A Project report on

PHISHING DETECTION

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering (Cyber Security)

Submitted by

G.CHARAN SAI
(20H51A6279)
A.RISHIKA
(20H51A6297)
G.DHEERAJ REDDY
(20H51A62A0)

Under the esteemed guidance of

Dr. R. Venkateswara Reddy
(Associate Professor & HOD)

Department of Computer Science and Engineering (Cyber Security)

CMR COLLEGE OF ENGINEERING & TECHNOLOGY
(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A+ Grade
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

Department of Computer Science and Engineering (Cyber Security)

EXPLORE TO INVENT

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Phishing Detection"** being submitted by GARIKAPATI CHARAN SAI (20H51A6279), AKULA RISHIKA (20H51A6297) , GARLAPATI DHEERAJ REDDY (20H51A62A0) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering (Cyber Security)** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Dr.R.Venkateswara Reddy
Associate Professor and HOD
Dept. of CSE(Cyber Security)

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Dr.Punyaban Patel, Professor** , Department of **Computer Science and Engineering (Cyber Security)** for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr.R.Venkateswara Reddy,** Head of the Department of **Computer Science and Engineering (Cyber Security)**, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr.Vijaya Kumar Koppula** , Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of CSE (Cyber Security) for their co-operation.

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

GARIKAPATI CHARAN SAI	20H51A6279
AKULA RISHIKA	20H51A6297
GARLAPATI DHEERAJ REDDY	20H51A62A0

DECLARATION

We hereby declare that results embodied in this Report of Project on "**PHISHING DETECTION**" are from work carried out by using partial fulfillment of the requirements for the award of B. Tech degree. We have not submitted this report to any other university/institute for the award of any other degree.

NAME	ROLL NO	SIGNATURE
G.CHARAN SAI	20H51A6279	
A.RISHIKA	20H51A6297	
G.DHEERAJ REDDY	20H51A62A0	

TABLE OF CONTENTS

List of Figures

ABSTRACT

Phishing detection is the process of identifying phishing attacks in their early stages, warning users and administrators, and ideally, mitigating the threat. Phishing is a type of cyberattack where attackers attempt to trick individuals into revealing sensitive information, such as login credentials or financial details, by impersonating a trustworthy entity via electronic communication, typically email. Identifying these phishing websites is typically a challenging task because phishing is mainly a semantics-based attack, that mainly focuses on human vulnerabilities. To address this issue, we proposed the Phishing Detection project. Our project is designed to enhance online security by detecting and promptly alerting users to potential phishing and malware threats. This comprehensive approach involves Grammar analysis using NLP libraries, Database comparison, Greek Alphabet Analysis and port forwarding detection using Python Libraries to assess the legitimacy of URL's.

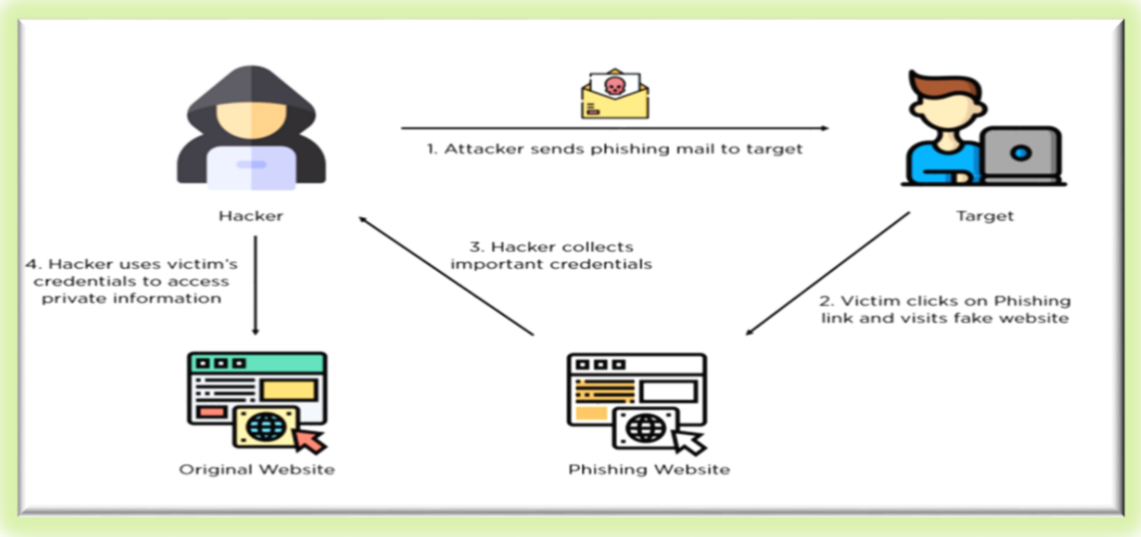

1. Attacker sends phishing mail to target
Hacker
Target
4. Hacker uses victim's credentials to access private information
3. Hacker collects important credentials
2. Victim clicks on Phishing link and visits fake website
Original Website
Phishing Website

