**Mini Project Report on**



**CYBERCRIME DETECTION**

**VIA ML/AI**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

Name University Roll No.

**(Deepanshu Gupta) (2019472)**

***Under the Mentorship of***

Dr. Priya Matta

**Associate Professor, Department of Computer Science and Engineering**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January 2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Cyber Crime Detection Via Machine Learning and Artificial Intelligence”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Priya Matta,** Associate Professor, Department of Computer Science and Engineering,Graphic Era (Deemed to be University), Dehradun.

Name University Roll No. Signature

**(Deepanshu Gupta) (2019472)**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Abstract | 1 |
| Chapter 2 | Introduction | 2-3 |
| Chapter 3 | Literature Survey | 4-5 |
| Chapter 4 | Methodology | 6-8 |
| Chapter 5 | Result and Discussion | 9-10 |
| Chapter 6 | Conclusion and Future Work | 11 |
|  | References | 12-13 |

**Chapter 1**

**Abstract**

With advancements in communication after the invention of the Internet. Cybercrime which can also be termed as a computer crime is any crime that involves networking in the computer over the internet. Cybercriminals are becoming more sophisticated and targeting both the public and private sectors. Therefore, additional layers of protection are required. This paper proposes an approach to phishing detection, one of the cybercrimes covered in this paper, and adds multiple layers in form of two-way authentication. Therefore, it can be utilized for identification and authentication and become a legitimate tool to prevent an individual from getting tricked.

1

**Chapter 2**

**Introduction**

Cybercrime, also known as computer crime, refers to any criminal activity that involves the use of computers or the internet. This can include activities such as hacking, identity theft, cyberstalking, and the spread of malicious software. Cybercrime can have serious consequences for both individuals and organizations, as sensitive information can be stolen or compromised, and systems can be made inoperable. To combat cybercrime, law enforcement agencies and private companies have developed a variety of tools and techniques to detect and prevent these types of activities. Additionally, individuals and organizations need to take steps to protect themselves, such as by using strong passwords and keeping software up to date. Cybercrime is carried out by individuals or organizations. [1]  
Most cybercrime is an attack on information about individuals, corporations, or the government.[2]

**1.1 Cyber Security**

Cyber security means protecting information, equipment, devices, computer, computer resources, communication devices, and information stored therein from unauthorized access, use, disclosure disruption, modification, or destruction.[3]

Crime such as spamming, passing on computer viruses, harassment, cyberstalking, and others have become common in our modern world. While these issues do not carry potential monetary loss, they are just as harmful in the possibility of losing files, information, and access to your computer. This is why Cyber Security is needed.[3]

**1.2 Why Cyber Security?**

Computer security is important because it allows users to protect their information on the network and in the system (right to privacy).[3]

It also helps in defending the computer system against different types of destructive technologies and protects the PC from damage (viruses, worms, bugs, and bacteria).[3]

It also helps monitor the network and protects it from different threats. So, we should use computer security solutions on some level to protect our data from different types of sniffing stolen problems.[3]

In general, Computer Security is vital for protecting the confidentiality, integrity, and availability of computer systems, resources, and data.[3]

Without confidentiality, trade secrets or personally identifying information can be lost. Without integrity, we cannot be sure that the data we have is the same data that was initially sent (i.e., Altered data).[3]

Without availability, we may be denied access to computing resources (i.e., A virus that disables your keyboard and mouse).[3]

**1.3 Introduction to Phishing**

Phishing can be defined as impersonating a valid site to trick users by stealing their data comprising usernames, passwords, account numbers, national insurance numbers, etc. [4] There are countless domains where phishing attacks can occur, online payment sector, webmail, financial institution, file hosting or cloud storage, and many others.[4]

The most vital objective of the proposed project is to verify the website's validity by capturing blacklisted URLs. The user can be notified if the blacklisted website is being accessed. The system involves features like capturing blacklisted websites, viewing the blacklisted website, displaying pop-up notifications, and also displaying email notifications.

