1 Feature 1 – Classification of URL:

The primary feature of this project is to classify the given URL as phishing or benign. Various classification algorithms are used to achieve this.

##### 7.1.1 Methodology:

###### 7.1.1.1 Data collection:

URL features of legitimate websites and phishing websites were collected. The data set consists of total 11,055 URLs which include 6,157 legitimate URLs and 4,898 phishing URLs. Legitimate URLs are labelled as “1” and phishing URLs are labelled as “-1”. The features that are present in the data set include:

- IP Address in URL
- Length of URL
- Using URL Shortening Services
- "@" Symbol in URL
- Redirection "/" in URL
- Prefix or Suffix "-" in Domain
- Having Sub Domain
- Length of Domain Registration
- Favicon
- Port Number
- HTTPS Token
- Request URL
- URL of Anchor
- Links in Tags
- SFH
- Email Submission
- Abnormal URL
- Status Bar Customization (on mouse over)
- Disabling Right Click
- Presence of Popup Window
- IFrame Redirection
- Age of Domain
- DNS Record
- Web Traffic
- Page Rank
- Google Index
- Links pointing to the page
- Statistical Report
- Result

Using IBM Cloud Storage this data is accessed throughout the project. The code written below is used to import the dataset.

```

import os, types

import pandas as pd

from botocore.client import Config

import ibm_boto3

def __iter__(self): return 0

# The following code accesses a file in your IBM Cloud Object
Storage. It includes your credentials.

# You might want to remove those credentials before you share the
notebook.

cos_client = ibm_boto3.client(service_name='s3',
                              ibm_api_key_id='',
                              ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
                              config=Config(signature_version='oauth'),
                              endpoint_url='https://s3.private.us.cloud-object-
storage.appdomain.cloud')

bucket = 'webphishingdetection-donotdelete-pr-icmjtvktnzli2s'
object_key = 'dataset_website.csv'

body = cos_client.get_object(Bucket=bucket,Key=object_key) ['Body']

# add missing_iter_method, so pandas accepts body as file-like
object

if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType(
__iter__, body )

data0 = pd.read_csv(body)

data0.head()

```

#### 7.1.1.2 Data pre-processing and Exploratory Data Analysis:

Few plots and graphs were drawn to find how the data is distributed and the how features are related to each other.

##### Univariate analysis:

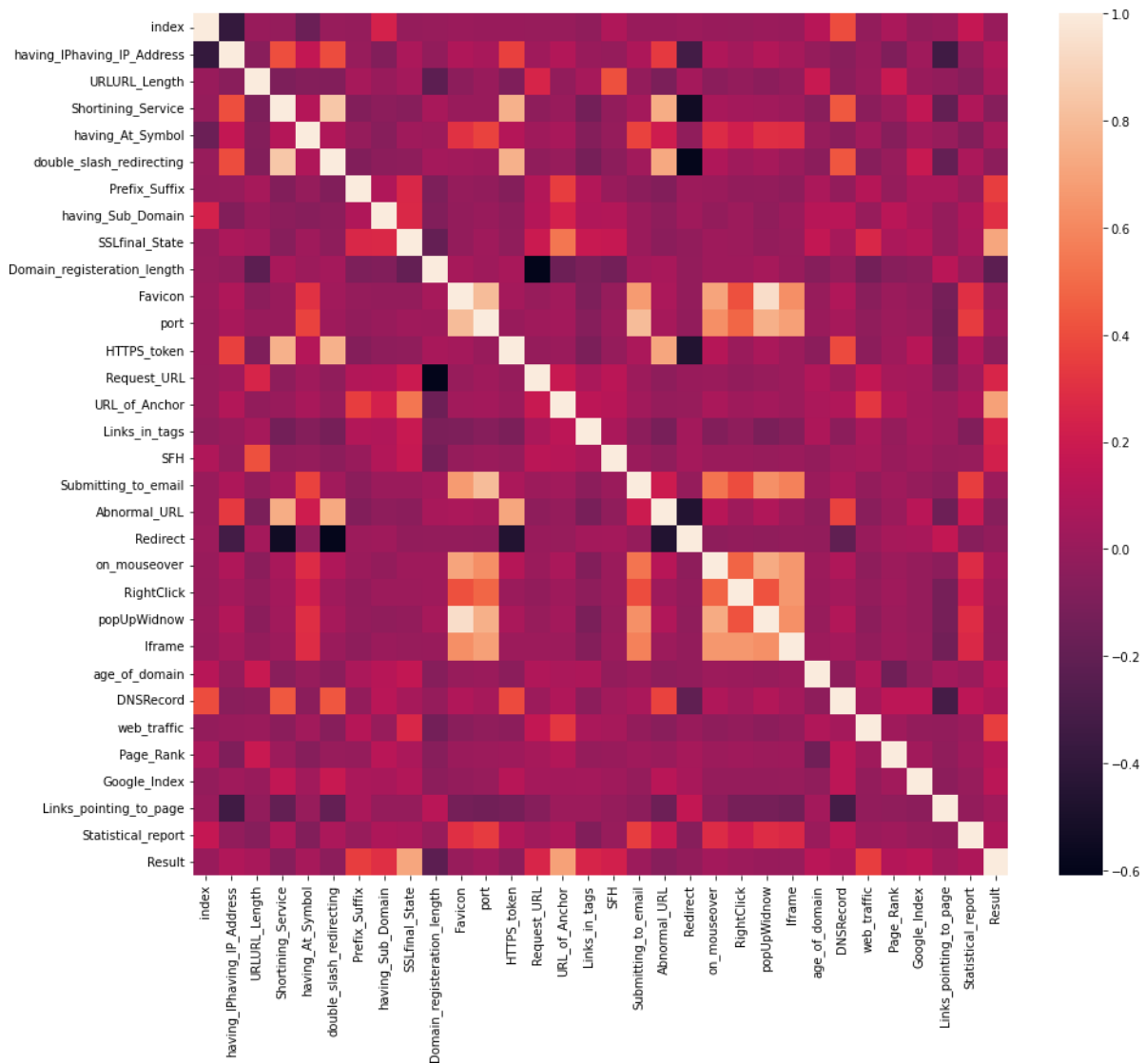
Univariate analysis provides an understanding in the characteristics of each feature in the data set. Different characteristics are computed for numerical and categorical data. For the numerical features characteristics are standard deviation, skewness, kurtosis, percentile, interquartile range (IQR) and range. For the categorical features characteristics are count, cardinality, list of unique values, top and freq.

```
data0.describe()
```

|       | index        | having_IPhaving_IP_Address | URLURL_Length | Shortining_Service | having_At_Symbol | double_slash_redirecting | Prefix_Suffix | having_Sub_Domain | SSLfinal_State | Domain_registration_length | popUpWinow   | iframe       | age_of_domain | DNSRecord    | web_traffic  | Page_Rank    | Google_Index |
|-------|--------------|----------------------------|---------------|--------------------|------------------|--------------------------|---------------|-------------------|----------------|----------------------------|--------------|--------------|---------------|--------------|--------------|--------------|--------------|
| count | 11055.000000 | 11055.000000               | 11055.000000  | 11055.000000       | 11055.000000     | 11055.000000             | 11055.000000  | 11055.000000      | 11055.000000   | 11055.000000               | 11055.000000 | 11055.000000 | 11055.000000  | 11055.000000 | 11055.000000 | 11055.000000 | 11055.000000 |
| mean  | 5538.000000  | 0.317295                   | 0.622198      | 0.732761           | 0.702928         | 0.741471                 | 0.724562      | 0.663252          | 0.220027       | 0.336711                   | 0.811188     | 0.819915     | 0.068719      | 0.311154     | 0.767797     | 0.481841     | 0.72         |
| std   | 3191.447947  | 0.948514                   | 0.766095      | 0.672988           | 0.711298         | 0.671011                 | 0.678139      | 0.817218          | 0.911982       | 0.941620                   | 0.789818     | 0.576704     | 0.998168      | 0.926209     | 0.827733     | 0.872289     | 0.69         |
| min   | 1.000000     | -1.000000                  | -1.000000     | -1.000000          | -1.000000        | -1.000000                | -1.000000     | -1.000000         | -1.000000      | -1.000000                  | -1.000000    | -1.000000    | -1.000000     | -1.000000    | -1.000000    | -1.000000    | -1.00        |
| 25%   | 2764.500000  | -1.000000                  | -1.000000     | 1.000000           | 1.000000         | 1.000000                 | 1.000000      | 1.000000          | 1.000000       | -1.000000                  | 1.000000     | 1.000000     | -1.000000     | -1.000000    | 0.000000     | -1.000000    | 1.00         |
| 50%   | 5538.000000  | 1.000000                   | -1.000000     | 1.000000           | 1.000000         | 1.000000                 | -1.000000     | 0.000000          | 1.000000       | 1.000000                   | 1.000000     | 1.000000     | 1.000000      | 1.000000     | 1.000000     | 1.000000     | 1.00         |
| 75%   | 8291.500000  | 1.000000                   | 1.000000      | 1.000000           | 1.000000         | 1.000000                 | 1.000000      | 1.000000          | 1.000000       | 1.000000                   | 1.000000     | 1.000000     | 1.000000      | 1.000000     | 1.000000     | 1.000000     | 1.00         |
| max   | 11055.000000 | 1.000000                   | 1.000000      | 1.000000           | 1.000000         | 1.000000                 | 1.000000      | 1.000000          | 1.000000       | 1.000000                   | 1.000000     | 1.000000     | 1.000000      | 1.000000     | 1.000000     | 1.000000     | 1.00         |

