International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Phishing URL and Website Detection using MI

[1]Dhiraj Dubhashe, [2]Aditya Kumavat, [3]Ishwari Date, [4]Krutika Ahire, [5]A.S. Kolhe

[1234]*Student Department of Information Technology*

[5]*HOD Department of Information Technology*

Matoshri Aasarabai Polytechnic, Eklahare, Nashik, Maharashtra, -422105

Abstract - Phishing attacks remain a significant threat in the digital landscape, with cybercriminals constantly developing sophisticated techniques to deceive users into revealing sensitive information. This study presents a robust framework for phishing URLs and website detection utilizing the XG-Boost (Extreme Gradient Boosting) algorithm, known for its superior performance and efficiency in classification tasks. The proposed system focuses on analyzing various features extracted from URLs and webpage content, including lexical, structural, and host-based attributes, to distinguish between legitimate and malicious sites. By leveraging the strengths of the XG-Boost algorithm, the framework aims to improve detection accuracy and reduce false positive rates, thereby enhancing user safety in online environments. The research involves a comprehensive evaluation of the XG-Boost model's performance against several benchmark datasets representative of real-world phishing scenarios. The model is trained on a diverse set of features, with hyperparameter tuning conducted to optimize its predictive capabilities. Results indicate that the XG-Boost based approach achieves a high classification accuracy while demonstrating robustness against imbalanced datasets common in phishing detection tasks. The findings underscore the effectiveness of machine learning techniques in cybersecurity applications and highlight the potential for implementing the XG-Boost algorithm in real-time phishing detection systems to safeguard users from online threats. The proposed system leverages a diverse set of URL-based and website features such as the presence of special characters, domain age, HTTPS usage, and the similarity of the website's content to known phishing sites. These features are collected from URLs and website characteristics to form a feature vector for each website. The XG-Boost algorithm is trained on a labeled dataset consisting of both phishing and legitimate websites, learning to identify patterns indicative of phishing activities. XG-Boost's gradient boosting framework allows for a more accurate classification by combining multiple weak decision trees into a strong predictive model, making it highly suitable for this task. To evaluate the effectiveness of the proposed system, extensive experiments are conducted using publicly available datasets containing thousands of phishing and legitimate URLs. The model's performance is measured in terms of accuracy, precision, recall, and F1 score. The results demonstrate that the XG-Boost-based model outperforms traditional machine learning algorithms, such as decision trees and support vector machines (SVM), in detecting phishing websites with high accuracy and minimal false positives. This highlights the superiority of XG-Boost in handling the imbalanced nature of phishing datasets, where phishing samples are often underrepresented compared to legitimate ones.

Key Words: XG-Boost (Extreme Gradient Boosting), Classifier, Features, Phishing, Train, Accuracy.

1. INTRODUCTION -

Phishing attacks have emerged as one of the most pervasive cyber threats, targeting individuals and organizations to steal sensitive information such as usernames, passwords, and financial details. These attacks often involve deceptive emails or websites that mimic legitimate sources, making it increasingly difficult for users to distinguish between genuine and malicious content. The rapid evolution of phishing techniques necessitates the development of effective detection mechanisms that can identify and mitigate these threats promptly. As a result, there is a growing demand for automated solutions capable of analyzing and classifying URLs and websites to enhance cybersecurity measures. Machine learning (ML) has gained prominence as a powerful tool for detecting phishing attempts due to its ability to learn from data patterns and make informed predictions. Among the various ML algorithms available, XG-Boost (Extreme Gradient Boosting) stands out for its superior performance in classification tasks, particularly in scenarios characterized by large datasets and complex feature interactions. XG-Boost's efficiency stems from its implementation of gradient boosting techniques, which optimize the model by minimizing errors through an ensemble of weak learners. This makes it particularly suitable for the dynamic and evolving nature of phishing attacks, where the characteristics of malicious websites can change rapidly. The detection of phishing URLs and websites involves analyzing various features that can indicate malicious intent. These features can be categorized into lexical features (e.g., the length of the URL, presence of special characters), structural features (e.g., the layout and design of the website), and host-based features (e.g., domain age, IP address). By extracting and leveraging these features, machine learning models can be trained to classify URLs and websites accurately.