Fig 1 - Phishing Process

CHAPTER 1

INTRODUCTION

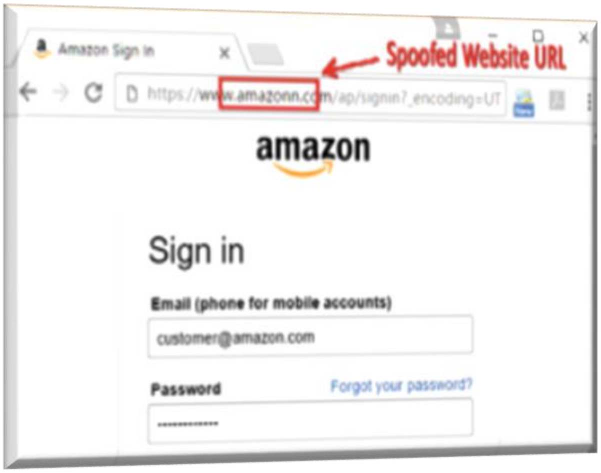
CHAPTER 1

INTRODUCTION

1.1 Problem Statement

In the rapidly advancing era of information technology, the ubiquity of online communication and transactions has exposed individuals and organizations to an escalating threat phishing attacks. Phishing, a deceptive practice wherein cybercriminals attempt to trick users into divulging sensitive information, poses a substantial risk to the confidentiality and integrity of personal and organizational data. As the techniques employed by malicious actors become increasingly sophisticated, traditional security measures are often insufficient in safeguarding against these evolving threats. Hence, there is a critical need for advanced and adaptive systems capable of detecting and mitigating phishing attempts in real-time. This paper introduces a comprehensive approach to phishing detection, combining state-of-the-art machine learning algorithms, behavioral analysis, and domain-specific heuristics. By integrating these elements, our proposed system aims to enhance the accuracy and efficiency of identifying phishing attempts across various platforms, such as emails, websites, and social media .

Our proposed system is designed to enhance online security by detecting and promptly alerting users to potential phishing and malware threats. This comprehensive approach involves Grammar analysis using NLP libraries, Database comparison, Greek Alphabet Analysis and port forwarding detection using Python Libraries.



https://www.amazonn.com/ap/signin?_encoding=UT

Sign in

Email (phone for mobile accounts)

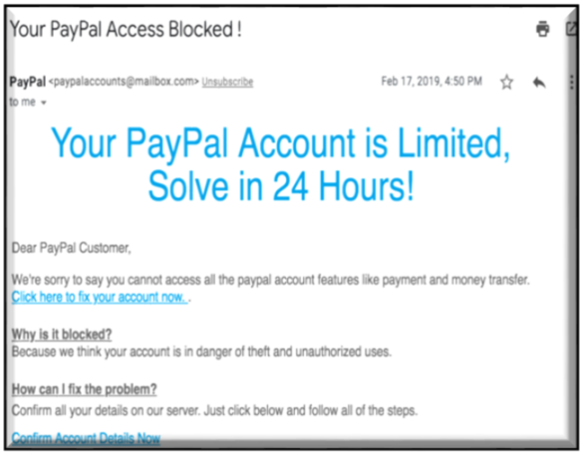
customer@amazon.com

Password

Forgot your password?

••••••••••••

Fig 2 - Phishing Website



Your PayPal Access Blocked !

PayPal <paypalaccounts@mailbox.com> Unsubscribe Feb 17, 2019, 4:50 PM

to me

Your PayPal Account is Limited,
Solve in 24 Hours!

Dear PayPal Customer,

We're sorry to say you cannot access all the paypal account features like payment and money transfer.
Click here to fix your account now. .

Why is it blocked?
Because we think your account is in danger of theft and unauthorized uses.

How can I fix the problem?
Confirm all your details on our server. Just click below and follow all of the steps.

