Phishing Simulation

Semester – 7
Capstone Project (01CT0715)

A PROJECT REPORT

Submitted by

Heet Chothani
92100133043

Sahil Patel

92100133009

BACHELOR OF TECHNOLOGY

in

Information and

Communication Technology

Marwadi University, Rajkot

December, 2024

Major Project-I (01CE0716)

Marwadi University

Faculty of Technology

Department of Information and Communication Technology

2024-25

CERTIFICATE

This is to certify that the project report submitted along with the project entitled Phishing Simulation has been carried out by **Heet Chothani(92100133043)** and **Sahil Patel(92100133009)** under our guidance in partial fulfilment for the degree of Bachelor of Technology inInformation and Communication Technology, 7th Semester of Marwadi University, Rajkot during the academic year 2024-25.

Prof. Chandrasinh Parmar

Head of the Department

Table of Contents

Abstract

1. **Introduction to the Problem:**
   - Discuss the significance of spam detection and its role in combating phishing.
   - Highlight the challenges posed by phishing messages and the importance of effective text classification.
2. **Objective:**
   - State the goal of the project, which is to build an efficient machine learning model for detecting phishing messages in a given dataset.
3. **Data Preprocessing and Cleaning:**
   - Mention the steps taken to clean and preprocess the dataset, such as removing unnecessary columns, renaming for clarity, handling duplicates, and transforming text data.
4. **Exploratory Data Analysis (EDA):**
   - Outline the steps of EDA, including character, word, and sentence analysis, along with visualizations like histograms, pie charts, and word clouds for spam and ham messages.
5. **Text Preprocessing:**
   - Detail the text transformation process: converting to lowercase, tokenization, removing stop words and special characters, and applying stemming to reduce words to their root forms.
6. **Feature Engineering:**
   - Explain the use of text vectorization techniques, including TF-IDF (max features: 3000), to convert textual data into numerical format suitable for machine learning.
7. **Model Selection and Training:**
   - Highlight the evaluation of multiple classifiers (e.g., Naive Bayes, SVM, Decision Trees, Random Forest, etc.) using metrics such as accuracy and precision.
   - Discuss advanced techniques like ensemble learning (voting and stacking classifiers).
8. **Performance Evaluation:**
   - Compare and visualize the performance of various algorithms, showing the trade-

offs between accuracy and precision.

9. **Deployment Readiness:**
    - Describe the creation of serialized models (vectorizer.pkl, model.pkl) and the use of Flask for building a web interface to deploy the phishing detection system.
10. **Conclusion and Future Scope:**
    - Summarize the key findings and performance of the chosen models.
    - Suggest potential improvements, such as handling imbalanced data or integrating real-time phishing simulations.

CHAPTER 1

Understanding Phishing Attacks

1. Introduction to Phishing

Phishing is one of the most widespread and insidious cyberattack methods. Malicious actors use deceptive tactics to trick individuals into divulging sensitive personal and financial information or granting unauthorized access to systems. Phishing exploits human psychology and technical vulnerabilities, making it a critical issue in the realm of cybersecurity.

1.1 Definition and Overview

The term "phishing" originates from the word "fishing," signifying the act of luring victims into a trap. Phishing attacks are typically carried out through emails, text messages, social media, or phone calls, often impersonating trusted organizations or individuals. The attackers may embed malicious links or attachments in their communications to extract information or install malware. While financial gain is the primary motivation, phishing is also used for espionage, data theft, and system compromise.

1.2 Types of Phishing Attacks

Phishing has evolved into various forms, each designed to exploit specific targets or vulnerabilities:

- **Email Phishing**: The most prevalent form, involving mass-distributed emails with fraudulent content.
- **Spear Phishing**: Highly targeted attacks, often using personal details about the victim to increase the likelihood of success.
- **Whaling**: A specialized form of spear phishing aimed at high-value targets such as CEOs, CFOs, and other executives.
- **Smishing and Vishing**: Phishing through SMS (smishing) or phone calls (vishing), often impersonating customer service or support teams.
- **Clone Phishing**: The attacker replicates legitimate communications, replacing links or attachments with malicious ones.
- **Pharming**: Victims are redirected to counterfeit websites, even if they type the correct URL into their browsers.

---

2. Techniques and Tactics Used in Phishing

Phishers deploy a mix of psychological manipulation and technical subterfuge to deceive their targets.

2.1 Social Engineering

The success of phishing often hinges on exploiting human behavior. Common social engineering tactics include:

- **Fear and Urgency**: Messages threatening dire consequences, such as account suspension or fraud, if immediate action is not taken.
- **Curiosity and Greed**: Offering enticing rewards, discounts, or exclusive opportunities to lure victims into clicking links or sharing details.
- **Impersonation of Authority**: Attackers pose as reputable entities, such as banks, employers, or government agencies, to exploit trust.

2.2 Technical Exploits

Phishers also employ advanced technical methods to enhance the credibility of their schemes:

- **Email and Domain Spoofing**: Creating fake email addresses and websites that closely resemble legitimate ones.
- **Malicious Attachments and Links**: Embedding harmful content that, when clicked or opened, can install malware or steal information.
- **Man-in-the-Middle (MITM) Attacks**: Intercepting data exchanges to steal login credentials or sensitive information.

---

3. Impact of Phishing Attacks

The repercussions of phishing attacks are far-reaching, affecting individuals, organizations, and society as a whole.

3.1 Personal Impact

Victims may suffer financial losses, identity theft, and emotional trauma. For instance, compromised bank accounts or credit card information can lead to unauthorized transactions, while restoring one’s identity can take years.

3.2 Organizational Impact

Phishing poses significant risks to businesses, including:

- **Data Breaches**: Exposure of sensitive corporate or customer data.
- **Financial Losses**: Costs associated with fraud, legal penalties, and operational downtime.
- **Reputational Damage**: Loss of customer trust and market credibility following a successful attack.

3.3 Societal Impact

Phishing undermines public confidence in digital communication and transactions, slowing technological adoption and increasing the demand for stringent cybersecurity measures.

4. Prevention and Mitigation Strategies

Combating phishing requires a combination of user education, technical defenses, and responsive measures.

4.1 User Education and Awareness

An informed user base is the first line of defense. Educational initiatives should focus on:

- Identifying suspicious email content and URL irregularities.
- Avoiding interaction with unsolicited messages.
- Verifying the legitimacy of requests through independent channels.

4.2 Technical Solutions

Organizations can deploy advanced technologies to deter phishing attempts:

- **Spam Filters**: Tools that detect and block phishing emails before they reach users.
- **Two-Factor Authentication (2FA)**: Adds an additional verification step beyond passwords.
- **SSL Certificates**: Encrypt communication between users and legitimate websites.

4.3 Incident Response and Recovery

Preparedness is key to minimizing the impact of phishing:

- **Monitoring and Reporting**: Establishing clear channels for users to report suspected phishing attempts.
- **Containment Protocols**: Rapid isolation of affected systems to prevent further spread.
- **Post-Incident Analysis**: Learning from attacks to improve defenses and prevent recurrence.

5. Conclusion

Phishing is a dynamic and persistent threat in the digital age. By exploiting both human vulnerabilities and technical weaknesses, attackers continue to refine their tactics, making awareness and adaptability essential. Through a combination of education, advanced technical safeguards, and proactive incident response strategies, individuals and organizations can significantly reduce their susceptibility to phishing, paving the way for a more secure digital environment.

CHAPTER 2

Detecting Phishing Attacks Using Code

1. Introduction to Automated Detection

To combat phishing effectively, automation plays a pivotal role. Writing scripts or applications that analyze emails, URLs, or messages for signs of phishing helps in identifying and neutralizing threats before they cause harm. This chapter outlines an approach using Python to detect phishing through email analysis.

1.1 Python Code for Detecting Phishing

Below is an example of Python code to analyze email content and detect potential phishing indicators:

```
import re


# List of suspicious keywords and phrases often found in phishing emails
suspicious_keywords = ["verify your account", "urgent action required", "click here", "login now"]


# Function to detect phishing in email content
def detect_phishing(email_content):
    # Check for suspicious keywords
    for keyword in suspicious_keywords:
        if keyword.lower() in email_content.lower():
            print(f"Warning: Suspicious keyword detected - '{keyword}'")

    # Check for suspicious links
    links = re.findall(r'http[s]?://\S+', email_content)
    for link in links:
        if "bit.ly" in link or "tinyurl" in link:  # Shortened URLs often indicate phishing
            print(f"Warning: Suspicious link detected - {link}")

    if not links:
        print("No links found in the email.")

# Example email content
email_sample = "Dear user, please verify your account by clicking here: http://bit.ly/phishing-example"
```

```
# Run the phishing detection function
detect_phishing(email_sample)
```

1.2 Code Explanation

- **Suspicious Keywords**: The script checks for predefined phrases commonly used in phishing attacks to create urgency or mislead users.
- **Suspicious Links**: By identifying links in the email content, the script highlights shortened URLs often associated with phishing.
- **Regex for Links**: The re.findall function extracts URLs from the email content, allowing detailed analysis.

