International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 08 | Aug 2024 www.irjet.net p-ISSN: 2395-0072

AI-Driven Phishing URL Detection

Ankit Das[1], Anushka Behere[2]

[1] *Student, Dept. of Computer Science Engineering (Cyber Security), Thakur College of Engineering and Technology, Mumbai, Maharashtra*
[2]*Student, Dept. of Computer Science Engineering (Cyber Security), Thakur College of Engineering and Technology, Mumbai, Maharashtra*

Abstract - *Phishing attempts are getting more complex and employ questionable URLs to trick people into giving out sensitive information. This study investigates how well AI and machine learning can identify various phishing efforts. We evaluated the efficacy of three machine learning models—XGBoost, SGD, and AdaBoost—in identifying malicious URLs by examining a UC Irvine dataset and examining characteristics such URL length, special characters, HTTPS usage, and the existence of suspicious keywords. XGBoost outperformed AdaBoost and SGD, according to our results, with the maximum accuracy of 99.95%. This illustrates how sophisticated machine learning techniques may be used to improve the identification of phishing attempts and emphasizes the necessity of ongoing model adaption and improvement in order to combat changing cyberthreats.*

Key Words: Phishing Attacks, Cybersecurity, AI, Machine Learning, XGBoost, SGD, AdaBoost, Feature Extraction, Accuracy, Precision, Recall, Adversarial Attacks.

1.INTRODUCTION

The internet's rapid expansion has completely changed how individuals obtain services and information, providing previously unheard-of levels of convenience and connectedness. However, the rise in digitalization has also given rise to a number of cyberthreats, the most prevalent and harmful of which is phishing. Phishing is a deceptive technique used by cybercriminals to pretend to be reputable companies and trick victims into divulging private information such as passwords, credit card details, and other sensitive data. Phishing assaults have changed throughout time, become more complex and difficult to identify.

Phishing attacks have evolved into new, sneakier forms in recent years, such as manipulating and changing website URLs. Attackers now use sophisticated techniques to produce spoof URLs that are almost identical to real ones, fooling even the most watchful internet users. Heuristic-based technique in [1] can identify newly created malicious web-sites in real-time by using signatures of known attack payloads. However, this approach would fail to detect novel attacks that result in zero-day exploits and signature detection is often evaded by attackers using change in patterns and obfuscation techniques. Obfuscation techniques used by the attacker to evade static detection in malicious URLs. Since obfuscation-based features have been widely used for phishing attacks [2,3], we also study the effect of the obfuscation techniques on different type of malicious URLs to determine which attack type is mostly affected with what kind of obfuscation technique.

The number of attacks has significantly increased as a result of this change in phishing strategies. Over 255 million phishing attacks were reported, according to a thorough analysis by SlashNext that examined billions of link-based URLs, attachments, and natural language messages via email, mobile, and browser channels. Unbelievably, since 2021, the frequency of these attacks has increased by 61% [4]. These results highlight a crucial point: the increasingly complex strategies used by hackers can no longer be defeated by outdated security solutions like firewalls, secure email gateways, and proxy servers. To further complicate detection attempts, attackers are now starting their assaults from trusted platforms, such as personal and professional messaging apps, in addition to compromised servers.

The difficulty of recognising fraudulent URLs is made more difficult by the ongoing development of phishing techniques. Cybercriminals constantly modify their tactics to evade detection, making it challenging for current security measures to stay up to date. Retrieving relevant data from URLs, such as length, the existence of particular protocols (HTTP/HTTPS), and the quantity of special characters, is necessary for effective detection [5,6]. However, to further impede the detecting process, attackers use techniques like URL obfuscation and the malicious exploitation of reliable websites. Furthermore, real-time detection systems have scalability issues, particularly for organisations with limited infrastructure, because they require a significant amount of computational power to evaluate URLs and the material that goes along with them.

The incorporation of artificial intelligence (AI) and machine learning (ML) into cybersecurity has become a viable approach to improve phishing detection skills in response to these issues. Large-scale datasets are used by AI-driven systems to find patterns and anomalies suggestive of phishing attempts, greatly increasing detection accuracy [7]. With the help of these technologies, one may keep one step ahead of cybercriminals by constantly learning about and adjusting to new threats.

The purpose of this research project is to investigate and assess the various AI-driven methods for phishing attack detection. The study will investigate several machine learning models and approaches through a thorough literature analysis in order to determine the best tactics for thwarting phishing assaults in the constantly changing field of cybersecurity.