Confirm Account Details Now

Fig 3 - Phishing Email

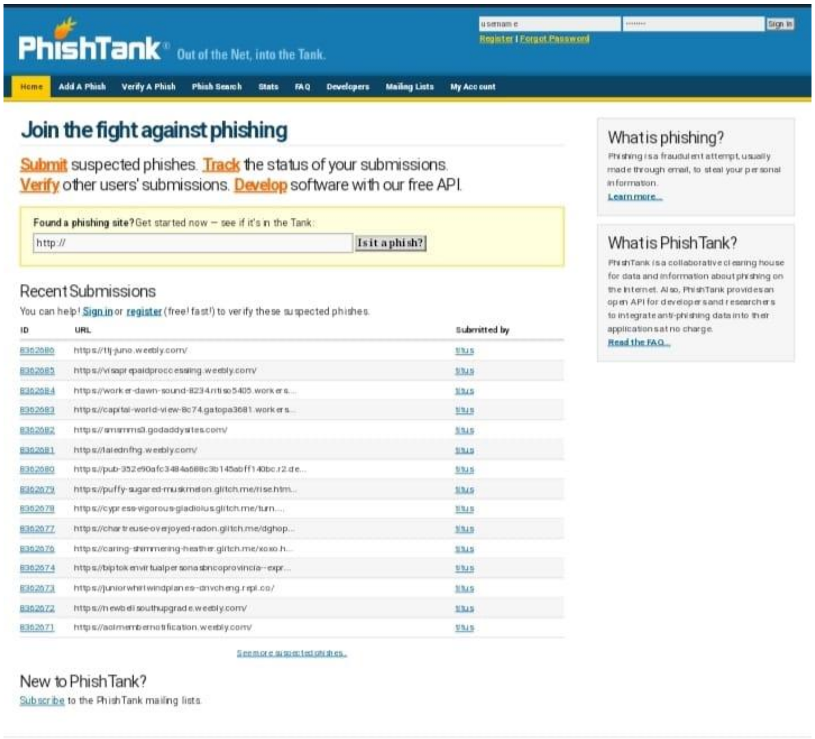
CHAPTER 2
Background Work

CHAPTER 2
BACKGROUND WORK

2.1 Existing Solutions:

1.PHISHTANK:

PhishTank is a website and web service that provides information about phishing sites. It offers a real-time dataset of phishing websites. PhishTank is not a piece of software and doesn't run on your computer. PhishTank is commonly used in academic studies. Preliminary experiments show that PhishTank can successfully identify 91.44% of phishing targets.



Join the fight against phishing

Submit suspected phishes. Track the status of your submissions. Verify other users' submissions. Develop software with our free API.

Found a phishing site? Get started now — see if it's in the Tank:

Recent Submissions

You can help! Sign in or register (free! fast!) to verify these suspected phishes.

ID	URL	Submitted by
8362686	https://ttj-juno.weebly.com/	titus
8362685	https://visaprepaidproccessing.weebly.com/	titus
8362684	https://worker-dawn-sound-8234.rntiso5405.workers...	titus
8362683	https://capital-world-view-8c74.gatopa3681.workers...	titus
8362682	https://smsmms3.godaddysites.com/	titus
8362681	https://laiednfhg.weebly.com/	titus
8362680	https://pub-352e90afc3484a688c3b145abff140bc.r2.de...	titus
8362679	https://puffy-sugared-muskmelon.glitch.me/rise.htm...	titus
8362678	https://cypress-vigorous-gladiolus.glitch.me/turn....	titus
8362677	https://chartreuse-overjoyed-radon.glitch.me/dghop...	titus
8362676	https://caring-shimmering-heather.glitch.me/xoxo.h...	titus
8362674	https://biptokenvirtualpersonasbncoprovincia--expr...	titus
8362673	https://juniorwhirlwindplanes-dnvcheng.repl.co/	titus
8362672	https://newbellsouthupgrade.weebly.com/	titus
8362671	https://aolmembernotification.weebly.com/	titus

New to PhishTank?

Subscribe to the PhishTank mailing lists.

What is phishing?

Phishing is a fraudulent attempt, usually made through email, to steal your personal information.

