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 12, April 2025
Impact Factor: 7.67

A Review on Phishing Detection using AI and ML

Prof. Pratiksha Prakash Pansare[1], Lokhande Ritesh Motiram[2], Thakre Prathmesh Sanjay[2], Zolekar Avishkar Bharat[2]

[1] Assistant Professor, [2]Students, Department of Computer Engineering,
Samarth College of Engineering and Management, Belhe, Junnar, Pune, Maharashtra, India

Abstract: *Phishing websites are a major threat to online security, aiming to deceive users into revealing confidential information by imitating legitimate websites. Detecting such fraudulent websites is crucial to safeguarding users from potential harm. This paper proposes an intelligent model for detecting phishing websites based on Extreme Learning Machine (ELM). Phishing websites exhibit various distinguishing features, and therefore, detecting them requires an appropriate set of URL features. Our model employs machine learning techniques to classify web pages as phishing or legitimate, utilizing a dataset containing phishing and legitimate URLs. The methodology involves preprocessing a dataset of URLs, followed by the extraction of features from four key categories: domain-based, address-based, abnormal behavior-based, and HTML/JavaScript-based features. These features are processed to generate values for each URL attribute, which are then analyzed using machine learning algorithms, including ELM, Random Forest, and Support Vector Machines (SVM). The system computes range and threshold values for URL attributes to aid in classification..*

Keywords: Phishing detection, Extreme Learning Machine (ELM), Machine learning, URL feature extraction, Phishing websites, Legitimate websites, Web security, Phishing classification

I. INTRODUCTION

The project focuses on the detection of phishing sites, which are malicious websites designed to deceive users into revealing sensitive information. Phishing attacks have become increasingly sophisticated, making it challenging for traditional rule-based and blacklist approaches to keep up with the evolving techniques. Therefore, there is a need for more advanced methods like machine learning to combat this growing threat.

Machine learning offers a promising approach to tackle the problem of phishing site detection. By leveraging machine learning algorithms, patterns and characteristics indicative of phishing activities can be learned and utilized to distinguish between legitimate and malicious websites. Features such as URL properties, website content, and user behavior can be analyzed to identify suspicious patterns and anomalies associated with phishing attacks

II. LITERATURE REVIEW(TABLE FORMAT)

Title	Authors	Years	Methodology
Phishing URL detection using machine learning methods	SK Hasane Ahammad a , Sunil D. Kale b , Gopal D. Upadhye b , Sandeep Dwarkanath Pande c,* , E Venkatesh Babu a , Amol V. Dhumane b , Mr. Dilip Kumar Jang Bahadur d	2022	involves applying machine learning techniques to analyze and classify URLs to detect phishing attempts.
Phishing Detection and Prevention using Chrome Extension	M. Amir Syafiq Rohmat Rose; Nurlida Nabila g; Nurzi Basir; Nur Fatin Mohd Rafie Hen Mohd Juana Zaizi; Madihah	2022	The methodology used in the proposed model involves the implementation of a self- destruct detection algorithm that employs supervised machine learning techniques, specifically focusing

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 12, April 2025
Impact Factor: 7.67

	Saudi		on URL-based web characteristics for phishing detection and prevention.
Phishing Site Detection Using Similarity of Website Structure.	Suleiman Y. Yerima, K. Mohammed Alzaylaee.	2021	It Proposes feature selection method are also used to increase the accuracy of classification model by selecting best feature & result .
An Intelligent System for Phishing Attack Detection and Prevention	N Megha, K R Remesh Babu Elizabeth Sherly	2020	Implementing a multi agent based architecture and ML classifier for detecting and rectifying web phishing attacks

III. PROPOSED SYSTEM

The proposed system for phishing website detection leverages machine learning techniques to classify websites as phishing or legitimate based on URL and webpage features. The system follows three main stages: data collection and preprocessing, feature extraction and selection, and phishing detection using machine learning models. It collects a dataset of phishing and legitimate URLs, preprocesses the data, and extracts key features such as URL length, suspicious keywords, domain age, SSL usage, and HTML content.

