Predictive Model Based on Machine Learning for Phishing Detection in URLs and Emails in Peruvian SMEs

Huamani Felix Romina Stephanie

Facultad de Ingeniería de Sistemas de Información, Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Perú

Giancarlo André Román Zamora

Facultad de Ingeniería de Sistemas de Información, Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Perú

Pedro Castañeda

Facultad de Ingeniería de Sistemas de Información, Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Perú

Abstract:

In today's digital environment, small and medium-sized enterprises in Peru are increasingly vulnerable to phishing attacks, representing a significant risk to the security of their customers and operations. This study implements an intelligent filtering system based on machine learning techniques to effectively detect and mitigate phishing attacks targeting these institutions. The XGBoost, LightGBM and Random Forest models were evaluated in terms of accuracy, sensitivity and specificity, with XGBoost standing out with an AUC of 0.99 and an accuracy of 97.8%. The system demonstrated robust performance, classifying malicious URLs with high effectiveness and minimizing false negatives, which is essential for real-time security. In addition, a continuity plan is proposed to ensure the smooth integration of the system in Peruvian SMEs. The developed solution offers a scalable tool that improves the cybersecurity of these companies, protecting the sensitive information of their customers.

Keywords:

Add-On Web, Artificial Intelligence, Machine Learning, Phishing.