DETECTING EMAIL ATTACKS
CONTAINS SHISCIOUS EMAIL CONTENT?
CONTAINS SUSPICIOUS KEAWODS?
'YES
YES'
CONTAINS SUSPICIOUS KEYWODS?
YES
CONTAINS SUSPICIOUS EMALTS?
?
YES
'YES
'NO
Y5
FLAG AS PHISHING
FANTAINS SUSPICIOUS PHISHING?
FLAG AS SUSPICIOUS EMILNG?
'YES
YES
SUSPICIOUS EMALILS?
NO
SAFE EMAIL
YES
FLAG AS PHISHING?
FLAG AS EHISING
SAFE
SAFE EMAIL
NO

1.3 Benefits of Automated Detection

Using such scripts has several advantages:

- **Scalability**: Automating the detection process allows organizations to analyze large volumes of emails efficiently.
- **Speed**: Quick identification minimizes the window for potential damage.
- **Customization**: The code can be tailored to include more advanced indicators and integrate with existing security systems.

1.4 Limitations and Future Enhancements

While basic scripts are effective for initial detection, sophisticated phishing attacks may bypass simple rules. Enhancements like:

- Incorporating machine learning models.
- Integrating with threat intelligence feeds.
- Using natural language processing (NLP) for deeper analysis. can further improve the system's accuracy and reliability.

CHAPTER 3

VIRUSTOTAL

1. Introduction to VirusTotal

VirusTotal is an online service that analyzes files, URLs, and other content to detect threats such as viruses, malware, and phishing attempts. It aggregates the results from over 70 antivirus engines and other security tools to provide comprehensive threat detection.

1.1 Overview

VirusTotal allows users to upload files and URLs for scanning, making it a valuable tool for both individual users and organizations. It helps identify whether a particular file or website is safe or contains malicious elements.

1.2 Functionality

The service provides detailed reports based on scans from multiple antivirus programs, allowing users to assess the safety of their files or URLs. VirusTotal also includes metadata analysis and behavior analysis to detect hidden threats.

2. Features and Capabilities of VirusTotal

VirusTotal offers several powerful features for detecting malware and phishing content through its free and premium services.

2.1 File and URL Scanning

- File Scanning: Users can upload files such as executables or documents to be scanned by more than 70 antivirus engines.
- URL Scanning: URLs are analyzed for phishing or malicious content, helping users avoid harmful websites.

2.2 Threat Detection and Analysis

- Aggregated Detection: VirusTotal compiles results from multiple antivirus vendors to provide a more thorough analysis.
- Behavioral and Metadata Analysis: In addition to signature-based detection, VirusTotal assesses the behavior and metadata of files to identify potentially harmful activity.

2.3 Community-Based Insights

- VirusTotal encourages user submissions and provides community-based feedback, enabling better identification of emerging threats.
- Users can access reports and contribute by flagging suspicious files or URLs, enhancing collaborative cybersecurity efforts.

3. Use Cases of VirusTotal

VirusTotal is widely used in both individual and organizational settings to improve security measures and prevent cyber threats.

3.1 Individual Use

For personal users, VirusTotal offers a quick and easy way to verify the safety of files and URLs. It helps protect individuals from malicious emails, downloads, and websites.

3.2 Organizational Use

Businesses leverage VirusTotal’s API to automate the scanning process for files, URLs, and network traffic, enhancing internal security protocols. VirusTotal also assists in identifying data breaches and detecting advanced persistent threats (APTs).

3.3 Security Research and Analysis

VirusTotal is a critical tool for cybersecurity researchers, providing them with access to vast amounts of data for threat analysis and malware research. It also allows researchers to track emerging threats by monitoring real-time reports and updates.

4. Advantages and Limitations of VirusTotal

VirusTotal provides many benefits, but there are some limitations to be aware of.

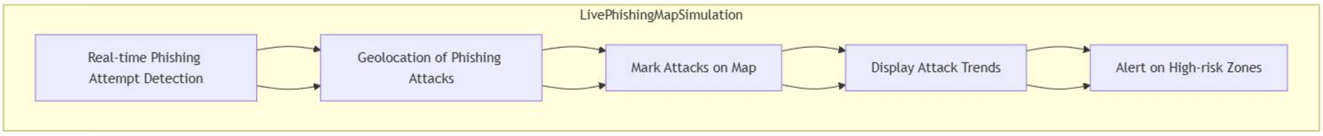
4.1 Advantages

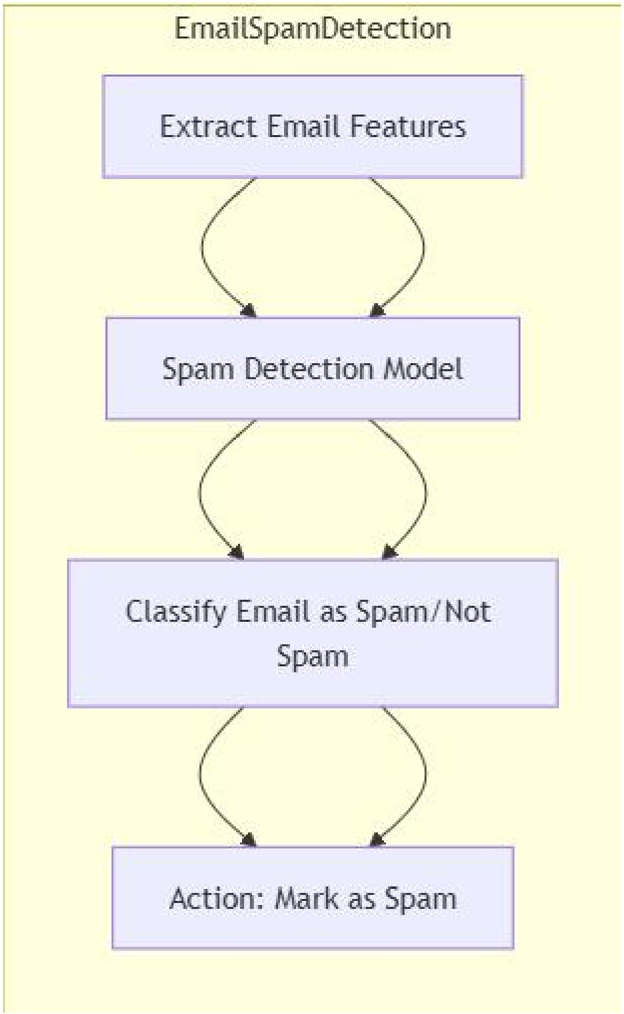
- Comprehensive Threat Detection: Multiple antivirus engines provide more accurate threat detection.
- Free Service: VirusTotal’s core services are available for free, making it accessible to everyone.
- User-Friendly Interface: The service is easy to use, with intuitive navigation for file and URL submissions.

