Phishing URL Detection Tool Using Machine Learning

Prepared By:
Rujuta Shetkar
Pratiksha Swami

Introduction

- Cybersecurity threats are increasing, and phishing remains a top method used by attackers to steal information.
- This project presents a web-based tool powered by machine learning to identify phishing URLs.
- The tool is designed to be lightweight, fast, and user-friendly.
- It can be used by individuals or integrated into larger systems.
- This presentation walks through the motivation, design, implementation, and results of the tool.

Problem Statement

- Over 90% of cyber attacks begin with phishing.
- Traditional blacklist-based detection methods are limited to known URLs and often miss new, obfuscated threats.
- Phishing sites can appear and disappear quickly, making static defense ineffective.
- There is a pressing need for a smarter, adaptive approach to detect phishing attempts.
- Our project aims to solve this problem using machine learning models trained on URL patterns.

Proposed Solution

- We propose a machine learning-based phishing URL detection tool accessible through a web interface.
- Our tool analyzes lexical features of a URL to predict whether it's legitimate or phishing.
- By using a trained model, we can detect phishing attempts even if the URL is new or modified.
- The system requires no external API calls or large-scale databases.
- It provides fast and accurate results with a simple user experience.
- The solution is scalable, easy to integrate, and highly adaptable to evolving threats.

Code/Tool Breakdown

- The tool consists of several stages starting with the user entering a URL.
- Next, a feature extraction module processes the URL to capture patterns such as length, symbols, domain type, etc.
- These features are passed to a trained machine learning model which predicts whether the URL is phishing or legitimate.
- The result is displayed instantly on the web interface.
- The backend is built using Flask, and the model is pre-loaded for fast prediction.
- This modular design ensures that the tool is lightweight and responsive.

Machine Learning Pipeline

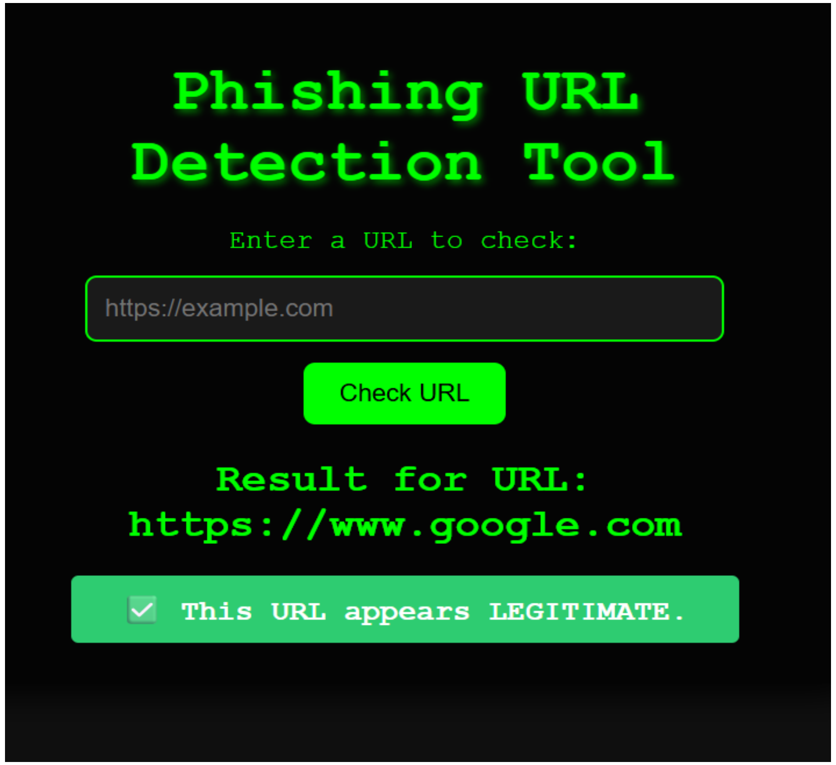
- The core of the system is the machine learning model trained on labeled phishing and legitimate URLs.
- We have used Random Forest algorithm for our experiment.
- Features include presence of IP addresses, url_length, use of hyphens, number of dots, suspicious extensions and phishing keywords.
- Data preprocessing and cleaning were essential to ensure consistent feature formats.
- The model was trained on a dataset with thousands of samples to ensure generalization

