ISBN : 978-93-342-8169-9

AN PHISING URL CLASSIFICATION AND DETECTION USING MACHINE LEARNING

Mr.M.Kamarajan Assistant Professor Computer Science and Engineering PSNA College of Engineering and Technology Dindigul, India m.kamarajan@psnacet.edu.in
Ajaykumar J Computer Science and Engineering PSNA College of Engineering and Technology Dindigul, India ajaykumarj21cs@psnacet.edu.in
Abivarman K Computer Science and Engineering PSNA College of Engineering and Technology Dindigul, India abivarmank21cs@psnacet.edu.in
Godwin rex C Computer Science and Engineering PSNA College of Engineering and Technology Dindigul, India godwinrexc21cs@psnacet.edu.in
Aribhaskar M Computer Science and Engineering PSNA College of Engineering and Technology Dindigul, India aribhaskarm21cs@psnacet.edu.in

Abstract– The phishing URL classification and detection system utilizes machine learning techniques to accurately distinguish between legitimate and malicious URLs, enhancing cybersecurity measures. The input dataset, consisting of labeled URLs, undergoes preprocessing steps such as data cleaning, feature extraction, and label encoding. Key features are derived from URL structure, domain characteristics, and content-based attributes. The dataset is then split into training and testing sets for model evaluation and prediction. The system employs a hybrid machine learning approach, integrating Random Forest (RF) and Multi-Layer Perceptron (MLP) algorithms for improved classification accuracy. Model performance is assessed using evaluation metrics such as accuracy, precision, recall, and F1-score. Additionally, the system provides real-time threat detection, automated alerts, and a reporting module to help users and organizations mitigate phishing attacks effectively.

Keywords–Phishing Detection, URL Classification, Machine Learning, Random Forest, MLP, Cybersecurity