**MARATHWADA MITRA MANDAL’S**

**COLLEGE OF COMMERCE**

202/A, Deccan Gymkhana, PUNE-411004

**A**

**PROJECT REPORT ON**

***Is it Phishy?***

Submitted in partial fulfillment of the requirements

For the award of the Degree of

**Bachelor of Science (Computer Science)**

**Under guidance of**

**Dr. Pranita Raskar**

Submitted by:

***Prasad Patil***

***Shubham Zende***

***Jay Bahirat***

***Aditya More***

**SAVITRIBAI PHULE** PUNE **UNIVERSITY**

**(2019-2020)**



**MARATHWADA MITRA MANDAL’S COLLEGE**

Bachelor of Business Administration (Computer Application)

302/A, Deccan Gymkhana, PUNE-411004

**CERTIFICATE**

**This is to certify that**

***Prasad Patil***

***Shubham Zende***

***Jay Bahirat***

***Aditya More***

**Have completed the project entitled**

***‘Is it Phishy?***

**in a partial fulfillment for the award of**

**Bachelor of Science (Computer Science)**

**Savitribai Phule Pune University**

**Academic Year 2021-2022**

**Internal Examiner External Examiner**

**Dr. Pranita Raskar Dr. Anpat S.M**

**Project Guide HOD**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **SR.NO** | **PARTICULARS** | **PAGE NO** |
| **1** | **INTRODUCTION** |  |
|  | 1.1 INTRODUCTION TO SYSTEM | 5 |
|  | 1.2 SCOPE OF SYSTEM | 6 |
|  | 1.3 PURPOSE OF SYSTEM | 7 |
| **2** | **SYSTEM ANALYSIS** |  |
|  | 2.1 FEASIBILITY STUDY | 8 |
| **3** | **REQUIREMENT SPECIFICATION** |  |
|  | 3.1 HARDWARE REQUIREMENTS | 9 |
|  | 3.2 SOFTWARE REQUIREMENTS | 9 |
| **4** | **SYSTEM DESIGN** |  |
|  | 4.1 ER DIAGRAM | 10 |
|  | 4.2 CLASS DIAGRAM | 11 |
|  | 4.1 USE CASE DIAGRAM | 12 |
| **5** | **RESULT** | 13 |
| **6** | **LIMITATIONS** | 18 |
| **7** | **CONCLUSION** | 19 |
| **8** | **FUTURE ENHANCEMENT** | 20 |
| **9** | **REFERENCES** | 21 |

**ACKNOWLEDGEMENT**

I express my sincere thanks to sir **Dr Devidas Golhar**, Principal of, MARATHWADA MITRA MANDAL COLLEGEOFCOMMERCE (MMCC).

I pay my deep sense of gratitude to **Dr. Sandip Anpat**, Head of Department, Department of Computer Science, to encourage us to the highest peak and to provide us the opportunity to prepare the project.

I feel to acknowledge my indebtedness and deep sense of gratitude to our guide **Dr. Pranita Raskar**, Faculty of Computer Science, whose valuable guidance and kind supervision given to us throughout the course which shaped the present work as its show.

**INTRODUCTION**

* **Introduction to System**
  + Phishing is a major threat to all Internet users and is difficult to trace or defend against since it does not present itself as obviously malicious in nature. In today's society, everything is put online and the safety of personal credentials is at risk. Phishing can be seen as one of the oldest and easiest ways of stealing information from people and it is used for obtaining a wide range of personal details.
  + In reality, phishing has become a complex and escalating threat to everyone’s Internet security. By gathering even, a small amount of information about a victim, the attacker can produce a personalized and believable email. These phishers are not easy to catch either, as most of them can hide the location of their servers and work in almost complete anonymity. Even a user with excellent security software can fall victim to a phishing attack, because for the most part they depend entirely on information typed into a form, not malware infection of a computer.
  + Phishing involves sending fraudulent emails to a target that appear to come from a creditable source. An individual or several individuals known as phishers or attackers orchestrate the attack. The individuals who are affected by the attack are called victims or targets. The goal of phishing is to gather sensitive data, such as login credentials or bank account details or install malware into the target’s system. Investigating such a complex attack is very challenging to the cybersecurity experts. Phishing attacks can be performed manually but to over- come the attack and to respond effectively to the attack requires a lot of time, intelligence and manpower. This may take days or even weeks to respond and analyze the attack in depth. Manual investigation has lot of dependency on the security analyst’s talents and tools available for investigation.
* **SCOPE OF SYSTEM**
  + Our goal is to make the internet a safer place through collaboration between members of the antivirus industry, researchers and end users of all kinds.
  + We are creating a platform through which many of the people will find it easy to know the website details like:
    - * 1. Whether the website is genuine or not
        2. Is it safe to access the website?
        3. Is it safe to make payments through that website?
        4. Is my storage data being accessed or deleted or encrypted (Ransomware attack) through that website?
        5. Details of website like Registrant Name, Address, Contact information, etc.

**PURPOSE OF SYSTEM**

The purpose of Phishing Domain Detection is detecting phishing domain names. Therefore, passive queries related to the domain name, which we want to classify as phishing or not, provide useful information to us. Some useful Domain-Based Features are given below.

* Its domain name or its IP address in blacklists of well-known reputation services?
* How many days passed since the domain was registered?
* Is the registrant name hidden?

**SYSTEM ANALYSIS**

* **FEASIBILITY STUDY: -**
  + Feasibility study is a crucial introduce the software development process. It enables the developer to possess an assessment of the merchandise being developed. It refers to the feasibility study of the merchandise in terms of outcomes of the merchandise, operational use and technical support required for implementing it.
  + Feasibility study should be performed on the idea of varied criteria and parameters. the varied feasibility studies are:
* Economic Feasibility

It refers to the advantages or outcomes we are deriving from the merchandise as compared to the entire cost we are spending for developing the merchandise. If the advantages are more or less equivalent because the older system, then it is not feasible to develop the merchandise

* Operational Feasibility

It refers to the feasibility of the merchandise to be operational. Some products may go alright at design and implementation but may fail within the real environment. It includes the study of additional human resource required and their technical expertise