The integration of XG-Boost into this process enhances the model's ability to handle imbalanced datasets, a common challenge in phishing detection, where legitimate instances often outnumber phishing instances. This study aims to develop a robust framework for phishing URLs and website detection utilizing the XG-Boost algorithm. The research focuses on collecting and preprocessing a comprehensive dataset of legitimate and phishing URLs, followed by feature extraction and model training. The effectiveness of the proposed model will be evaluated against established benchmarks, with a focus on accuracy, precision, recall, and F1-score. By demonstrating the capabilities of the XG-Boost algorithm in phishing detection, this research contributes to the ongoing efforts to enhance cybersecurity and protect users from malicious online threats. Among various machine learning algorithms, XG-Boost (Extreme Gradient Boosting) has gained considerable attention for its impressive performance in classification tasks. XG-Boost is a highly efficient and scalable gradient- boosting framework that has demonstrated state-of-the-art results in many fields, including image classification, fraud detection, and text mining. Its ability to handle large datasets and complex features, along with its ability to perform well in imbalanced datasets, makes it an ideal candidate for phishing detection. Phishing URLs and website detection based on machine learning can be approached by extracting several features from URLs and websites that are indicative of phishing.

These features include attributes like domain age, the presence of HTTPS, the length of the URL, the presence of special characters, domain registration information, and URL obfuscation techniques. By evaluating these characteristics, machine learning models can learn to classify websites as either legitimate or phishing. Using XG-Boost, a robust ensemble method, the model can automatically assign weights to these features and combine them into a predictive framework, enhancing its ability to identify phishing URLs with high accuracy .The first step in developing a phishing detection system using XG-Boost involves feature extraction. For URL-based detection, features are derived directly from the URL structure, such as the presence of certain keywords, the number of characters, and domain- related attributes like the reputation and registration history. For website detection, additional features such as visual elements (e.g., the presence of SSL certificates) and the structure of the website itself are considered.

2. Problem Statement -

Among various machine learning algorithms, XG-Boost (Extreme Gradient Boosting) has gained considerable attention for its impressive performance in classification tasks. XG-Boost is a highly efficient and scalable gradient-boosting framework that has demonstrated state-of- the-art results in many fields, including image classification, fraud detection, and text mining. Its ability to handle large datasets and complex features, along with its ability to perform well in imbalanced datasets, makes it an ideal candidate for phishing detection. Phishing URLs and website detection based on machine learning can be approached by extracting several features from URLs and websites that are indicative of phishing. These features include attributes like domain age, the presence of HTTPS, the length of the URL, the presence of special characters, domain registration information, and URL obfuscation techniques. By evaluating these characteristics, machine learning models can learn to classify websites as either legitimate or phishing. Using XG-Boost, a robust ensemble method, the model can automatically assign weights to these features and combine them into a predictive framework, enhancing its ability to identify phishing URLs with high accuracy .The first step in developing a phishing detection system using XG-Boost involves feature extraction. For URL- based detection, features are derived directly from the URL structure, such as the presence of certain keywords, the number of characters, and domain-related attributes like reputation and registration history. For website detection, additional features such as visual elements (e.g., the presence of SSL certificates) and the structure of the website itself are considered.

These features are then used to train an XGBoost classifier, which applies boosting techniques to iteratively improve the accuracy of predictions by minimizing errors over successive iterations .The XG-Boost algorithm works by building a series of decision trees in a sequential manner, where each new tree tries to correct the errors made by previous ones.