**Chapter 3**

**Literature Survey**

**2.1 Related Work**

In emerging technology industry which deeply influences today’s security problems has given a non-ease of mind to some employers and home users. Cyber security employees are currently searching for trustworthy and steady detection techniques for phishing websites detection.[5] Phishing has become one of the highest 3 most current forms of law-breaking in line with recent reports, and both frequencies of events and user susceptibleness have enlarged in recent years, more combining the danger of economic damage. [6]

Phishing is often done by email spoofing or texting, and it typically guides the user to enter points of interest at a fake website that looks and feels the same. The primary key feature is to allow the user to inquire whether visited websites are original or fake. This paper proposes a security tool called Detecting Phishing Website Using Machine Learning.

**2.2 Literature Review**

The current situation that is the majority of the population has been fooled into giving their details to hackers without noticing it. According to the Anti-Phishing Working Group (APWG), there are at least 47, 324 phishing attacks, and a top-ten American bank estimates that at least US$300 is lost for every hour that a phishing site remains up. [7] Machine learning is the science of obtaining computers to act while not being expressly programmed. [8] Machine Learning was implemented to develop this proposed system. Machine learning techniques identify phishing URLs and typically assess a URL based on some feature or set of features extracted from it. [9]

**Phishtank** was proposed to inspect once a link has been posted on the section given. This allows users to keep on track of faked websites. They can copy and paste the link to identify whether the site they will access is safe or not. A search engine displayed on the Phishtank website is to be used as the first method. Using its API will be the second method. The limitation of this project is there was no facility for displaying pop-ups and email notifications once the user had accessed the blacklisted website. [10]

**PhishZoo** was proposed to evaluate a new method for web phishing detection based on profiles of complex sites’ appearance and content. These profiles are kept in a local folder and are either synchronized against the newly loaded sites at the time of loading or against risky sites, for instance, links in email offline. The limitation of this project is there was no facility for displaying pop-ups and email notifications once the user had accessed the blacklisted website. [11]

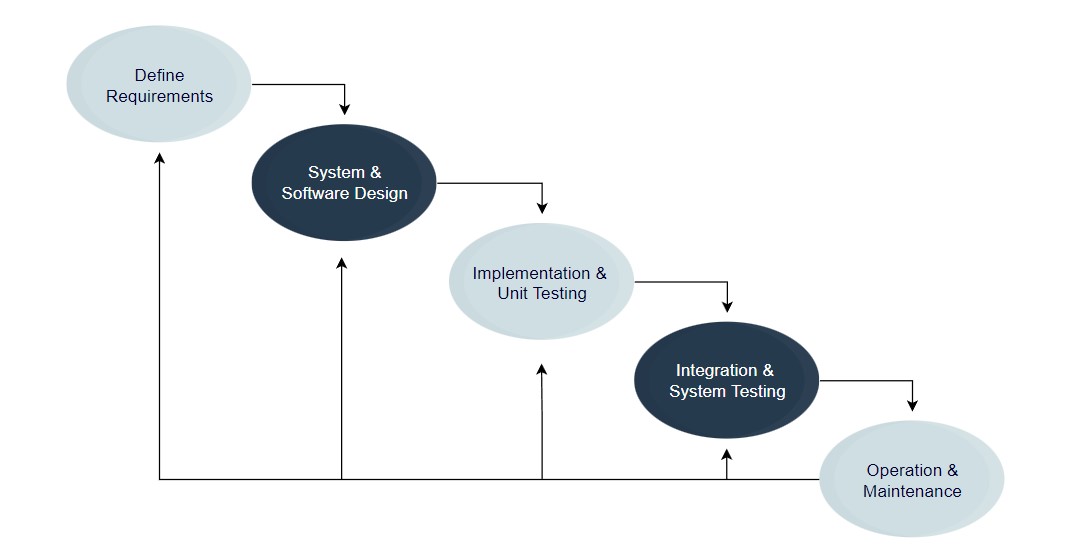
**GoldPhish** was proposed to perceive and report phishing sites. This was done by using optical character recognition (OCR) to recite the text from an image of the page precisely from the company logo, grasping the top hierarchical areas from a search engine, and comparing them with the current website. The limitation of this project is there was no facility for displaying pop-ups and email notifications once the user had access to the blacklisted website. [12]