2. LITERATURE REVIEW

A crucial component of cybersecurity is phishing URL detection, which shields users from phony websites that aim to steal sensitive data like credit card numbers and login credentials. Machine learning (ML) and deep learning approaches have advanced throughout time, leading to changes in traditional detection systems. However, these systems remain vulnerable to adversarial attacks, where malicious actors deliberately manipulate URLs to evade detection. This literature review examines the current state of adversarial analysis in phishing URL detection, focusing on various methodologies, their effectiveness, and resilience against adversarial examples.

Elsadig et al. [8] provides a unique URL phishing detection algorithm based on BERT feature extraction and deep learning. Elsadig et al. [8] explain that using BERT allows the model to capture contextual information from URLs, which improves detection accuracy. According to Elsadig et al. [8], a natural language processing (NLP) algorithm was applied to the unique data column, yielding a large number of usable data features in the form of relevant text information. Elsadig et al. [8] shows that a deep convolutional neural network method was utilized to detect phishing URLs, achieving an accuracy of 96.66% in the experimental results.

Blum et al. [9] uses online learning in order to perform URL classification. Their work uses a similar set of lexical features. However, the Blum et al.'s [9] work totally discards the use of host-based features. Their classifier achieves an accuracy of around 97% if quality training data can be provided. The new work extends Blum et al.'s [9] work by incorporating host-based features. Host-based characteristics are proven to be reliable markers for identifying phishing URLs [10].

Another significant contribution is Karim et al. [11] which proposes a hybrid approach combining multiple ML techniques to detect phishing URLs. Karim et al. [11] explains that the system integrates feature-based methods and machine learning classifiers to enhance detection accuracy. While Karim et al. [11] notes that this hybrid model demonstrates improved robustness compared to single-method approaches, it remains vulnerable to adversarial sample.

A critical examination of adversarial attacks against phishing detection systems is provided in the study by Shirazi et al. [12]. Shirazi et al. [12] focuses on how adversarial samples can be used to undermine the effectiveness of phishing URL detectors. Shirazi et al. [12] demonstrates that by carefully crafting adversarial URLs, attackers can significantly reduce the detection accuracy of even the most advanced models. This research by Shirazi et al. [12] highlights the need for developing robust adversarial defenses to enhance the resilience of phishing detection systems.

In 2023, the authors Sasi et al. [13] introduced a novel approach towards detecting phishing URLs employing a generative adversarial network (GAN) with a variational autoencoder (VAE) as the generator and a transformer model with self-attention as the discriminator. Sasi et al. [13] reports that the model is effective, achieving an impressive accuracy of 97.75% in the results.

A Google Scholar search shows 100+ research papers from 2021 to 2024 investigating how adversarial attacks can compromise phishing URL detection systems. These studies explore various adversarial techniques and their impact on ML models, highlighting the significant threat posed by adversarial examples. However, very few of these papers propose effective solutions with good accuracy to mitigate these attacks, indicating a substantial gap in the current research.

While significant progress has been made in phishing URL detection through advanced ML and deep learning techniques, the challenge of adversarial attacks remains prevalent. In the future, research should focus more on strengthening detection systems' and creating strong adversarial defenses. One potential solution to lessen these risks is to incorporate adversarial training, in which models are trained with adversarial cases. This is the approach we will be implementing in our research paper.

3. METHODOLOGY

The basic method for phishing detection involves collecting a dataset of URLs, extracting relevant features that may indicate malicious intent, and then using machine learning models to classify the URLs as either legitimate or phishing. Traditional approaches often rely on lexical analysis, content inspection, and host-based features to make these determinations.

3.1 Dataset:

The quality of the prediction of a ML algorithm is strongly related to the quality of its training set. The Machine Learning approach requires a supervised learning algorithm, and therefore, the samples will need to be labelled as either legitimate or phishing [14].

We used a publicly available dataset from UC Irvine to identify phishing URLs. It has over 11000 records [15].