This boosting process improves the model's accuracy and makes it particularly effective in distinguishing between phishing and legitimate websites. The trained model outputs the probability of a website being phishing or legitimate, with a higher probability indicating a higher likelihood of being a phishing site. The rise in online activities has unfortunately paralleled an increase in cyber threats, particularly phishing attacks. Phishing occurs when attackers create fraudulent websites and URLs that mimic legitimate ones, tricking users into revealing sensitive information like usernames, passwords, and credit card details. These attacks not only compromise individual security but also pose significant risks to businesses and organizations. As phishing tactics become more sophisticated, traditional detection methods struggle to keep pace, highlighting the need for more advanced and automated solutions. This project seeks to address the growing threat of phishing by developing a robust detection system using the XG-Boost machine learning algorithm. XG-Boost, known for its high performance in classification tasks, is particularly well-suited for detecting phishing URLs due to its ability to handle complex datasets and its efficiency in learning from large amounts of data. By extracting relevant features from URLs—such as length, the presence of special characters, domain registration information, and

SSL certificate status—the model will be trained to distinguish between legitimate and phishing websites. The core objective of this project is to enhance cybersecurity by providing an accurate and efficient tool for phishing detection.

Literature Review –

1. In their 2022 paper, "A Deep Learning-Based Framework for Phishing Website Detection," L. Tang and Q. H. Mahmoud propose an innovative framework that leverages deep learning techniques to enhance the detection of phishing websites. The authors highlight the increasing sophistication of phishing attacks and the limitations of traditional detection methods. Their framework utilizes a combination of Convolutional Neural Networks (CNNs) and Long Short Term Memory (LSTM) networks to analyze features from URLs and website content effectively. The proposed model demonstrates superior performance in detecting phishing sites compared to existing techniques, achieving higher accuracy and lower false positive rates. The study underscores the potential of deep learning in improving cybersecurity measures against phishing threats and advocates for further research in this area to adapt to evolving attack strategies.[1]
2. In the 2021 paper "Internet of Things (IoTs) Security: Intrusion Detection using Deep Learning," O.K. Sahingoz, U. Cekmez, and A. Buldu address the critical security challenges faced by Internet of Things (IoT) systems, particularly focusing on intrusion detection. The authors present a deep learning-based approach that utilizes various neural network architectures to effectively identify and classify intrusion attempts in IoT networks. Through extensive experimentation, the study evaluates the performance of different models, demonstrating that deep learning techniques significantly enhance detection accuracy and reduce false alarm rates compared to traditional methods. The findings emphasize the importance of implementing robust intrusion detection systems to protect IoT environments, advocating for the integration of advanced machine learning techniques in future security frameworks to counteract evolving threats.[2]
3. In their 2021 paper, "Generating Rules to Detect Phishing Websites Using URL Features," A. Awasthi and N. Goel propose a novel approach for phishing website detection by leveraging features extracted from URLs. The authors emphasize the significance of URL characteristics as indicators of potential phishing attempts. They develop a set of rules based on these features to enhance the detection process, demonstrating that simple yet effective heuristics can significantly improve the identification of malicious sites. The study provides a comprehensive analysis of various URL features and their relevance to phishing detection, showing promising results in terms of accuracy and efficiency. The authors advocate for the integration of their rule-based approach into existing security systems to bolster protection against phishing threats, highlighting its applicability in real-world scenarios.[3]
4. In the 2021 paper "Improving the Phishing Website Detection using Empirical Analysis of Function Tree and its Variants," O. Abdullateef et al. explore innovative methods to enhance the detection of phishing websites through the use of Function Tree (FT) structures and their variants. The authors conduct a thorough empirical analysis to evaluate the effectiveness of FT models in identifying phishing threats based on various website features. Their research demonstrates that incorporating FT and its modified versions leads to improved detection rates compared to traditional methods. By systematically analyzing the strengths and weaknesses of different FT configurations, the study highlights the potential of these models in addressing the dynamic nature of phishing attacks. The authors conclude that leveraging advanced data structures like function trees can significantly enhance the accuracy and reliability of phishing detection systems, encouraging further exploration of their applications in cybersecurity. [4]