The capability of AI to avoid cyber-attacks is promoted through AI reasoning. The knowledge offered to AI systems has been found to have significant results in avoiding attacks. The development of Machine learning technologies has also promoted Cyber security. [13] This is achieved from a system learning in real-time, ensuring the user's safety. AI in Cyber security has automated features that act at a fast response time, ensuring safety. The systems are therefore able to detect key threats and create ways through which these attacks could be avoided. The response time promotes this type of security compared to humans. [14]

**Chapter 4**

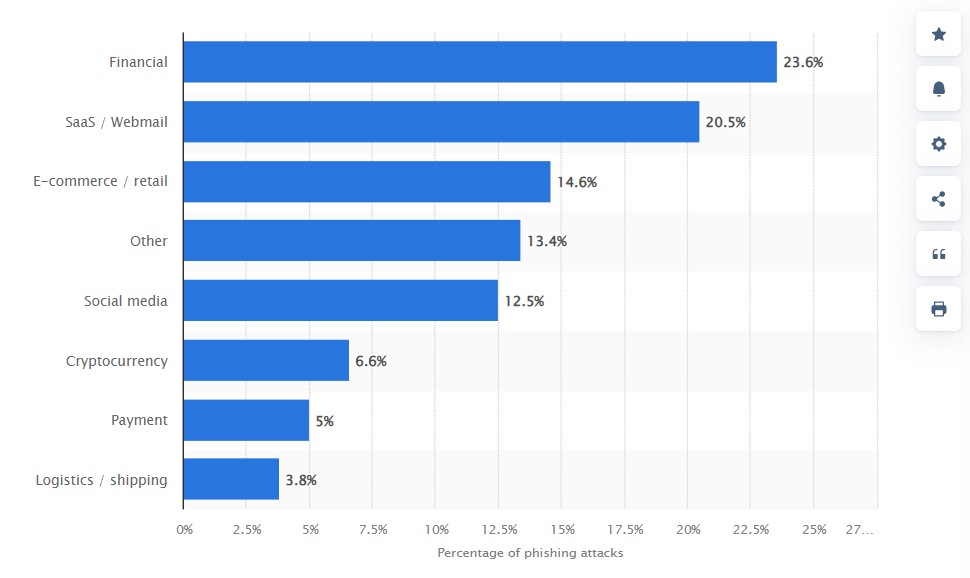
**Methodology**

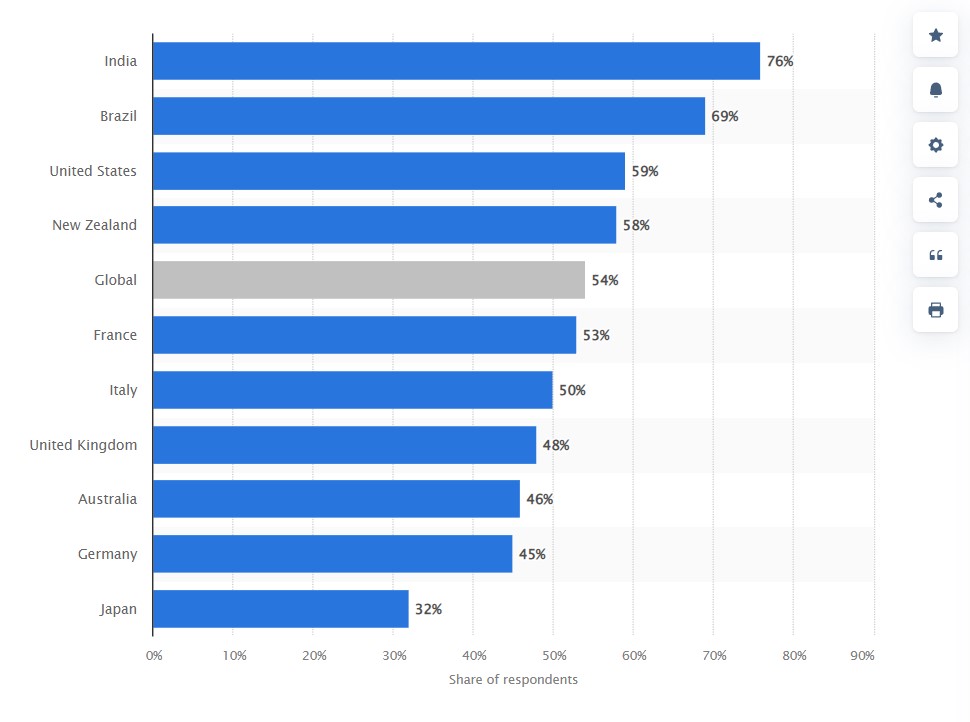
1. A small-scale survey was done to see which cybercrime has been experienced the most in the region.
2. The processing model was decided. (Figure 3.1)