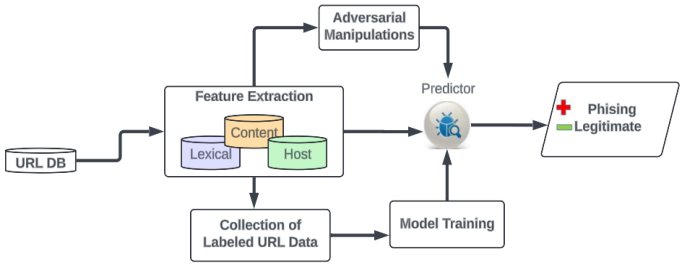

Fig -1: Framework for Malicious URL Detection using Machine Learning

3.2 Feature Extraction:

We extracted various features from the URLs to help distinguish between legitimate and phishing URLs.

These features fall into three categories:

3.2.1 Lexical Features

Lexical features are the literary properties of a URL, such as hostname length, URL length, symbols contained in the URL, and so on. Lexical features have grown in popularity in machine learning due to their lightweight computation, safety, and excellent classification accuracy [3]. Because many malicious URLs have a brief life cycle, lexical features remain available even when dangerous web pages are unavailable. [16]

Features related to Length: Addition of variables in the URL makes length longer. [18, 20]. such as, Length of URL (URL_Length), domain (Domain_Length)

- *URL length*: Returns the URL's overall length. Longer URLs are used to trick consumers or conceal harmful content, raising suspicions. Increase in URL length may be a sign of increased phishing activity.

- *Domain length*: Short or excessively long domain names can be suspicious. Phishing URLs often use domain names that are either very short (to mimic popular websites) or very long (to include extra words and avoid detection).

Type ratio: The proportion of different URL types based on length. A higher type ratio suggests that a particular URL stands out in length compared to others in its category. If a phishing URL has a significantly different length ratio, it could signal an anomaly, increasing the suspicion level.

$$Type\ ratio = \frac{Url\_length}{\sum Url\_length\ (group\ by\ category)} \times 100$$

Features related to count of Symbol, Digit and Letters: The frequency of characters in the URL are calculated in the form of letters, tokens and symbol [16, 17, 18]. These characters are categorized and counted from these components of URLs

- *Special Characters*: Count of characters such as '@', '?', '-', '=', '.', '#', '%', '+', '$', '!', '*', ',', '//', and '/' in the URL. An increased number of special characters can be a red flag, as attackers often insert these characters to confuse users or security filters. Phishing URLs e.g. have more dots compared to benign ones [19].

- *Digit Count*: URLs with many digits are often generated programmatically and might be used in phishing attacks. A higher digit count can indicate an attempt to mask the URL's true nature or make it appear legitimate, which could increase the likelihood of it being a phishing URL.

- *Letter Count*: A higher letter count typically suggests a more descriptive or legitimate-looking URL. However, excessively long URLs with many letters might also be used in phishing attempts to disguise the true destination, especially if combined with suspicious patterns.

URL entropy: Measures the randomness or unpredictability of the URL. High entropy indicates a more random and complex URL, which is often a sign of phishing attempts, as attackers might use encoded or obfuscated URLs to bypass security mechanisms.

3.2.2 Content Features

Has_HTTPS: Verifies if HTTPS is used by the URL. Although HTTPS is frequently connected to secure websites, phishing websites can also use HTTPS to look authentic. While HTTPS by itself does not ensure security, its absence may be a reliable sign of a phishing effort. A URL that isn't secured using HTTPS is more likely to raise a red flag.

Has_Shortening_Service: Determines if a service for shortening URLs is being used. Phishing attempts often utilize URL shorteners (like bit.ly) to hide the true destination. Shortening service-enabled URLs may be reported as possibly dangerous since they conceal the entire URL path, making it harder for visitors to determine the actual destination.

Has_IP_Address: Checks if the URL contains an IP address. URLs with raw IP addresses instead of domain names are often associated with phishing and other malicious activities. The use of an IP address can be a significant red flag, as legitimate websites typically use domain names. A URL containing an IP address is more likely to be identified as phishing.

Has_javascript_Code: Identifies if JavaScript code is embedded in the URL. It is unusual for a URL to contain JavaScript code, which might be used to run malicious scripts, reroute users, or carry out other undesirable operations. If the URL contains JavaScript code, there is a significant chance that it will be classified as phishing.

Has_Text_Encoding: Checks if the URL has text encoding parameters. Text encoding in URLs can be used to obfuscate content or hide malicious intent. URLs with encoded components are often harder to interpret and can indicate an attempt to bypass security filters, making them more suspicious.