## Bivariate analysis:

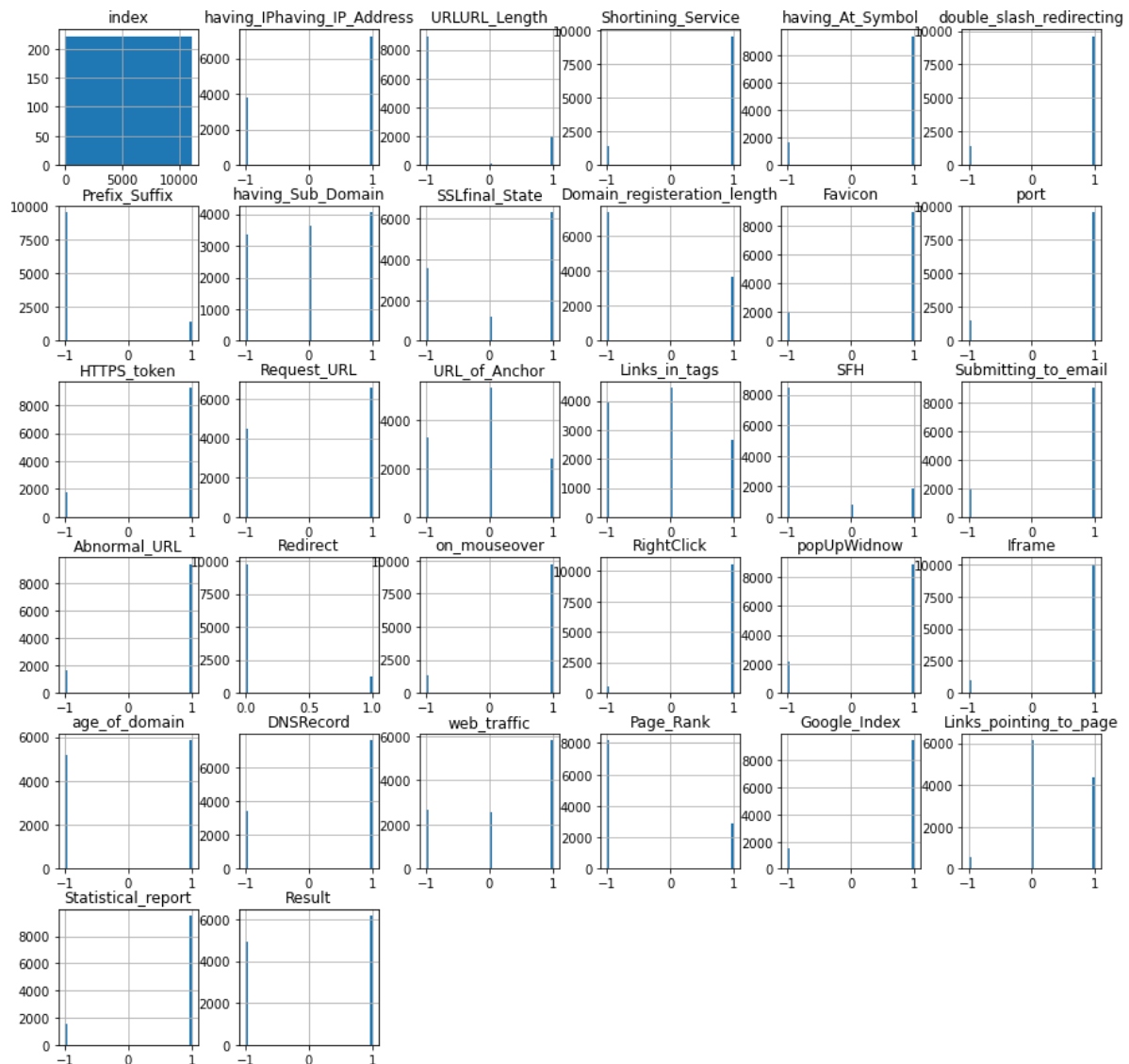
```
plt.figure(figsize=(15,13))
sns.heatmap(data0.corr())
plt.show()
```



From this correlation matrix, it is evident that there is no correlation with many features. So, it is crucial to eliminate these features.

## Multivariate analysis:

```
data0.hist(bins = 50,figsize = (15,15))  
plt.show()
```



From data distribution graph and correlation matrix, we can conclude that the following features do not have much impact on the result:

- having\_Sub\_Domain
- Domain\_registration\_length
- Favicon
- Request\_URL
- URL\_of\_Anchor

- Links\_in\_tags
- Submitting\_to\_email
- Redirect
- web\_traffic
- Page\_Rank
- Google\_Index
- Links\_pointing\_to\_page

All the above features will not be included in further processing.

```
#Removing the features which do not have much impact on Result
data=data0.iloc[:, [1,2,3,4,5,6,12,20,21,22,23,24,25,30,31]]
data.head()
```

### Checking for null values:

This dataset doesn't contain any null values.

```
#checking the data for null or missing values
data.isnull().sum()
```

```
having_IPhaving_IP_Address      0
URLURL_Length                  0
Shortining_Service              0
having_At_Symbol                0
double_slash_redirecting        0
Prefix_Suffix                  0
HTTPS_token                     0
on_mouseover                    0
RightClick                     0
popUpWidnow                     0
Iframe                          0
age_of_domain                   0
DNSRecord                      0
Statistical_report              0
Result                          0
dtype: int64
```

### 7.1.1.3 Model building:

From the dataset above, it is clear that this is a supervised machine learning task. There are two major types of supervised machine learning problems, called classification and regression.

This data set comes under classification problem, as the input URL is classified as phishing (-1) or legitimate (1). The supervised machine learning models (classification) considered to train the dataset in this notebook are:

- XGBoost
- Decision Tree
- Random Forest
- Support Vector Machines

### **XGBoost:**

XGBoost is one of the most popular machine learning algorithms these days. XGBoost stands for eXtreme Gradient Boosting. Regardless of the type of prediction task at hand; regression or classification. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

```
#XGBoost Classification model

from xgboost import XGBClassifier

import warnings
warnings.filterwarnings("ignore", category=UserWarning)

# instantiate the model
xgb = XGBClassifier(learning_rate=0.4,max_depth=7,verbosity = 0)

#fit the model
xgb.fit(X_train, y_train)

#predicting the target value from the model for the samples
y_test_xgb = xgb.predict(X_test)
y_train_xgb = xgb.predict(X_train)

#computing the accuracy of the model performance
acc_train_xgb = accuracy_score(y_train,y_train_xgb)
acc_test_xgb = accuracy_score(y_test,y_test_xgb)

print("XGBoost: Accuracy on training Data:
{:.3f}".format(acc_train_xgb))

print("XGBoost : Accuracy on test Data: {:.3f}".format(acc_test_xgb))
```



### Decision Tree Classifier:

Decision trees are widely used models for classification and regression tasks. Essentially, they learn a hierarchy of if/else questions, leading to a decision. Learning a decision tree means learning the sequence of if/else questions that gets us to the true answer most quickly.

In the machine learning setting, these questions are called tests (not to be confused with the test set, which is the data we use to test to see how generalizable our model is). To build a tree, the algorithm searches over all possible tests and finds the one that is most informative about the target variable.

```
# Decision Tree model
from sklearn.tree import DecisionTreeClassifier

# instantiate the model
tree = DecisionTreeClassifier(max_depth = 5)

# fit the model
tree.fit(X_train, y_train)

#predicting the target value from the model for the samples
y_test_tree = tree.predict(X_test)
y_train_tree = tree.predict(X_train)

#computing the accuracy of the model performance
acc_train_tree = accuracy_score(y_train, y_train_tree)
acc_test_tree = accuracy_score(y_test, y_test_tree)

print("Decision Tree: Accuracy on training Data:
{:.3f}".format(acc_train_tree))

print("Decision Tree: Accuracy on test Data:
{:.3f}".format(acc_test_tree))
```

### Random Forest Classifier:

Random forests for regression and classification are currently among the most widely used machine learning methods. A random forest is essentially a collection of decision trees, where each tree is slightly different from the others. The idea behind random forests is that each tree might do a relatively good job of predicting, but will likely overfit on part of the data.

If we build many trees, all of which work well and overfit in different ways, we can reduce the amount of overfitting by averaging their results. To build a random forest model, you need to decide on the number of trees to build (the `n_estimators` parameter of `RandomForestRegressor` or `RandomForestClassifier`). They are very powerful, often work well without heavy tuning of the parameters, and don't require scaling of the data.

```

# Random Forest model
from sklearn.ensemble import RandomForestClassifier

# instantiate the model
forest = RandomForestClassifier(max_depth=5)

# fit the model
forest.fit(X_train, y_train)

#predicting the target value from the model for the samples
y_test_forest = forest.predict(X_test)
y_train_forest = forest.predict(X_train)

#computing the accuracy of the model performance
acc_train_forest = accuracy_score(y_train,y_train_forest)
acc_test_forest = accuracy_score(y_test,y_test_forest)

print("Random forest: Accuracy on training Data:
{:.3f}".format(acc_train_forest))

print("Random forest: Accuracy on test Data:
{:.3f}".format(acc_test_forest))

```

### Support Vector Machines:

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

```

#Support vector machine model
from sklearn.svm import SVC

# instantiate the model
svm = SVC(kernel='linear', C=1.0, random_state=12)

#fit the model
svm.fit(X_train, y_train)

#predicting the target value from the model for the samples
y_test_svm = svm.predict(X_test)
y_train_svm = svm.predict(X_train)

```

```

#computing the accuracy of the model performance
acc_train_svm = accuracy_score(y_train,y_train_svm)
acc_test_svm = accuracy_score(y_test,y_test_svm)

print("SVM: Accuracy on training Data: {:.3f}".format(acc_train_svm))
print("SVM : Accuracy on test Data: {:.3f}".format(acc_test_svm))

```

### 7.1.2 User interface:

The user opens the site and inputs a URL to check its legitimacy. Necessary features are extracted from this URL and predictions are made.

#### 7.1.2.1 Feature extraction:

We will extract the 13 features that we used to train our model.

##### IP Address in URL:

Checks for the presence of IP address in the URL. URLs may have IP address instead of domain name. If an IP address is used as an alternative of the domain name in the URL, we can be sure that someone is trying to steal personal information with this URL.

If the domain part of URL has IP address, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```

def having_IPhaving_IP_Address(self):
    try:
        ipaddress.ip_address(self.url)
        return -1
    except:
        return 1

```

##### Length of URL:

Computes the length of the URL. Phishers can use long URL to hide the doubtful part in the address bar. In this project, if the length of the URL is greater than or equal 54 characters then the URL classified as phishing otherwise legitimate.

If the length of URL  $\geq 54$ , the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```

def URLURL_Length(self):
    if len(self.url) < 54:
        return 1
    else:
        return -1

```

### Using URL Shortening Services:

URL shortening is a method on the “World Wide Web” in which a URL may be made considerably smaller in length and still lead to the required webpage. This is accomplished by means of an “HTTP Redirect” on a domain name that is short, which links to the webpage that has a long URL.

If the URL is using Shortening Services, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def Shortening_Service(self):
    shortening_services =
    r"bit\.ly|goo\.gl|shorte\.st|go2l\.ink|x\.co|ow\.ly|t\.co|tinyurl|tr\.im|is\.gd|cli\.gs|" \
    r"yfrog\.com|migre\.me|ff\.im|tiny\.cc|url4\.eu|twit\.ac|su\.pr|twurl\.nl|snipurl\.com|" \
    r"short\.to|BudURL\.com|ping\.fm|post\.ly|Just\.as|bkite\.com|snipr\.com|fic\.kr|loopt\
    us|" \
    r"doiop\.com|short\.ie|kl\.am|wp\.me|rubyurl\.com|om\.ly|to\.ly|bit\.do|t\.co|lnkd\.in|
    db\.tt|" \
    r"qr\.ae|adf\.ly|goo\.gl|bitly\.com|cur\.lv|tinyurl\.com|ow\.ly|bit\.ly|ity\.im|q\.gs|is\.gd|
    " \
    r"po\.st|bc\.vc|twitthis\.com|u\.to|j\.mp|buzurl\.com|cutt\.us|u\.bb|yourls\.org|x\.co|" \
    r"prettylinkpro\.com|scrnch\.me|filoops\.info|vzturl\.com|qr\.net|1url\.com|tweez\.me|v
    \.gd|" \
    r"tr\.im|link\.zip\.net"
    match=re.search(shortening_services,self.url)
    if match:
        return -1
    else:
        return 1
```

### "@" Symbol in URL:

Checks for the presence of '@' symbol in the URL. Using "@" symbol in the URL leads the browser to ignore everything preceding the "@" symbol and the real address often follows the "@" symbol.

If the URL has '@' symbol, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def having_At_Symbol(self):
    if "@" in self.url:
        return -1
    else:
        return 1
```

### Redirection "/" in URL:

Checks the presence of "/" in the URL. The existence of "/" within the URL path means that the user will be redirected to another website. The location of the "/" in URL is computed. We find that

if the URL starts with “HTTP”, that means the “//” should appear in the sixth position. However, if the URL employs “HTTPS” then the “//” should appear in seventh position.