**Figure 4.1[16] Processing Model**

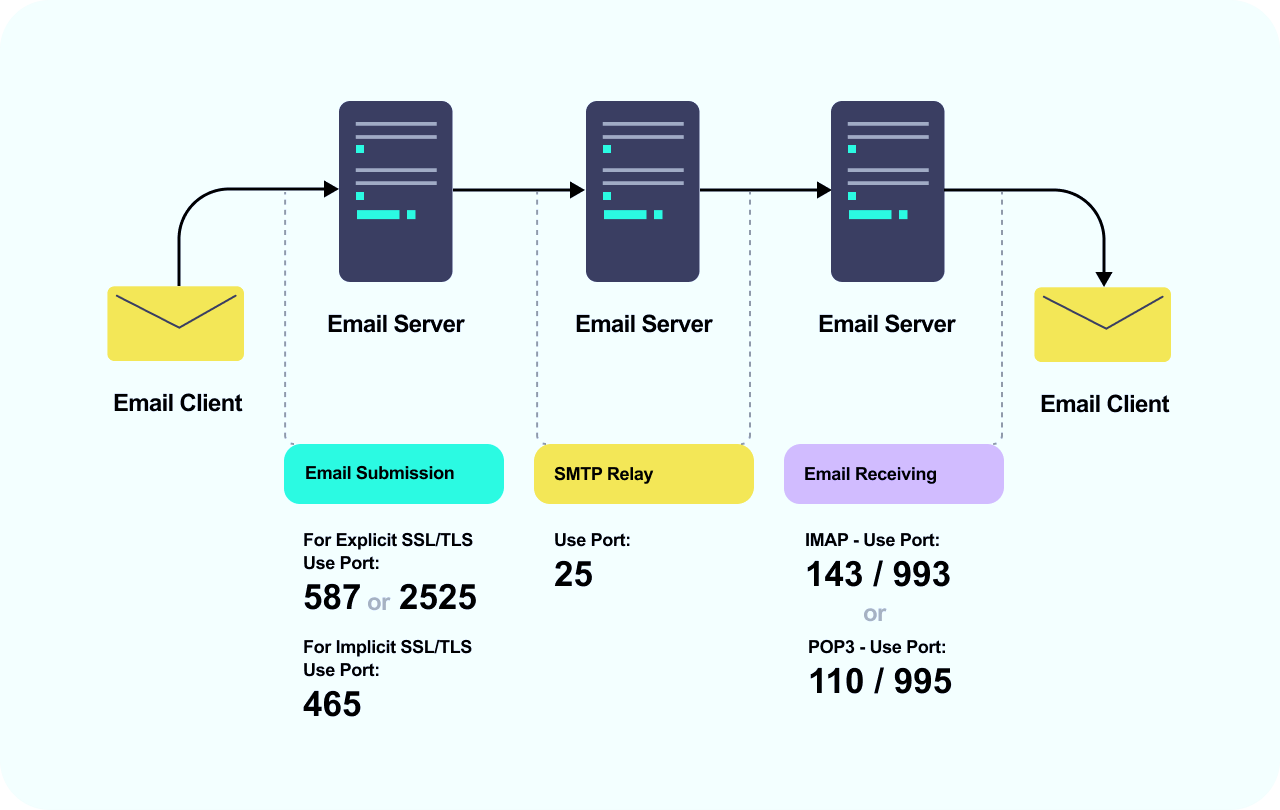
1. Graphical Data was researched on phishing for example- Ways (Methods) to do phishing, Number of cases registered all over the world till 2020. (Figures 3.2 & 3.3)
2. A Web model was developed based on the statics shown and the results got through the survey done earlier.