Contains_suspicious_words: Identifies the presence of suspicious words like 'login', 'secure', 'update', 'confirm', 'invoice', 'post', 'important', and 'required'. Since these keywords are frequently used to deceive visitors into thinking the URL is affiliated with a reputable service that needs to be acted upon, their existence is a strong indicator of phishing.

3.2.3 Host-Based Features

The motivation behind using URL host-based features comes from Ma et al.'s work [21]. They obtained significant metadata for a URL from the Domain Name System (DNS), including A (the URL's IP address), MX (the IP address of the mail exchanger), NS (the IP address of the name server), and PTR (pointer) records.

Num subdomains: determines how many subdomains are there in the URL. Since attackers often employ several subdomains to generate complicated URLs that look legitimate at first glance, having a high number of subdomains can be a sign of phishing. Subdomains like "login.bank.security.example.com" are one way that a phishing URL could imitate a reliable domain.

Age of Domain: This feature was calculated from WHOIS information, measures the time since the domain was registered. Phishing sites often use newly registered domains to avoid detection, so a shorter domain age can be indicative of a phishing attempt.

In our study, we used a dataset from UC Irvine and focused on extracting a comprehensive set of features from each URL. These features were categorized into three main groups: lexical, content, and host-based features. The dataset was split into training and testing sets with an 80:20 ratio to build and evaluate the models effectively.

We introduced a new approach by calculating a unique feature called the type-ratio, which represents the proportion of different URL types based on length, providing an additional layer of analysis.

Additionally, we used a combination of traditional special character counts and more advanced features like URL entropy and presence suspicious words, which looked for specific keywords commonly associated with phishing attempts.

We trained three machine learning models XGBoost, SGD, and AdaBoost on these features to enhance detection accuracy.

4. RESULT AND DISCUSSION

We evaluated the performance of three machine learning models for phishing detection in our research - XGBoost, SGD, and AdaBoost. The accuracy of these models was used to gauge their performance.

Table -1: Model Result

MODEL	ACCURACY	PRECISION	RECALL	F1
XG Boost Classifier	.9995	1.00	.9991	.9996
Ada Boost Classifier	.9934	.9957	.9991	.9974
SGD Classifier	.7069	.9957	.9991	.9974

XGBoost achieved an impressive accuracy of 99.95%, proved its potent capability to identify phishing URLs with high precision. This high performance can be attributed to XGBoost's robust gradient boosting framework, which effectively handles complex patterns and interactions within the feature set.

AdaBoost also performed well, with an accuracy of 99.34%. AdaBoost's ensemble approach, which combines multiple weak classifiers into a strong predictor, contributed to its effective classification of URLs, although it slightly lagged behind XGBoost.

The accuracy of 70.69% for SGD (Stochastic Gradient Descent) was noticeably lower. There are various reasons for this decreased performance. Because SGD is a more straightforward linear model, it might not be able to handle the dataset's complicated feature interactions. It might not be as good at identifying the subtle patterns linked to phishing URLs as XGBoost and AdaBoost, which are more sophisticated and designed to deal with such complexity. Furthermore, the model's sensitivity to feature scaling and parameter tweaking may have an impact on its performance; therefore, for best outcomes, rigorous optimization is necessary. Additionally, the model's performance might be influenced by its sensitivity to feature scaling and parameter tuning, which requires careful optimization for optimal results.

Overall, while XGBoost and AdaBoost demonstrated excellent accuracy in detecting phishing URLs, SGD's simpler approach and its limitations in handling complex features led to a lower accuracy. This highlights the importance of selecting appropriate models based on the complexity of the problem and the nature of the data.

5. CONCLUSIONS

In this paper, we investigated how well machine learning methods work to identify phishing URLs, a vital aspect for new age cybersecurity. Our results show that using sophisticated algorithms can greatly improve the precision and effectiveness of phishing detection systems. We reduced the likelihood that users would be exposed to phishing attacks by precisely identifying bad URLs using a combination of feature extraction techniques and machine learning models.

The findings highlight how crucial it is to continuously train and modify models in order to stay up to date with the ever-evolving strategies used by hackers. Our research also emphasizes the necessity of large-scale datasets that capture the latest phishing patterns in order to strengthen detection systems' resilience.

The incorporation of machine learning in phishing URL identification is a viable approach to augmenting online security, given the ongoing sophistication of cyber-attacks. In an increasingly digital society, future research should concentrate on improving existing models, investigating hybrid techniques, and creating real-time detection systems that can successfully counter phishing threats. By giving these initiatives top priority, we can help make the internet a safer place for both individuals and businesses.

