ML Based Phishing Defense WebApp

Bhong Tejashree[1], Gandhi Deepa[2], Pawar Prerana[3], More Priyanka[4], Prof. P. S. Jadhav[5]
[1,2,3,4] *Dept. of Computer Science Engineering, SMSMPITR, Akluj, India*
[5]*Guide, Dept. of Computer Science Engineering, SMSMPITR, Akluj, India*

***Abstract*—Phishing attacks are a major cybersecurity threat that deceive users into visiting fake websites and stealing sensitive information such as passwords and banking details. This research presents a Python-based *ML Based Phishing Defense WebApp* developed to detect whether a given URL is safe or phishing-related. The system analyzes various URL features including HTTPS usage, domain structure, suspicious keywords, and redirection patterns to identify malicious links. By offering instantaneous deployment via an intuitive frontend, the developed application empowers non-technical operators to evaluate link integrity dynamically. *Index Terms* - Phishing, Cyber security, Python, URL Detection, Web Security, Malicious Links.**

I. INTRODUCTION

A. Introduction

Among modern digital threats, phishing remains highly prevalent due to its reliance on social engineering tactics that manipulate targets into interacting with fraudulent web portals and exposing credential assets. Cybercriminals use malicious links, fake login pages, and deceptive messages to perform these attacks. As internet usage and online transactions continue to increase, detecting phishing websites has become an important challenge for ensuring user security and privacy.

The proposed *ML Based Phishing Defense WebApp* is developed using Python and Flask to identify whether a given URL is legitimate or phishingrelated. The system analyzes different URL-based and content-based features such as HTTPS usage, suspicious keywords, special characters, login forms, and external links to detect malicious websites. The application also supports QR code phishing detection using OpenCV, providing users with a simple and effective platform for real-time phishing analysis and online protection.

B. Scope of Project

The scope of the *ML Based Phishing Defense WebApp* is to develop a cybersecurity solution that helps users identify phishing and malicious websites using Python and Flask technologies. The system focuses on analyzing URLs and webpage content to detect suspicious activities based on features such as HTTPS usage, URL length, special characters, suspicious keywords, login forms, external links, and redirection behavior. The project also includes QR code phishing detection using OpenCV by extracting and analyzing URLs from uploaded QR images.

The application is designed to provide real-time phishing analysis through a simple and user-friendly interface, making it useful for students, internet users, organizations, and cyber security awareness platforms. The project demonstrates the practical implementation of phishing detection techniques and can be further enhanced by integrating machine learning algorithms, browser extensions, real-time threat intelligence APIs, and advanced security analysis methods for improved phishing prevention.

C. Objectives of the Project

1. To develop a Python-based and *ML Based Phishing Defense WebApp* for detecting phishing and malicious URLs.
2. To analyze URL-based features such as HTTPS usage, URL length, suspicious keywords, and special characters for phishing detection.
3. To implement content-based analysis by detecting login forms, external links, iframes, and redirection behavior in websites.
4. To integrate QR code phishing detection using OpenCV for analyzing URLs hidden inside QR codes.
5. To provide real-time phishing analysis and security alerts through a simple and user-friendly web interface.

6. To improve cyber security awareness and help users avoid online fraud and phishing attacks.

II. LITERATURE REVIEW

1. Said Salloum, Tarek Gaber, Sunil Vadera, and Khaled Shaalan authored *A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques*. The paper reviews research studies on phishing email detection using NLP and machine learning methods, highlighting important algorithms, datasets, and evaluation techniques used between 2006 and 2022.

2. Trevor Wooda, Vitor Basto-Fernandes, Eerke Boitenc, and Iryna Yevseyeva presented *AntiPhishing Defences and Their Application to Before-the-click Phishing Email Detection*. The study discusses different anti-phishing defense techniques and explains how AI, machine learning, and heuristic methods can detect phishing emails before users interact with them.

3. Dr. K.N.S. Lakshmi, Suvvari Pavan Kumar, Vudutala Srihari, Kadavati Manohar, and Pappala Srinidhi proposed *Phish Catcher: Client-Side Defense against Web Spoofing Attacks Using Machine Learning*. The paper introduces a Chrome extension that uses a random forest algorithm to detect phishing and spoofed websites with high accuracy and fast response time.

4. V. Suganya conducted *A Review on Phishing Attacks and Various Anti Phishing Techniques*. The paper explains different types of phishing attacks and reviews several anti-phishing methods used to improve cybersecurity and protect users from online fraud.