If the “//” is anywhere in the URL apart from after the protocol, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def double_slash_redirecting(self):
    pos = self.url.rfind('//')
    if pos > 6:
        if pos > 7:
            return -1
        else:
            return 1
    else:
        return 1
```

#### **Prefix or Suffix “-” in Domain:**

Checking the presence of ‘-’ in the domain part of URL. The dash symbol is rarely used in legitimate URLs. Phishers tend to add prefixes or suffixes separated by (-) to the domain name so that users feel that they are dealing with a legitimate webpage.

If the URL has ‘-’ symbol in the domain part of the URL, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def Prefix_Suffix(self):
    if '-' in urlparse(self.url).netloc:
        return -1
    else:
        return 1
```

#### **HTTPS Token:**

Checks for the presence of “http/https” in the domain part of the URL. The phishers may add the “HTTPS” token to the domain part of a URL in order to trick users.

If the URL has “http/https” in the domain part, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def HTTPS_token(self):
    domain = urlparse(self.url).netloc
    if 'https' in domain:
        return -1
    else:
        return 1
```

### Status Bar Customization (on mouse over):

Phishers may use JavaScript to show a fake URL in the status bar to users. To extract this feature, we must dig-out the webpage source code, particularly the “onMouseOver” event, and check if it makes any changes on the status bar.

If the response is empty or onmouseover is found then, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def on_mouseover(self):
    try:
        if re.findall("", self.response.text):
            return -1
        else:
            return 1
    except:
        return -1
```

### Disabling Right Click:

Phishers use JavaScript to disable the right-click function, so that users cannot view and save the webpage source code. This feature is treated exactly as “Using onMouseOver to hide the Link”. Nonetheless, for this feature, we will search for event “event.button==2” in the webpage source code and check if the right click is disabled.

If the response is empty or onmouseover is not found then, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def RightClick(self):
    if self.response == "":
        return -1
    else:
        if re.findall(r"event.button ?== ?2", self.response.text):
            return 1
        else:
            return -1
```

### Presence of Popup Window:

Pop up windows are another option used by phishers to redirect users to other pages. They display attractive ads to lure the user to click the link. Nonetheless, for this feature, we will search for event “alert” in the webpage source code and check if it is present.

If the response is empty or alert is not found then, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def popUpWidnow(self):
    try:
        if re.findall(r"alert\(", self.response.text):
            return 1
        else:
            return -1
    except:
        return -1
```

### **IFrame Redirection:**

IFrame is an HTML tag used to display an additional webpage into one that is currently shown. Phishers can make use of the “iframe” tag and make it invisible i.e. without frame borders. In this regard, phishers make use of the “frameBorder” attribute which causes the browser to render a visual delineation.

If the iframe is empty or response is not found then, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```
def Iframe(self):
    try:
        if re.findall(r"<iframe>|<frameBorder>", self.response.text):
            return 1
        else:
            return -1
    except:
        return -1
```

### **Age of Domain:**

This feature can be extracted from WHOIS database. Most phishing websites live for a short period of time. The minimum age of the legitimate domain is considered to be 12 months for this project. Age here is nothing but difference between creation and expiration time.

If age of domain > 12 months, the value of this feature is -1 (phishing) else 1 (legitimate).

```
def age_of_domain(self):
    creation_date = self.domain_name.creation_date
    expiration_date = self.domain_name.expiration_date
    if (isinstance(creation_date, str) or isinstance(expiration_date, str)):
        try:
            creation_date = datetime.strptime(creation_date, '%Y-%m-%d')
            expiration_date = datetime.strptime(expiration_date, '%Y-%m-%d')
        except:
            return -1
    if ((expiration_date is None) or (creation_date is None)):
        return -1
    elif ((type(expiration_date) is list) or (type(creation_date) is list)):
```

```

    return -1
else:
    ageofdomain = abs((expiration_date - creation_date).days)
    if ((ageofdomain/30) < 6):
        return -1
    else:
        return 1

```

### DNS Record:

For phishing websites, either the claimed identity is not recognized by the WHOIS database or no records founded for the hostname.

If the DNS record is empty or not found then, the value assigned to this feature is -1 (phishing) or else 1 (legitimate).