Web Interface Screenshot

- Here are some screenshots of the tool in action.
- The home page allows the user to input a URL for checking.
- Once submitted, the result is displayed.
- The interface is clean and intuitive, requiring no technical knowledge to use.
- Users can quickly verify links before clicking, reducing their risk.
- The tool works on both desktop and mobile devices

Screenshot

Legitimate



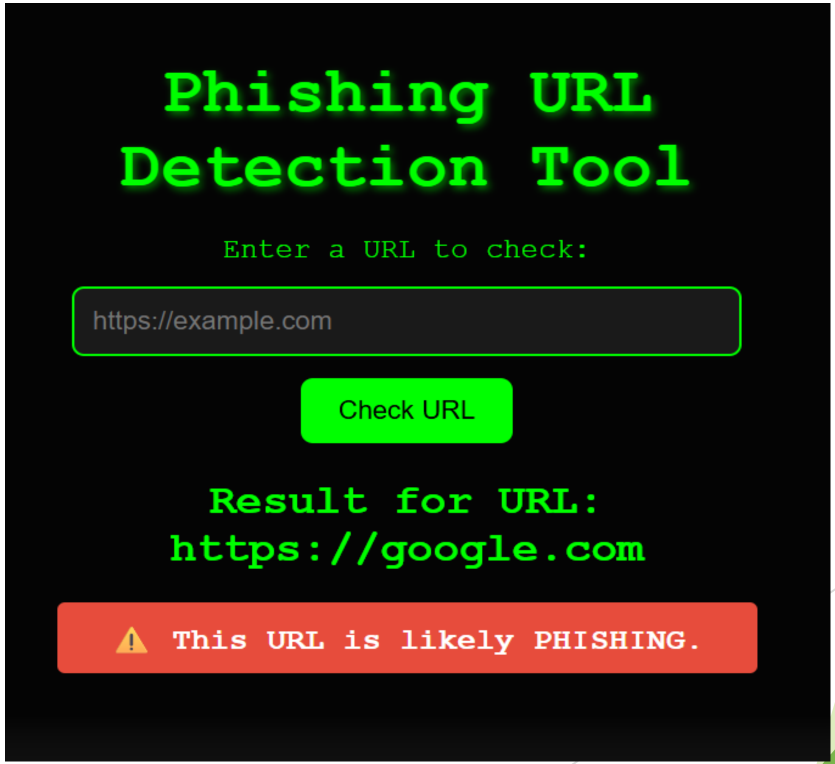
Phishing URL Detection Tool

Enter a URL to check:

Result for URL: https://www.google.com

✅ This URL appears LEGITIMATE.

Phishing



Phishing URL Detection Tool

Enter a URL to check:

Result for URL: https://google.com

⚠️ This URL is likely PHISHING.

Real-World Use Case

- Imagine an employee at a company receives a suspicious-looking email.
- Instead of guessing, they paste the link into our tool to check if it's safe.
- This quick action prevents them from falling into a phishing trap.
- The tool can be integrated into enterprise portals or browser extensions.
- It serves as a preventive layer in a larger cybersecurity strategy.
- This makes it valuable for both individuals and organizations.

Future Enhancement

- While the tool is functional, there is room for future enhancement.
- We plan to expand the dataset with more real-world phishing URLs.
- Image-based phishing detection using screenshots is another next step.
- Integration with real-time threat intelligence feeds can improve accuracy.
- Deploying it as a Chrome extension will offer instant alerts in-browser.
- We also aim to allow multilingual support for global reach.

Thank You!