* Technical Feasibility

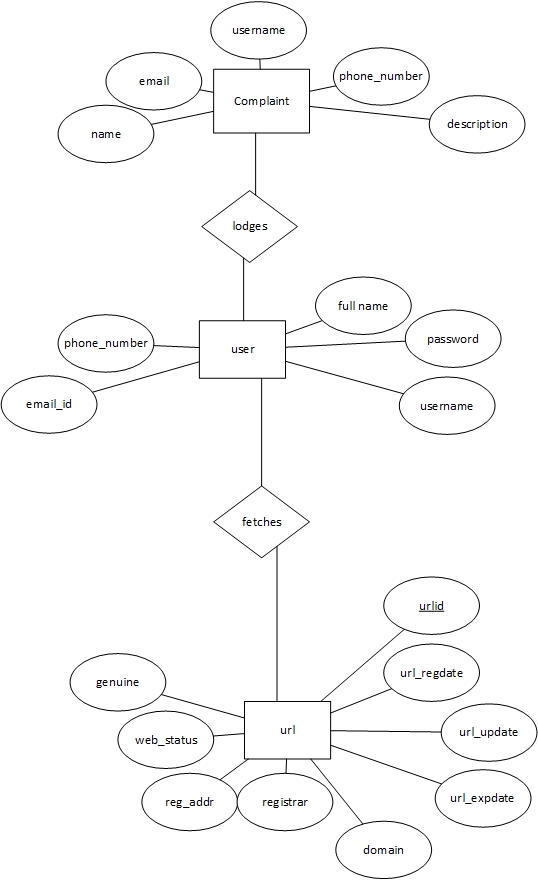
It refers as to if the software that's available within the market fully supports this application. It studies the pros and cons of using particular software for the event and its feasibility. It also studies the extra training needed to tend to the people to form the appliance work.

**REQUIREMENT SPECIFICATION**

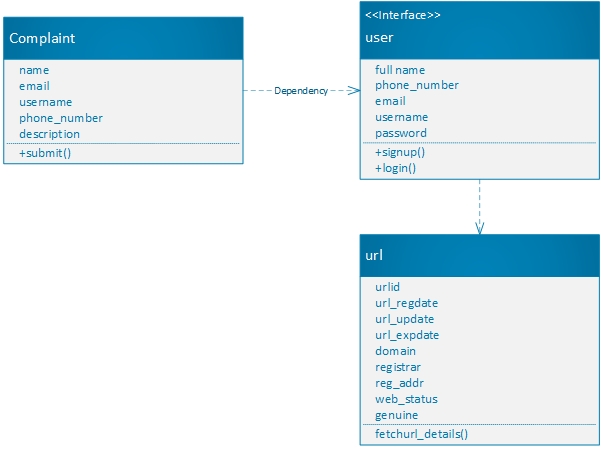
* **HARDWARE REQUIREMENTS**
  + Minimum Intel Pentium Processor
  + 1GB RAM
  + 20GB Storage
* **SOFTWARE REQUIREMENTS**
  + Operating System – Windows / MAC / Linux
  + Browsers Recommended – Google Chrome / Firefox / Safari