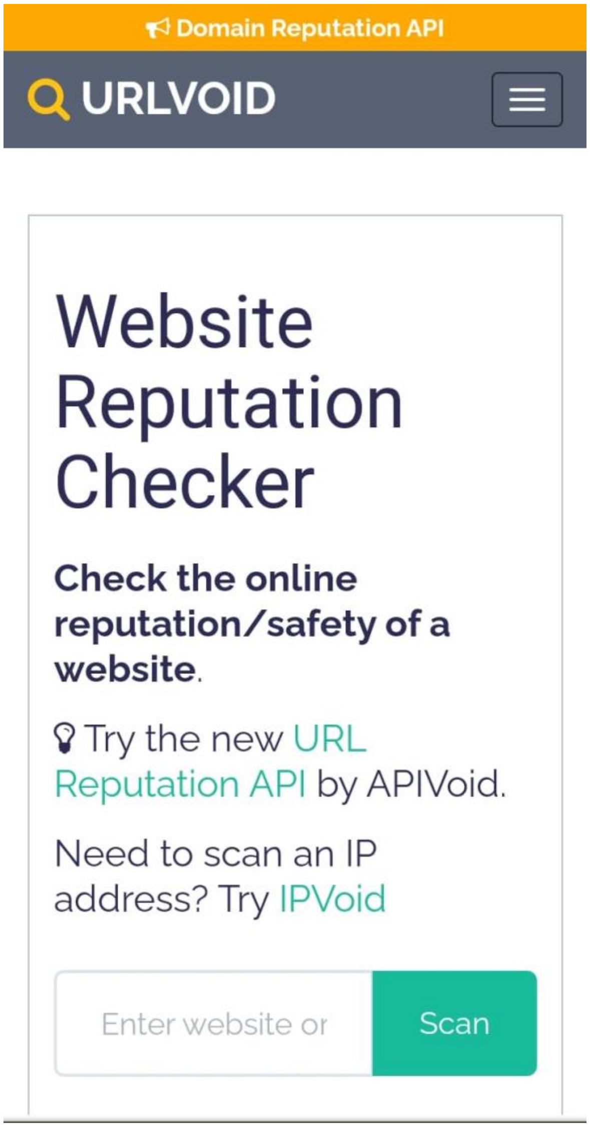
What is PhishTank?

PhishTank is a collaborative clearing house for data and information about phishing on the Internet. Also, PhishTank provides an open API for developers and researchers to integrate anti-phishing data into their applications at no charge.

Fig 4 - PhishTank Website

2. URL VOID:

URLVoid was a website that provided a free online tool for website reputation and safety analysis. Users could enter a website URL, and URLVoid would check the site against multiple databases to provide information about its safety, potential threats, and reputation. The service was useful for individuals and organizations to assess the security of a website before visiting it.



Domain Reputation API

URLVOID

Website Reputation Checker

Check the online reputation/safety of a website.

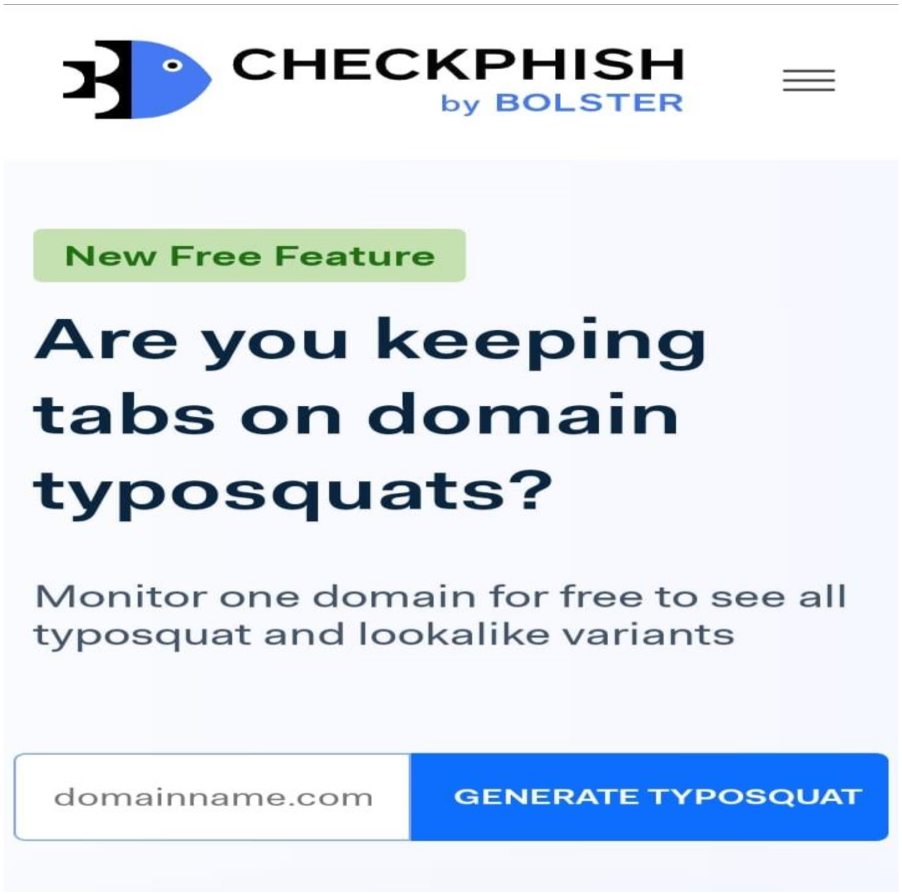
Try the new URL Reputation API by APIVoid.