5. In the 2021 paper "An Efficient Multistage Phishing Website Detection Model Based on the CASE Feature Framework: Aiming at the Real Web Environment," Dong-Jie Liu, Guang-Gang Geng, Xiao Bo Jin, and Wei Wang present a comprehensive multistage model designed to detect phishing websites effectively within real-world web environments. The authors introduce the CASE feature framework, which integrates multiple features related to the content, access, structure, and engagement ofwebsites to improve detection accuracy. Through rigorous experimentation, the model demonstrates superior performance in identifying phishing attempts compared to existing detection methods. The study emphasizes the importance of adapting detection mechanisms to the dynamic nature of the web, proposing that a multistage approach can effectively filter out phishing sites while minimizing false positives. The authors conclude that their model represents a significant advancement in phishing detection technology, advocating for its implementation in practical

cybersecurity solutions to safeguard users from online threats.[5]

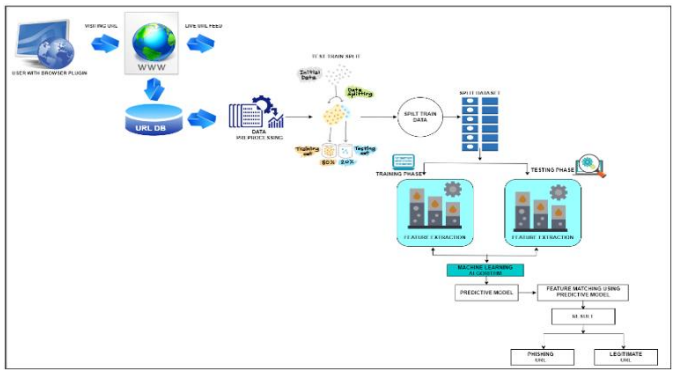
6. In the 2021 paper "Phishing Websites Detection via CNN and Multihead Self-Attention on Imbalanced Datasets," Xi Xiao et al. address the challenge of detecting phishing websites in the context of imbalanced datasets, which often hinder the performance of traditional detection methods. The authors propose a novel approach that combines Convolutional Neural Networks (CNNs) with a multihead self-attention mechanism to enhance feature extraction and improve classification accuracy. By effectively focusing on critical features and mitigating the impact of imbalanced data, their model demonstrates significant improvements in detection rates compared to existing techniques. The study provides extensive experimental results, showcasing the model's robustness in identifying phishing sites while maintaining a low false positive rate. The authors conclude that integrating CNNs with self-attention mechanisms offers a promising direction for advancing phishing detection methodologies, particularly in challenging data scenarios, and highlight the need for continued innovation in this critical area of cybersecurity.[6]

7. In the 2021 paper "Stop-Phish: An Intelligent Phishing Detection Method Using Feature Selection Ensemble," A.V. Ramana, K.L. Rao, and R. S present a novel approach to phishing detection that emphasizes the importance of feature selection to enhance model accuracy. The authors introduce the Stop- Phish method, which employs an ensemble of feature selection techniques to identify the most relevant features for distinguishing between legitimate and phishing websites. Through comprehensive experiments, the study demonstrates that the Stop Phish model significantly outperforms traditional detection methods, achieving higher accuracy and lower false positive rates. The authors highlight the critical role of effective feature selection in improving the performance of machine learning models in cybersecurity applications. The paper concludes that the proposed approach not only enhances phishing detection capabilities but also provides insights into the practical implementation of intelligent systems for safeguarding users against online threats.

8. In the 2020 paper "Detection of Phishing Websites by Using Machine Learning-Based URL Analysis," M. Korkmaz, O. K. Sahingoz, and B. Diri explore the effectiveness of machine learning techniques in identifying phishing websites through detailed URL analysis. The authors emphasize the importance of URL features as indicators of potential phishing attempts and propose a machine-learning framework that leverages these features to classify websites accurately. The study evaluates various algorithms, including decision trees, support vector machines, and neural networks, assessing their performance in detecting phishing sites. Results indicate that the proposed machine learning model can effectively distinguish between legitimate and malicious URLs, achieving promising accuracy levels. The authors conclude that machine learning-based URL analysis offers a practical solution for enhancing phishing detection systems and recommend further research to improve model robustness and adaptability in dynamic web environments.[8]

