JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Phishing Website Detection Using Machine Learning

Ms. Tumula Suma[*1], Mr.CH.Srinivas Reddy[*2]

[1]MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
[2]Assistant Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.

Abstract:

Phishing website detection, utilizing machine learning algorithms, stands as a crucial defense mechanism against online scams and fraudulent activities, aiming to protect users from malicious endeavors. This abstract offers an overview of the applied methodology in detecting phishing websites through machine learning techniques. Phishing, characterized as a deceptive practice, involves attackers creating fraudulent websites that closely mimic legitimate ones, deceiving users into divulging sensitive information such as login credentials, financial data, or personal details.In response to this pervasive threat, researchers and cybersecurity experts have turned to machine learning as a potent solution. Various machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Logistic Regression, etc., are applied to a feature-rich dataset. This dataset encompasses attributes like URL structure, content analysis, SSL certificate details, and more. The machine learning algorithm learns from labeled data, paving the way for the development of a predictive model proficient in distinguishing between phishing and legitimate websites.Phishing attacks have escalated into a significant cybersecurity threat, posing substantial risks to individuals and organizations on a global scale. This research introduces an innovative approach to detecting phishing websites, leveraging the capabilities of machine learning algorithms. The primary goal is to establish a robust and adaptive system, exhibiting high accuracy in identifying fraudulent websites. The envisioned outcome is the creation of a defense mechanism that effectively safeguards users from falling victim to deceptive online practices.

Keywords: Phishing Detection, Machine Learning Algorithms, Cybersecurity, Fraudulent Websites, Predictive Model, Feature-Rich Dataset, Online Security.

1. INTRODUCTION

In the era of widespread internet usage and increasing dependence on online platforms, the escalation of cyber threats has emerged as a paramount concern for both individuals and organizations. Among the array of cybercrimes, phishing attacks have evolved as one of the most prevalent and deceptive methods employed by malicious actors. Phishing entails the creation of fraudulent websites that closely emulate legitimate ones, with the nefarious aim of deceiving unsuspecting users into divulging sensitive information, including login credentials, financial data, and personal details.

Conventional methods of phishing detection, such as rule-based heuristics and signature-based approaches, have proven insufficient in keeping pace with the rapidly evolving tactics employed by attackers. As phishing techniques become more sophisticated and elusive, there arises a pressing need for smarter and more adaptive solutions to effectively combat this ever-growing threat.

In recent years, machine learning has showcased its prowess in addressing complex challenges across diverse domains, notably in the realm of cybersecurity. Machine learning algorithms possess the capability to autonomously learn from extensive datasets, identify intricate patterns, and make data-driven predictions. These inherent attributes position machine learning as an ideal candidate for creating intelligent phishing website detection systems.

The primary objective of this research is to implement a phishing website detection system utilizing machine learning algorithms. By harnessing the power of machine learning, our aim is to construct a robust and proactive defense mechanism capable of identifying and blocking phishing websites in real-time, thus mitigating potential harm to users before it occurs. Through this endeavor, we envision contributing to the advancement of cybersecurity measures in the ever-evolving landscape of online threats.

