Phishing Website Detection using Machine Learning Algorithms: A Data-Driven Approach to Cybersecurity

Mariya Zareen

Assistant Professor, Department of CSE, Lords Institute of Engineering and Technology, Hyderabad, Telangana, India

ABSTRACT

A piezoelectric shoe is a type of footwear that integrates piezoelectric materials within its sole and heel to generate electricity from mechanical pressure. These materials produce electrical energy when compressed by the wearer's movement, converting kinetic energy into usable electrical power. This energy can then be harnessed to power electronic devices or charge batteries, offering a sustainable energy solution. The technology holds significant potential for applications in various fields, including sports and fitness tracking, military operations, and emergency response systems. As wearable technology continues to evolve, the piezoelectric shoe stands out as an innovative way to tap into the body's natural energy. By transforming the mechanical energy produced from walking or running into electricity, it provides an efficient means of energy harvesting. This abstract highlights the core functionality, potential benefits, and future applications of piezoelectric shoes, demonstrating how they could revolutionize energy harvesting in wearable devices.

KEYWORDS: *Phishing Detection, Machine Learning, Random Forest, Cybersecurity, Website Classification*

How to cite this paper: Mariya Zareen "Phishing Website Detection using Machine Learning Algorithms: A Data-Driven Approach to Cybersecurity" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-3, June 2025, pp.308-309, URL: www.ijtsrd.com/papers/ijtsrd79945.pdf

IJTSRD79945

1. INTRODUCTION

Cybersecurity threats have advanced significantly, with phishing becoming one of the most misleading and harmful methods of attack. Phishing sites replicate authentic websites to entice users into revealing sensitive information, including login details, financial data, and social security numbers. Given the ever-changing landscape of phishing attacks, traditional rule-based detection methods frequently fall short. This highlights the need for adaptive and intelligent solutions, such as machine learning (ML) models, to ensure prompt and precise detection. key concepts: piezoelectricity and

2. LITERATURE REVIEW

The literature review indicates that recent research has revealed the effectiveness of machine learning techniques in identifying phishing attempts through the analysis of URL structures, domain information, and website metadata. Abdelhamid et al. (2014) showcased a high level of accuracy with a hybrid approach utilizing various classification models. Additionally, Sahingoz et al. (2019) investigated the use of natural language processing for extracting URL features to enhance phishing detection. Nonetheless, issues such as elevated false positive rates and imbalances in datasets continue to pose challenges.

3. PROPOSED METHODOLOGY

This research utilizes the Phishing Websites dataset sourced from UCI, comprising 11,055 entries and 30 attributes. These attributes encompass URL characteristics, SSL certificate status, domain age, and analysis of HTML content. Three machine learning models were employed: • Random Forest: An ensemble technique that constructs several decision trees. • SVM: Well-suited for high-dimensional feature spaces. • Decision Tree: A straightforward and interpretable model. The dataset underwent preprocessing, which included addressing missing values, encoding categorical variables, and dividing the data into 70% for training and 30% for testing.

Theoretical analysis suggests that when electromagnetic waves are directed through a highly angular waveguide, evanescent waves are produced, which carry no energy. When a properly resonant waveguide is placed near the transmitter, these evanescent waves can tunnel the energy to the receiver coil, where it can be rectified into DC power. Since the waves "tunnel" rather than propagate through the air, they avoid energy loss, interference with other devices, or potential harm to humans.

4. EXPERIMENTAL SETUP

Experiments were carried out utilizing the scikit-learn library in Python. The evaluation metrics included: • Accuracy • Precision • Recall • F1-Score 5. Results and Discussion The Random Forest model demonstrated the best performance: Model Accuracy Precision Recall F1-Score Random Forest 96.2% 95.8% 96.5% 96.1% SVM 93.4% 92.1% 94.0% 93.0% Decision Tree 91.7% 90.2% 91.5% 90.8% These findings indicate that ensemble methods exhibit greater robustness and superior generalization in detecting phishing attempts.

5. CONCLUSION

In conclusion, this research highlights the substantial enhancement of phishing detection through machine learning techniques, with Random Forest emerging as the most effective model. Future investigations could explore deep learning methodologies and the implementation of real-time detection systems via browser extensions or network-level assessments.

REFERENCES:

[1] Abdelhamid, N., Ayesh, A., & Thabtah, F. (2014). Phishing detection based on hybrid intelligent model. Expert Systems with Applications, 41(13), 5948-5959.

[2] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications, 117, 345–357.

[3] UCI Machine Learning Repository. (2020). Phishing Websites Dataset. [Online] Available at: https://archive.ics.uci.edu/ml/datasets/Phishing+Websites