```

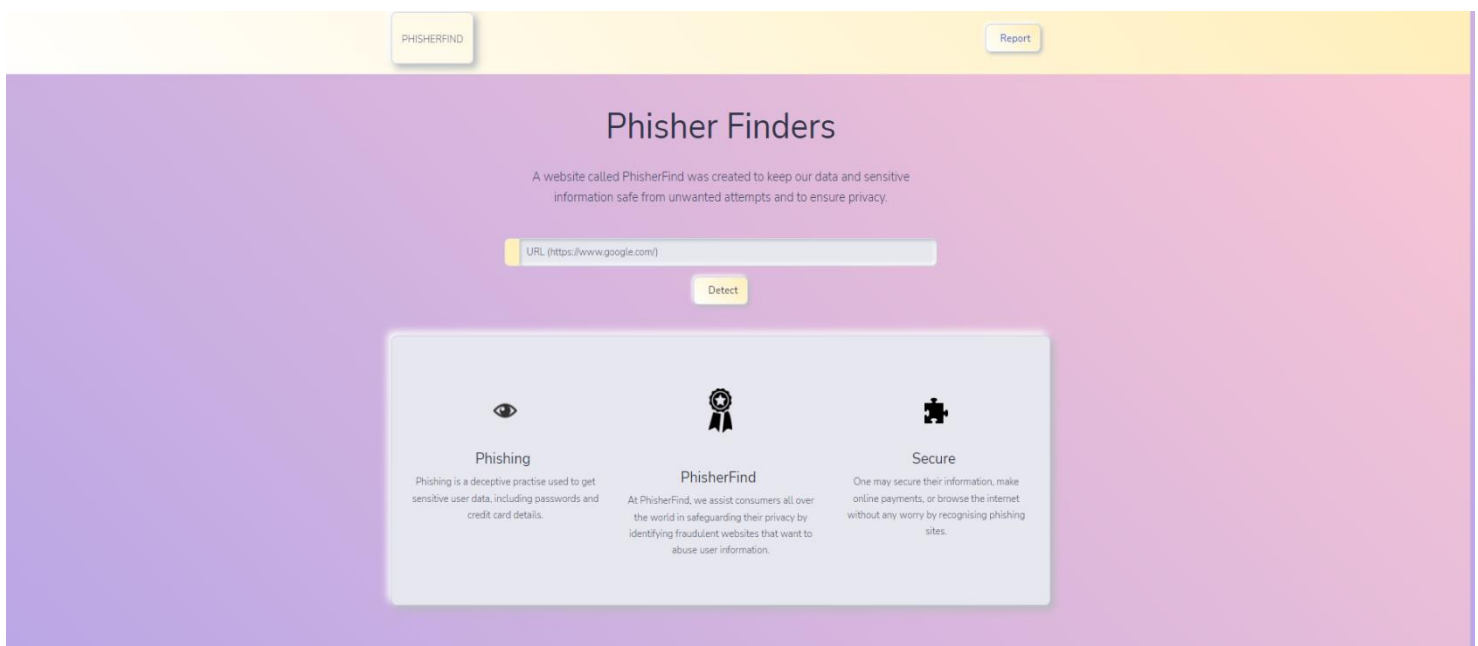
dns = -1
try:
    self.domain_name = whois.whois(urlparse(url).netloc)
except:
    dns = 1

```

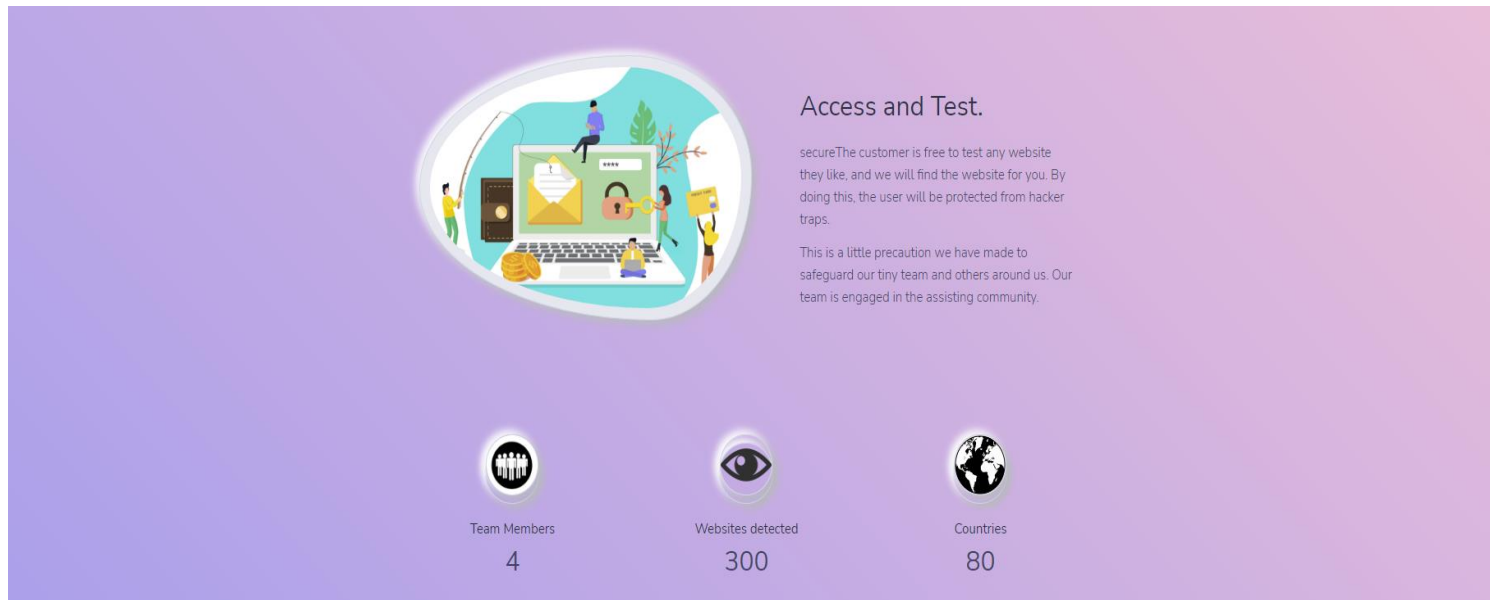
### 7.1.2.2 Dashboard:

The home page of our site “PhisherFind” contains all basic features of the site and a form to get input(URL) from the user.

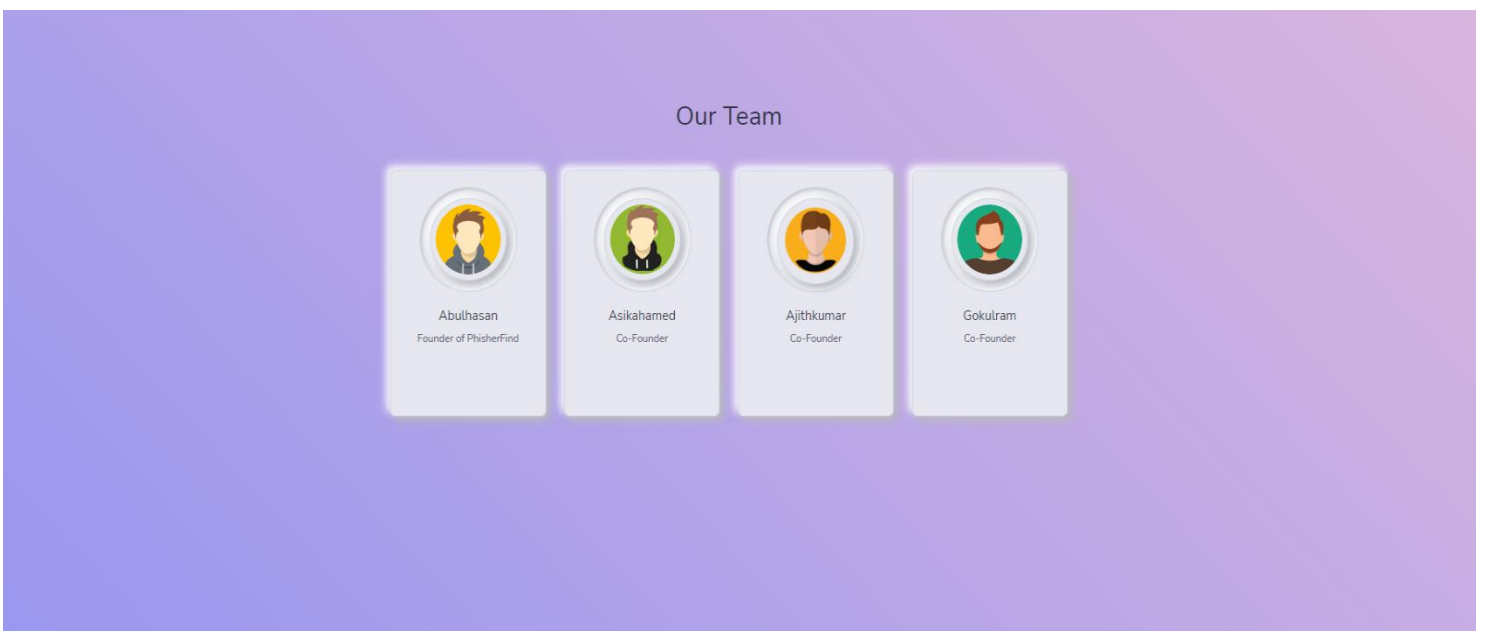
**Fig 7.1 Home page**



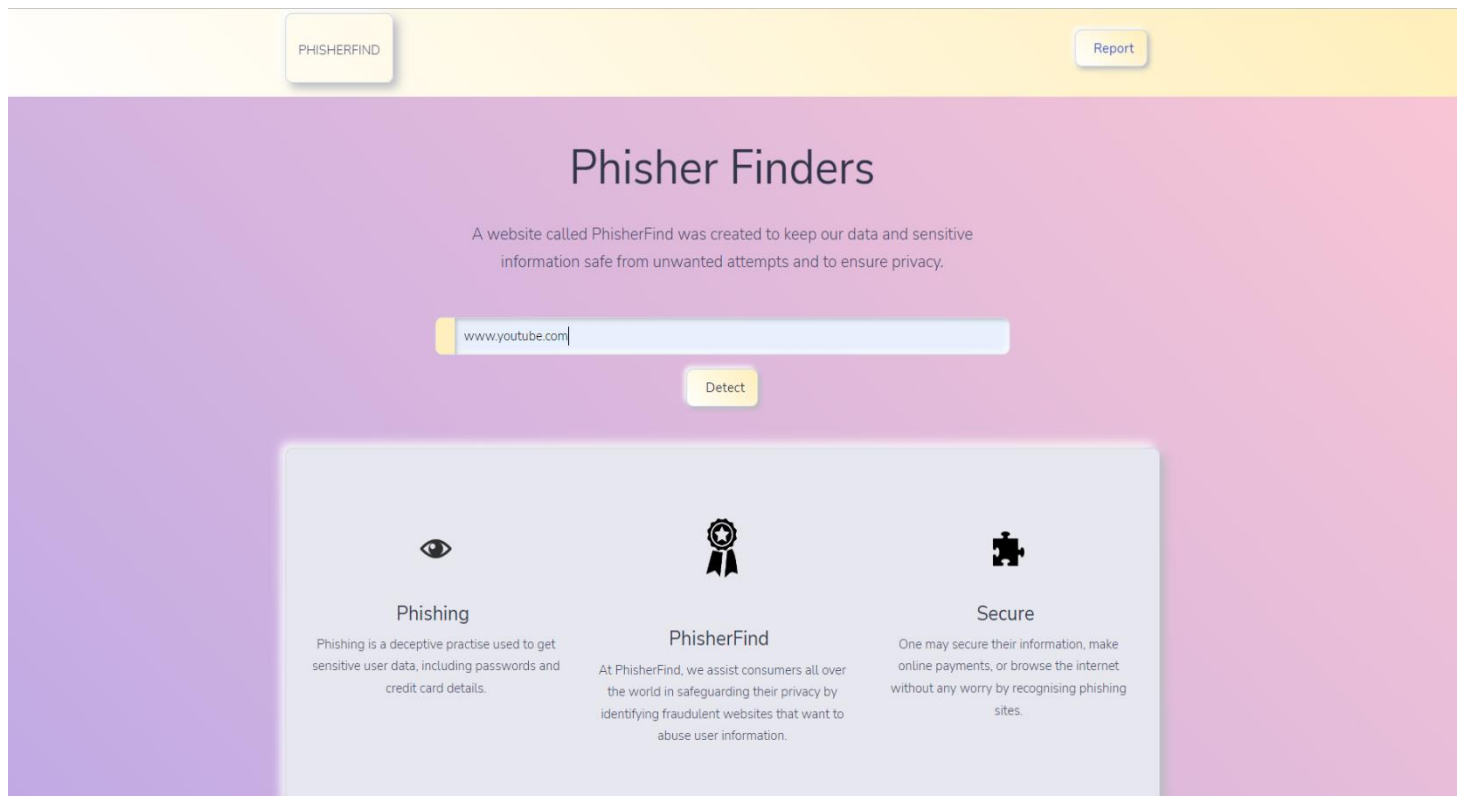




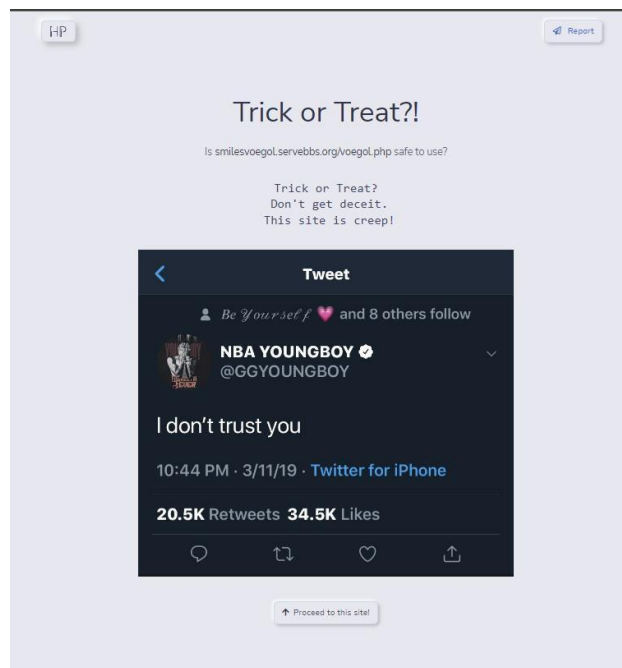
**Fig 7.2 Services section**



**Fig 7.5 Legitimate URL result displayed to the user**



**Fig 7.6 Phishing URL input from user**



**Fig 7.7 Phishing URL result displayed to the user**

## 7.2 Feature 2 – Report:

Report page of our site allows users to provide feedback or ask queries to us. It is a platform to connect with the users of our site. The details provided by the user are stored in a database and is accessible by the admin. The report section consists of a basic form with inputs like name, email and query message. After submitting there is a simple response page displayed to the user to confirm their submission.

**Fig 7.8 Report page**

PHISHERFIND Home

# We are here to help you

Queries? Contact us now

We are here to deliver better results for our clients.

- Unlimited scope
- Identify critical vulnerabilities, malware, exploits and misconfigurations
- Prioritize vulnerabilities remediation
- Detect suspicious activity and stop attacks

Your Name

e.g. Annie

Identify critical vulnerabilities, malware, exploits and misconfigurations

Prioritize vulnerabilities remediation

Detect suspicious activity and stop attacks

Your Name

e.g. Annie

Your Email

example@company.com

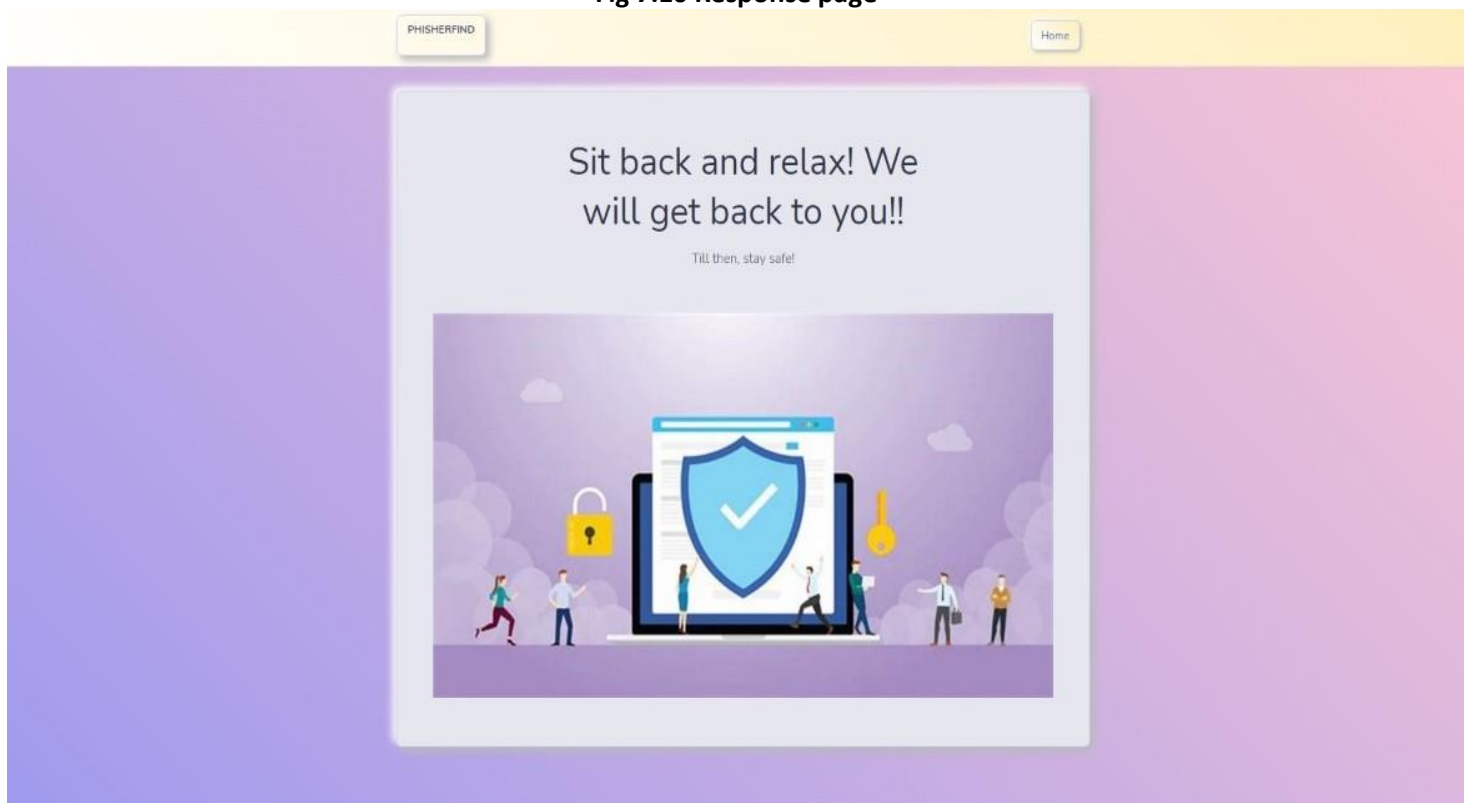
Your Query

Type your query here...

POST







**Fig 7.9 Sample user input to the report section**

**Fig 7.10 Response page**



### 7.3 Database schema:

MySQL is used to create a database to store the inputs from the “Report” page of the website. A table named “responses” is created under the database named “report” with 3 columns named name, email and query. Every time a user submits the report form, the table gets updated with those values.

|                          | # | Name         | Type        | Collation          | Attributes | Null | Default | Comments | Extra | Action   |
|--------------------------|---|--------------|-------------|--------------------|------------|------|---------|----------|-------|--|
| <input type="checkbox"/> | 1 | <b>name</b>  | varchar(50) | utf8mb4_general_ci |            | No   | None    |          |       |  Change  Drop <a href="#">More</a> |
| <input type="checkbox"/> | 2 | <b>email</b> | varchar(50) | utf8mb4_general_ci |            | No   | None    |          |       |  Change  Drop <a href="#">More</a> |
| <input type="checkbox"/> | 3 | <b>query</b> | text        | utf8mb4_general_ci |            | No   | None    |          |       |  Change  Drop <a href="#">More</a> |

**Fig 7.11 Database schem**



## CHAPTER 8

### TESTING

#### 8.1 Test Cases:

| Test case ID     | Feature Type | Component   | Test Scenario  | Steps To Execute  | Test Data  | Expected Result   | Actual Result       | Status |
|------------------|--------------|-------------|--|---|--|---|---------------------|--------|
| DashBoard_TC_OO1 | Functional   | Home Page   | Verify user is able to enter the URL in the form                             | 1. Open PhisherFind website<br>2. Enter a URL and click submit  | <a href="https://google.com/">https://google.com/</a>  | Result of classification will be displayed  | Working as expected | Pass   |
| DashBoard_TC_OO2 | UI           | Home Page   | Verify the UI elements in the form   | 1. Enter URL and click go<br>2. The services and teams' sections are visible<br>3. Enter a URL and click submit                         | <a href="https://google.com/">https://google.com/</a>  | Application should show below UI elements:<br>a. input form<br>b. submit button<br>c. services<br>d. team | Working as expected | Pass   |
| DashBoard_TC_OO3 | Functional   | Home page   | Verify user is able to see an alert when nothing is entered in the textbox   | 1. Enter URL and click go<br>2. Enter nothing and click submit<br>3. An alert is displayed to provide proper input                      |  | Alert of incomplete input   | Working as expected | Pass   |
| DashBoard_TC_OO4 | Functional   | Home page   | Verify user is able to see the result when URL is entered in the textbox     | 1. Enter URL and click go<br>2. Enter any URL and click submit<br>3. The result of the classification is displayed.                     | <a href="https://google.com/">https://google.com/</a>  | Result of classification will be displayed  | Working as expected | Pass   |
| Report_TC_OO1    | Functional   | Report page | Verify user is able to enter their name, email and query message in the form | 1. Enter URL and click go<br>2. Click on report button<br>3. Enter Valid name, email and query in the form<br>4. Click on submit button | <b>Name:</b> Alex<br><b>Email:</b> <a href="mailto:alex123@gmail.com">alex123@gmail.com</a><br><b>Query:</b> Hey! I need to check if a website is legitimate | Details are stored in the database  | Working as expected | Pass   |



## 8.2 User Acceptance Testing:

### Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

| Resolution     | Severity 1 | Severity 2 | Severity 3 | Severity 4 | Subtotal |
|----------------|------------|------------|------------|------------|----------|
| By Design      | 10         | 4          | 2          | 3          | 20       |
| Duplicate      | 1          | 0          | 3          | 0          | 4        |
| External       | 2          | 3          | 0          | 1          | 6        |
| Fixed          | 11         | 2          | 4          | 20         | 37       |
| Not Reproduced | 0          | 0          | 1          | 0          | 1        |
| Skipped        | 0          | 0          | 1          | 1          | 2        |
| Won't Fix      | 0          | 5          | 2          | 1          | 8        |
| Totals         | 24         | 14         | 13         | 26         | 77       |

### Test Case Analysis:

This report shows the number of test cases that have passed, failed, and untested

| Section             | Total Cases | Not Tested | Fail | Pass |
|---------------------|-------------|------------|------|------|
| Print Engine        | 5           | 0          | 0    | 5-   |
| Client Application  | 51          | 0          | 0    | 51   |
| Security            | 2           | 0          | 0    | 2    |
| Outsource Shipping  | 3           | 0          | 0    | 3    |
| Exception Reporting | 9           | 0          | 0    | 9    |
| Final Report Output | 4           | 0          | 0    | 4    |
| Version Control     | 2           | 0          | 0    | 2    |

## CHAPTER 9

### RESULTS

#### 9.1 Performance metrics:

The median efficiency is used to assess each categorization model's effectiveness. The final item will appear in the way it was envisioned. Graphical representations are used to depict information during classification. The percentage of predictions made using the testing dataset is used to gauge accuracy. By dividing the entire number of forecasts even by properly predicted estimates, it is simple to calculate. The difference between actual and anticipated output is used to calculate accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP = True Positives, TN = True Negatives, FN = False Negatives and FP = False Positives.

Thus, accuracy for all the four used models were calculated and ranked. XGBoost performed better than other models.

|   | ML Model      | Train Accuracy | Test Accuracy |
|---|---------------|----------------|---------------|
| 3 | XGBoost       | 0.913          | 0.905         |
| 0 | Decision Tree | 0.898          | 0.894         |
| 1 | Random Forest | 0.893          | 0.886         |
| 2 | SVM           | 0.886          | 0.883         |

Fig 9.1 Performance metrics

## CHAPTER 10

### ADVANTAGES & DISADVANTAGES

#### ADVANTAGES:

- **Increases user alertness to phishing risks** Whenever the user navigates into the website and provide the URL of the website that needs to be verified for legitimacy, the system detects phishing sites by applying a machine learning algorithm which implements classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy which in turn helps the customers to eliminate the risks of cyber threat and protect their valuable corporate or personal data.
- **Users will also be able to pose any query to the admin through the report page designed**  
Our system is also provided with an option for the clients to report to the administrator which helps them to ask their questions significantly improving their experience on our site.

#### DISADVANTAGES:

- Not a generalized model
- Huge number of rules
- Needs feed continuously

## **CHAPTER 11**

### **CONCLUSION**

Phishing detection is now an area of great interest among the researchers due to its significance in protecting privacy and providing security. There are many methods to perform phishing detection. Our system aims to enhance the detection method to detect phishing websites using machine learning technology. We achieved a high detection accuracy, and the results show that the classifiers give better performance when we use more data as training data.

In future, hybrid technology will be implemented to detect phishing websites more accurately.

## **CHAPTER-12**

### **FUTURE SCOPE**

In future we intend to build an add-ons for our system and if we get a structured dataset of phishing, we can perform phishing detection much faster than any other technique. We can also use a combination of any two or more classifiers to get maximum accuracy. We plan to explore various phishing techniques which use Network based features, Content based features, Webpage based features and HTML and JavaScript features of web pages which will improve the performance of the system. In particular, we extract features from URLs and pass it through the various classifiers.

## CHAPTER 13

### APPENDIX

#### 13.1 Source code:

##### app.py

