IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 13, April 2025

Impact Factor: 7.67

Malicious URL Analysis with Real Time Blocking System

Sham P[1], Giritharan S[2], Kishore S[3], Manikandan AR[4]

Department of Computer Science and Engineering[1-4]

Mahendra institute of Engineering and Technology, Salem, India

Abstract: *This research presents a machine learning-based framework aimed at detecting phishing websites in real-time. Phishing is a growing concern in the cybersecurity domain, threatening user data and privacy globally. Our system employs advanced ML models including Ada boost to analyze web content and domain-level features. The use of Selenium provides real-time protection by blocking phishing sites. This work contributes a practical and scalable solution to enhance online safety by integrating automation and intelligent detection. Ethical practices and data transparency are upheld throughout the development lifecycle, ensuring responsible AI use in cybersecurity*

Keywords: machine learning

I. INTRODUCTION

Phishing attacks are increasingly sophisticated and threaten millions of users each year. They exploit human behavior by masquerading as legitimate websites to steal personal information such as usernames, passwords, and credit card details. The growing dependency on digital platforms has amplified the reach of such threats. Traditional detection methods, which rely on blacklists and heuristic rules, are often reactive and unable to identify new forms of phishing attacks. This underscores the need for intelligent, data-driven approaches capable of adapting to emerging threats in real time. Machine learning (ML) offers promising avenues for automating detection by analyzing patterns and anomalies within URLs, website structures, and domain metadata. In this paper, we outline our comprehensive ML-based framework, which is designed to improve detection rates and offer immediate protection through browser automation.

II. LITERATURE REVIEW

Various studies have been conducted to explore the capabilities of ML in phishing detection. Amani Alswailem et al. used Random Forest and achieved high classification accuracy by selecting the most relevant features. IEEE Access proposed using login page URLs to simulate real-world phishing cases more effectively. Adarsh Mandadi et al. examined traditional classifiers like Decision Trees and Random Forests, while Mahmoud Atari employed XGBoost and achieved 97% accuracy. Other researchers have explored deep learning approaches, such as RNN-GRU models, which can model temporal dependencies in phishing behavior. Feature extraction techniques varied across studies, from lexical analysis of URLs to analyzing HTML structures and domain-level attributes. The common thread across all research is the reliance on large datasets and the importance of updating models regularly to counteract evolving phishing techniques. Our work builds upon these studies by integrating practical real-time blocking mechanisms and expanding feature sets.

III. RESEARCH METHODOLOGY

The methodology followed in this study involves data acquisition, feature engineering, model training, and deployment. A curated dataset of legitimate and phishing URLs is gathered from various open-source repositories and verified sources. Feature engineering includes extracting lexical features (length, special characters), domain-related features (age, WHOIS), and content features (HTML tag frequency, anchor tags). The dataset is divided into training and testing sets with proper cross-validation to ensure the model's generalization capability. We implement and compare multiple algorithms, including Random Forests for robustness, Gradient Boosting for high accuracy, Deep Neural Networks for learning complex patterns, and KNN for simplicity and efficiency. Hyperparameter tuning is conducted using grid search. Post-training, the models are evaluated using accuracy, precision, recall, and F1-score. Selenium is integrated to act upon model predictions in real time, automating the process of warning users and blocking malicious content. The methodology ensures a balance between theoretical modeling and real-world application.

TABLE 1 PERFOMENCE RESULT

Algorithm	Accuracy	precision	recall	F1 score
logistic regression	**94.2%**	**93.5%**	**94.8%**	**94.1%**
decision tree	**92.7%**	**91.0%**	**93.0%**	**92.0%**
adaptive boost	**96.3%**	**95.5%**	**96.9%**	**96.2%**