  
**Figure 4.2[15] Ways (Methods) to do phishing**



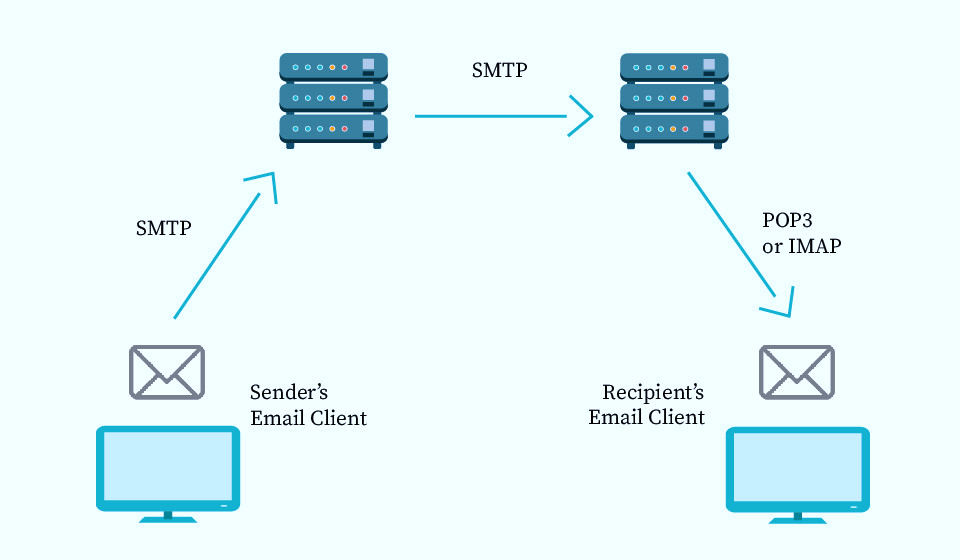
**Figure 4.3[15] Number of cases registered all over the world till 2020**

1. SMTP is the industry standard protocol for sending email. If you’re looking to send email, then you’ll use SMTP instead of IMAP. An SMTP relay service can help you send email without having to build your own SMTP server.
2. IMAP is one of the most common protocols for receiving email. IMAP syncs messages across all devices.
3. POP3 is another protocol for receiving email on a single device. Using POP3 means that your email will be accessible offline and deleted from the server.



**Figure 4.4 Relation between SMTP, IMAP and POP3**

1. The SMTP client will connect to the SMTP server. The email is transferred using that connection. The client and the server terminate the connection. When a user checks for a new email, the client makes a connection to the POP3 or IMAP server. The email client then provides the server with its username and password for authentication.