```
from flask import Flask, render_template, request
import numpy as np
import pandas as pd
from sklearn import metrics
import warnings
import pickle
warnings.filterwarnings('ignore')
from features import FeatureExtraction
from flask_mysql import MySQL
import requests

# NOTE: you must manually set API_KEY below using information retrieved
# from your IBM Cloud account.
API_KEY = ""

token_response = requests.post('https://iam.cloud.ibm.com/identity/token',
data={"apikey":
    API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' +
mltoken}

app = Flask(__name__)

app.config['MYSQL_HOST'] = 'localhost'
app.config['MYSQL_USER'] = 'root'
app.config['MYSQL_PASSWORD'] = ''
app.config['MYSQL_DB'] = 'report'

mysql = MySQL(app)
```

```

xgb = pickle.load(open("XGBoostClassifier.pkl", "rb"))

@app.route("/", methods=["GET", "POST"])
def home():
    if request.method == "POST":

        url = request.form["url"]
        obj = FeatureExtraction(url)

        x = np.array(obj.getFeaturesList()).reshape(1,13)
        print(x)
        t=obj.getFeaturesList()
        print("t")
        print(t)

        # NOTE: manually define and pass the array(s) of values to be
        scored in the next line
        payload_scoring = {"input_data": [{"fields":
        [['f0','f1','f2','f3','f4','f5','f6','f7','f8','f9','f10','f11','f12']],
        "values": t}]}

        response_scoring = requests.post('https://us-
        south.ml.cloud.ibm.com/ml/v4/deployments/859ae568-d692-4958-9dbe-
        60431a8a0918/predictions?version=2022-11-11', json=payload_scoring,
        headers={'Authorization': 'Bearer ' + mltoken})

        print("Scoring response")
        print(response_scoring.json())

        y_pred =xgb.predict(x) [0]
        print(y_pred)
        y_pro_phishing = xgb.predict_proba(x) [0,0]
        print(y_pro_phishing)
        y_pro_non_phishing = xgb.predict_proba(x) [0,1]
        print(y_pro_non_phishing)

        if(y_pro_phishing*100<60):
            msg="Trick or Treat?\n Look at this tweet.\n This site is
            elite!\n"

```

```

        flag=1
    else:
        msg="Trick or Treat?\n Don't get deceit.\n This site is
creep!\n"
        flag=-1

    return render_template('result.html', msg=msg, url=url, val=flag)

return render_template("index.html")