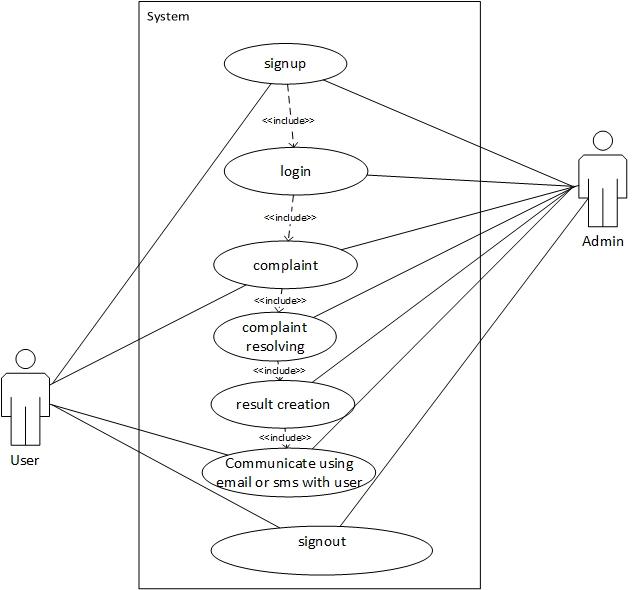
**SYSTEM DESIGN**

**ENTITY RELATIONSHIP –**