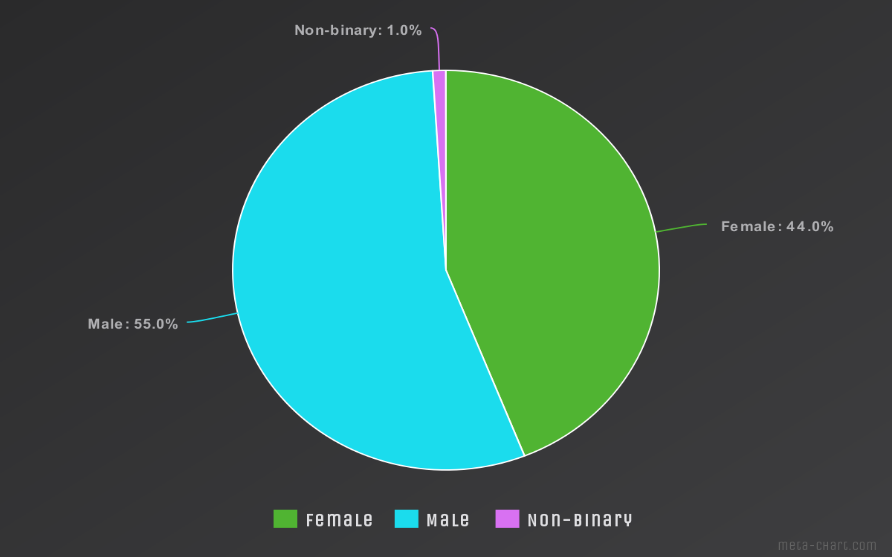
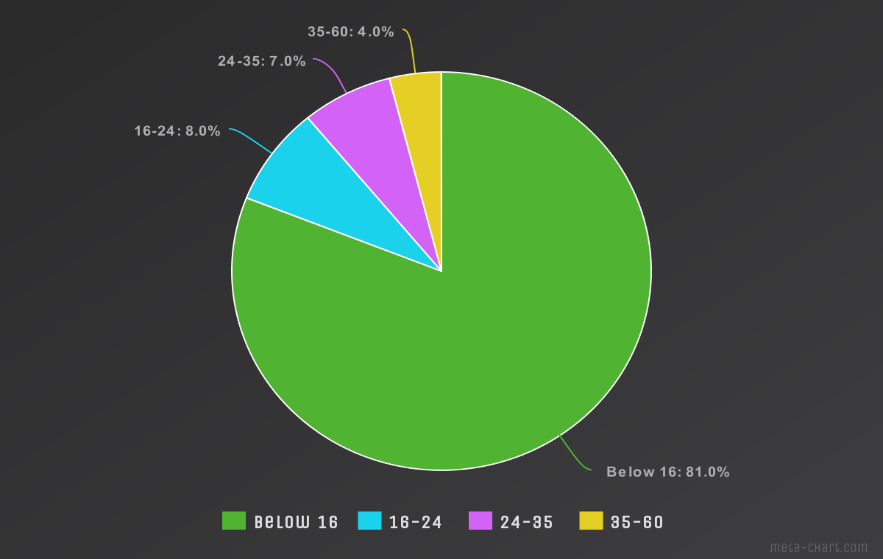