@app.route("/report", methods=["GET", "POST"])
def report():
    if request.method == 'GET':
        return render_template("contact.html")

    if request.method == 'POST':
        name = request.form['name']
        email = request.form['email']
        query = request.form['query']
        cursor = mysql.connection.cursor()
        cursor.execute(''' INSERT INTO responses
VALUES(%s,%s,%s)''', (name,email,query))
        mysql.connection.commit()
        cursor.close()
        return render_template("alert.html")

if __name__ == '__main__':
    app.run(debug=True)

```

### **features.py**

```

from urllib.parse import urlparse
import ipaddress
import re
import requests

```

```

import whois

from datetime import datetime

class FeatureExtraction:

    features=[]

    def __init__(self,url):

        self.features=[]

        self.url = url


        #Address bar based features

        self.features.append(self.having_IPhaving_IP_Address())

        self.features.append(self.URLURL_Length())

        self.features.append(self.Shortining_Service())

        self.features.append(self.having_At_Symbol())

        self.features.append(self.double_slash_redirecting())

        self.features.append(self.Prefix_Suffix())

        self.features.append(self.HTTPS_token())


        # HTML & Javascript based features

        try:

            self.response = requests.get(url)

        except:

            self.response = ""


        self.features.append(self.on_mouseover())

        self.features.append(self.RightClick())

        self.features.append(self.popUpWidnow())

        self.features.append(self.Iframe())


        #Domain based features

        dns = -1

        try:

            self.domain_name = whois.whois(urlparse(url).netloc)

        except:

```



```

        dns = 1

    self.features.append(1 if dns == 1 else self.age_of_domain())
    self.features.append(dns)

# 1.UsingIp
def having_IPhaving_IP_Address(self):
    #print("IP")
    try:
        ipaddress.ip_address(self.url)
        print("IP")
        return -1
    except:
        print("IP except")
        return 1

# 2.longUrl
def URLURL_Length(self):
    #print("Length")
    if len(self.url) < 54:
        return 1
    else:
        return -1

# 3.shortUrl
def Shortening_Service(self):
    #print("short")
    shortening_services =
r"bit\.ly|goo\.gl|shorte\.st|go2l\.ink|x\.co|ow\.ly|t\.co|tinyurl|tr\.im|is
\.gd|cli\.gs|" \

r"yfrog\.com|migre\.me|ff\.im|tiny\.cc|url4\.eu|twit\.ac|su\.pr|twurl\.nl|s
nipurl\.com|" \

r"short\.to|BudURL\.com|ping\.fm|post\.ly|Just\.as|bkite\.com|snipr\.com|fi
c\.kr|loopt\.us|" \

```

```

r"doiop\.com|short\.ie|kl\.am|wp\.me|rubyurl\.com|om\.ly|to\.ly|bit\.do|t\.
co|lnkd\.in|db\.tt|" \

r"qr\.ae|adf\.ly|goo\.gl|bitly\.com|cur\.lv|tinyurl\.com|ow\.ly|bit\.ly|ity
\.im|q\.gs|is\.gd|" \

r"po\.st|bc\.vc|twitthis\.com|u\.to|j\.mp|buzurl\.com|cutt\.us|u\.bb|yourls
\.org|x\.co|" \

r"prettylinkpro\.com|scrnch\.me|filoops\.info|vzturl\.com|qr\.net|lurl\.com
|tweez\.me|v\.gd|" \

        r"tr\.im|link\.zip\.net"

match=re.search(shortening_services,self.url)

if match:

    return -1

else:

    return 1

# 4.Symbol@
def having_At_Symbol(self):

    #print("at")

    if "@" in self.url:

        return -1

    else:

        return 1

# 5.Redirecting//
def double_slash_redirecting(self):

    #print("//")

    pos = self.url.rfind('//')

    if pos > 6:

        if pos > 7:

            return -1

        else:

            return 1

    else:

        return 1

```

```

# 6.prefixSuffix
def Prefix_Suffix(self):
    #print("prefix")
    if '-' in urlparse(self.url).netloc:
        return -1
    else:
        return 1

#HTTPS token
def HTTPS_token(self):
    #print("https")
    domain = urlparse(self.url).netloc
    if 'https' in domain:
        return -1
    else:
        return 1

def on_mouseover(self):
    #print("mouse")
    try:
        if re.findall("", self.response.text):
            return -1
        else:
            return 1
    except:
        return -1

def RightClick(self):
    #print("right")
    if self.response == "":
        return -1
    else:
        if re.findall(r"event.button ?== ?2", self.response.text):

```

```

        return 1
    else:
        return -1

# 11. UsingPopupWindow
def popUpWidnow(self):
    #print("popup")
    try:
        if re.findall(r"alert\(", self.response.text):
            return 1
        else:
            return -1
    except:
        return -1

# 12. IframeRedirection
def Iframe(self):
    #print("iframe")
    try:
        if re.findall(r"<iframe>|<frameBorder>]", self.response.text):
            return 1
        else:
            return -1
    except:
        return -1

# 13.Survival time of domain: The difference between termination time
and creation time (Domain_Age)
def age_of_domain(self):
    #print("age")
    creation_date = self.domain_name.creation_date
    expiration_date = self.domain_name.expiration_date
    if (isinstance(creation_date,str) or
isinstance(expiration_date,str)):
        try:

```

```

        creation_date = datetime.strptime(creation_date,'%Y-%m-%d')
        expiration_date = datetime.strptime(expiration_date,"%Y-%m-%d")
    except:
        return -1

    if ((expiration_date is None) or (creation_date is None)):
        return -1

    elif ((type(expiration_date) is list) or (type(creation_date) is
list))):
        return -1

    else:
        ageofdomain = abs((expiration_date - creation_date).days)
        if ((ageofdomain/30) < 6):
            return -1
        else:
            return 1

def getFeaturesList(self):
    print(self.features)
    return self.features

```

## index.html

```

<!DOCTYPE html>
<html lang="en">

<head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
    <!-- Primary Meta Tags -->
    <title>PhisherFind</title>
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-
to-fit=no">
    <meta name="title" content="PhisherFind">
    <meta name="author" content="PhisherFind">

    <!-- Favicon -->

```

```

<link rel="apple-touch-icon" sizes="180x180"
href="../static/assets/img/favicon/apple-touch-icon.png">

<link rel="icon" type="image/png" sizes="32x32"
href="../static/assets/img/favicon/favicon-32x32.png">

<link rel="icon" type="image/png" sizes="16x16"
href="../static/assets/img/favicon/favicon-16x16.png">

<link rel="manifest" href="../static/assets/img/favicon/site.webmanifest">


<!-- Fontawesome -->

<link type="text/css" href="../static/vendor/@fontawesome/fontawesome-free/css/all.min.css" rel="stylesheet">


<!-- Pixel CSS -->

<link type="text/css" href="../static/css/neumorphism.css"
rel="stylesheet">


</head>


<body>

    <header class="header-global">

        <nav id="navbar-main" aria-label="Primary navigation" class="navbar navbar-main navbar-expand-lg navbar-theme-primary headroom navbar-light">

            <div class="container position-relative">

                <a class="navbar-brand shadow-soft py-2 px-3 rounded border border-light mr-lg-4" href="/">

                </a>

                <div class="navbar-collapse collapse" id="navbar_global">

                    <div class="navbar-collapse-header">

                        <div class="row">

                            <div class="col-6 collapse-brand">

                                <a href="/" class="navbar-brand shadow-soft py-2 px-3 rounded border border-light">

                                </a>

```

```

        </div>

        <div class="col-6 collapse-close">
            <a href="#navbar_global" class="fas fa-times"
data-toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" title="close" aria-
label="Toggle navigation"></a>
        </div>
    </div>
</div>
</div>
</div>
<div class="d-flex align-items-center">
    <a href="/report" target="_blank" class="btn btn-primary
text-secondary d-none d-md-inline-block mr-3"><i class="far fa-paper-plane
mr-2"></i> Report</a>

    <button class="navbar-toggler ml-2" type="button" data-
toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" aria-label="Toggle
navigation">

        <span class="navbar-toggler-icon"></span>
    </button>
</div>
</div>
</nav>
</header>
<main>
    <!-- Hero -->
    <div class="section section-header pb-7">
        <div class="container">
            <div class="row justify-content-center">
                <div class="col-12 col-lg-8 text-center">
                    <h1 class="display-2 mb-4">We are PhisherFind</h1>
                    <p class="lead mb-5">PhisherFind is a website
designed to protect our data and sensitive information from malicious
attacks and maintain confidentiality. </p>
                    <form action="/" method="post" class="lead mb-5">
                        <div class="form-group">
                            <div class="input-group">
                                <div class="input-group-prepend">

```

```

                                <span class="input-group-
text"><span class="fa fa-link"></span></span>
                                </div>
                                <input class="form-control" id="url"
name="url" placeholder="URL (https://www.google.com/)" type="text" aria-
label="url-input" required>
                                </div>
                                </div>
                                <button type="submit" class="btn btn-
primary"><span class="fa fa-user-secret mr-2"></span>Detect</button>
                                </form>
                                </div>
                                </div>
                                </div>
                                <!-- End of Hero section -->
                                <section class="section section-lg pt-0">
                                    <div class="container">
                                        <div class="row">
                                            <div class="col">
                                                <div class="card bg-primary shadow-soft border-
light p-4">
                                                    <div class="row">
                                                        <div class="col-12 col-lg-4 px-md-0 mb-4
mb-lg-0">
                                                            <div class="card-body text-center bg-
primary py-5">
                                                                <div class="icon icon-shape shadow-
inset border-light rounded-circle mb-3">
                                                                    <span class="far fa-
eye"></span>
                                                                </div>
                                                                <!-- Heading -->
                                                                <h2 class="h4 mr-2">
                                                                    Phishing
                                                                </h2>
                                                                <!-- Text -->
                                                                <p class="mb-0">Phishing is a
fraudulent activity that is done to steal confidential user information
such as credit card numbers and passwords.</p>

```



```

        </div>
    </div>
    <div class="col-12 col-lg-4 px-md-0 mb-4
mb-lg-0">
        <div class="card-body text-center bg-
primary py-5">
            <div class="icon icon-shape shadow-
inset border-light rounded-circle mb-3">
                <span class="fas fa-
medal"></span>
            </div>
            <!-- Heading -->
            <h2 class="h4 mr-2">
                PhisherFind
            </h2>
            <!-- Text -->
            <p class="mb-0">At PhisherFind, we
help users across the world to protect their data by detecting the
fraudulent sites intended to misuse user's data.</p>
        </div>
    </div>
    <div class="col-12 col-lg-4 px-md-0">
        <div class="card-body text-center bg-
primary py-5">
            <div class="icon icon-shape shadow-
inset border-light rounded-circle mb-3">
                <span class="fas fa-puzzle-
piece"></span>
            </div>
            <!-- Heading -->
            <h2 class="h4 mr-2">
                Secure
            </h2>
            <!-- Text -->
            <p class="mb-0">By detecting the
phishing sites, one can secure his/her informations and make online
payments or surf the internet without any fear.</p>
        </div>
    </div>

```

```

        </div>
    </div>
</div>
</div>
</div>
</div>
</section>
<!-- Section -->
<section class="section section-lg pt-0">
    <div class="container">
        <div class="row align-items-center justify-content-around">
            <div class="col-md-6 col-xl-6 mb-5">
                <div class="card bg-primary shadow-soft border-
light organic-radius p-3">
                    