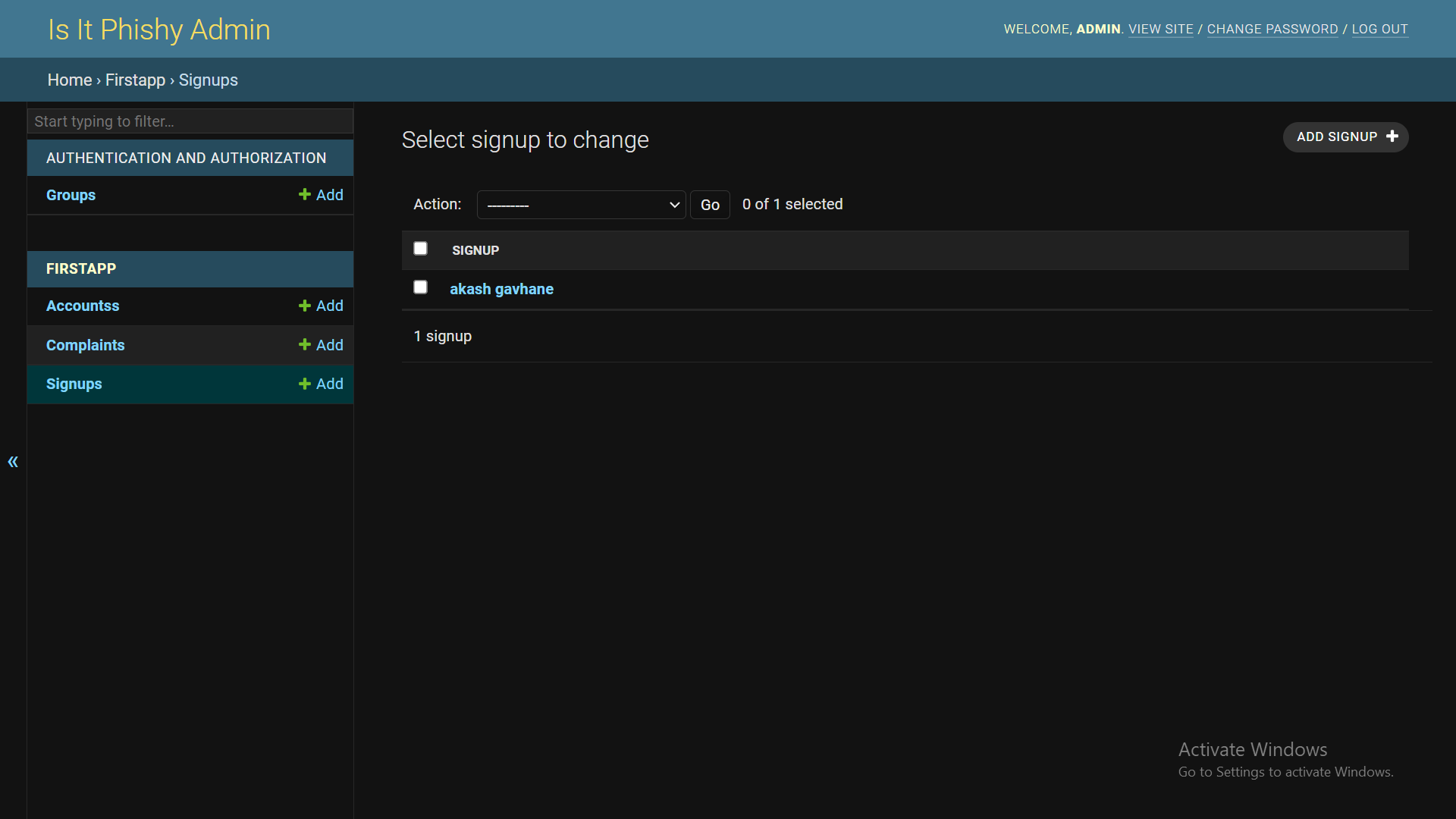
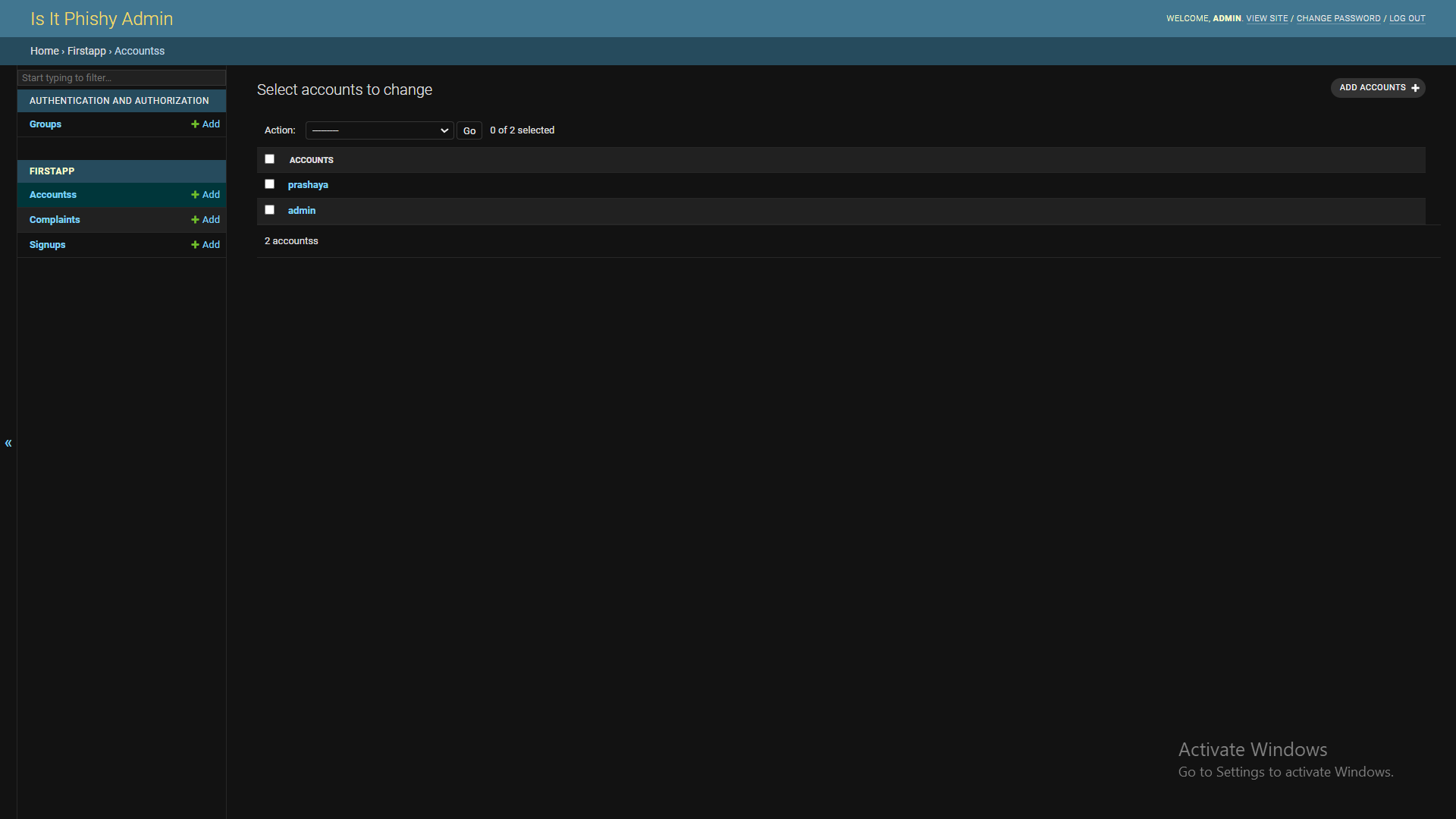
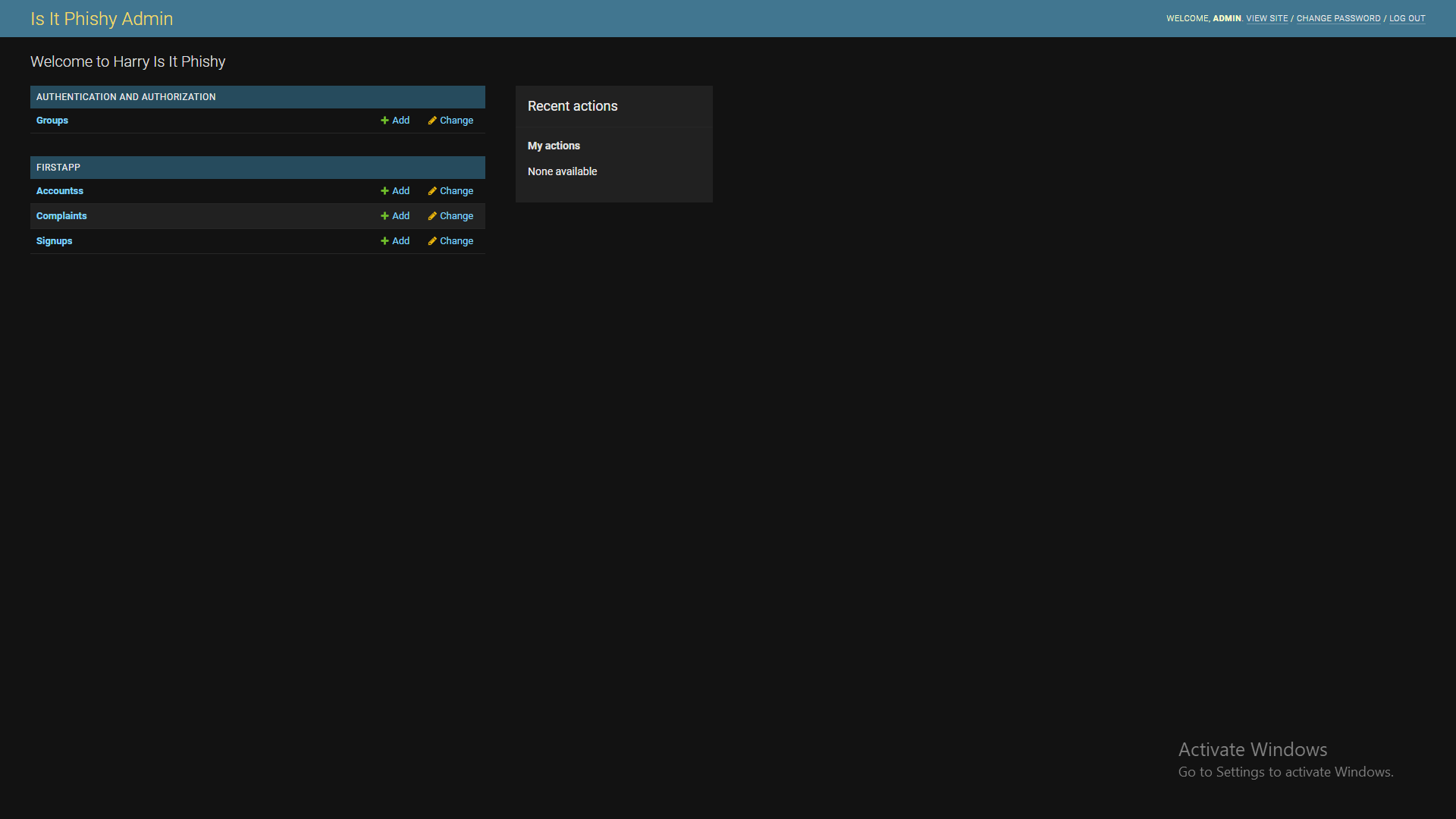
