**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**

**Asst.Prof. Dr. Pranita Raskar**

Submitted by:

***Aditya More***

***Prasad Patil***

***Shubham Zende***

***Jay Bahirat***

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

**(2021-2022)**



**MARATHWADA MITRA MANDAL’S COLLEGE**

Bachelor of Science (Computer Science)

302/A, Deccan Gymkhana, PUNE-411004

**CERTIFICATE**

**This is to certify that**

***Aditya More***

***Prasad Patil***

***Shubham Zende***

***Jay Bahirat***

**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**

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

**Project Guide HOD**

**ACKNOWLEDGEMENT**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

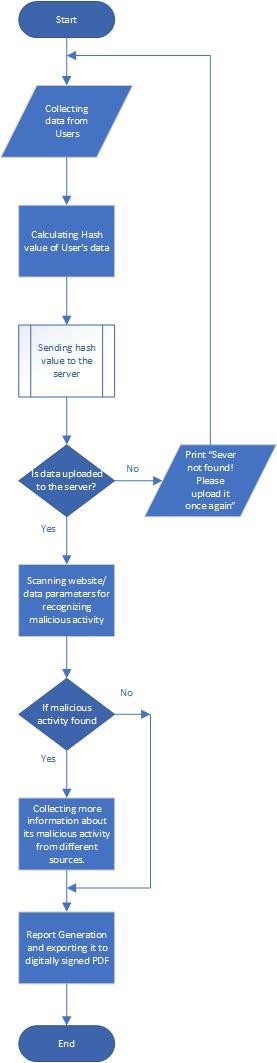
I am highly indebted to Dr. Pranita Raskar madam for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

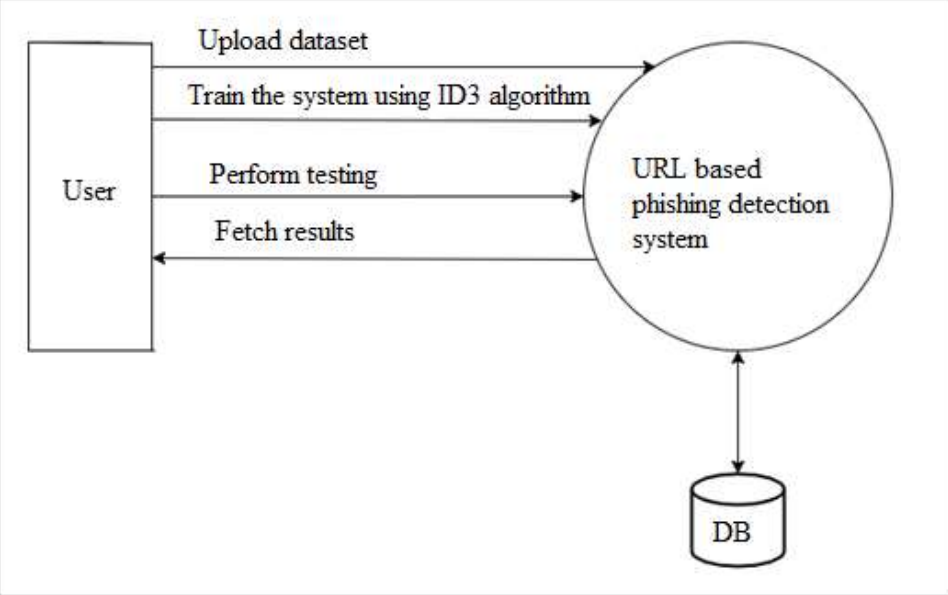
I would like to express my gratitude towards my Team members for their kind co-operation and encouragement which help me in completion of this project.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