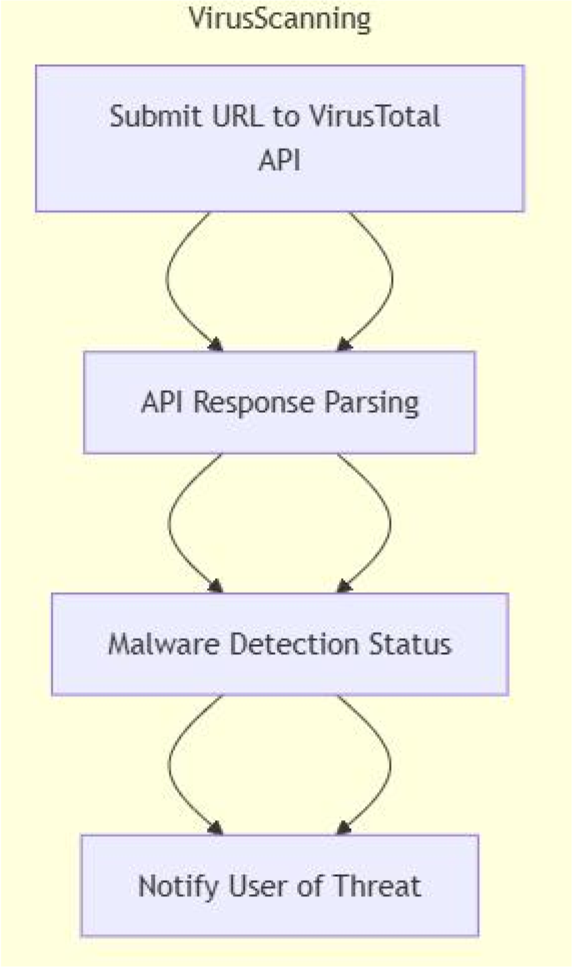
4.2 Limitations

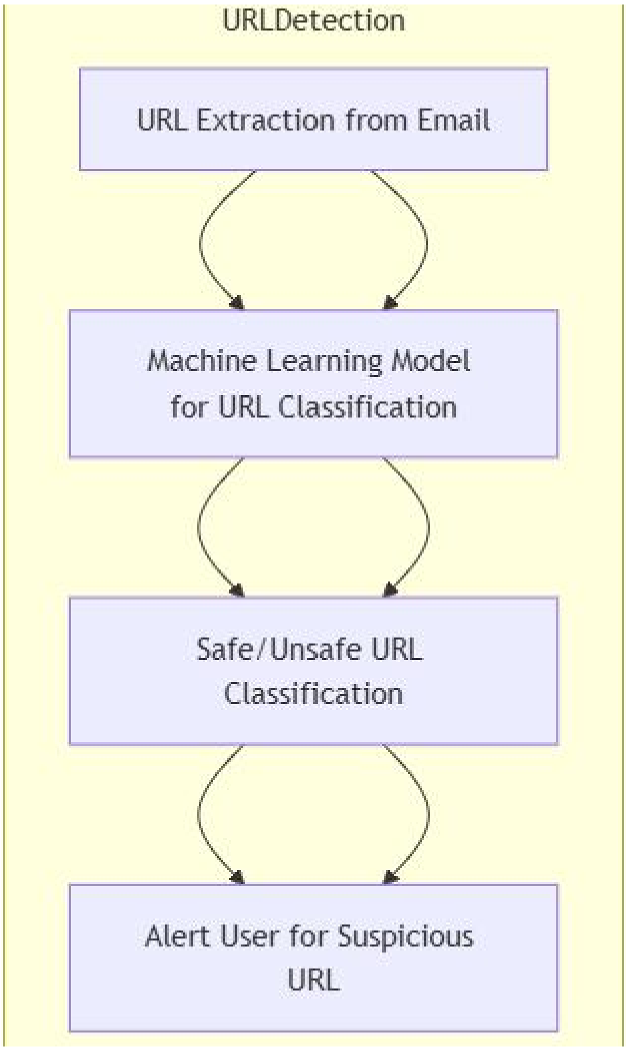
- Limited File Size: Free users are limited by the maximum file size for uploads (usually up to 256MB).
- False Positives: Like any automated service, there may be occasional false positives or negatives.
- Dependence on Antivirus Definitions: VirusTotal's scans are only as good as the antivirus engines it aggregates, meaning that new or unrecognized threats may go undetected.

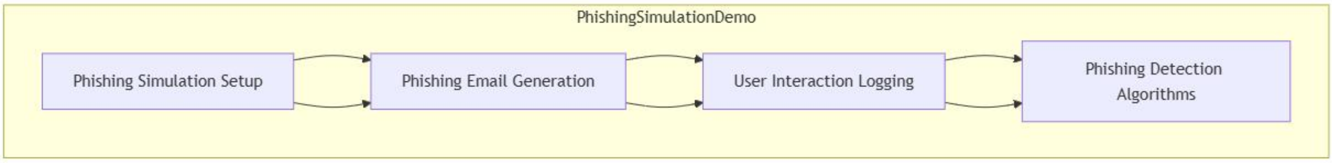
Flow Chart:


LivePhishingMapSimulation
Real-time Phishing Attempt Detection
Geolocation of Phishing Attacks
Mark Attacks on Map
Display Attack Trends
Alert on High-risk Zones

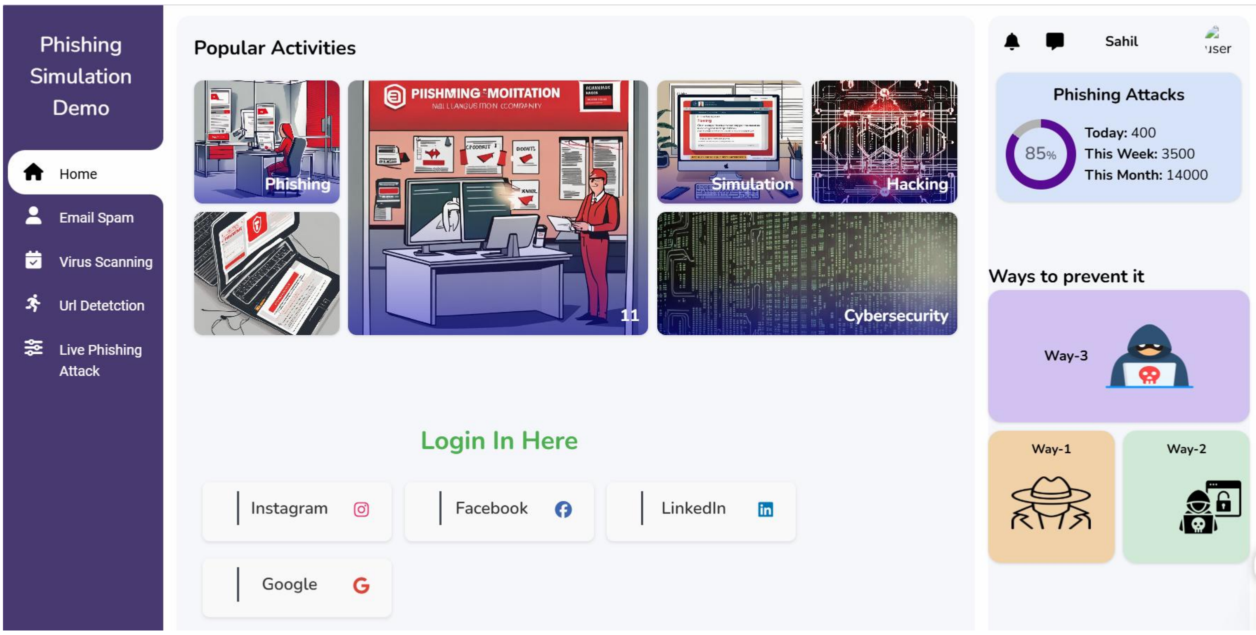

EmailSpamDetection
Extract Email Features
Spam Detection Model
Classify Email as Spam/Not Spam
Action: Mark as Spam

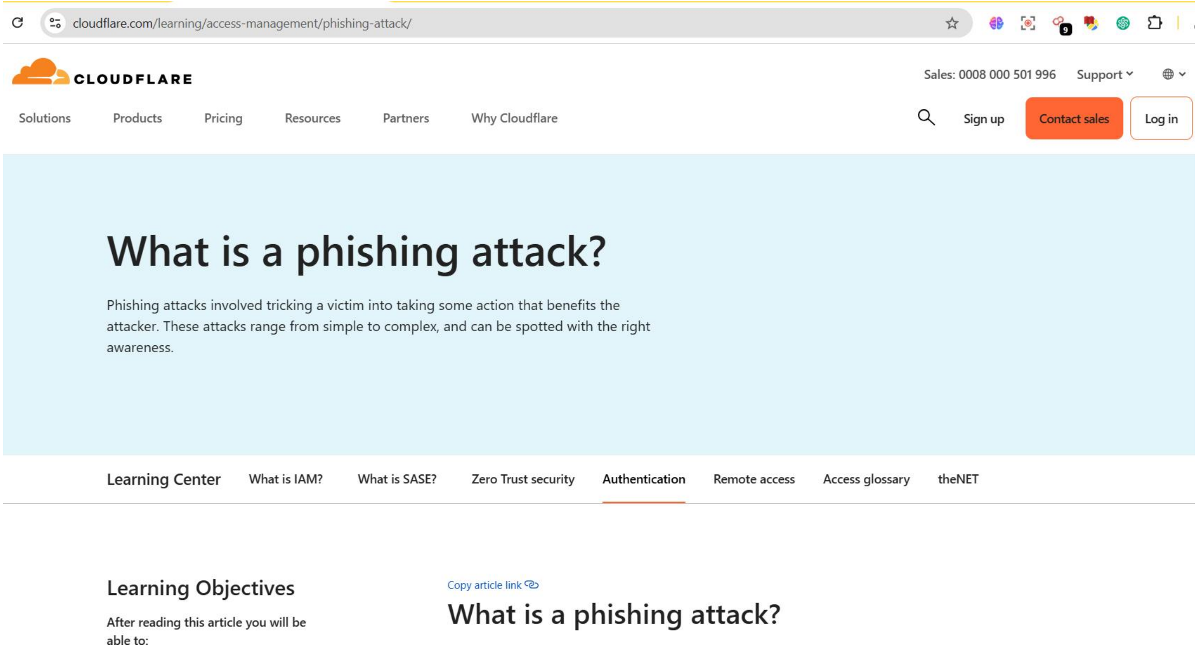

VirusScanning
Submit URL to VirusTotal API
API Response Parsing
Malware Detection Status
Notify User of Threat


URLDetection
URL Extraction from Email
Machine Learning Model for URL Classification
Safe/Unsafe URL Classification
Alert User for Suspicious URL


PhishingSimulationDemo
Phishing Simulation Setup
Phishing Email Generation
User Interaction Logging
Phishing Detection Algorithms

Project Images





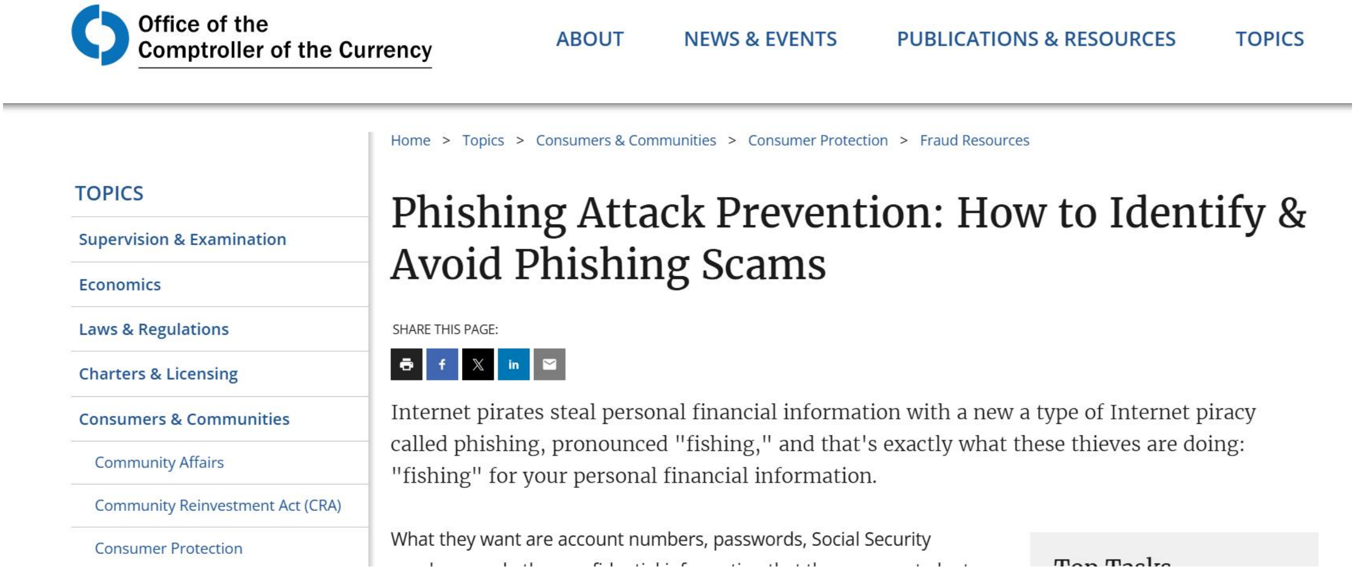
What is a phishing attack?