Need to scan an IP address? Try IPVoid

Fig 5- URL Void Website

3. CHECK PHISH:

CheckPhish is a free, AI-powered scanner that detects phishing and fraudulent URLs. It uses machine learning to detect digital threats, including look-alike domains and fake sites. CheckPhish can be used right from your inbox. CheckPhish is a community offering from Bolster. It's used by some of the largest companies in the world, including nearly every Fortune 500 company.



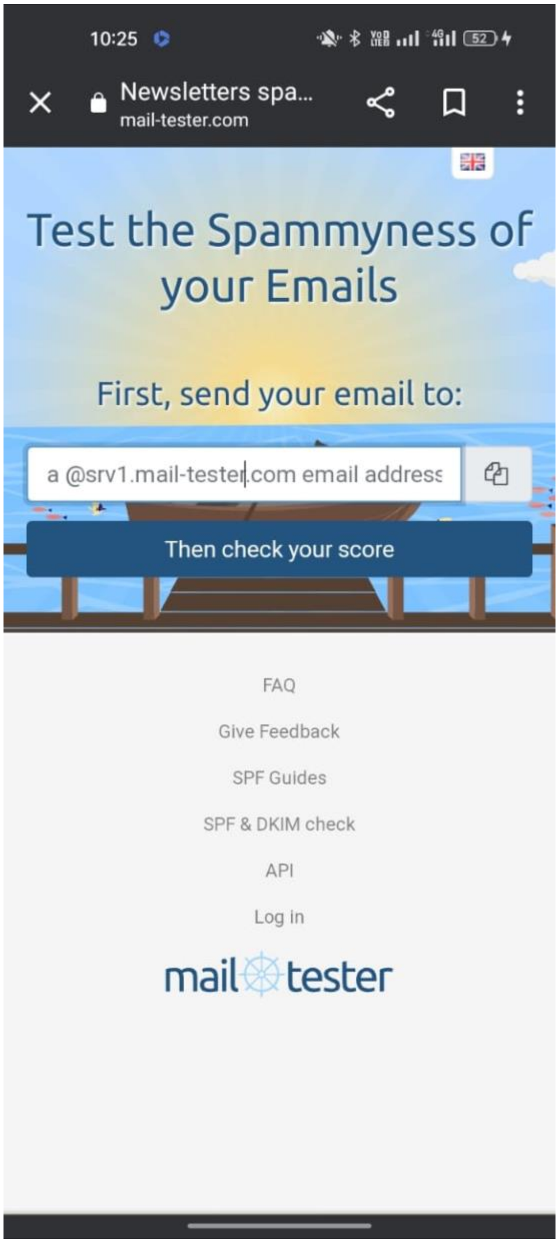
Are you keeping tabs on domain typosquats?

Monitor one domain for free to see all typosquat and lookalike variants

Fig 6- CheckPhish Website

4.MAIL TESTER:

Mail Tester is an email verification tool that can check emails from any database with up to 99% accuracy. It can identify duplicate and bounced emails, remove spam traps and emails with high-risk keywords, and verify email domains and syntax.



10:25

Newsletters spa...
mail-tester.com

Test the Spammyness of your Emails

First, send your email to:

a @srv1.mail-tester.com email address

Fig 7-Mail Tester Website

CHAPTER 3

PROPOSED SOLUTION

CHAPTER 3

PROPOSED SOLUTION

3.1 Proposed Solution

Our proposed system for phishing attacks is a website with comprehensive and multi-layered solution designed to enhance online security and protect users from phishing and malware threats. This comprehensive approach involves several layers of analysis to assess the legitimacy of URLs. The multi-layered approach includes URL analysis, Port forwarding detection, Database Comparison, Greek alphabet analysis and Homographic attack detection. The URL which we want to check has to be pasted in the website and the following analysis is done and result is displayed along with the accuracy.

The steps employed for the phishing detection is given below:

1. Loading the data
2. Familiarizing with data & EDA
3. Visualizing the data
4. Splitting the data
5. Training the data
6. Comparison of Model
7. Conclusion