The system applies three machine learning algorithms: Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Random Forest (RF), to classify the URLs. It evaluates models using metrics like accuracy, precision, recall, and F1-score, ensuring high detection performance. Once trained, the system can perform real-time phishing detection by classifying new URLs and providing a phishing risk prediction. The system also includes a feedback loop for continuous model retraining, ensuring it adapts to evolving phishing techniques.

The main components include URL feature extraction, model training, real-time classification, and a user interface for URL input. The system is scalable, accurate, and flexible, with the ability to handle large datasets and integrate new features. Overall, the system provides an efficient and adaptive solution for protecting users from phishing attacks.

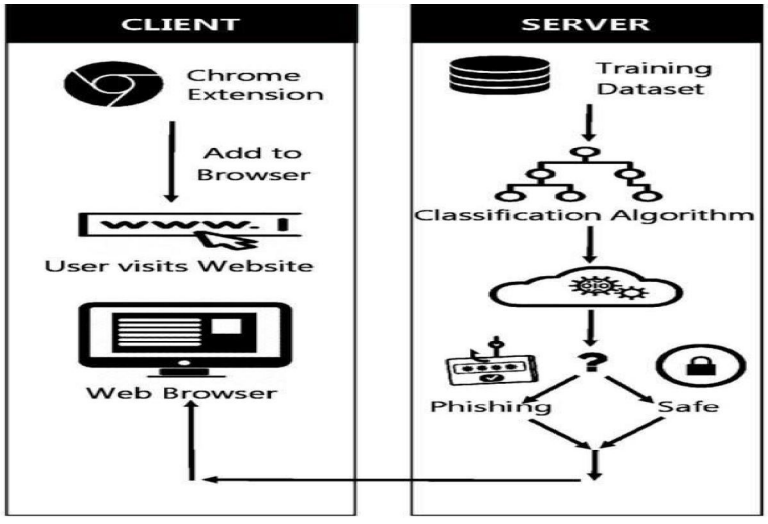

Figure 1. System Architecture

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 12, April 2025
Impact Factor: 7.67

IV. SYSTEM REQUIREMENTS

SOFTWARE REQUIREMENT:

- IDE: Python Idle and Jupyter notebook
- Platform used: Microsoft Windows

HARDWARE REQUIREMENT:

- Processor: Pentium Processor Core 2 Duo or Higher
- Hard Disk: Min 250 GB
- Processor Speed: 3.2 Ghz or faster Processor

V. SCOPE

The scope of smart waste segregation dustbins includes:

- Real-time Phishing Detection
- Enhanced Accuracy and Precision
- Scalable Solutions for Large-Scale Systems
- Fraudulent Email Detection
- Protection Against Zero-Day Attacks
- Integration with Existing Security Systems
- Reduction of False Positives and Negatives
- Personalized Security for End Users

VI. CONCLUSION

Phishing site detection projects have emerged as a promising solution to combat these attacks by identifying malicious websites that are designed to steal sensitive information from users.

While phishing site detection projects have their advantages and disadvantages, they can be a valuable tool in preventing users from falling victim to phishing attacks.

In conclusion, phishing site detection projects can play a critical role in protecting users and organizations from the devastating impact of phishing attacks.

Thus it is essential to continue developing and implementing new technologies and strategiesto stay ahead of cybercriminals and keep users' sensitive information safe.

REFERENCES

[1] Y. Zhang, et al., "A Survey of Machine Learning Techniques for Phishing Detection," IEEE Access, 2021.

[2] A. Gupta, et al., "Phishing Website Detection Using Random Forest Algorithm," Proceedings of the International Conference on Cyber Security and Cloud Computing, 2019.

[3] S. Patel, et al., "Phishing Detection Using Extreme Learning Machine and Feature Engineering," International Journal of Computer Applications, 2020.

[4] P. S. Srinivas, et al., "Detecting Phishing Websites Using Support Vector Machines and Decision Trees," International Journal of Computer Science & Engineering, 2018.