Phishing attacks involved tricking a victim into taking some action that benefits the attacker. These attacks range from simple to complex, and can be spotted with the right awareness.

Learning Objectives

After reading this article you will be able to:

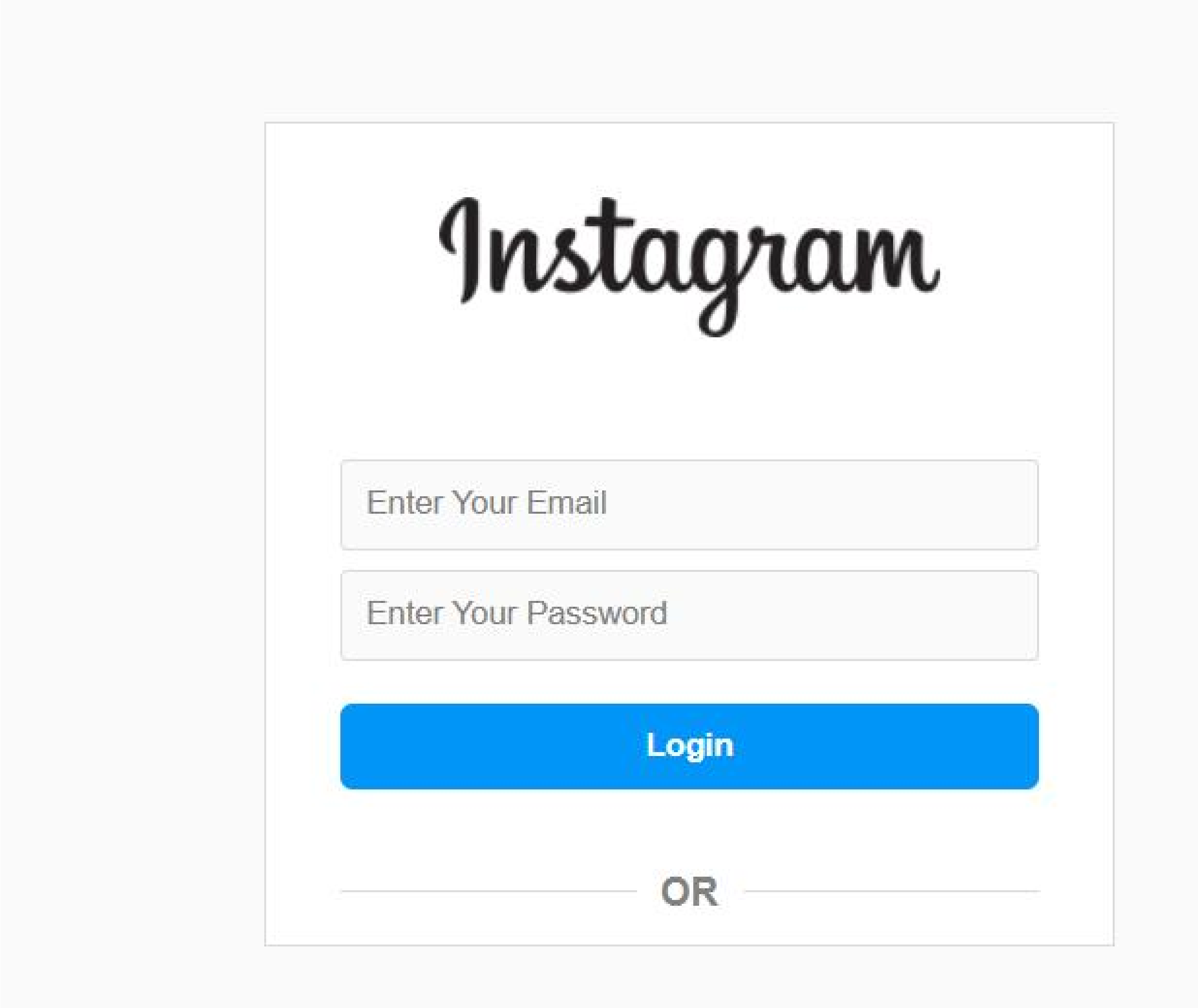
What is a phishing attack?