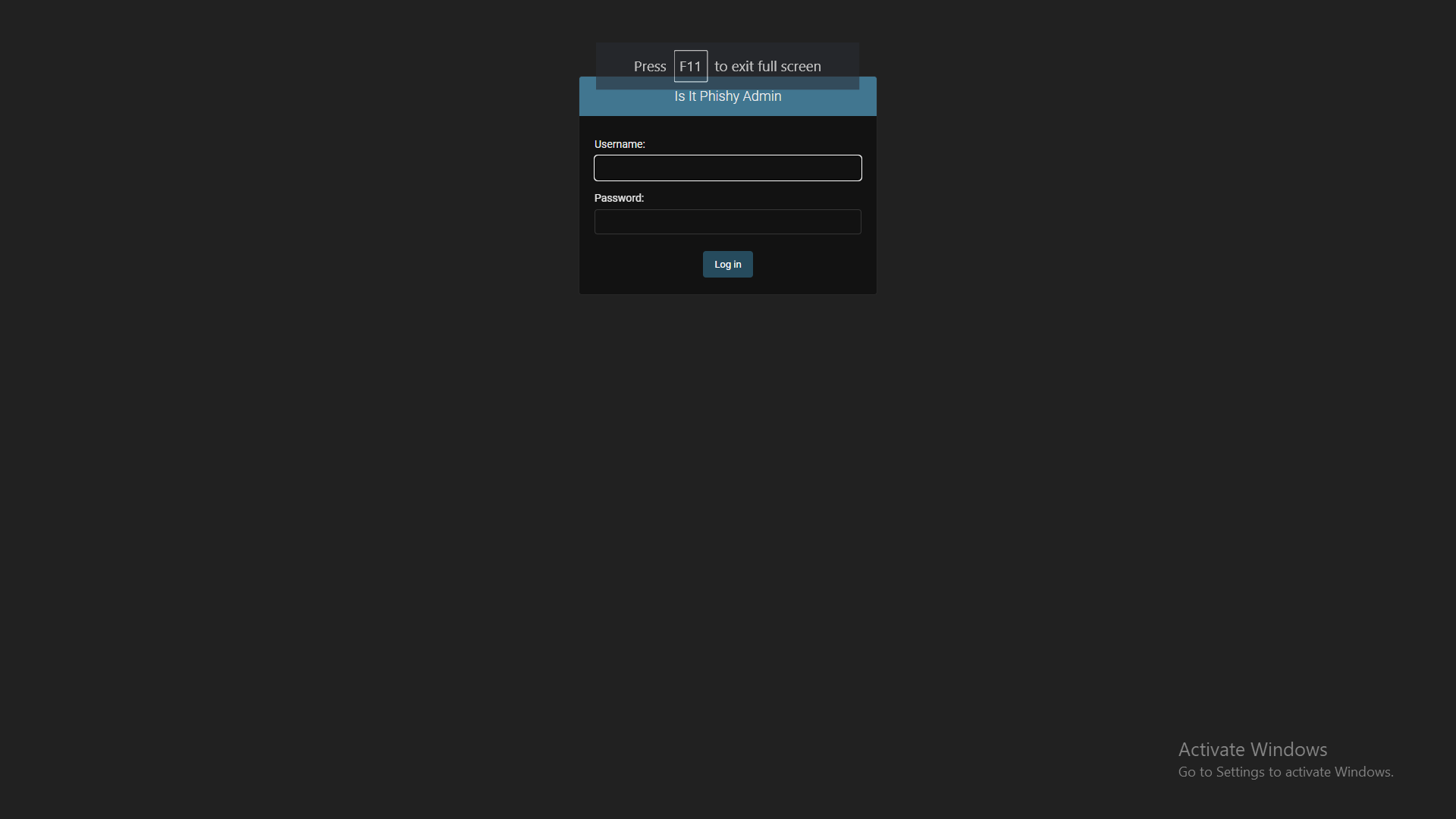
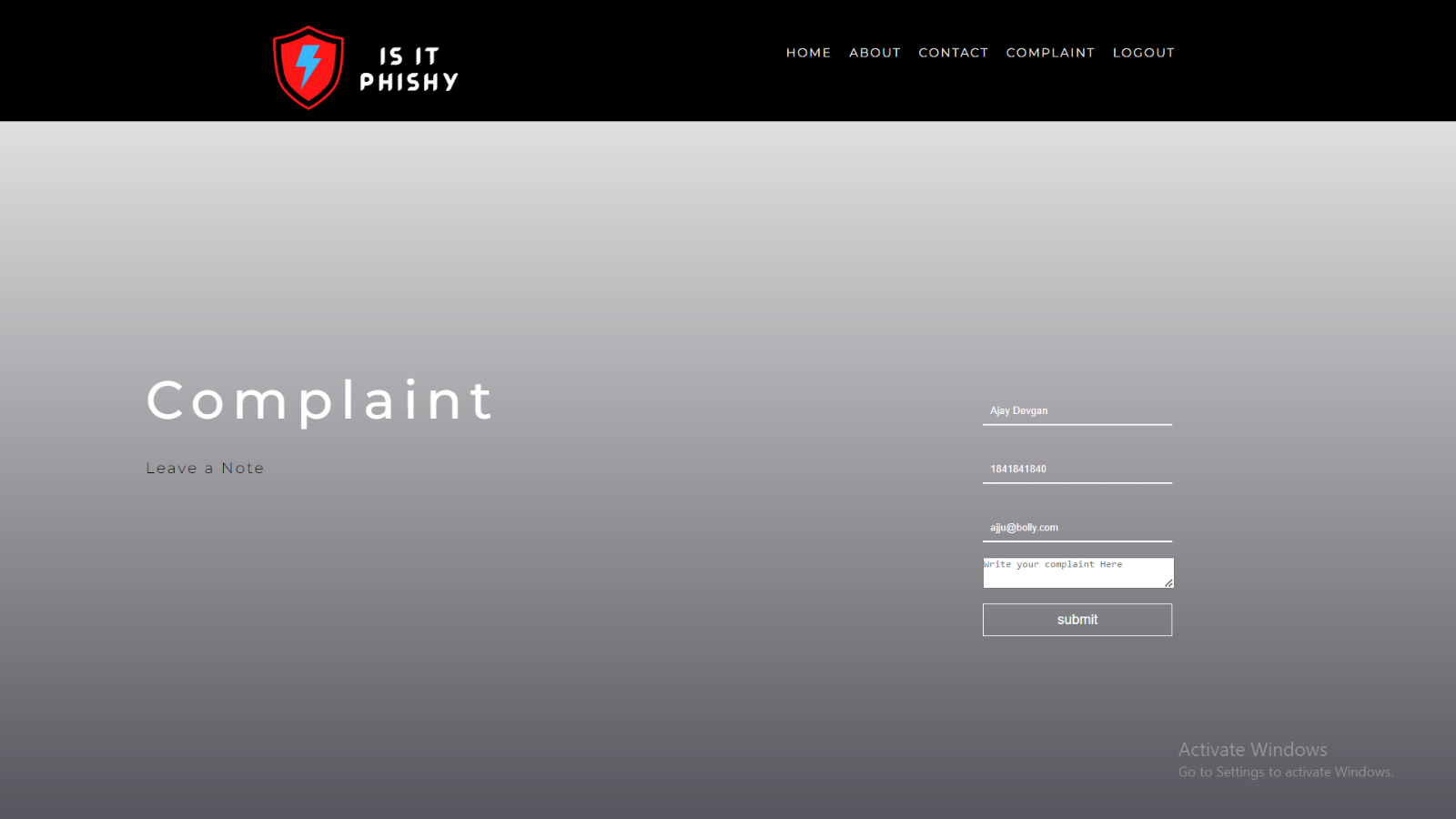