REFERENCES

[1] Abdelhamid, N., Aladdin, A., Thabtah, F.: Phishing detection based associative classification data mining. Expert Syst. Appl. 41(3), 5948–5959 (2014)

[2] Le, A., Markopoulou, A., Faloutsos, M.: PhishDef: URL names say it all. In: Proceedings IEEE, INFOCOM. IEEE (2011)

[3] Garera, S., Provos, N., Chew, M., Rubin, A.: A framework for detection and measurement of phishing attacks. In: Proceedings of the ACM workshop on Recurring malcode, pp. 1–8. ACM (2007)

[4] SlashNext- the state of phishing report (2023). Retrieved fromhttps://slashnext.com/wpcontent/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf

[5] Doyen Sahoo, Chenghao Liu, and Steven C.H. Hoi. Malicious URL Detection using Machine Learning: A Survey. https://doi.org/10.1145/nnnnnnn.nnnnnnn (2019)

[6] Iwendi, C.; Jalil, Z.; Javed, A.R.; Reddy, G.T.; Kaluri, R.; Srivastava, G.; Jo, O. KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against Cyber-Attacks. IEEE Access (2020)

[7] Arathi Krishna V, Anusree A, Blessy Jose, Karthika Anilkumar, Ojus Thomas Lee, 2021, Phishing Detection using Machine Learning based URL Analysis: A Survey, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCREIS (2021)

[8] Elsadig M, Ibrahim AO, Basheer S, Alohali MA, Alshunaifi S, Alqahtani H, Alharbi N, Nagmeldin W. Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction. Electronics. (2022)

[9] Blum, A., Wardman, B., Solorio, T., and Warner, G., "Lexical Feature Based Phishing URL Detection Using Online Learning". in AISec '10 Proceedings of the 3rd ACM workshop on Artificial Intelligence and Security, Illinois, USA, ACM New York, NY, USA, pp. 54-60 (2010)

[10] MA, J., Saul, L.K., Savage, S., and Voelker, G.M., "Learning to Detect Malicious URLs". in ACM Transactions on Intelligent Systems and Technology. 2, 3, Article 30 April, ACM New York, NY, USA, pp. 1245-1254 (2011)

[11] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari and S. R. K. Joga, "Phishing Detection System Through Hybrid Machine Learning Based on URL," in IEEE Access, vol. 11, pp. 36805-36822 (2023)

[12] Shirazi, H., Bezawada, B., Ray, I., Anderson, C. Adversarial Sampling Attacks Against Phishing Detection. In: Foley, S. (eds) Data and Applications Security and Privacy XXXIII. DBSec 2019. Lecture Notes in Computer Science(), vol 11559. Springer (2019)

[13] Sasi, Jishnu Kaitholikkal, and Arthi Balakrishnan. "Generative adversarial network-based phishing URL detection with variational autoencoder and transformer." *Int J Artif Intell* 13.2: 2165-2172 (2024)

[14] Christou, O., Pitropakis, N., Papadopoulos, P., McKeown, S. and Buchanan, W., "Phishing URL Detection Through Top-level Domain Analysis: A Descriptive Approach". In Proceedings of the 6th International Conference on Information Systems Security and Privacy (ICISSP), pages 289-298 (2020)

[15] Dataset from https://archive.ics.uci.edu/datasets

[16] Chu, W., et al.: Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs. In: IEEE International Conference on Communications (ICC) (2013)

[17] Thomas, K., et al.: Design and evaluation of a real-time URL spam filtering service. In: Proceeding of the IEEE Symposium on Security and Privacy (SP) (2011)

[18] Abdelhamid, N., Aladdin, A., Thabtah, F.: Phishing detection based associative classification data mining. Expert Syst. Appl. 41(3), 5948 5959 (2014)

[19] Xiang, G., et al.: CANTINA+: a feature-rich machine learning framework for detecting phishing web sites. ACM Trans. Inf. Syst. Secur. (TISSEC) 14(2), 21 (2011)

[20] Xu, L., et al.: Cross-layer detection of malicious websites. In: Proceedings of the Third ACM Conference on Data and Application Security and Privacy. ACM (2013)

[21] Ma, J., Saul, L., Savage, S., and Voelker, G., "Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs". in KDD'09 Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, ACM New York, NY, USA, pp. 1245-1254 (2019)