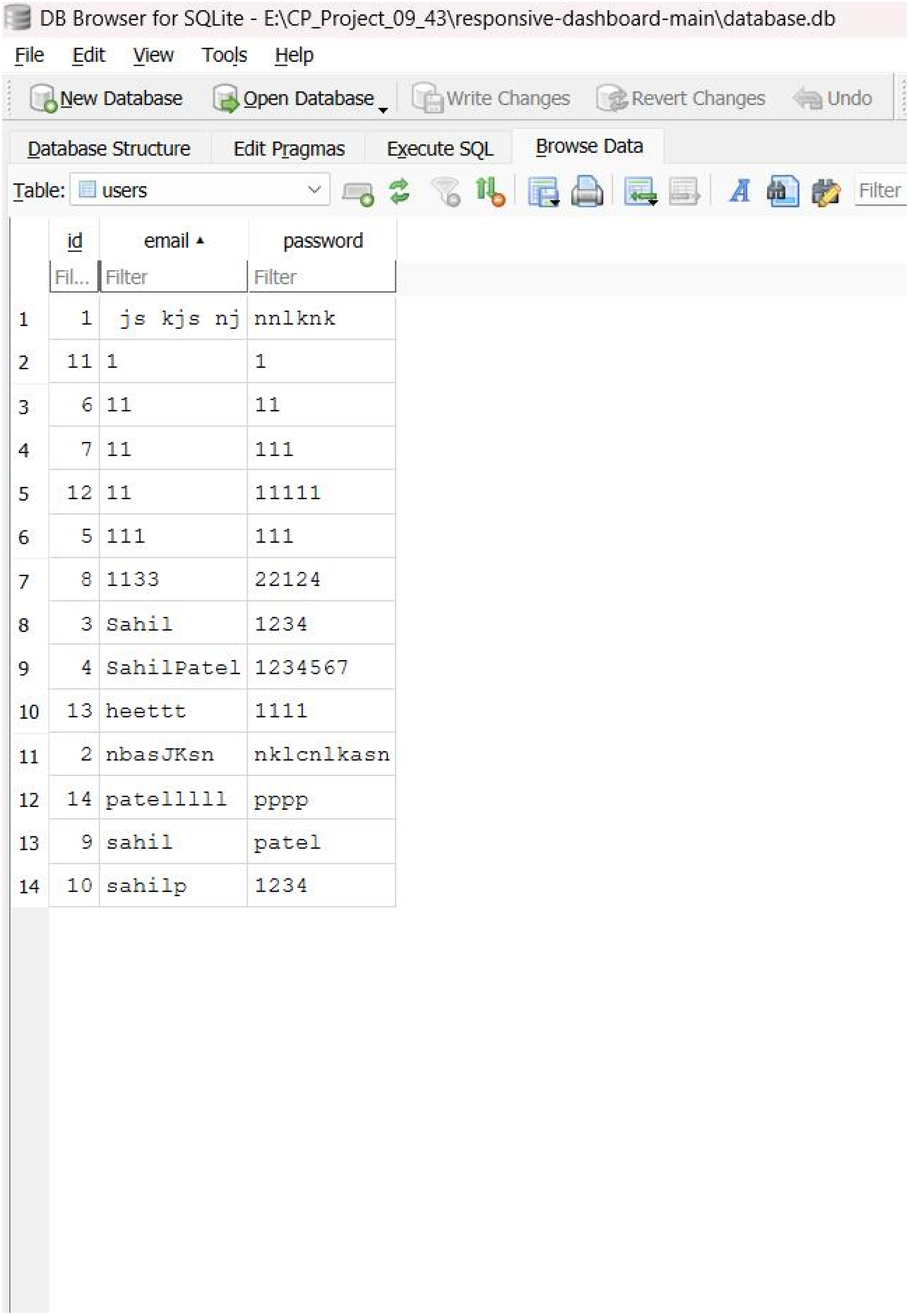


Phishing Attack Prevention: How to Identify & Avoid Phishing Scams

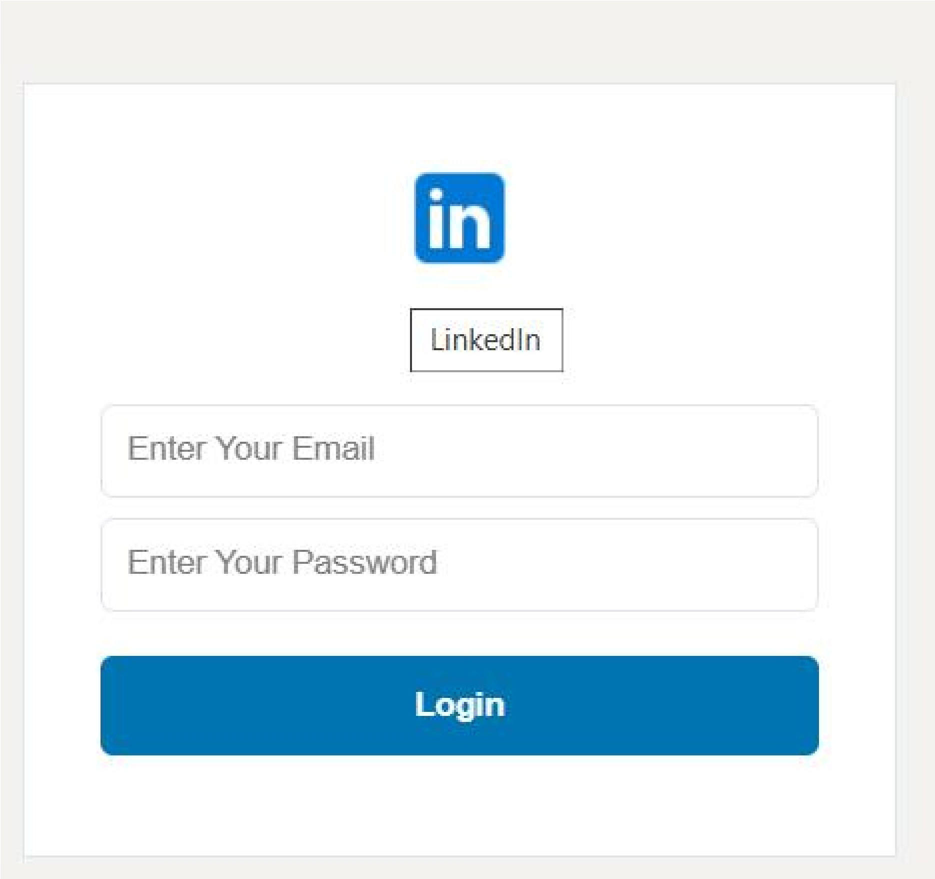
Internet pirates steal personal financial information with a new a type of Internet piracy called phishing, pronounced "fishing," and that's exactly what these thieves are doing: "fishing" for your personal financial information.

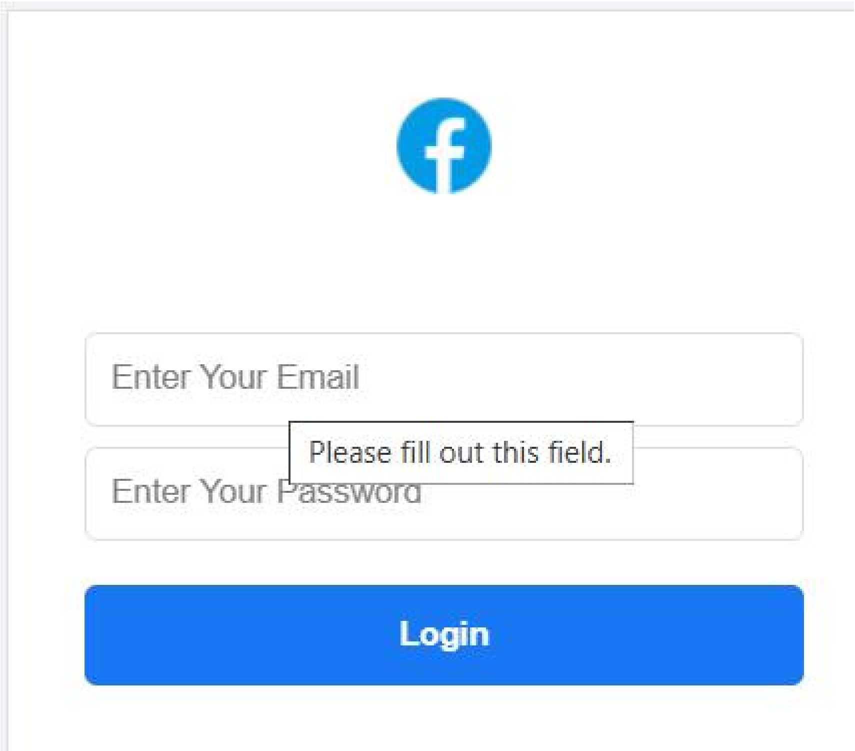
What they want are account numbers, passwords, Social Security



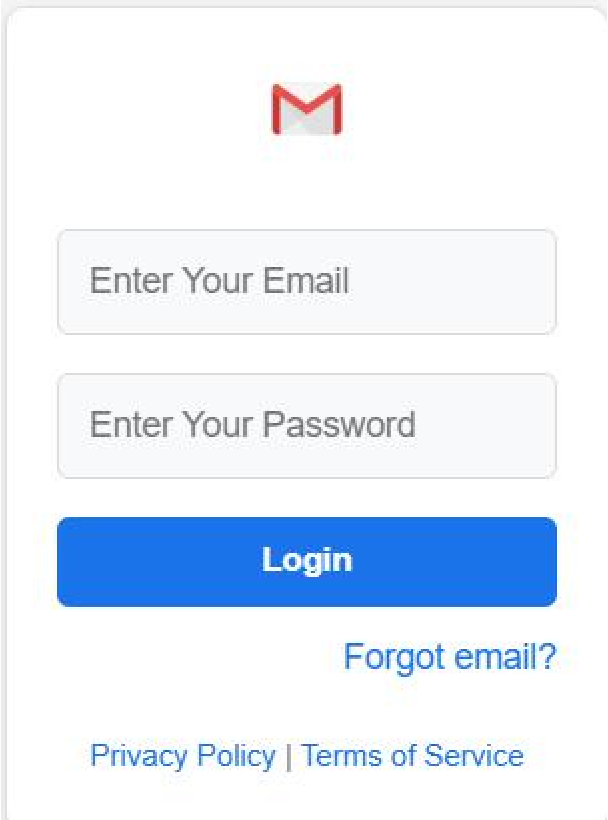


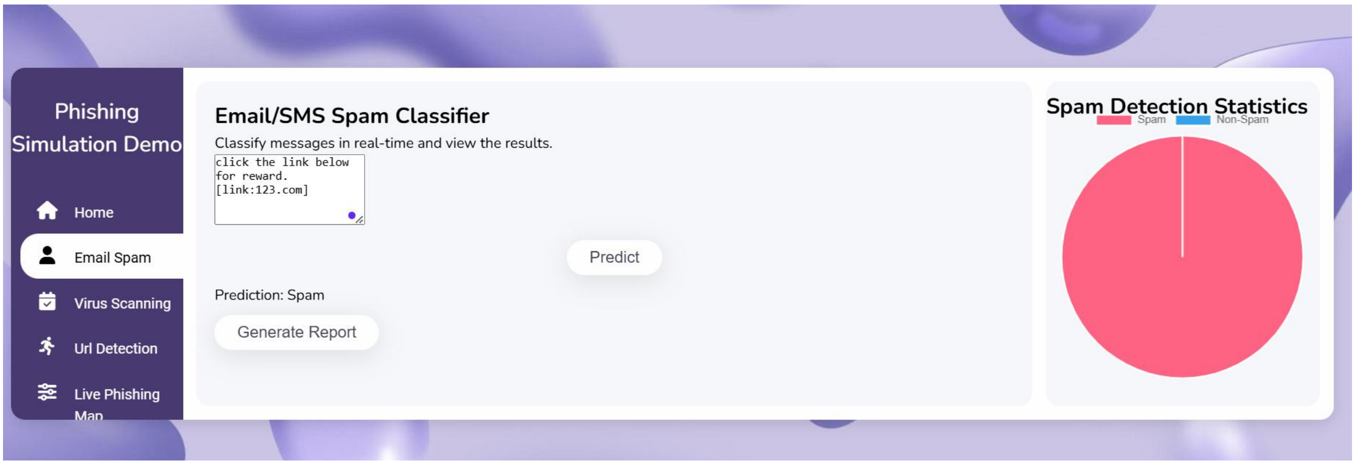
	id	email ▲	password
	Fil...	Filter	Filter
1	1	js kjs nj	nnlknk
2	11	1	1
3	6	11	11
4	7	11	111
5	12	11	11111
6	5	111	111
7	8	1133	22124
8	3	Sahil	1234
9	4	SahilPatel	1234567
10	13	heettt	1111
11	2	nbasJKsn	nklcnlkasn
12	14	patelllll	pppp
13	9	sahil	patel
14	10	sahilp	1234





Please fill out this field.