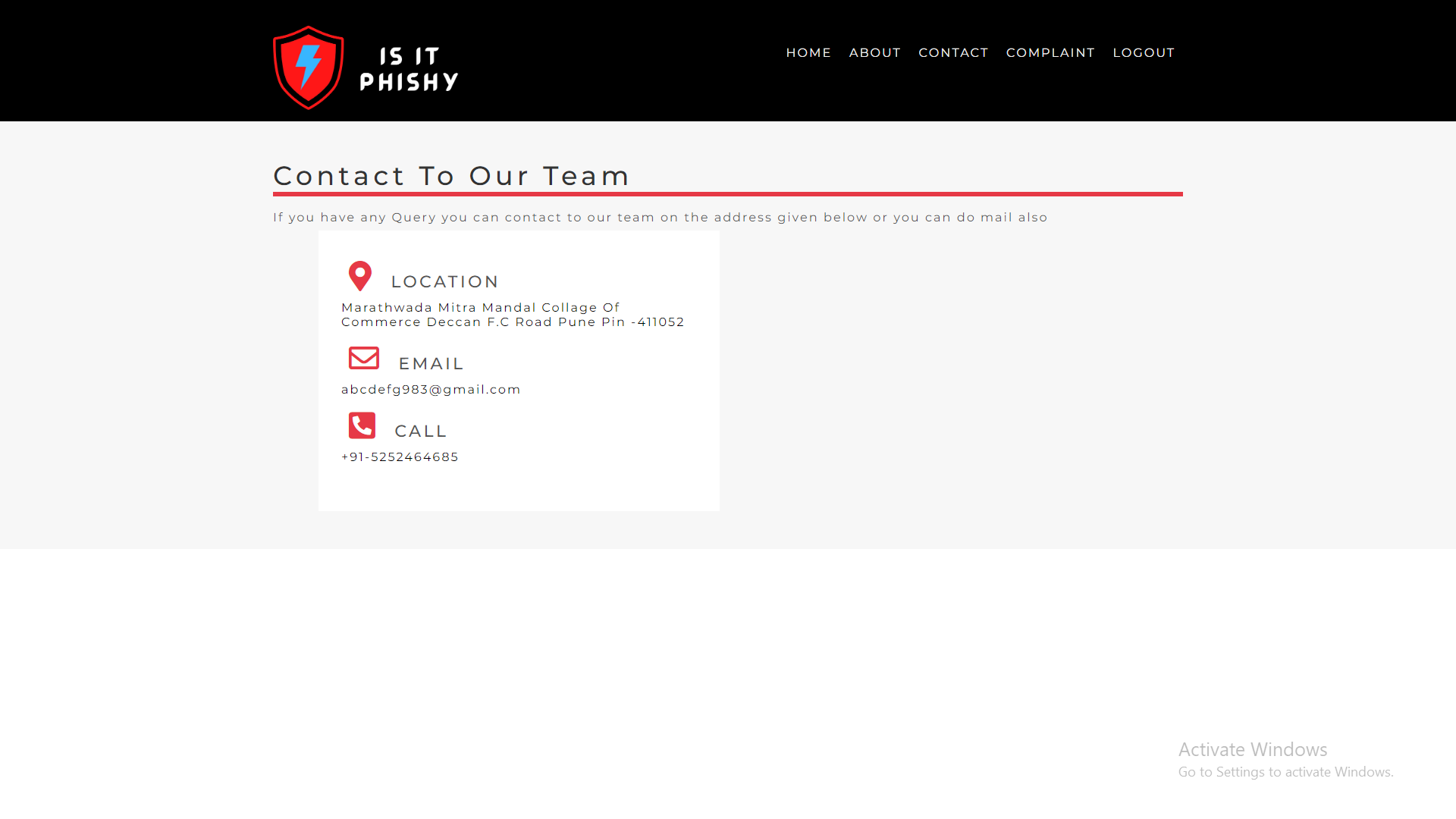
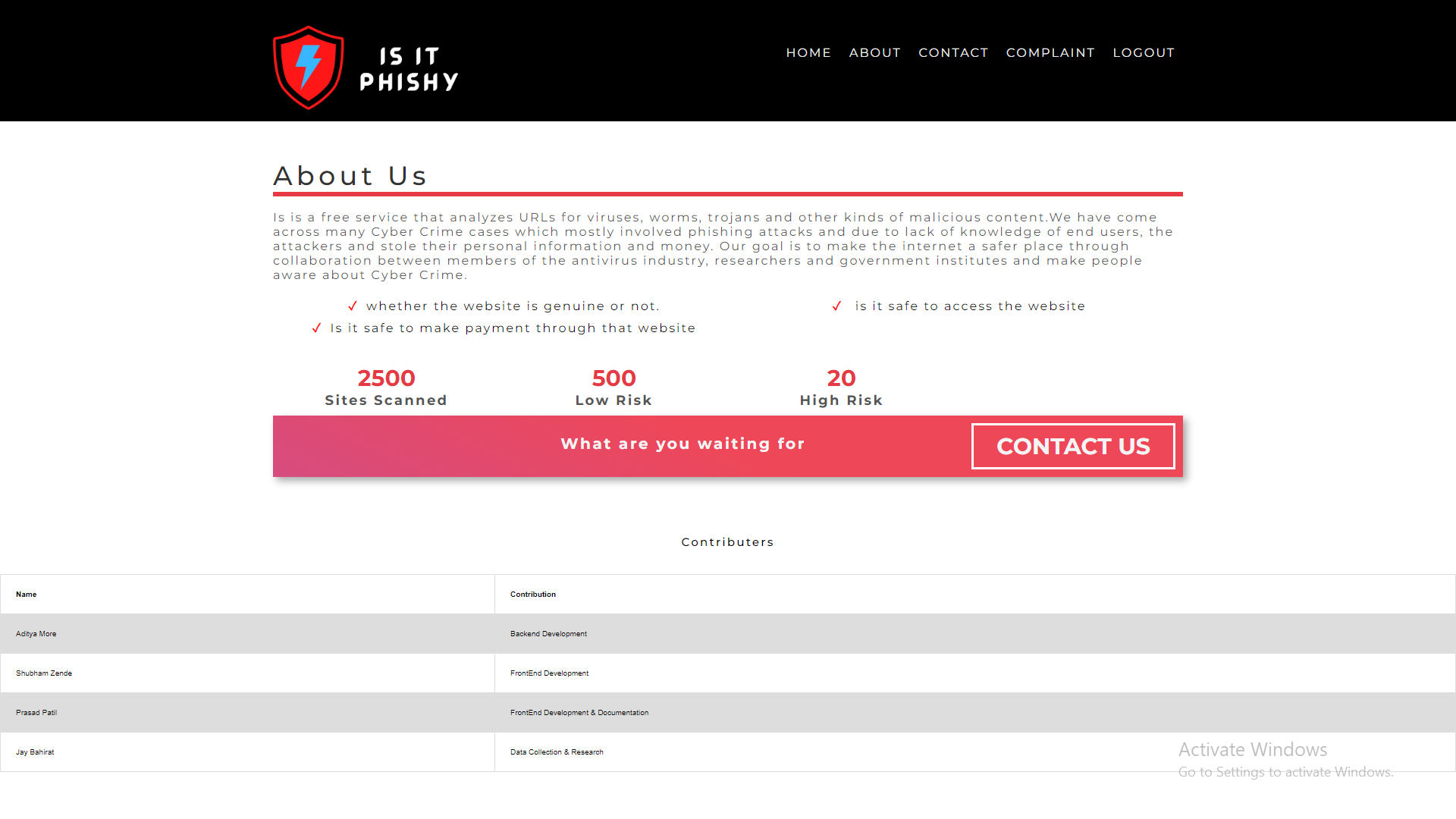
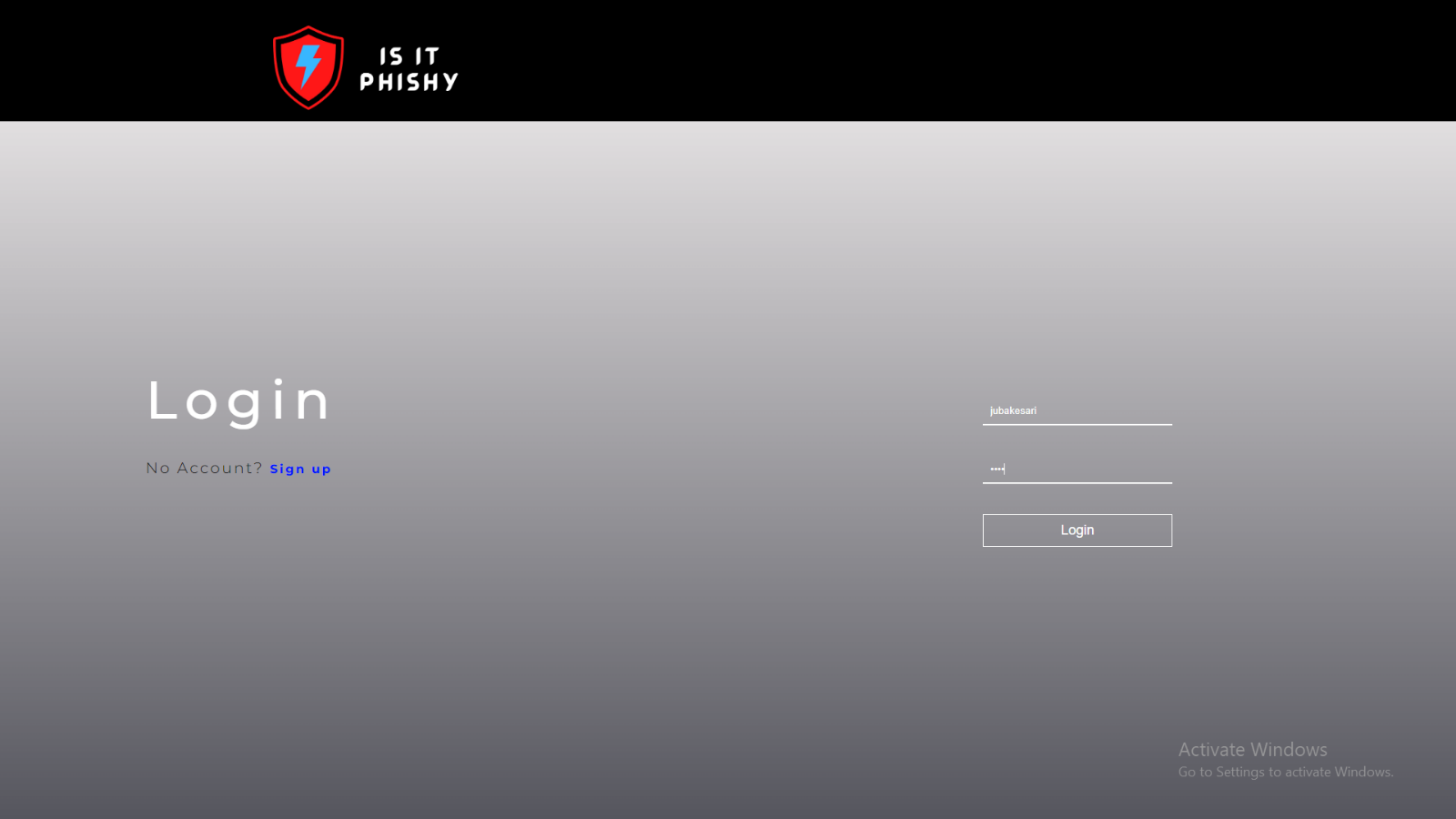