9. In the paper "Detecting Phishing Website Using Machine Learning," presented at the 16th IEEE International Colloquium on Signal Processing and its Applications (CSPA 2020), Mohammed Hazim Alkawaz, Stephanie Joanne Steven, and Asif Iqbal Hajamydeen investigate the application of machine learning techniques for identifying phishing websites. The authors propose a machine learning-based framework that analyzes various website features to classify URLs as either legitimate or phishing. Through extensive experimentation, the study evaluates several algorithms, including logistic regression, decision trees, and ensemble methods, comparing their effectiveness in phishing detection. The results demonstrate that the machine learning models can achieve high accuracy rates while effectively reducing false positives. The authors conclude that machine learning presents a promising approach for enhancing phishing detection capabilities, advocating for its integration into existing cybersecurity frameworks to provide more robust protection against online threats.[9] 10. In the paper "Detecting Phishing Websites," published in the International Research Journal of Engineering and Technology in February 2020, Sagar Patil, Yogesh Shetye, and Nilesh Shendage explore various methodologies for identifying phishing websites effectively. The authors discuss the growing prevalence of phishing attacks and the challenges associated with traditional detection methods. They propose a framework that incorporates multiple techniques, including feature extraction from URLs and website content, combined with machine learning algorithms to enhance detection accuracy. The study evaluates several classification methods, analyzing their performance in detecting phishing sites based on different feature sets. Results indicate that the proposed approach significantly improves detection rates while minimizing false positives. The authors conclude that an integrated detection framework is essential for addressing the evolving nature of phishing threats, recommending further research into the implementation of more advanced techniques for real-time detection in practical applications.[10]

Proposed System –



This is the most common form of phishing attack, where a cybercriminal poses as a well- known entity, domain, or organization to trick victims into revealing sensitive information such as login credentials, passwords, bank account details, and credit card information. These attacks are typically unsophisticated, lacking personalization or customization for individual targets. For example, phishing emails containing malicious URLs are sent in bulk to a large number of recipients. Given the high volume, attackers rely on the probability that some users will open these emails, click on the links, or download infected attachments. The primary tactics used in this type of phishing are deception and impersonation. These emails often create a sense of urgency or panic, prompting victims to act quickly. Common subject lines may include warnings like "Your account has been hacked—change your password immediately!" or "Your bill is overdue— pay now to avoid a fine!" Once users interact with these messages or click the links, significant damage can occur, leading to financial loss, compromised private information, and damaged reputations .Addressing these threats requires effective mitigation strategies. Traditionally, phishing website detection has relied on blacklists maintained by services like Phish Tank, which identify known malicious sites. However, blacklist-based methods have limitations— they may not cover all phishing sites and are ineffective against newly generated or evolving phishing websites. To overcome these challenges, machine learning techniques have emerged as powerful tools for detecting and classifying phishing websites, providing more robust and adaptive security solutions. The system architecture for phishing URLs and website detection using the XGBoost algorithm comprises several key components that work together to ensure the effective identification of malicious threats. At the core, the architecture includes a data collection module that gathers a diverse dataset of URLs, encompassing both legitimate and phishing sites. This data is then processed through a feature extraction module, which identifies relevant characteristics such as URL length, use of HTTPS, presence of suspicious keywords, and domain age. Once the features are extracted, they are fed into the XGBoost model, which is trained on historical data to learn patterns indicative of phishing attempts. The trained model is deployed within a real- time detection framework, where new URLs are continuously analyzed against the learned patterns. Finally, the system includes a user interface that provides alerts and reports, enabling users to take immediate action against detected phishing threats, thereby enhancing overall online security. The system architecture for phishing URLs and website detection using the XGBoost algorithm is designed to provide a comprehensive solution for identifying malicious websites. At the core of the architecture is the data collection module,. which is responsible for gathering a wide array of URLs. This dataset includes both legitimate websites and phishing sites, sourced from publicly available databases, web crawlers, and known phishing repositories.