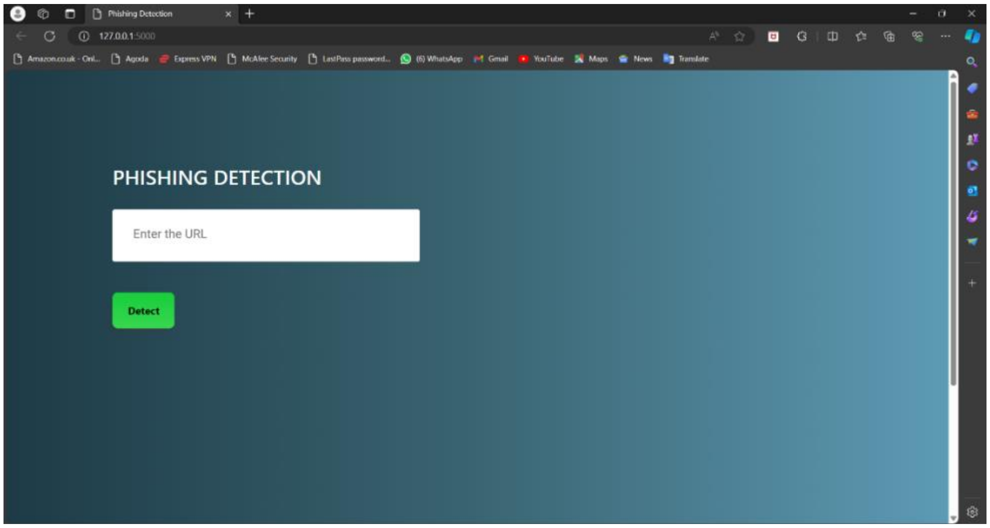


Fig 8 – Phishing Detection Website

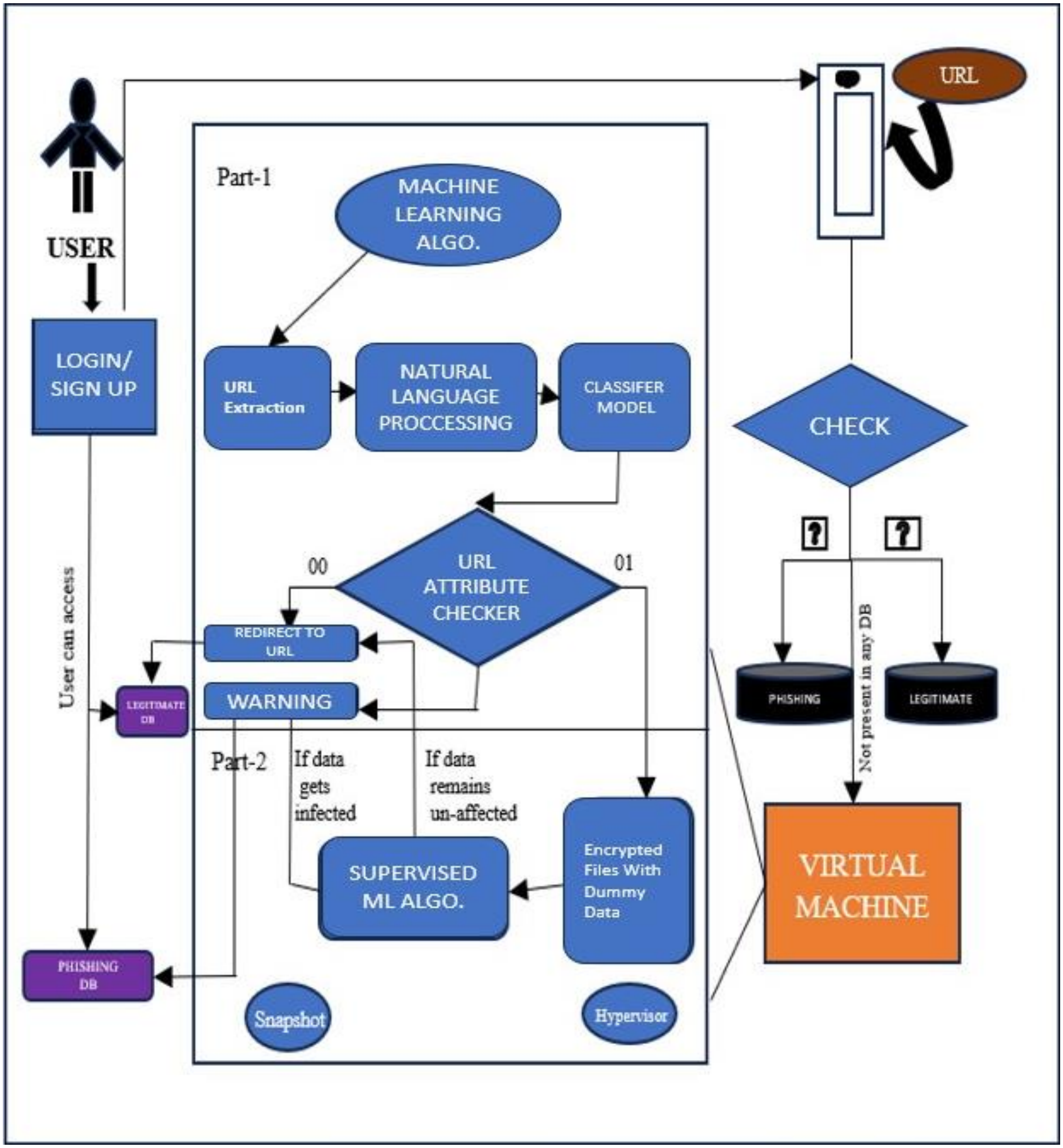


Fig 9- Block Diagram

CHAPTER 4

SYSTEM REQUIREMENTS & TOOLS

CHAPTER 4
SYSTEM REQUIREMENTS & TOOLS

4.1 Software Requirements:

- **Operating System:** Windows 10, 11

Fig 10- Windows

- **Coding Language :** Python

Fig 11 - Python

- **Technologies :**

Fig 12-Scikit Learn ,Flask, pandas ,NumPy

- **Editor :** VS Code ,Python 3.1