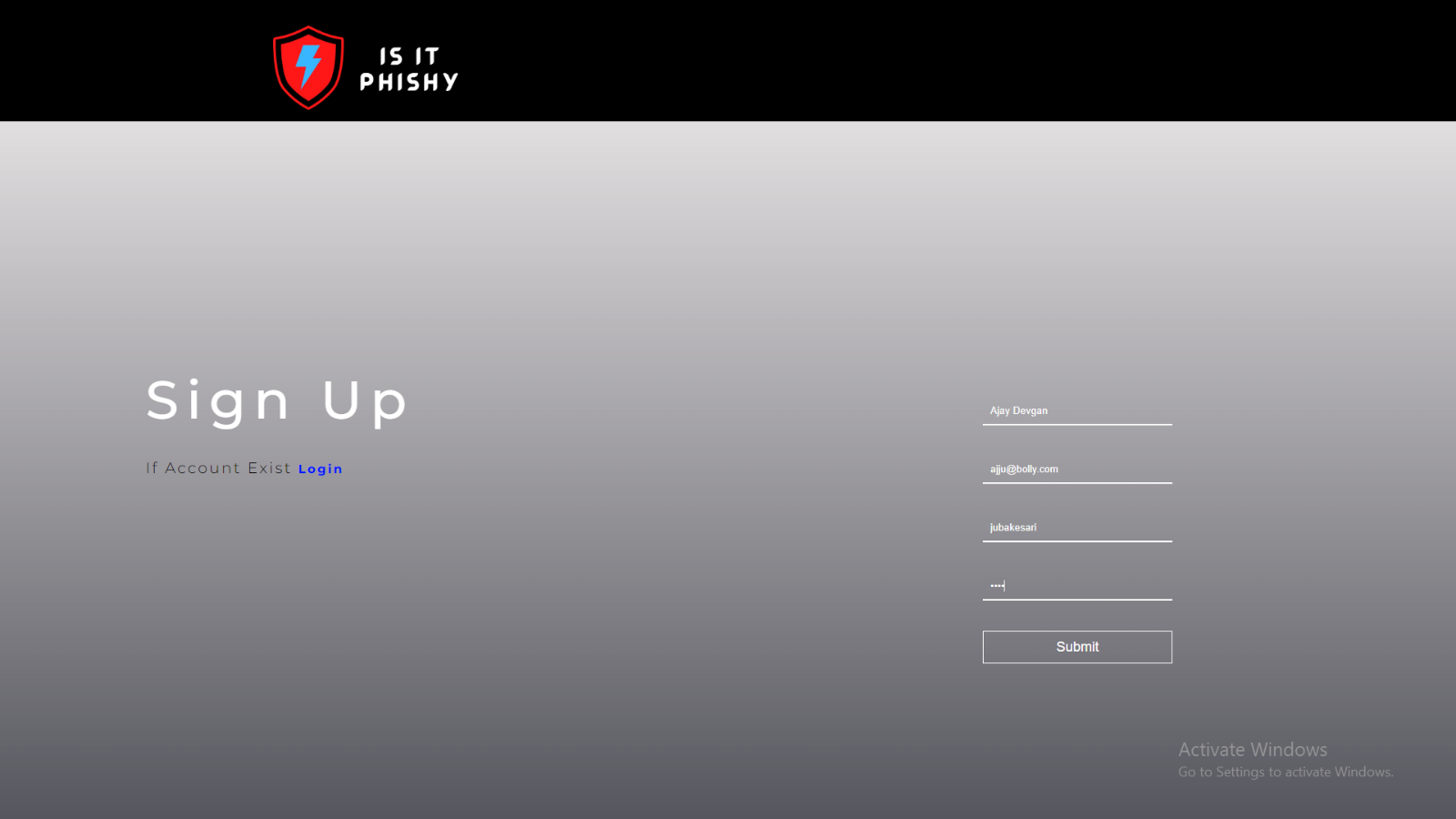
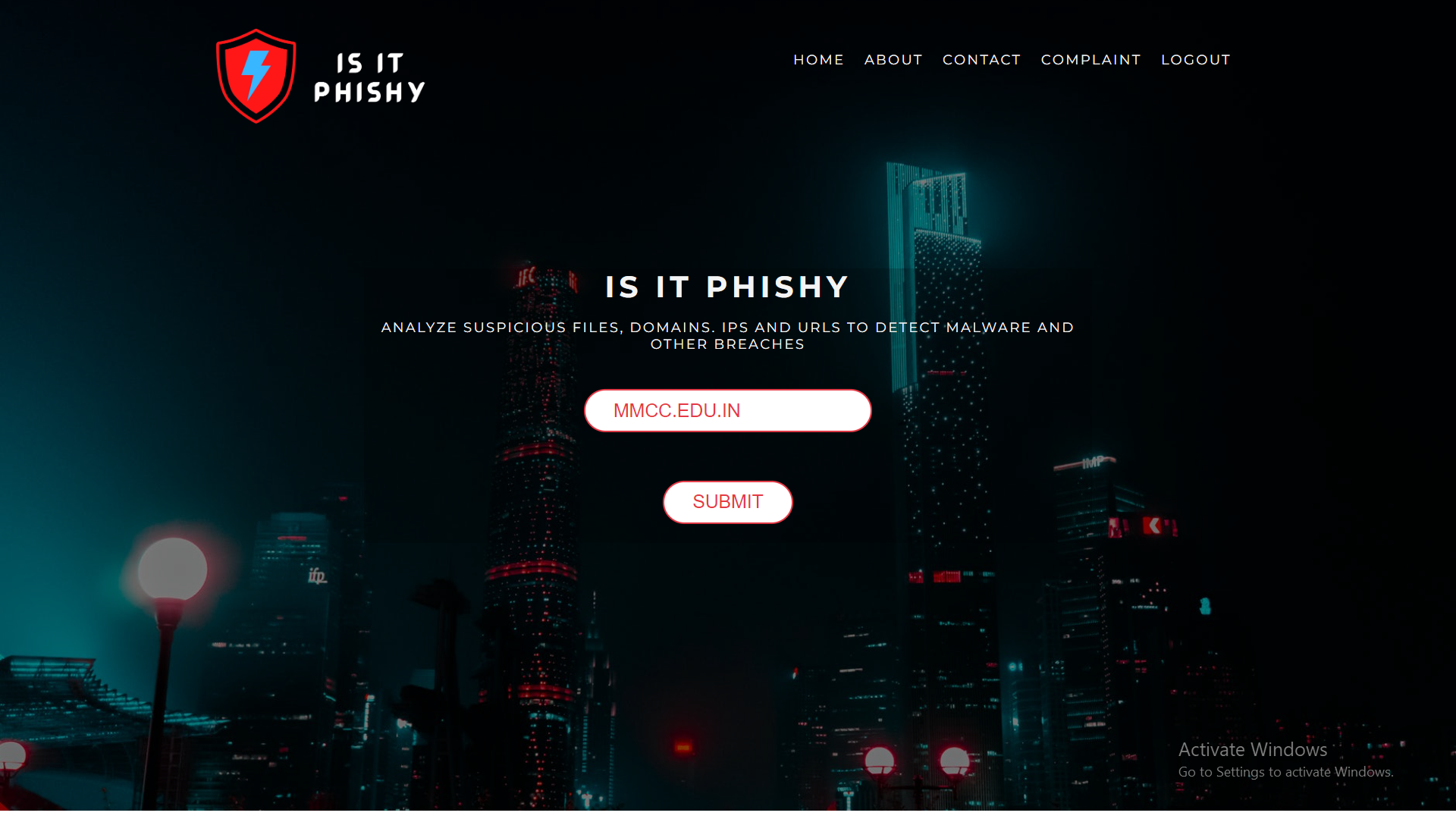