                </div>
            </div>
            <div class="col-md-6 col-xl-5 text-center text-md-
left">
                <h2 class="h1 mb-4">Enter and Test.</h2>
                <p class="lead">The user can enter any site that
they want to test and we'll detect the website for you. This will ensure
the user don't fall into the traps of hackers .</p>
                <p class="lead">Our small team is active in the
helping community and this is a small step taken by us to protect ourselves
and the people surrounding us.</p>
                
            </div>
        </div>
    </div>
</div>
</section>
<!-- End of section -->
<!-- Section -->
<section class="section section-lg pt-0">
    <div class="container">
        <div class="row">
            <div class="col-12 col-sm-4 col-lg-4 text-center">

```

```

<!-- Visit Box -->
<div class="icon-box mb-4">
    <div class="icon icon-shape shadow-soft border
border-light rounded-circle mb-4">
        <span class="far fa-smile-beam"></span>
    </div>
    <h3 class="h5">Team Members</h3>
    <span class="counter-slow display-3 text-gray
d-block">4</span>
</div>
<!-- End of Visit Box -->
</div>
<div class="col-12 col-sm-4 col-lg-4 text-center">
    <!-- Call Box -->
    <div class="icon-box mb-4">
        <div class="icon icon-shape shadow-soft border
border-light rounded-circle mb-4">
            <span class="far fa-eye"></span>
        </div>
        <h3 class="h5">Websites detected</h3>
        <span class="counter display-3 text-gray d-
block">300</span>
    </div>
    <!-- End of Call Box -->
</div>
<div class="col-12 col-sm-4 col-lg-4 text-center">
    <!-- Email Box -->
    <div class="icon-box mb-4">
        <div class="icon icon-shape shadow-soft border
border-light rounded-circle mb-4">
            <span class="fas fa-globe-europe"></span>
        </div>
        <h3 class="h5">Countries</h3>
        <span class="counter-slow display-3 text-gray
d-block">80</span>
    </div>
    <!-- End of Email Box -->

```

```

        </div>
    </div>
</div>
</section>
<!-- End of section -->
<!-- Section -->
<section class="section section-lg pt-0">
    <div class="container">
        <div class="row justify-content-center mb-5">
            <h2 class="h1">Our Team</h2>
        </div>
        <div class="row">
            <div class="col-12 col-md-6 col-lg-3">
                <!-- Profile Card -->
                <div class="card bg-primary shadow-soft border-
light text-center py-4 mb-5">
                    <div class="profile-image shadow-inset border
border-light bg-primary rounded-circle p-3 mx-auto">
                        
                    </div>
                    <div class="card-body">
                        <h3 class="h5 mb-2">Abulhasan</h3>
                        <span class="h6 font-weight-normal text-
gray mb-3">Founder of PhisherFind</span>
                        <ul class="list-unstyled d-flex justify-
content-center my-3">
                            <li>
                                <a
href="https://www.instagram.com/ig_hasan " target="_blank" aria-
label="instagram social link" class="icon icon-xs icon-facebook mr-3">
                                    <span class="fab fa-
instagram"></span>
                                </a>
                            </li>
                            <li>

```

```

                                <a
href="https://github.com/abulhasan-salam/abulhasan-salam"
target="_blank" aria-label="githubsocial link" class="icon icon-xs
icon-dribbble mr-3">

                                <span class="fab fa-
github"></span>

                                </a>

                                </li>

                                </ul>

                                </div>

                                </div>

                                <!-- End of Profile Card -->

                                </div>

                                <div class="col-12 col-md-6 col-lg-3">

                                    <!-- Profile Card -->

                                    <div class="card bg-primary shadow-soft border-
light text-center py-4 mb-5">

                                        <div class="profile-image shadow-inset border
border-light bg-primary rounded-circle p-3 mx-auto">

                                            </div>

                                            <div class="card-body">

                                                <h3 class="h5 mb-2">Asikahamed</h3>

                                                <span class="h6 font-weight-normal text-
gray mb-3">Co-Founder</span>

                                                <ul class="list-unstyled d-flex justify-
content-center my-3">

                                                    <li>

                                                        <a
href="https://www.instagram.com/ashik_says " target="_blank" aria-
label="instagram social link" class="icon icon-xs icon-facebook mr-3">

                                                            <span class="fab fa-
instagram"></span>

                                                        </a>

                                                    </li>

                                                    <li>

                                                        <a
href="https://github.com/AsikahamedA" target="_blank" aria-
label="githubsocial link" class="icon icon-xs icon-dribbble mr-3">

```

```

                                <span class="fab fa-
github"></span>

                                </a>
                                </li>
                                </ul>
                                </div>
                                </div>
                                <!-- End of Profile Card -->
                                </div>
                                <div class="col-12 col-md-6 col-lg-3">
                                    <!-- Profile Card -->
                                    <div class="card bg-primary shadow-soft border-
light text-center py-4 mb-5">
                                        <div class="profile-image shadow-inset border
border-light bg-primary rounded-circle p-3 mx-auto">
                                            
                                        </div>
                                        <div class="card-body">
                                            <h3 class="h5 mb-2">Ajithkumar</h3>
                                            <span class="h6 font-weight-normal text-
gray mb-3">Co-Founder</span>
                                            <ul class="list-unstyled d-flex justify-
content-center my-3">
                                                <li>
                                                    <a
href="https://instagram.com/" target="_blank" aria-label="instagram
social link" class="icon icon-xs icon-facebook mr-3">
                                                        <span class="fab fa-
instagram"></span>
                                                    </a>
                                                </li>
                                                <li>
                                                    <a href="https://github.com/Ajith-
29" target="_blank" aria-label="github social link" class="icon icon-xs
icon-dribbble mr-3">
                                                        <span class="fab fa-
github"></span>
                                                    </a>

```

```

        </li>
    </ul>
</div>
</div>
<!-- End of Profile Card -->
</div>
<div class="col-12 col-md-6 col-lg-3">
    <!-- Profile Card -->
    <div class="card bg-primary shadow-soft border-
light text-center py-4 mb-5">
        <div class="profile-image shadow-inset border
border-light bg-primary rounded-circle p-3 mx-auto">
            
        </div>
        <div class="card-body">
            <h3 class="h5 mb-2">Gokulram</h3>
            <span class="h6 font-weight-normal text-
gray mb-3">Co-Founder</span>
            <ul class="list-unstyled d-flex justify-
content-center my-3">
                <li>
                    <a
href="https://instagram.com/" target="_blank" aria-label="instagramsocial
link" class="icon icon-xs icon-facebook mr-3">
                        <span class="fab fa-
instagram"></span>
                    </a>
                </li>
                <li>
                    <a
href="https://github.com/Gokulram" target="_blank" aria-label="github
social link" class="icon icon-xs icon-dribbble mr-3">
                        <span class="fab fa-
github"></span>
                    </a>
                </li>
            </ul>

```





```

<script src="../../static/vendor/bootstrap-datepicker/dist/js/bootstrap-
datepicker.min.js"></script>

<script
src="../../static/vendor/waypoints/lib/jquery.waypoints.min.js"></script>

<script      src="../../static/vendor/jarallax/dist/jarallax.min.js"></script>

<script
src="../../static/vendor/jquery.counterup/jquery.counterup.min.js"></script>

<script src="../../static/vendor/jquery-
countdown/dist/jquery.countdown.min.js"></script>

<script src="../../static/vendor/smooth-scroll/dist/smooth-
scroll.polyfills.min.js"></script>

<script src="../../static/vendor/prismjs/prism.js"></script>

<script async defer src="https://buttons.github.io/buttons.js"></script>

<!-- Neumorphism JS -->
<script src="../../static/assets/js/neumorphism.js"></script>
</body>

</html>

```

### result.html

```

<!DOCTYPE html>
<html lang="en">

<head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
    <!-- Primary Meta Tags -->
    <title>PhisherFind</title>
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-
to-fit=no">
    <meta name="title" content="PhisherFind">
    <meta name="author" content="PhisherFind">

    <!-- Favicon -->

```

```

<link rel="apple-touch-icon" sizes="180x180"
href="../static/assets/img/favicon/apple-touch-icon.png">

<link rel="icon" type="image/png" sizes="32x32"
href="../static/assets/img/favicon/favicon-32x32.png">

<link rel="icon" type="image/png" sizes="16x16"
href="../static/assets/img/favicon/favicon-16x16.png">

<link rel="manifest" href="../static/assets/img/favicon/site.webmanifest">


<!-- Fontawesome -->

<link type="text/css" href="../static/vendor/@fontawesome/fontawesome-free/css/all.min.css" rel="stylesheet">


<!-- Pixel CSS -->

<link type="text/css" href="../static/css/neumorphism.css"
rel="stylesheet">


</head>


<body>

    <header class="header-global">

        <nav id="navbar-main" aria-label="Primary navigation" class="navbar navbar-main navbar-expand-lg navbar-theme-primary headroom navbar-light">

            <div class="container position-relative">

                <a class="navbar-brand shadow-soft py-2 px-3 rounded border border-light mr-lg-4" href="/">

                </a>

                <div class="navbar-collapse collapse" id="navbar_global">

                    <div class="navbar-collapse-header">

                        <div class="row">

                            <div class="col-6 collapse-brand">

                                <a href="/" class="navbar-brand shadow-soft py-2 px-3 rounded border border-light">

                                </a>

```

```

        </div>

        <div class="col-6 collapse-close">
            <a href="#navbar_global" class="fas fa-times"
data-toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" title="close" aria-
label="Toggle navigation"></a>
        </div>
    </div>
</div>

<div class="d-flex align-items-center">
    <a href="/report" target="_blank" class="btn btn-primary
text-secondary d-none d-md-inline-block mr-3"><i class="far fa-paper-plane
mr-2"></i> Report</a>

    <button class="navbar-toggler ml-2" type="button" data-
toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" aria-label="Toggle
navigation">

        <span class="navbar-toggler-icon"></span>
    </button>
</div>
</div>
</nav>
</header>

<main>

    <!-- Result -->
    <div class="section section-header pb-7">
        <div class="container">
            <div class="row justify-content-center">
                <div class="col-12 col-lg-8 text-center">
                    <h1 class="display-2 mb-4">Trick or Treat?!</h1>
                    <p class="lead">Is <a href= {{ url }}
target="_blank">{{ url }}</a> safe to use?</p>
                    <br>
                    <h3><pre>{{ msg }}</pre></h3>
                    <br>

```



```

<!-- Vendor JS -->

<script src="../../static/vendor/onscreen/dist/on-screen.umd.min.js"></script>

<script
src="../../static/vendor/nouislider/distribute/nouislider.min.js"></script>

<script src="../../static/vendor/bootstrap-datepicker/dist/js/bootstrap-
datepicker.min.js"></script>

<script
src="../../static/vendor/waypoints/lib/jquery.waypoints.min.js"></script>

<script      src="../../static/vendor/jarallax/dist/jarallax.min.js"></script>

<script
src="../../static/vendor/jquery.counterup/jquery.counterup.min.js"></script>