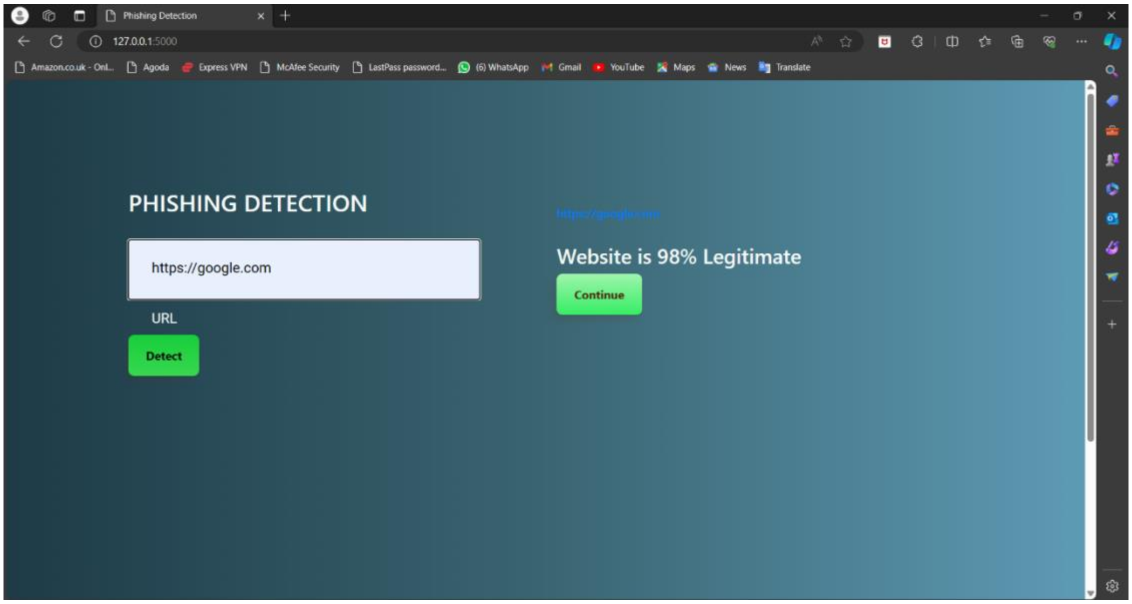
Fig 13- VS Code

Fig 14- Python 3.1

4.2 Hardware Requirements:

- **System :** Any Processor of 3rd,4th or 5th Generation
- **Hard Disk :** Min 50Gb of vacant space
- **RAM :** MIN 4GB

CHAPTER 5
IMPLEMENTATON


Phishing Detection
127.0.0.1:5000
Amazon.co.uk - Onl... Agoda Express VPN McAfee Security LastPass password... (6) WhatsApp Gmail YouTube Maps News Translate
Detect
Continue

