A Tool for Automating Pre-Penetration Testing Reconnaissance

# Abstract

This project introduces a web-based tool designed to automate the initial phase of penetration testing, known as reconnaissance. Traditionally, this step is manual, time-consuming, and resource-heavy. Our solution leverages web scraping and specialized APIs to streamline data collection, enabling security analysts and developers to focus on addressing vulnerabilities rather than spending hours gathering information.

# Introduction

Cybersecurity threats are constantly evolving, making it critical to secure web systems. Penetration testing plays a key role by simulating attacks to uncover weaknesses before malicious actors exploit them. The first step, reconnaissance, involves collecting as much information as possible about the target system. Currently, this process is mostly manual, which slows down security assessments. Our goal is to develop a web-based tool that automates this step, saving time and improving efficiency.

# Problem Statement

Currently, there is no effective tool that automates the classification of web-based systems based on security needs. Manual processes are inefficient, time-intensive, and prone to human error.

# Objectives

- Develop a web-based tool to classify publicly available web systems.  
- Automate data collection for classification.  
- Generate clear, easy-to-understand reports for individual and grouped systems.

# Scope

In Scope:  
- Building a web application for security asset classification.  
- Automating data collection using web scraping and public APIs.  
- Assessing security requirements using the CIA Triad and Exposure Factor.  
- Generating reports in PDF and CSV formats.  
  
Out of Scope:  
- Hardware deployment or integration with external devices.  
- Native mobile application development.

# Proposed Solution

Our solution is a web-based application that automates security classification using a two-step approach: data collection and analysis. We will use Open-Source Intelligence (OSINT) techniques to gather data from public sources without direct interaction with the target system. Combining web scraping with security-focused APIs like Shodan will provide a comprehensive security profile.  
  
Data Collection: Tools like Jsoup will extract HTML data, while APIs will supply additional security insights.  
Classification: The system will apply the CIA Triad (Confidentiality, Integrity, Availability) and calculate an Exposure Factor to quantify risk.  
  
Technologies:  
- Backend: Python with Flask or Django  
- Frontend: HTML, CSS, JavaScript  
- Libraries: For web scraping, data handling, and report generation  
- OS: Platform-independent, browser-based

# Stakeholders

Primary: Security testers, analysts, and developers  
Secondary: Organizations and academic institutions

# Expected Outcomes

- A functional prototype of the automated classification tool  
- Reduced time and effort for reconnaissance  
- Clear, detailed reports for security assessments  
- Demonstration of automation's impact on security operations

# References

[1] N. Samant, "AUTOMATED PENETRATION TESTING," Master's Project, San Jose State University, 2011.  
[2] V. Singh and K. Jaiswal, "An AI-Based Approach for Automating Penetration Testing," International Journal of Research Publication and Reviews, 2023.  
[3] V. Kumar and S. Singh, "Osint Automation Application," Journal of Physics: Conference Series, 2022.