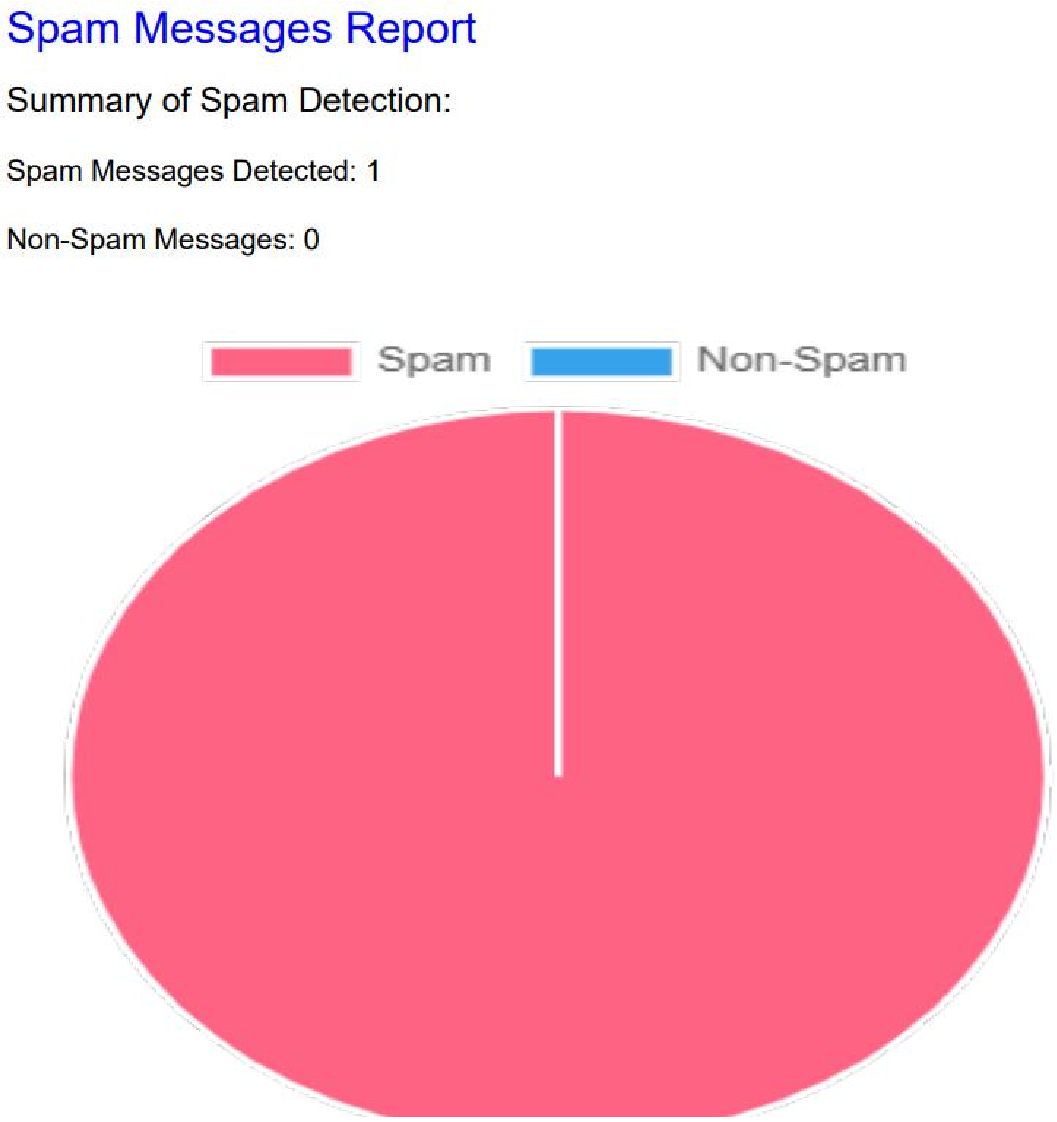


Email/SMS Spam Classifier

Classify messages in real-time and view the results.

```
click the link below
for reward.
[link:123.com]
```

Prediction: Spam

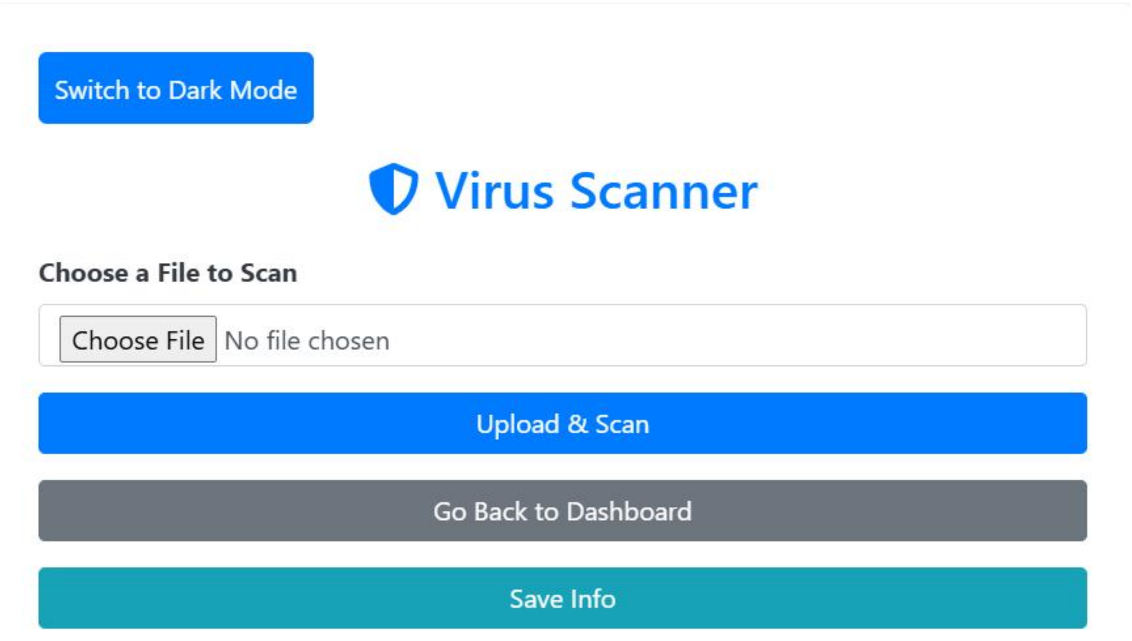


Spam Messages Report

Summary of Spam Detection:

Spam Messages Detected: 1

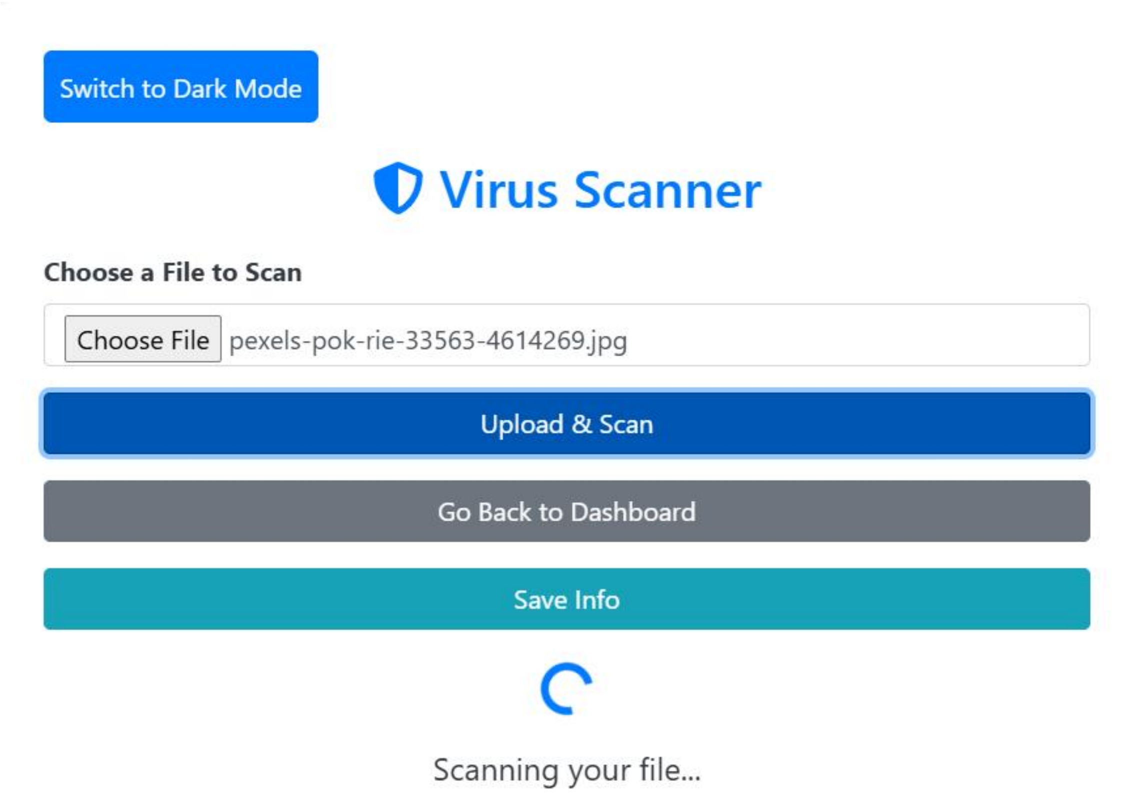
Non-Spam Messages: 0



Virus Scanner

Choose a File to Scan

No file chosen

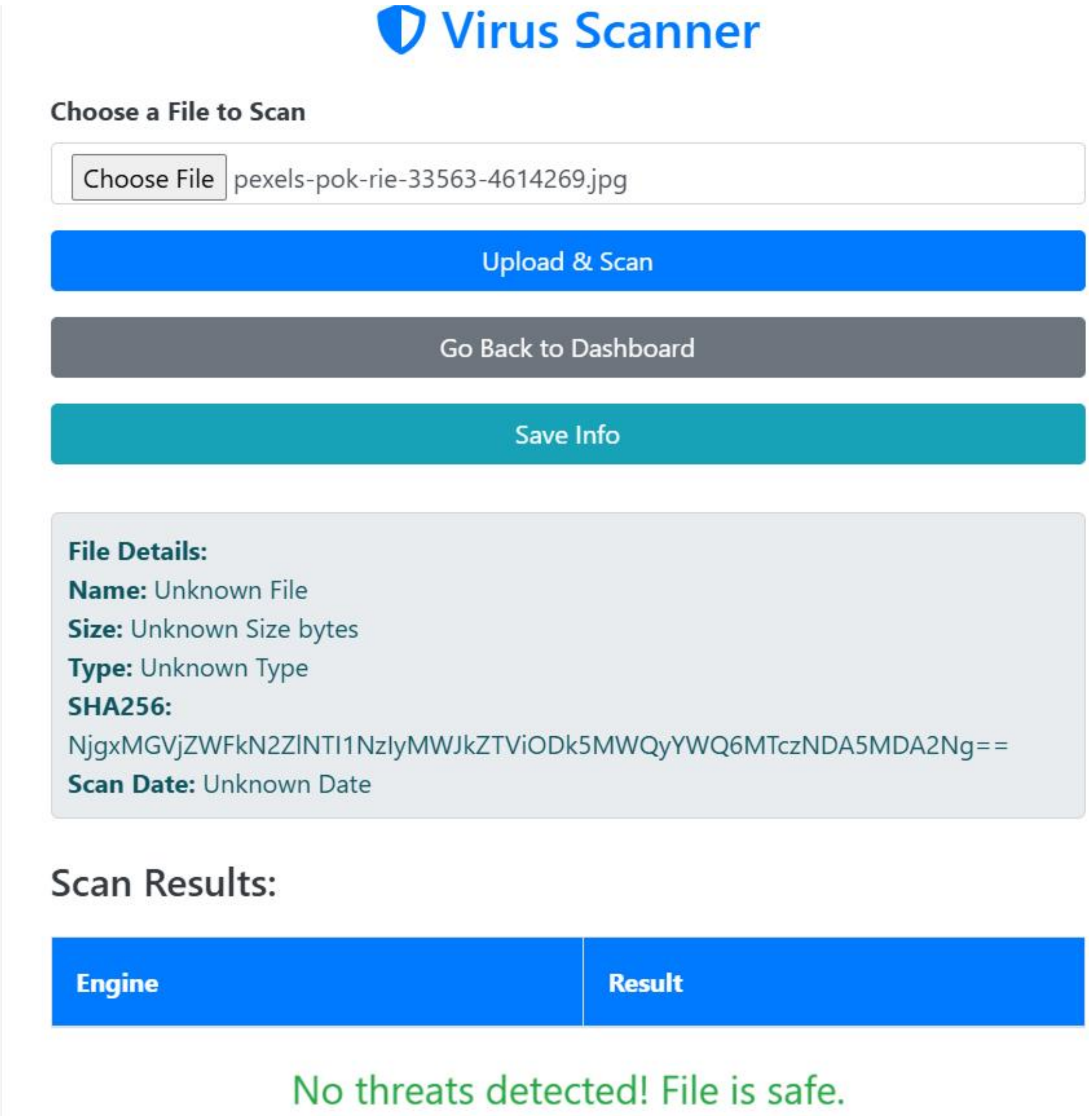


Virus Scanner

Choose a File to Scan

pexels-pok-rie-33563-4614269.jpg

Scanning your file...



Choose a File to Scan

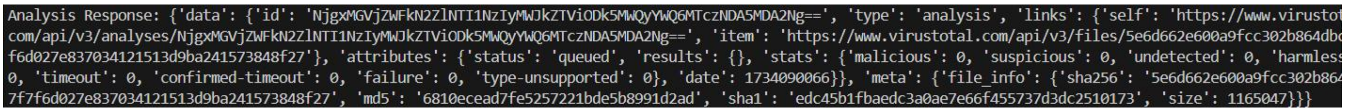
pexels-pok-rie-33563-4614269.jpg

File Details:
Name: Unknown File
Size: Unknown Size bytes
Type: Unknown Type
SHA256:
NjgxMGVjZWFkN2ZlNTI1NzIyMWJkZTViODk5MWQyYWQ6MTczNDA5MDA2Ng==
Scan Date: Unknown Date

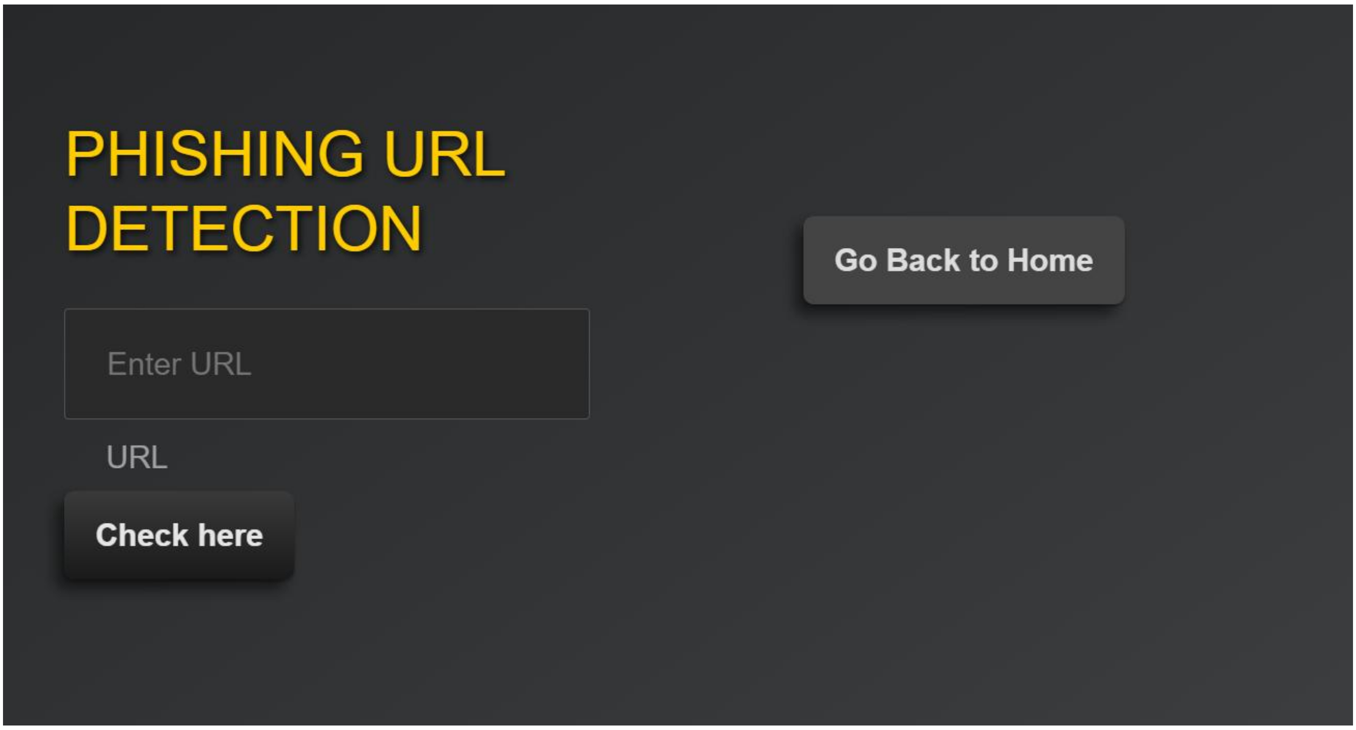
Scan Results:

