PhishSCAN

System Documentation

Date: 11 September 2025

Contact: support@phishscan.com | +61 2 9123 4567

Address: Level 12, 1 Market Street, Sydney, NSW 2000, Australia

Documentation

PhishSCAN is a comprehensive phishing URL detection system developed to support individuals, businesses, and educational institutions in the ongoing battle against cybercrime. Unlike traditional security tools that rely solely on static blacklists, PhishSCAN incorporates a hybrid methodology that blends multiple layers of analysis. These include blacklist verification, heuristic analysis of URL structures, SSL validation, and a machine learning model enhanced by explainable AI. The design philosophy behind PhishSCAN is accessibility. The system is delivered both as a standalone web application and a browser extension. The web platform enables batch testing and detailed analysis, while the extension provides on-the-go protection directly in the user's browsing experience. This dual approach makes the system suitable for academic demonstration, enterprise training, and personal use. From a technical perspective, the workflow begins with a user submitting a URL to the system. The backend immediately performs preprocessing steps such as tokenizing the URL, extracting domain age, checking for SSL presence, and flagging suspicious substrings (for example, "login-", "secure-", or use of uncommon TLDs). These extracted features are then transformed into a numerical vector, which the trained machine learning model evaluates to determine the likelihood of phishing. The decision is enriched with SHAP values, which provide explainability by highlighting which features influenced the model's judgment the most. The result is returned in real-time as one of three categories: Legitimate, Suspicious, or Phishing. In addition, a confidence percentage is displayed, enabling users to make informed choices about whether to trust or avoid a website. This transparency is critical, especially in academic contexts where explainability is as important as accuracy. Another distinguishing factor is the session history component. PhishSCAN allows users to review previously scanned URLs in the same session, providing continuity in case of repeated testing. This feature also highlights how classifications may change over time, reflecting the evolving nature of phishing campaigns. While PhishSCAN is highly effective as a prototype, it also acknowledges limitations. It does not guarantee perfect detection and should be viewed as an educational and supportive tool rather than a replacement for enterprise-grade solutions. Future improvements may include larger training datasets, integration with browser-native safe browsing APIs, and advanced clustering of phishing domains to identify campaigns earlier. In conclusion, PhishSCAN demonstrates how modern detection techniques can be combined into a user-friendly platform that both educates and protects. Its architecture balances technical rigor with accessibility, ensuring that security is not reserved solely for specialists but extended to anyone concerned about safe browsing.