2. LITERATURE SURVEY

1. Mahajan Mayuri Vilas, Kakade Prachighanshamsawant, Purva Jaypralash, and Pawar Shila:
   - Title: "Detection of Phishing Website Using Machine Learning Approach"
   - Approach: Employed Extreme Learning Machine (ELM) with 30 different primary components characterized using Machine Learning (ML).
   - Detection Methods: Evaluated phishing URLs using HTTPS, assessed URL components, determined website authority and introductions, and verified website veracity.
2. Malak Aljabri and Samiha Mirza:
   - Title: "Phishing Attacks Detection using Machine Learning and Deep Learning Models"
   - Approach: Selected highest correlated features from two datasets, combining content-based, URL, and domain-based features.
   - Results: Conducted a performance comparison of multiple ML models, highlighting Random Forest (RF) as the most effective in classifying phishing websites.
3. Adarsh Mandadi and Saikiran Boppana:
   - Title: "Detection of Phishing Websites using Machine Learning Algorithms"
   - Approach: Utilized SVM, Neural Networks, Random Forest, Decision Tree, XGBoost, and other ML algorithms to categorize user-received URLs.
   - Results: Achieved accuracy rates of 87.0% and 82.4% for Random Forest and Decision Tree classifiers, respectively.
4. Hemali Sampat, Manisha Saharkar, Ajay Pandey, and Hezal Lopes:
   - Title: "Detection of Phishing Websites using Machine Learning"
   - Proposed System: Combined Classification and Association algorithms with the WHOIS protocol for faster and more effective phishing website detection.
   - Results: Reduced inaccuracy rate by 30%, enhancing the system's efficiency.
5. Safa Alrefaai, Ghina Özdemir, and Afnan Mohamed:
   - Title: "Machine Learning-Based Detection of Phishing Websites"
   - Approach: Used Kaggle data with 86 features and 11,430 URLs, trained using Decision Tree, Random Forest, XGBoost, Multilayer Perceptrons, K-Nearest Neighbors, Naive Bayes, AdaBoost, and Gradient Boosting.
   - Results: Employed XGBoost for effective detection of phishing websites in a machine learning context.

3. EXISTING SYSTEM

In the existing system, plant diseases are detected and classified manually by experts, which is a time-consuming process. Also, the accuracy of manual detection and classification depends on the performance of the

person doing it. There are also some existing automated systems, but they require alarge amount of training data and also produce low accuracy results in detecting plant disease. Plants are considered as energy supply to mankind. Plant diseases can affect the agriculture which can be resulted in to huge loss on the crop yield. Therefore, leaf diseases detection plays a vital role in agricultural field. However, it requires large manpower, more processing time and extensive knowledgeand skills about plant diseases. Hence, machine learning comes in play in the detection of diseases in plant leaves as it analyzes the data from various areas, and classifies it into one of the predefined set ofclasses. The features and properties like color, intensity and dimensions of the plant leaves are considered as a major fact for classification and the various types of plant diseases and different classification techniques in machine learning that are used for identifying diseases in different plants