Algorithm Used for Proposed System

1. XGBoost: XGBoost, short for Extreme Gradient Boosting, is an advanced machine learning algorithm widely recognized for its efficiency and high performance. It belongs to a class of gradient boosting algorithms and is renowned for its ability to optimize both the speed and accuracy of predictive models. By building an ensemble of decision trees through boosting, XGBoost progressively refines weak classifiers and combines them to form a strong prediction model. The algorithm's core principle is to minimize errors by adding new trees that correct the mistakes made by previous ones, thus improving the overall accuracy of the model. This process makes XGBoost particularly well-suited for complex tasks such as phishing URLs and website detection.In the context of phishing URLs and website detection, XGBoost is highly effective at identifying malicious patterns in URLs and websites. The algorithm is trained on a dataset that includes both legitimate and phishing websites, using various features extracted from the URLs. These features can include URL length, presence of special characters (such as '@' or '%'), domain age, the status of HTTPS encryption, and unusual patterns in the URL structure. By analyzing these features, XGBoost can classify websites as either benign or phishing, ensuring that malicious websites attempting to steal user data are accurately identified.One of the key challenges in phishing detection is the imbalance in the dataset, where phishing sites are often much fewer than legitimate ones. XGBoost excels in handling such imbalanced data by using techniques like weighted loss functions and custom sampling methods. This ensures that the model pays more

attention to the minority class (phishing websites) without being overwhelmed by the majority class (legitimate websites). This capability makes XGBoost particularly suitable for phishing detection, where the detection of rare but potentially harmful phishing sites is critical.Another strength of XGBoost is its built-in regularization mechanisms, such as L1 (Lasso) and L2 (Ridge) regularization, which help prevent overfitting. In phishing detection, overfitting can occur when a model becomes too complex and starts to memorize the training data, failing to generalize well to new, unseen data. XGBoost's regularization techniques ensure that the model remains robust and generalized, even when trained on large datasets. This contributes to maintaining high performance across a range of phishing detection scenarios, reducing the risk of false positives or negatives.The iterative nature of XGBoost, where each subsequent tree aims to improve the model's weaknesses, results in highly accurate predictions. In the case of phishing detection, this means the model can reliably distinguish between legitimate and phishing websites with high precision and recall. By providing strong predictive capabilities, XGBoost plays a significant role in enhancing cybersecurity. The model's ability to quickly process large datasets, handle imbalanced data, and provide consistent results makes it an indispensable tool for real-time phishing URL and website detection, ultimately helping to protect users from cyber threats and preventing data theft.

Application

1. Detecting DDoS (Distributed Denial of Service) attacks in networks by analyzing traffic patterns and identifying abnormal spikes in requests.
2. Identifying multiple packet drop sources in a network, which can indicate issues such as network congestion or potential malicious activity.
3. Detecting network intrusions in military computers, ensuring the security of sensitive and classified information.
4. Detecting organizational intrusions networks, in large helping safeguard critical infrastructure and sensitive corporate data.
5. Email clients and spam filters: Machine learning can enhance spam detection, improving the filtering of unwanted emails and phishing attempts.
6. Web browsers: Use machine learning to identify phishing websites and block harmful URLs to protect users while browsing.
7. Online platforms: Detect fraudulent activities, such as fake accounts or phishing attempts, by analyzing patterns in user behavior.
8. Endpoint protection: Protect individual devices (e.g., laptops, desktops, and mobile devices) from malware and phishing attacks by analyzing file behaviors and system calls.
9. Network security: Monitor network traffic for signs of unauthorized access or malicious activities, ensuring secure data transfers and communications.
10. Mobile applications: Detect security threats like phishing links, malware, or suspicious activities within mobile apps, safeguarding personal information. 11. Cyber threat intelligence platforms: Use machine learning models to analyze threat data and predict potential attacks, enabling proactive cybersecurity measures. 12. Fraud prevention in financial services: Machine learning models can detect unusual financial transactions, potentially preventing fraud and securing financial assets. 13. Security awareness training: ML can personalize security training content based on user behaviors and patterns, helping employees better understand and prevent cyber threats. 14. Real-time threat detection: Machine learning can be used to continuously analyze network traffic in real-time, providing alerts on potential threats or breaches. 15. Behavioral abnormal analysis: user Detect behavior in organizations, such as unauthorized access or suspicious login patterns, to prevent insider threats. 16. Automated phishing website detection: Identify and block phishing sites by analyzing URL structures, website content, and metadata in real time. 17. Intrusion detection in IoT devices: Protect Internet of Things (IoT) devices from external attacks by using ML models to detect anomalies in communications. 18. Anti-bot protection: Detect and mitigate bot traffic on websites and platforms by analyzing patterns in automated interactions .