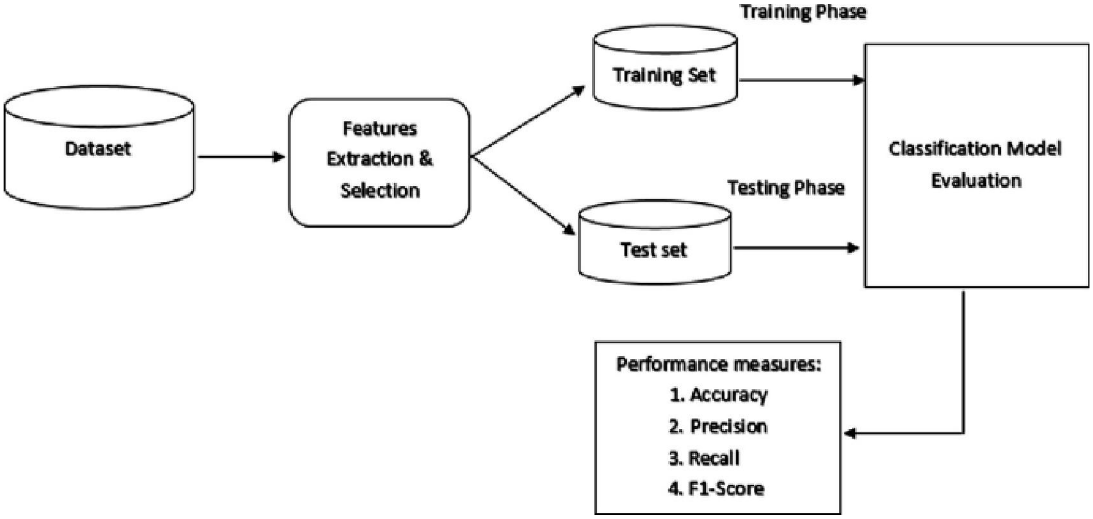

Fig.1. malicious detecting model

IV. EXPERIMENTAL RESULTS

The models achieved competitive accuracy with Random Forest and Gradient Boosting showing over 95% accuracy and precision. Deep Neural Networks displayed improved performance in detecting subtle phishing patterns but required more computational resources.

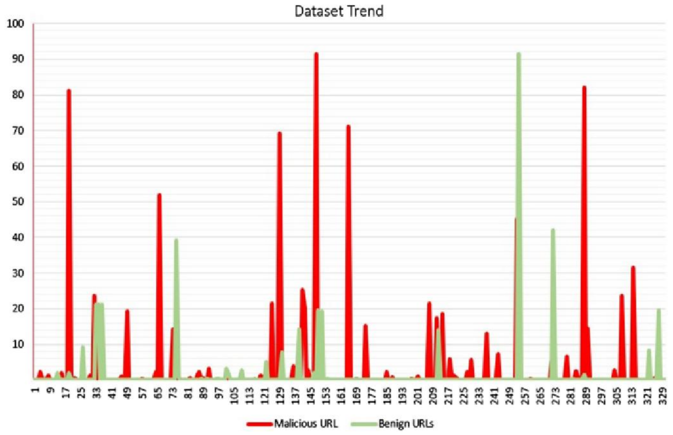

FIG.2.THE DISTRIBUTION OF VALUES

Ada boost offered fast predictions with slightly lower accuracy, making it suitable for lightweight systems. Cross-validation confirmed that the models are not overfitting and generalize well to unseen data. Selenium successfully blocked phishing sites detected by the model in real time during test simulations. The integration of automated blocking reduces human response time and increases system reliability. This experiment validates the use of ML models in live environments where threat detection and action must occur within seconds.

V. CONCLUSION AND FUTURE WORK

This research demonstrates that machine learning, combined with real-time automation tools like Selenium, provides an effective solution for phishing website detection. By creating a robust dataset and carefully selecting features, we improved the model's performance and ensured real-world applicability. The solution has immediate uses in browser security extensions, corporate email gateways, and enterprise-level firewalls. Looking forward, our system can be expanded to cover other cyber threats such as malware, ransomware links, and social engineering attacks. Enhancing our model with deep learning approaches like Transformers, and integrating threat intelligence feeds, can further boost performance. Real-time updates, user feedback loops, and cloud-based deployment are other directions that promise scalability and adaptability.

REFERENCES

[1]. Amani Alswailem et al., 'Detecting Phishing Websites Using Machine Learning', ICCAIS, 2019.

[2]. 'Phishing URL Detection: A Real-Case Scenario Through Login URLs', IEEE Access.

[3]. Adarsh Mandadi et al., 'Phishing Website Detection Using Machine Learning', I2CT, 2022.

[4]. Huaping Yuan et al., 'Detecting Phishing Websites and Targets', ICPR, 2018.

[5]. Joby James et al., 'Detection of phishing URLs using ML techniques', ICCC, 2013.

[6]. Mahmoud Atari et al., 'ML-Based Approach for Detecting Phishing URLs', IDSTA, 2022.

[7]. Moulana Mohammed et al., 'Phishing Detection Using ML Algorithms', ICSSIT, 2022.

[8]. 'Deep Learning-Based Framework for Phishing Detection', IEEE Access.

[9]. Mohammed Nazim Feroz et al., 'Phishing URL Detection Using URL Ranking', Big Data, 2015

DOI: 10.48175/IJARSCT-26046