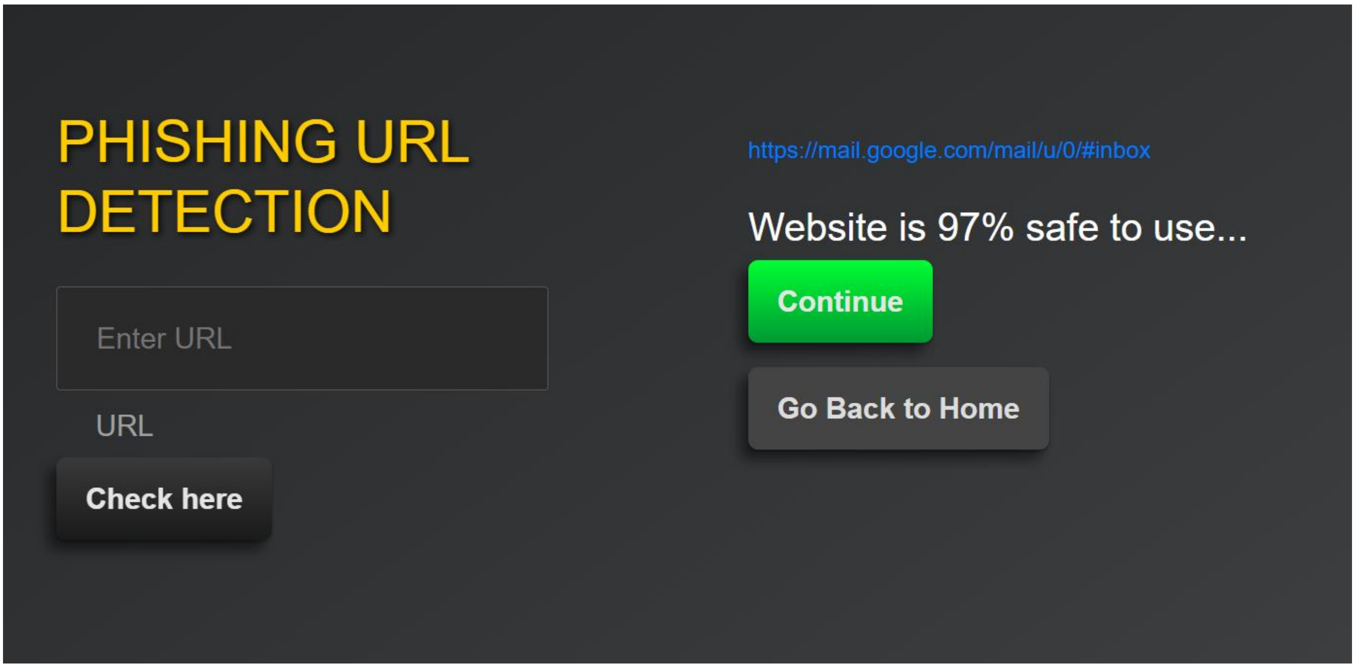
Engine	Result

No threats detected! File is safe.



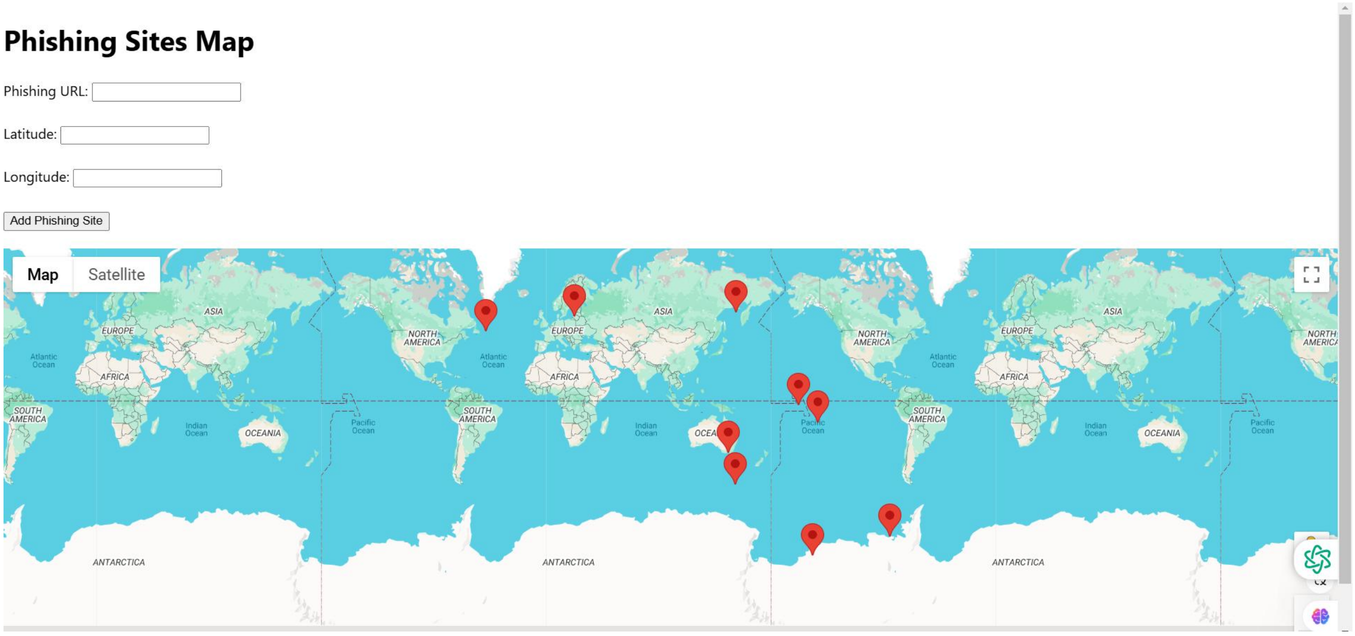
```
Analysis Response: {'data': {'id': 'NjgxMGVjZWFkN2ZlNTI1NzIyMWJkZTViODk5MWQyYWQ6MTczNDA5MDA2Ng==', 'type': 'analysis', 'links': {'self': 'https://www.virustot
com/api/v3/analyses/NjgxMGVjZWFkN2ZlNTI1NzIyMWJkZTViODk5MWQyYWQ6MTczNDA5MDA2Ng==', 'item': 'https://www.virustotal.com/api/v3/files/5e6d662e600a9fcc302b864dbc
f6d027e837034121513d9ba241573848f27'}, 'attributes': {'status': 'queued', 'results': {}, 'stats': {'malicious': 0, 'suspicious': 0, 'undetected': 0, 'harmless
0, 'timeout': 0, 'confirmed-timeout': 0, 'failure': 0, 'type-unsupported': 0}, 'date': 1734090066}}, 'meta': {'file_info': {'sha256': '5e6d662e600a9fcc302b864
7f7f6d027e837034121513d9ba241573848f27', 'md5': '6810ecead7fe5257221bde5b8991d2ad', 'sha1': 'edc45b1fbaedc3a0ae7e66f455737d3dc2510173', 'size': 1165047}}}
```





https://mail.google.com/mail/u/0/#inbox

Website is 97% safe to use...

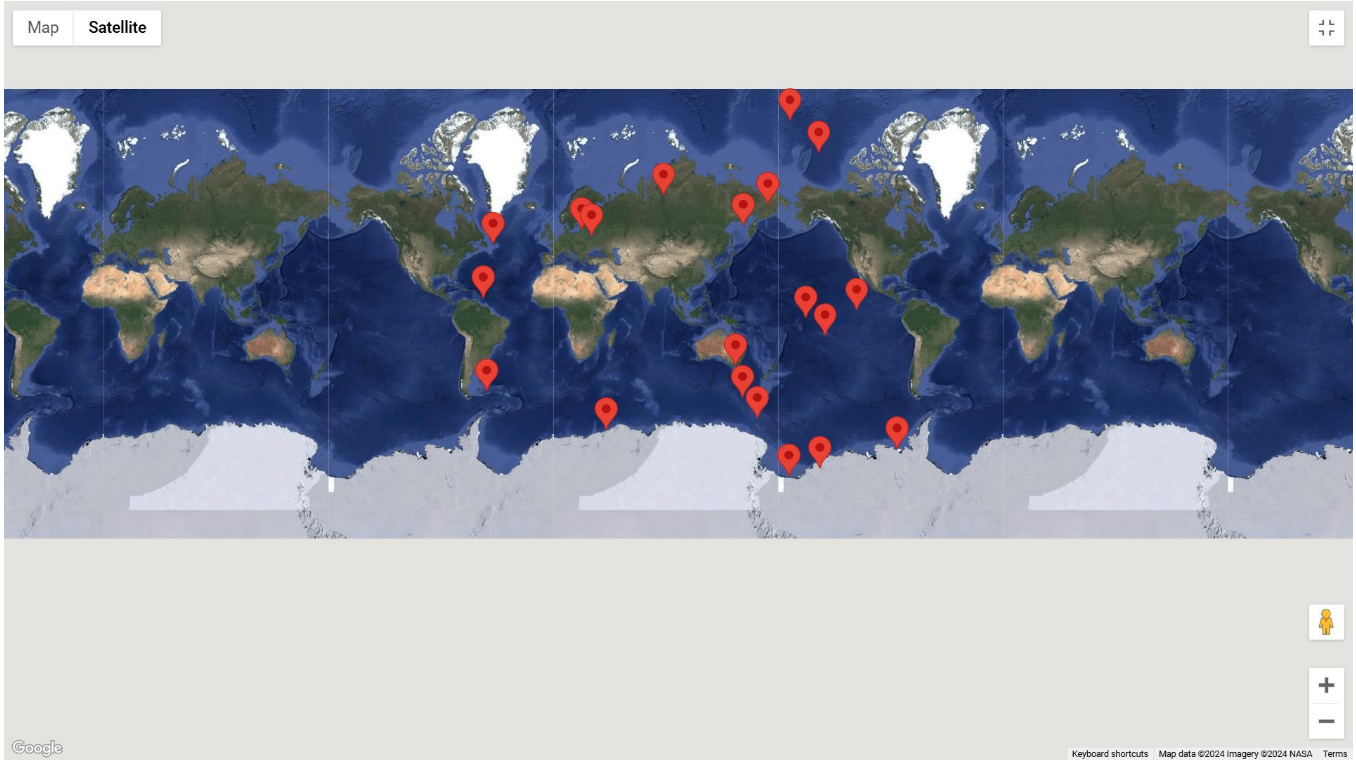


Phishing Sites Map

Phishing URL:

Latitude:

Longitude:



Map
Satellite
Google
Keyboard shortcuts
Map data ©2024 Imagery ©2024 NASA
Terms