**CLASS DIAGRAM**



**USE CASE -**

**RESULTS**

****

**LIMITATION**

* Our Anti-Phish has the following limitations and challenges. while WHOIS information is available for most of the top-level domains (TLD) such as (com, org, net, biz, info, pl, jp, uk, nz, . . .), but nevertheless some of the TLDs might not be accessible through general WHOIS which might be challenging especially in real-time deployment scenarios. One mitigation approach is to reach out to other regional Domain Name Service (DNS) servers to get information about these domains.
* Some of the collected information from the trusted third-party services might not be organized nor easy to extract what we want autonomously. For example, WHOIS information usually provides the age of the domain, but this piece of information might be presented under many tags such as "Registration Time, Registered Date, Commencement Date, Changed Date, Registered On, Created On, etc.".
* During our experiments, we had to go through a large number of samples to identify those tags. However, we expect there might be more variations of these tags which might be challenging on a real-time basis. One mitigation strategy might be to employ a more sophisticated unsupervised information extraction mechanism over unstructured data to be able to extract the dates from unseen tags.
* Most of the third-parties trusted entities may charge for the usage of collecting their information such as Google Ranking. They might also have imposed a daily limit (e.g., 10,000, 20,000, etc.) on these services. Therefore, one would need to consider these challenges and obtain a custom access package to deploy Anti-Phish in real-time.

**CONCLUSION**

* Phishing attacks remain one of the major threats to individuals and organizations to date. Ashighlighted in the article, this is mainly driven by human involvement in the phishing cycle. Often phishers exploit human vulnerabilities in addition to favoring technological conditions (i.e., technical vulnerabilities). It has been identified that age, gender, internet addiction, user stress, and many other attributes affect the susceptibility to phishing between people. In addition to traditional phishing channels (e.g., email and web), new types of phishing mediums such as voice and SMS phishing are on the increase. Furthermore, the use of social media-based phishing has increased in use in parallel with the growth of social media. Concomitantly, phishing has developed beyond obtaining sensitive information and financial crimes to cyber terrorism, hacktivism, damaging reputations, espionage, and nation-state attacks. Research has been conducted to identify the motivations and techniques and countermeasures to these new crimes, however, there is no single solution for the phishing problem due to the heterogeneous nature of the attack vector. This article has investigated problems presented by phishing and proposed a new anatomy, which describes the complete life cycle of phishing attacks. This anatomy provides a wider outlook for phishing attacks and provides an accurate definition covering end-to-end exclusion and realization of the attack.
* Although human education is the most effective defense for phishing, it is difficult to remove the threat completely due to the sophistication of the attacks and social engineering elements. Although, continual security awareness training is the key to avoid phishing attacks and to reduce its impact, developing efficient anti-phishing techniques that prevent users from being exposed to the attack is an essential step in mitigating these attacks. To this end, this article discussed the importance of developing anti-phishing techniques that detect/block the attack.

**FUTURE ENHANCEMENT**

* In future if we get structured dataset of phishing we can perform phishing detection much faster than any other technique. In future we can use a combination of any other two or more classifier to get maximum accuracy. We also plan to explore various phishing techniques that uses Lexical features, Network based features, Content based features, Webpage based features and HTML and JavaScript features of web pages which can improve the performance of the system. In particular, we extract features from URLs and pass it through the various classifiers.
* **REFERENCES**
  + [**https://www.ijert.org/detection-of-url-based-phishing-attacks-using-machine-learning**](https://www.ijert.org/detection-of-url-based-phishing-attacks-using-machine-learning)
  + [**https://cybersecurity.springeropen.com/track/pdf/10.1186/s42400-020-00059-1.pdf**](https://cybersecurity.springeropen.com/track/pdf/10.1186/s42400-020-00059-1.pdf)
  + [**https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full**](https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full)
  + [**https://arxiv.org/pdf/2204.00985.pdf**](https://arxiv.org/pdf/2204.00985.pdf)