III. PROBLEM STATEMENT

The rapid expansion of web-based communication and digital transactions has led to a surge in sophisticated social engineering threats, particularly phishing scams. Attackers routinely deploy deceptive strategies, such as masking malicious domains and hiding dangerous links inside QR codes, to exploit unsuspecting users. Traditional security measures often fail to intercept newly generated malicious links in real time, leaving average internet users vulnerable to data theft and financial fraud. To address this gap, there is a critical need for an accessible, automated application that evaluates URL structures, page content, and QR codes instantly. The proposed ML Based Phishing Defense WebApp addresses this vulnerability by utilizing a multi-layered verification system to detect threats and protect users before they interact with malicious elements

IV. METHODOLOGY

Phishing Attack Lifecycle (Figure 1)

The first image illustrates the complete lifecycle of a phishing attack, explaining how cybercriminals deceive users step by step. It begins with “The Lure,” where an attacker sends a deceptive message through email, SMS, or phone calls designed to create urgency or fear. Next is “The Hook,” in which the victim is manipulated into taking action, such as clicking a malicious link or opening an infected attachment. This action leads to “The Catch,” where sensitive information like passwords, financial data, or system access is stolen. The diagram highlights the final consequences of phishing attacks, including identity theft, financial loss, and malware infection, emphasizing how simple user interaction can result in serious cyber security breaches.

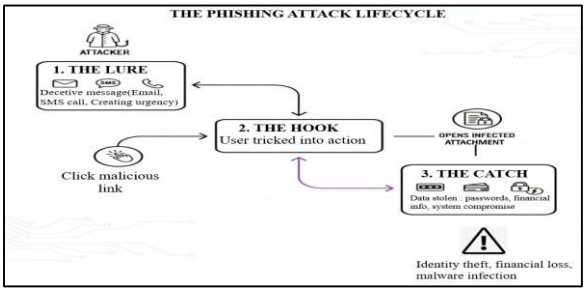

Figure 1:Phishing Attack Lifecycle

Phishing Defense Web Application Solution (Figure 2)

The second image presents a solution workflow using a phishing defense web application designed to prevent cyber attacks. The process starts when a user encounters a suspicious email, message, QR code, or social media link. Instead of directly clicking it, the user submits the link to the web application for analysis. The system then performs multiple security checks such as visual scanning, typo squatting detection, domain age and WHOIS lookup, link mismatch analysis, and heuristic scoring. Based on these evaluations, the application provides a verdict—

either warning the user that the link is dangerous or confirming that it appears safe while advising caution. The diagram promotes proactive cyber security awareness by encouraging users to verify links before interacting with them.

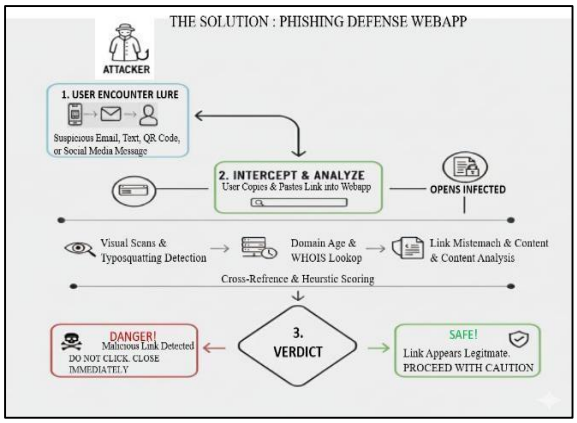

Figure 2: Phishing Defense Web Application

Project Output:

i. ML Based Phishing Defense WebApp UI:

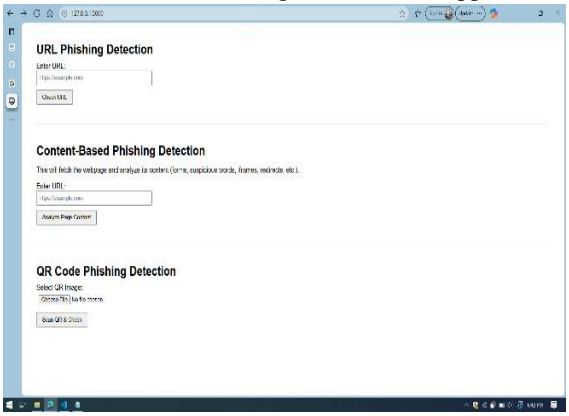

Figure 3: ML Based Phishing Defense WebApp UI

ii. URL Based Module:

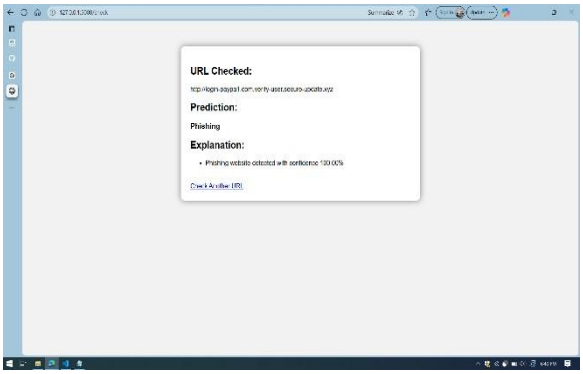

Figure 4: URL Based Module

Content Based Module:

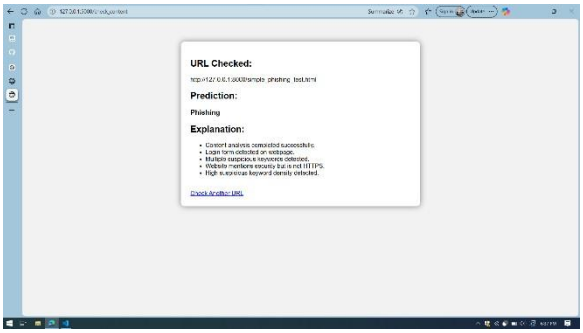

Figure 5: Content Based Module

iii. QR Code Based Module:

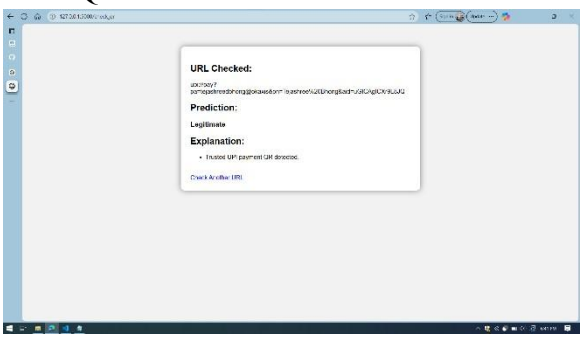

Figure 6: QR Code Based Module

V. PUBLICATION PRINCIPLES

The publication of this research work adheres to established academic ethics and standards to guarantee the originality and credibility of the proposed web application. Every methodology, diagnostic workflow, and architectural feature detailed in this study has been carefully documented to preserve research integrity. Plagiarism is strictly avoided by ensuring all text is original and that external literature is properly credited. All sample data utilized during system testing is managed securely to uphold user privacy. This study intends to provide a meaningful contribution to the domains of Web Security and Threat Detection by offering an efficient, Python-based framework capable of mitigating modern phishing risks.

VI. CONCLUSION

[4]. A Review on Phishing Attacks and Various Anti Phishing Techniques. V. Suganya [1] The *ML Based Phishing Defense WebApp* successfully demonstrates how Python and Flask can be used to develop an effective cyber security solution for

detecting phishing websites and malicious links. The system analyzes various URLbased and content-based features such as HTTPS usage, suspicious keywords, login forms, external links, and redirection behavior to identify phishing attempts. It also supports QR code phishing detection using OpenCV, providing an additional layer of security against modern cyber threats.

VII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to everyone who supported and guided us throughout the completion of the project titled “ML Based Phishing Defense WebApp.” We are especially thankful to our project guide and faculty members for their valuable guidance, encouragement, and continuous support during the development of this work. We also extend our gratitude to our institution for providing the necessary resources and learning environment to successfully complete the project. Finally, we would like to thank our friends and family members for their motivation, cooperation, and encouragement throughout the project journey.

REFERENCES

[1] A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. Said Salloum[1], Tarek Gaber[2], Sunil Vadera[3], Khaled Shaalan[4].
[2] Anti-Phishing Defences and Their Application to Before-the-click Phishing Email Detection. Trevor Wooda[1], Vitor Basto-Fernandesb[2], Eerke Boitenc[3], Iryna Yevseyeva[4].
[3] Phish Catcher: Client-Side Defense against Web Spoofing Attacks Using Machine Learning. Dr. K. N. S. Lakshmi[1] ,Suvvari Pavan Kumar[2] ,Vudutala Srihari[3] ,Kadavati Manohar [4] ,Pappala Srinidhi [5].