<script src="../../static/vendor/jquery-
countdown/dist/jquery.countdown.min.js"></script>

<script src="../../static/vendor/smooth-scroll/dist/smooth-
scroll.polyfills.min.js"></script>

<script src="../../static/vendor/prismjs/prism.js"></script>

<script async defer src="https://buttons.github.io/buttons.js"></script>

<!-- Neumorphism JS -->

<script src="../../static/assets/js/neumorphism.js"></script>

<script>
    let x= '{{val}}'
    if(x==1){
        document.getElementById('phishing').style.display = 'none';
    }else{
        document.getElementById('legitimate').style.display = 'none';
    }
</script>
</body>

</html>

```

## contact.html

```
<!DOCTYPE html>
```

```

<html lang="en">

<head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
    <!-- Primary Meta Tags -->
    <title>PhisherFind - Report</title>
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="title" content="PhisherFind - Report">
    <meta name="author" content="PhisherFind">

    <!-- Favicon -->
    <link rel="apple-touch-icon" sizes="120x120"
href="../static/assets/img/favicon/apple-touch-icon.png">
    <link rel="icon" type="image/png" sizes="32x32"
href="../static/assets/img/favicon/favicon-32x32.png">
    <link rel="icon" type="image/png" sizes="16x16"
href="../static/assets/img/favicon/favicon-16x16.png">
    <link rel="manifest" href="../static/assets/img/favicon/site.webmanifest">
    <link rel="mask-icon" href="../static/assets/img/favicon/safari-pinned-tab.svg" color="#ffffff">
    <meta name="msapplication-TileColor" content="#ffffff">
    <meta name="theme-color" content="#ffffff">

    <!-- Fontawesome -->
    <link type="text/css" href="../static/vendor/@fontawesome/fontawesome-free/css/all.min.css" rel="stylesheet">

    <!-- Pixel CSS -->
    <link type="text/css" href="../static/css/neumorphism.css"
rel="stylesheet">

</head>

<body>
    <header class="header-global">
        <nav id="navbar-main" aria-label="Primary navigation" class="navbar
navbar-main navbar-expand-lg navbar-theme-primary headroom navbar-light">

```

```

<div class="container position-relative">
    <a class="navbar-brand shadow-soft py-2 px-3 rounded border
border-light mr-lg-4" href="/">
        
        
    </a>
    <div class="navbar-collapse collapse" id="navbar_global">
        <div class="navbar-collapse-header">
            <div class="row">
                <div class="col-6 collapse-brand">
                    <a href="/" class="navbar-brand shadow-soft py-
2 px-3 rounded border border-light">
                        
                    </a>
                </div>
                <div class="col-6 collapse-close">
                    <a href="#navbar_global" class="fas fa-times"
data-toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" title="close" aria-
label="Toggle navigation"></a>
                </div>
            </div>
        </div>
    </div>
    <div class="d-flex align-items-center">
        <a href="/" class="btn btn-primary text-secondary d-none d-
md-inline-block mr-3"><i class="fa fa-home"></i> Home</a>
        <button class="navbar-toggler ml-2" type="button" data-
toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" aria-label="Toggle
navigation">
            <span class="navbar-toggler-icon"></span>
        </button>
    </div>
</div>
</nav>

```

```

</header>

<main>

  <!-- Section -->

  <section class="section section-lg">

    <div class="container">

      <div class="row align-items-center justify-content-center">

        <div class="col-md-12 col-lg-12 mb-5">

          <!-- Contact Card -->

          <div class="card bg-primary shadow-soft border-
light p-2 p-md-3 p-lg-5">

            <div class="card-header">

              <div class="row justify-content-center">

                <div class="col-12">

                  <h1 class="display-2 mb-3">We are
here to help you</h1>

                  <p class="lead">Queries? Contact us
now</p>

                </div>

              </div>

              <div class="card-body shadow-soft text-
center border border-light rounded">

                <p>We are here to deliver better
results for our clients.</p>

                <ul class="list-group mb-4">

                  <li class="list-group-
item"><strong>Unlimited scope</strong></li>

                  <li class="list-group-
item"><strong>Identify critical vulnerabilities, malware, exploits and
misconfigurations</strong></li>

                  <li class="list-group-
item"><strong>Prioritize vulnerabilities remediation</strong></li>

                  <li class="list-group-
item"><strong>Detect suspicious activity and stop attacks</strong></li>

                </ul>

              </div>

            </div>

          </div>

        </div>

      </div>

    </div>

  </div>

```



```

<div class="row mb-5">
    <div class="col-md-4 col-lg-4 text-
center">

    </div>

    <div class="col-md-4 col-lg-4 text-
center">

    </div>
</div>
</div>
<div>
    <form class="col-12 col-md-8 mx-auto"
action="/report" method = "POST" >
        <!-- Form -->
        <div class="form-group">
            <label for="nameInputIcon2">Your
Name</label>

            <div class="input-group mb-4">
                <div class="input-group-prepend">
                    <span class="input-group-
text"><span class="far fa-user-circle"></span></span>
                </div>
                <input class="form-control"
name="name" id="nameInputIcon2" placeholder="e.g. Alice" type="text" aria-
label="contact name input">
            </div>
        </div>
        <!-- Form -->
        <div class="form-group">
            <label for="emailInputIcon2">Your
Email</label>

            <div class="input-group mb-4">
                <div class="input-group-prepend">
                    <span class="input-group-
text"><span class="far fa-envelope"></span></span>
                </div>

```

```

                                <input class="form-control"
name="email" id="emailInputIcon2" placeholder="example@company.com"
type="email" aria-label="contact email input">

                                </div>

                                </div>

                                <!-- Form -->

                                <div class="form-group">

                                    <label
for="exampleFormControlTextarea2">Your Query</label>

                                    <textarea class="form-control"
name="query" placeholder="Type your query here..."
id="exampleFormControlTextarea2" rows="4"></textarea>

                                </div>

                                <div class="form-group text-center">

                                    <button type="submit" class="btn btn-
primary">POST</button>

                                </div>

                                <!-- End of Form -->

                                </form>

                                </div>

                                <!-- End of Contact Card -->

                                </div>

                                </div>

                                </div>

                                </section>

                                <!-- End of section -->

                                </main>

<footer class="d-flex pb-5 pt-6 pt-md-7 border-top border-light bg-
primary">

    <div class="container">

        <div class="row">

            <div class="col">

                <!-- <a href="https://themesberg.com" target="_blank"
class="d-flex justify-content-center">

```



```

<!-- Neumorphism JS -->
<script src="../../static/assets/js/neumorphism.js"></script>
</body>

</html>

```

## alert.html

```

<!DOCTYPE html>
<html lang="en">

<head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
    <!-- Primary Meta Tags -->
    <title>PhisherFind - Report</title>
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="title" content="PhisherFind - Report">
    <meta name="author" content="PhisherFind">

    <!-- Favicon -->
    <link rel="apple-touch-icon" sizes="120x120"
href="../../static/assets/img/favicon/apple-touch-icon.png">

    <link rel="icon" type="image/png" sizes="32x32"
href="../../static/assets/img/favicon/favicon-32x32.png">

    <link rel="icon" type="image/png" sizes="16x16"
href="../../static/assets/img/favicon/favicon-16x16.png">

    <link rel="manifest" href="../../static/assets/img/favicon/site.webmanifest">

    <link rel="mask-icon" href="../../static/assets/img/favicon/safari-pinned-tab.svg" color="#ffffff">

    <meta name="msapplication-TileColor" content="#ffffff">
    <meta name="theme-color" content="#ffffff">

    <!-- Fontawesome -->
    <link type="text/css" href="../../static/vendor/@fontawesome/fontawesome-free/css/all.min.css" rel="stylesheet">

    <!-- Pixel CSS -->

```

```

<link type="text/css" href="../../../static/css/neumorphism.css"
rel="stylesheet">

</head>

<body>

    <header class="header-global">

        <nav id="navbar-main" aria-label="Primary navigation" class="navbar
navbar-main navbar-expand-lg navbar-theme-primary headroom navbar-light">

            <div class="container position-relative">

                <a class="navbar-brand shadow-soft py-2 px-3 rounded border
border-light mr-lg-4" href="/">

                </a>

                <div class="navbar-collapse collapse" id="navbar_global">

                    <div class="navbar-collapse-header">

                        <div class="row">

                            <div class="col-6 collapse-brand">

                                <a href="/" class="navbar-brand shadow-soft py-
2 px-3 rounded border border-light">

                                </a>

                            </div>

                            <div class="col-6 collapse-close">

                                <a href="#navbar_global" class="fas fa-times"
data-toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" title="close" aria-
label="Toggle navigation"></a>

                            </div>

                        </div>

                    </div>

                </div>

            </div>

            <div class="d-flex align-items-center">

                <a href="/" class="btn btn-primary text-secondary d-none d-
md-inline-block mr-3"><i class="fa fa-home"></i> Home</a>

```

```

        <button class="navbar-toggler ml-2" type="button" data-
toggle="collapse" data-target="#navbar_global" aria-
controls="navbar_global" aria-expanded="false" aria-label="Toggle
navigation">

            <span class="navbar-toggler-icon"></span>

        </button>

    </div>

</div>

</nav>
</header>

<main>

    <!-- Section -->

    <section class="section section-lg">

        <div class="container">

            <div class="row align-items-center justify-content-center">

                <div class="col-md-12 col-lg-12 mb-5">

                    <!-- Contact Card -->

                    <div class="card bg-primary shadow-soft border-
light p-2 p-md-3 p-lg-5">

                        <div class="card-header">

                            <div class="row justify-content-center">

                                <div class="col-12">

                                    </div>

                                <div class="col-12 col-md-8 text-center
mb-5">

                                    <h1 class="display-2 mb-3">Sit back
and relax! We will get back to you!!</h1>

                                    <p class="lead">Till then, stay
safe!</p>

                                </div>

                            </div>

                        </div>

                    </div>

                </div>

            <!-- End of Contact Card -->

        </div>

    </div>

</div>

```



```

<script src="../../static/vendor/bootstrap-datepicker/dist/js/bootstrap-
datepicker.min.js"></script>

<script
src="../../static/vendor/waypoints/lib/jquery.waypoints.min.js"></script>

<script      src="../../static/vendor/jarallax/dist/jarallax.min.js"></script>

<script
src="../../static/vendor/jquery.counterup/jquery.counterup.min.js"></script>

<script src="../../static/vendor/jquery-
countdown/dist/jquery.countdown.min.js"></script>

<script src="../../static/vendor/smooth-scroll/dist/smooth-
scroll.polyfills.min.js"></script>

<script src="../../static/vendor/prismjs/prism.js"></script>


<script async defer src="https://buttons.github.io/buttons.js"></script>


<!-- Neumorphism JS -->
<script src="../../static/assets/js/neumorphism.js"></script>
</body>

</html>

```

### 13.2 GitHub & project demo link:

GitHub link: [IBM-EPBL/IBM-Project-38311-1660377839: Web Phishing Detection \(github.com\)](https://github.com/IBM-EPBL/IBM-Project-38311-1660377839)

Demo link: [PhisherFind - Web Phishing Detector - YouTube](https://www.youtube.com/watch?v=PhisherFind)