

Machine Learning Based URL Risk Analyzer and Website Blocker

Submitted By: Paresh Kulkarni

Project Description:

This project implements an AI-driven URL analyzer that predicts whether a website is potentially malicious using Machine Learning. It also includes a system-level website blocker using the Linux `/etc/hosts` mechanism. A Flask-based user interface allows URL scanning, blocking, and unblocking operations.

Technologies Used: Python, Flask, Machine Learning (Logistic Regression), scikit-learn, Linux hosts file.

Introduction

Cybersecurity threats like phishing and malicious websites continue to increase. This project helps detect suspicious URLs and block harmful domains using AI and system-level controls.

Machine Learning Model

A Logistic Regression model is trained on lexical URL features such as length, digits, dots, suspicious keywords, and presence of IP addresses. The model predicts the probability of a URL being malicious.

Flask Web Application

A lightweight Flask app provides a modern UI. Users can check URLs, view risk score, and block/unblock domains.

/etc/hosts Based Blocking

The project safely modifies `/etc/hosts` using a separate helper script with START/END markers.

Conclusion

This system demonstrates the use of ML in cybersecurity and provides a practical website-blocking mechanism.