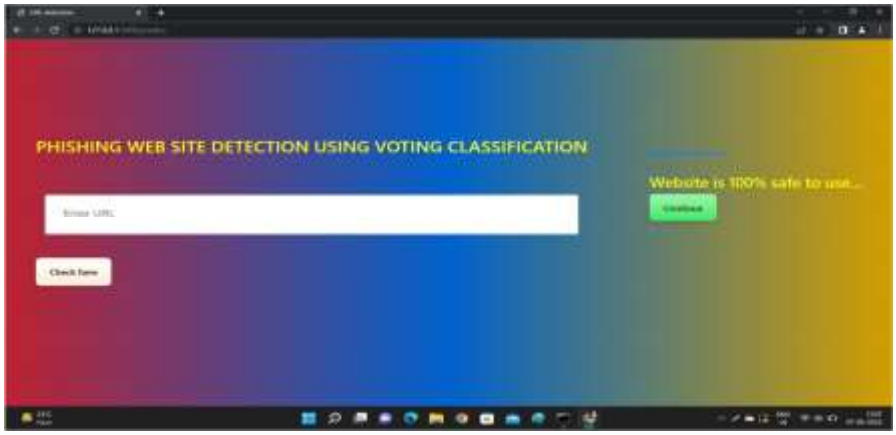
4. PROPOSED SYSTEM

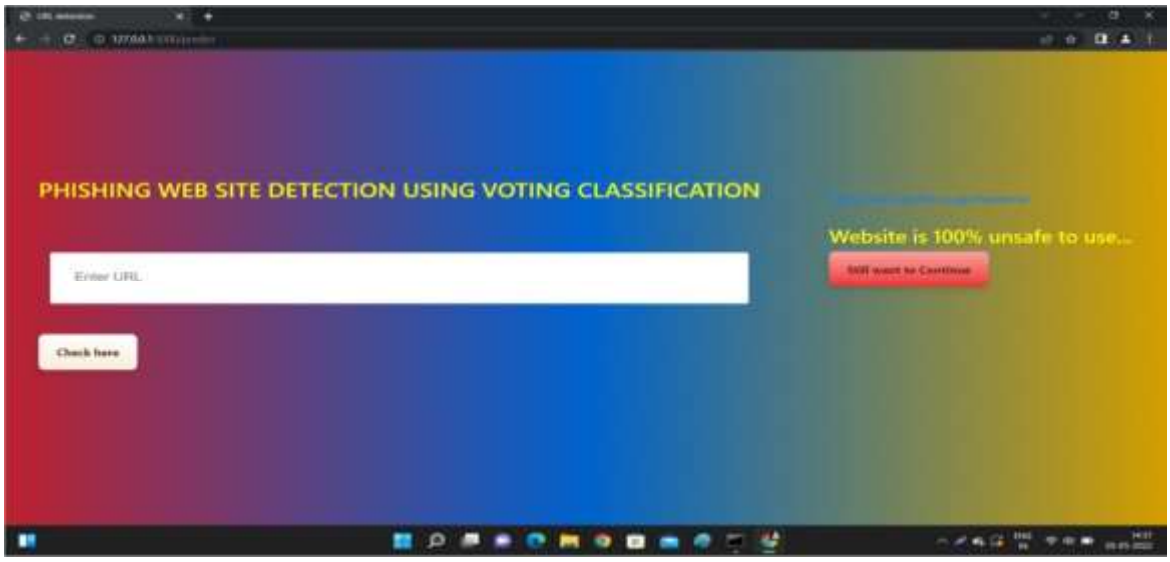
In this study, we implement different classification algorithm like Support vector machine, Random forest, K Nearest Neighbors, Naïve Bayes, Decision Tree, logistic regression, Gradient Boosting based classification was performed for the following 30 features extracted from the websites. In the proposed system once the user identify phishing website, the site is accessible, or the user is Informed that the website is not genuine Procedural steps for solving the classification problem presented is as follows:

- Collect dataset containing phishing and legitimate websites from the open source platforms.
- Write a code to extract the required features from the URL database.
- Analyse and preprocess the dataset.
- Divide the dataset into training and testing sets.
- Run selected machine learning model on the dataset.
- Write a code for displaying the evaluation result considering accuracy, f1_score, recall, precision metrices,
- Compare the obtained results for trained models and specify which is better.

5. EXPERIMENTAL RESULTS

Home Page:





6. CONCLUSION

In conclusion, this research underscores the importance of harnessing machine learning algorithms to effectively address the continually evolving challenge of phishing website detection. The continuous updating of datasets and the meticulous refinement of the feature extraction process are pivotal aspects that empower the system to adapt to emerging phishing techniques. This adaptability enhances the system's capacity to provide robust protection to users within the dynamic cybersecurity landscape. By leveraging the capabilities of machine learning, the proposed approach not only fortifies current defenses against phishing threats but also establishes a foundation for proactive and adaptive measures. The iterative nature of dataset updates and feature enhancement ensures that the system remains vigilant and resilient in the face of evolving cyber threats, ultimately contributing to a more secure online environment for users.

References

[1] M. M. Vilas, K. P. Ghansham, S. P. Jaypralash, and P. Shila, "Detection of Phishing Website Using Machine Learning Approach," 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India, 2019, pp. 384-389, doi: 10.1109/ICEECCOT46775.2019.9114695.

[2] M. Aljabri and S. Mirza, "Phishing Attacks Detection using Machine Learning and Deep Learning Models," 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2022, pp. 175-180, doi: 10.1109/CDMA54072.2022.00034.

1.

[3] H. Sampat, M. Saharkar, A. Pandey, and H. Lopes, "Detection of Phishing Website Using Machine Learning," 2018 International Research Journal of Engineering and Technology (IRJET), 2018, e-ISSN: 2395-0056, p-ISSN: 23950072.

[4] S. Alrefaai, G. Özdemir, and A. Mohamed, "Detecting Phishing Websites Using Machine Learning," 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2022, pp. 1-6, doi: 10.1109/HORA55278.2022.9799917.

[5] S. Pandiyan, P. Selvaraj, V. K. Burugari, J. B. P, K. P, "Phishing attack detection using Machine Learning," Measurement: Sensor Volume 24, 2022, 100476, ISSN: 2665-9174.