**Figure 4.5 Working Processor of SMTP, IMAP, and POP3**

**Chapter 5**

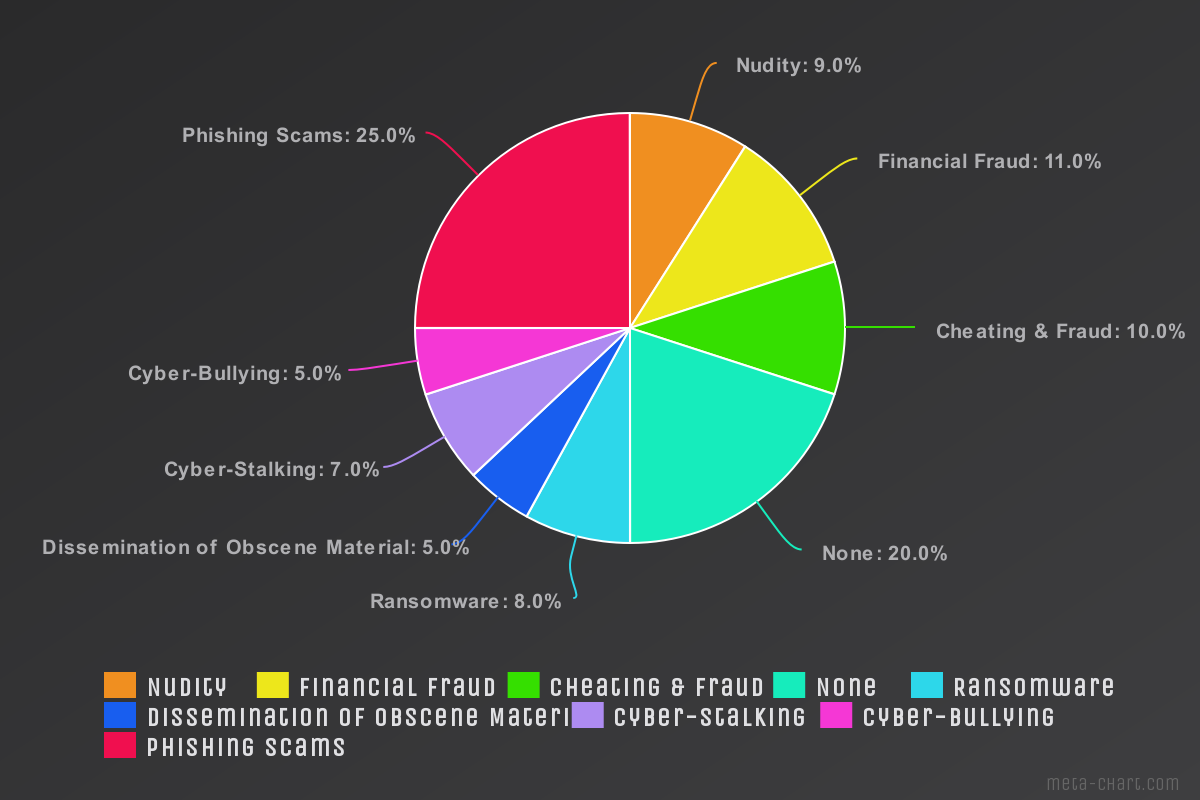
**Result and Discussion**

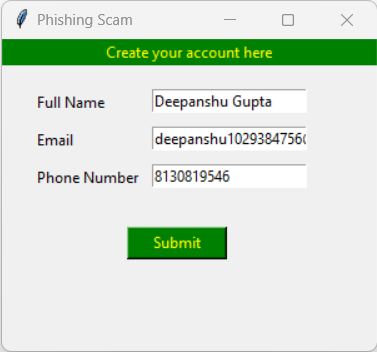
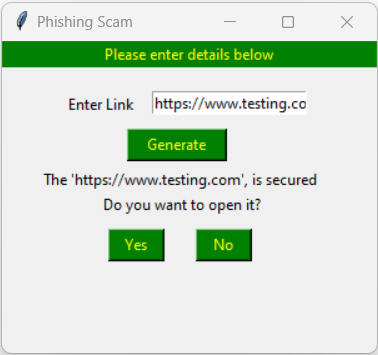
The results from the small-scale survey.

** **

**Figure 5.1 Gender Distribution Chart**

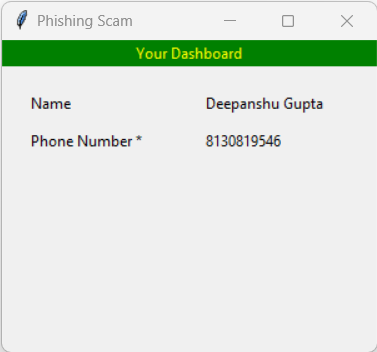
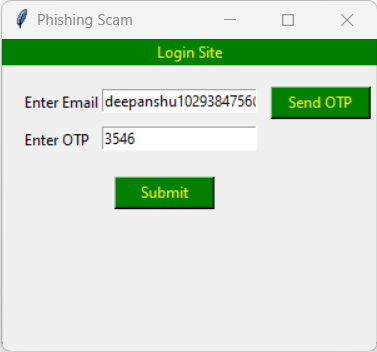
**Figure 5.2 Age Distribution Chart**