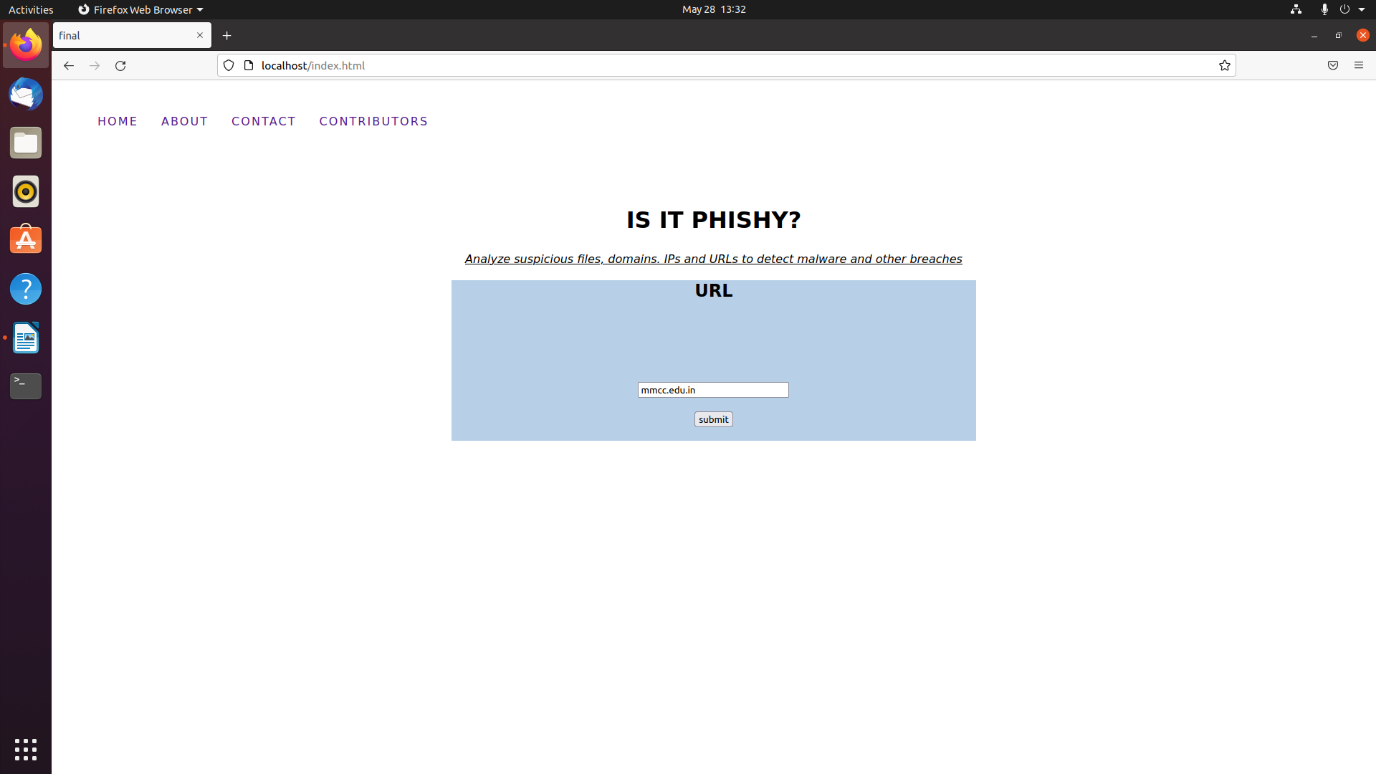
**INDEX**

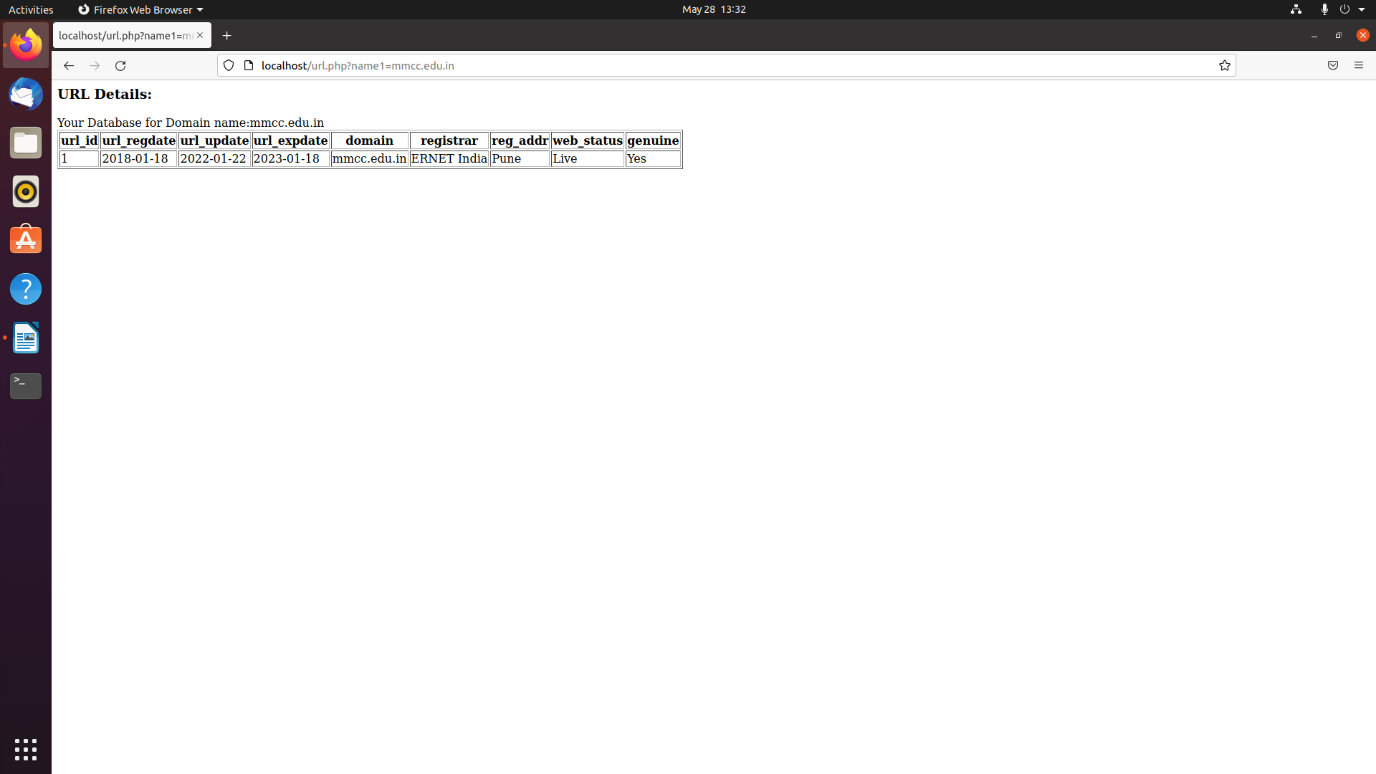
|  |  |  |
| --- | --- | --- |
| **SR.NO** | **PARTICULARS** | **PAGE NO** |
| **1** | **INTRODUCTION** | 5 |
|  | 1.1 INTRODUCTION TO SYSTEM |  |
|  | 1.2 PROBLEM STATEMENT |  |
|  | 1.3 SCOPE |  |
|  | 1.3 PROPOSE SYSTEM |  |
| **2** | **REQUIREMENT SPECIFICATION** | 6 |
|  | 2.1 HARDWARE REQUIREMENTS |  |
|  | 2.2 SOFTWARE REQUIREMENTS |  |
| **3** | **SYSTEM DESIGN** | 7 |
|  | 3.1 FLOW CHART |  |
|  | 3.2 DATA FLOW DIAGRAM |  |
| 4 | **OUTPUT** | 9 |
| 5 | **LIMITATIONS** | 10 |
| 6 | **CONCLUSION** | 11 |
| 7 | **FUTURE ENHANCEMENT** | 12 |
| 8 | **REFERENCES** | 13 |

* **INTRODUCTION**
  + **INTRODUCTION TO SYSTEM**
    - Phishing is one of the most severe cyber-attacks where researchers are interested to find a solution. In phishing, attackers lure end-users and steal their personal information. To minimize the damage caused by phishing, it must be detected as early as possible. There are various phishing attacks like juicing and so on. There are various phishing detection techniques based on a whitelist, black-list, content-based, URL-based, visual similarity, and machine-learning. In this paper, we discuss phishing attacks, and detection for detecting phishing sites. Challenges in phishing detection techniques are also given.
  + **PROBLEM STATEMENT**
    - Phishing attacks use email or fraudulent websites to try to trick you into providing personal or financial information to compromise an account or steal money by posing as a trustworthy entity. They may claim there’s a problem with payment information or that they’ve noticed activity on an account and ask you to click on a link or attachment and provide personal information.
  + **SCOPE**
    - 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.
  + **PROPOSE 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.
* **REQUIREMENT SPECIFICATION**
  + **HARDWARE REQUIREMENTS**
    - Minimum Intel Pentium Processor
    - 1GB RAM
    - 20GB Storage
  + **SOFTWARE REQUIREMENTS**
    - Browser with active internet connection for results.
* **SYSTEM DESIGN**
  + - **Flow Chart -**

**Data Flow Diagram -**

**OUTPUT**

****

****

* **LIMITATIONS**
  + 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.
* **CONCLUSIO****N**
  + 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)