Ministry of Higher Education and Scientific Research
University of Babylon
College of Information Technology
Department of Information Security
Study: (Morning)

Classification of Spam URL Links Using Machine Learning

A Graduate Project Submitted to the department of Information Security of the College of Information Technology, University of Babylon, in Partial Fulfillment of the Requirements for the Bachelor's degree in the Information Security of Information Technology.

By

Suhaila Majeed Zaki Alhujaymat

Supervised by

Asst.Lect.Rasha Hussein

2023-2024

Abstract

The Internet is used by billions of users every day because it offers fast and free communication tools and platforms. Nevertheless, with this significant increase in usage, huge amounts of spam are generated every second, which wastes internet resources and, more importantly, users' time. Now-a-days we get the message which contains "content" and "link."With the rapid growthof internet users, people are using them for illegal and unethical conducts, phishing, and fraud. Malicious Web sites are a cornerstone of Internet criminal activi ties. As a result, there has been broad interest in developing sys tems to prevent the end user from visiting such sites ,This research project delves into the critical realm of spam URL classification, aiming to bolster cybersecurity measures and safeguard internet users against malicious online activities. The study focuses on employing two powerful machine learning algorithms, Support Vector Machine (SVM) and Random Forest, to distinguish between legitimate and spam URLs. A vast database containing a diverse array of URLs was meticulously curated for this research endeavor. feature extraction techniques were employed to capture relevant characteristics intrinsic to URLs, enabling effective distinctions between legitimate and spam instances. The SVM and Random Forest algorithms were then trained and evaluated using the processed dataset, employing standard evaluation metrics such as precision, recall, F1-score, and support. The results yielded valuable insights into the strengths and weaknesses of each algorithm in the context of spam URL classification. We verify that the model designed in this paper has the highest accuracy (97%) in detecting malicious URL through these experiences.