PHISHING DETECTION

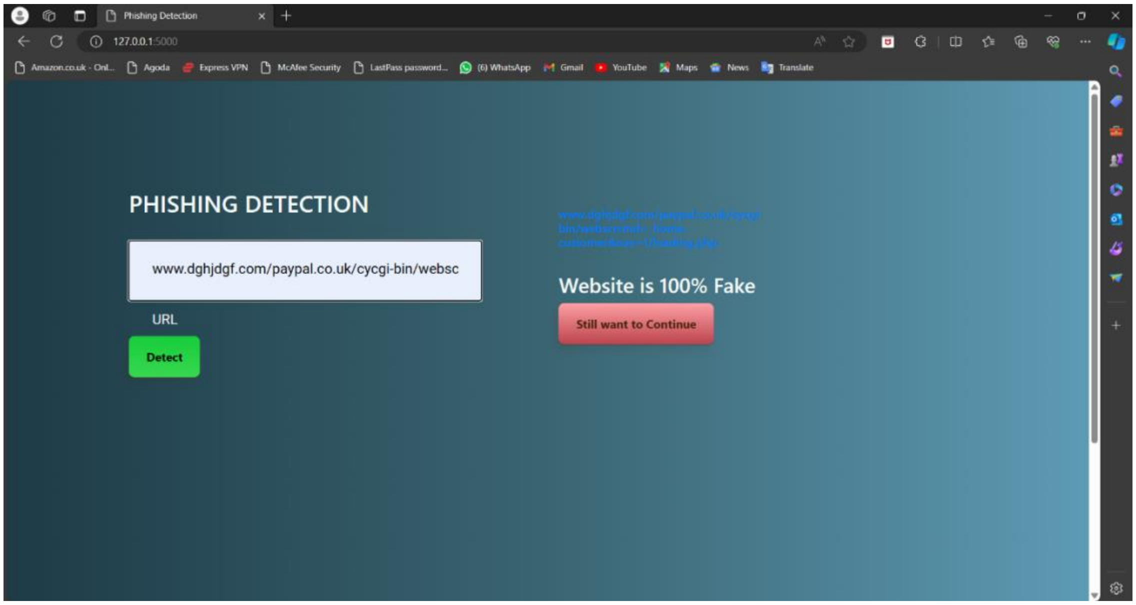
https://google.com

URL

https://google.com

Website is 98% Legitimate

Fig 15 – Result of Legitimate URL


Phishing Detection
127.0.0.1:5000
Amazon.co.uk - Onl... Agoda Express VPN McAfee Security LastPass password... (6) WhatsApp Gmail YouTube Maps News Translate
Detect
Still want to Continue

PHISHING DETECTION

www.dghjdgf.com/paypal.co.uk/cycgi-bin/websc

URL

www.dghjdgf.com/paypal.co.uk/cycgi-bin/webscrcmd=_home-customer&nav=1/loading.php

Website is 100% Fake

Fig 16- Result of Phishing URL

CHAPTER 6
CONCLUSION

CHAPTER 5

CONCLUSION

5.1 Conclusion :

Our project is a major advancement in online security, countering phishing and malware threats with a multi-layered approach. We employ grammar analysis, database comparison, port forwarding detection, and Greek alphabet analysis to thoroughly assess URLs. Our system detects semantic anomalies and evolving cybercriminal techniques, promptly alerting users. We aim to significantly reduce phishing and malware success rates, creating a safer digital environment. Our unwavering commitment to online security empowers users to navigate the internet with confidence.

Detecting and mitigating phishing attacks is a critical aspect of ensuring online security and protecting sensitive information. In conclusion, the evolving nature of phishing threats necessitates a multi-faceted approach that combines advanced technological solutions, user education, and ongoing vigilance. While technological tools such as email filtering, website analysis, and machine learning algorithms play a crucial role in identifying and preventing phishing attempts, they are most effective when complemented by user awareness and education programs. Continuous improvement of phishing detection technologies is essential, as threat actors constantly adapt and develop new methods to bypass existing defenses. Collaborative efforts within the cybersecurity community, information sharing, and staying abreast of the latest phishing trends contribute to a more robust defense against evolving threats.

Furthermore, the integration of artificial intelligence and machine learning into phishing detection systems enhances the ability to identify subtle patterns and anomalies indicative of phishing attempts. As these technologies continue to advance, they hold the promise of providing more accurate and real-time threat detection, reducing false positives, and enhancing overall security posture.

CHAPTER 6
REFERENCES

CHAPTER 6

REFERENCES

6.1 References:

[1].https://zvelo.com/phishing-detection-in-depth/

[2].https://github.com/emalderson/ThePhish

[3].https://www.csoonline.com/article/569867/9-top-anti-phishing-tools-and-services.html

[4].https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8504731/

[5].https://medium.com/@samarth.gawande19/nlp-based-phishing-url-detection-f2f644aadab

[6].https://bolster.ai/blog/phishing-threat-intelligence

[7].https://getcomputeractive.co.uk/protect-your-tech/fake-urls-with-cyrillic-letters

[8].https://intezer.com/blog/incident-response/url-analysis-phishing-part-1/#:~:text=Some%20examples%20could%20be%3A%201%20cnn-news.com%20%28added%20%E2%80%9C-news%E2%80%9D%29,flarecloud.com%20%28inverted%20the%20word%20order%2C%20instead%20of%20Cloudflare%29

[9].https://github.com/sublime-security/emailrep.io