Conclusion

In conclusion, the application of the XGBoost algorithm for phishing URLs and website detection marks a crucial advancement in the field of cybersecurity. The ability to analyze and classify URLs based on key features such as length, domain information, and the use of HTTPS allows for efficient detection of malicious websites. The strength of XGBoost lies in its high accuracy and ability to handle large datasets, making it a valuable tool in the fight against phishing attacks that increasingly target users online.This approach offers a scalable solution capable of adapting to new and evolving phishing threats. With the continuous development of machine learning techniques, the detection system can be enhanced to recognize new patterns in phishing behavior. As phishing tactics become more sophisticated, the integration of real-time data and feature refinement will enable more accurate and timely identification of threats, ensuring that the system remains effective in dynamic online environments. The real-time detection framework proposed in this system enhances user protection by providing immediate alerts and actions against detected phishing attempts. This proactive measure significantly reduces the risk of sensitive data theft and the associated financial losses, safeguarding both individuals and organizations. As the digital landscape expands, ensuring a secure online environment

through advanced detection mechanisms is essential to maintaining user trust and privacy.The project emphasizes the critical role of machine learning in strengthening cybersecurity frameworks. By continuously improving phishing detection models, we move closer to creating a safer internet. The integration of XGBoost in phishing URLs and website detection not only improves the detection process but also fosters a more secure and trustworthy digital experience for all users.

Reference

1. L. Tang and Q. H. Mahmoud, "A Deep Learning Framework for Detecting Phishing Websites," IEEE Access, vol. 10, pp. 1509-1521, 2022.

2. O. K. Sahingoz, U. Cekmez, and A. Buldu, "Internet of Things (IoTs) Security: Intrusion Detection using Deep Learning," Journal of Web Engineering, vol. 20, no. 6, pp. 1721-1760, 2021.

3. A. Awasthi and N. Goel, "Generating Rules to Detect Phishing Websites Using URL Features," 2021 1st Odisha International Conference on Electrical Power Engineering Communication and Computing Technology (ODICON), pp. 1-9, 2021.

4. O. Abdullateef et al., "Improving the phishing website detection using empirical analysis of Function Tree and its variants," Heliyon, vol. 7, no. 7, 2021.

5. Dong-Jie Liu, Guang-Gang Geng, Xiao-Bo Jin, and Wei Wang, "An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment," Computers Security, vol. 110, pp. 102421, 2021.

6. Xi Xiao, Wentao Xiao, Dianyan Zhang, Bin Zhang, Guangwu Hu, Qing Li, et al., "Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets," Computers Security, vol. 108, pp. 102372, 2021.

7. A.V. Ramana, Rao, K.L. Rao, and R. S, "Stop-Phish: an intelligent phishing detection method using feature selection ensemble," Soc. Netw. Anal. Min., vol. 11, no. 110, 2021.

8. M. Korkmaz, O. K. Sahingoz, and B. Diri, "Detection of Phishing Websites by Using Machine Learning-Based URL Analysis," 2020 11th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-7, 2020.

9. Mohammed Hazim Alkawaz, Stephanie Joanne Steven, and Asif Iqbal Hajamydeen, "Detecting Phishing Website Using Machine Learning," 16th IEEE International Colloquium on Signal Processing its Applications (CSPA 2020), 28-29 Feb. 2020. Sagar Patil, Yogesh Shetye, and Nilesh Shendage, "Detecting Phishing Websites," International Research Journal of Engineering and Technology, vol. 07, no. 02, Fe