**Figure 5.3 Cyber-crimes Experienced Chart**



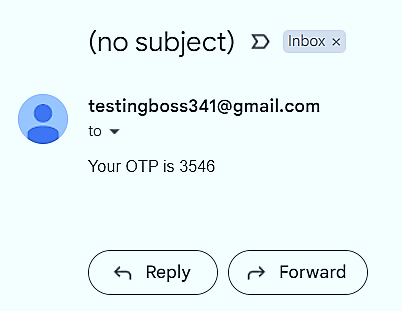
**Figure 5.4 Sample User Interface to check URL (Step 1)**

**Figure 5.5 Sample User Interface to create account (Step 2)**



**Figure 5.6 Sample User Interface for 2-steps Verification (Step 3)**

**Figure 5.7 Sample User Interface to login account (Step 4)**



**Figure 5.8 OTP received on Email Address**

With this data and statistical data collected altogether, decided to take phishing as the main concern in this whole report. Cybercrimes like cyber-stalking, cyber-bullying, cheating, all kinds of fraud, ransomware, etc. are connected to phishing in direct-indirect ways.

**Chapter 6**

**Conclusion and Future Work**

**Conclusion**

With all the data and research, it won’t be an issue to say that phishing is one crime that has a wide area and methods to be implemented and cover most parts of cybercrime in the country.

Our work gives you the method to fight against phishing websites spread all over the internet and also the two-way authentication method to prevent using important words and numbers that can be used against us in any way possible.

**Future work**

The Model is limited to only one cybercrime and also it is based on the web application, in the future, wants to cover most of the cybercrimes with one solution and wants to make it an inbuilt function that can be used as pop-ups whenever it senses any malicious activity.

**References**

[1] Mariam M.H.Alansari, Zainab Aljazzaf, Muhammad Sarfraz: “On Cybercrime and cyber security”.

[2] Wadha Abdullah Al-Khater, Somaya Al-maadead Abdulghani Ali Ahmed, Muhammad Khurram Khan: “Comprehensive Review of Cyber Crime Detection Techniques”, August 5, 2020.

[3] Ms. M Lakshmi Prasanthi, Tata A S K Ishwarya: “Cyber Crime: Prevention & Detection”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, March 2015.

[4]Mohammed Hazim Alkawaz, Stephanie Joanne Steven, Asif Iqbal Hajamydeen: “Detecting Phishing Website Using Machine Learning”, 2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020), 28-29 Feb. 2020, Langkawi, Malaysia.

[5] Rishikesh Mahajan (2018) “Phishing Website Detection using Machine Learning Algorithms”

[6] David G. Dobolyi, Ahmed Abbasi (2016) “PhishMonger: A Free and Open Source Public Archive of Real-World Phishing Websites”

[7] Jalil Nourmohammadi Khiarak (2017) “What is Machine Learning”

[8] Sadia Afroz, Rachel Greenstadt (2018) “PhishZoo: An Automated Web Phishing Detection Approach Based on Profiling and Fuzzy Matching”

[9] Arun Kulkarni, Leonard L. Brown (2019) “Phishing Websites Detection using Machine Learning”

[10] Rohan Saraf, Mayur Khatri, Mona Mulchandani (2014) “Phish Tank-A Phishing Detection Tool”

[11] Sadia Afroz, Rachel Greenstadt (2017) “PhishZoo: Detecting Phishing Websites By Looking at Them”

[12] Matthew Dunlop, Stephen Groat, David Shelly (2010) " GoldPhish: Using Images for Content-Based Phishing Analysis”

[13] D.Glăvan, "Detection of phishing attacks using the anti-phishing framework", Scientific Bulletin of Naval Academy, vol., no. 1, pp. 208-212, 2020. Available: 10.21279/1454- 864x-20-i1-028.

[14] Ansari, Meraj Farheen; Sharma, Pawan Kumar; and Dash, Bibhu (2022) "Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training," International Journal of Smart Sensor and Adhoc Network: Vol. 3: Iss. 3, Article 6.

[15] <https://www.statista.com>

[16] <https://www.educative.io/blog/software-process-model-types>