Phishing Website Detection Using Machine Learning: A Review

Marwa Abd Al Hussein Qasim (✉)

College of Computer Science & Information Technology, Basrah University, Iraq
itpg.marwa.qasim@uobasrah.edu.iq

Dr. Nahla Abbas Flayh
College of Computer Science & Information Technology, Basrah University, Iraq
nahla.flayh@uobasrah.edu.iq

Abstract— Phishing, a form of cyber-attack in which perpetrators employ fraudulent websites or emails to Deceive individuals into divulging sensitive information such as passwords or financial data, can be mitigated through various machine-learning algorithms for website detection.

These algorithms, including decision trees, support vector machines, and Random Forest, analyze multiple website features, such as URL structure, website content, and the presence of specific keywords or patterns, to ascertain the likelihood of a website being a phishing site.

This comprehensive review elucidates the concept of phishing website detection and the diverse techniques employed while summarizing previous studies, their outcomes, and their contributions. Overall, machine learning algorithms serve as a potent tool in the identification of phishing websites, thereby safeguarding users against falling prey to such malicious attacks.

Keywords— Phishing Detection, Machine learning, Phish Tank

I. Introduction

In contemporary times, a substantial portion of the population is well aware of the utilization of the Internet for a multitude of purposes, including online banking, shopping, bill payments, and mobile device recharges. However, users engaging in these online activities often face a plethora of security concerns, ranging from cybercrime and spam to fraud and cyber terrorism, with phishing being just one among the various types of cybercrimes that are commonly perpetrated [1].

The objective of machine learning, which is a subfield of artificial intelligence, is to create systems that can improve and learn without explicit programming through experience [2].

In the field of machine learning, there are two distinct types of learning methodologies, namely supervised and unsupervised learning [3]. In supervised learning, the training dataset is composed of previous instances where both the input and output values are known and provided as labeled data. [4].

One approach to detecting phishing websites utilizing machine learning involves the utilization of supervised learning, where the training dataset exclusively comprises labeled data [5].

The process involves training a model with a dataset that encompasses both phishing and legitimate websites, enabling the model to acquire characteristics for distinguishing between the two types. Subsequently, the trained model can be employed to classify new websites as either phishing or legitimate, based on the learned features obtained from the training dataset. Notable features that can be leveraged for detecting phishing websites include the presence of specific words or phrases in the website's content or URL, the structure of the website's URL, and the overall layout and design of the website. Additionally, other features such as the presence of SSL certificates or the age of the domain may also prove valuable in the detection of phishing websites [6].

There exist multiple phishing detection techniques that utilize approaches such as white-listing, black-listing, content-based analysis, URL-based analysis, visual-similarity analysis, and machine-learning algorithms[7].

To effectively train a machine learning model to detect phishing websites, it is imperative to utilize a substantial and diverse dataset that encompasses both phishing and legitimate websites. Additionally, the trained model should be thoroughly evaluated and tested on a separate dataset to ascertain its accuracy and reliability in accurately discerning between phishing and legitimate websites [3].

In the following section, the concept of detecting phishing websites and their techniques will be elucidated. The challenges encountered in detecting phishing websites will also be discussed, as well as a summary of previous studies. Finally, a conclusion will be discussed.

2. Phishing Website Detection

The detection of phishing websites entails the identification of websites that are intentionally created to deceive users into revealing their personal information, typically by imitating legitimate websites. These fraudulent websites often replicate the appearance and functionality of genuine websites, posing a considerable challenge in discerning between authentic and deceptive sites.

There are several techniques used in phishing website detection, including:

A. URL Analysis: this is a formalized process that involves scrutinizing the structural components of a website's URL to identify any aberrations or inconsistencies that may signal a phishing endeavor. For instance, phishing

websites may employ URLs that closely mimic legitimate ones but contain subtle deviations, such as misspelled words or extra subdomains, Machine learning (ML)-based phishing URL detectors function as an initial line of defense aimed at safeguarding users and organizations against falling prey to phishing attacks[8]

B. Content Analysis: which involves a meticulous examination and scrutiny of various elements within websites, including text, images, and links, has the potential to detect malicious intentions associated with a website. As a result, cybersecurity management's technical approaches must incorporate automated detection mechanisms aimed at thwarting phishing attacks [9]. The primary purpose of content analysis is to identify and analyze potentially suspicious elements that may indicate the presence of phishing attacks. This technique diligently examines the content and structure of websites with the aim of detecting and mitigating potential phishing attacks by identifying any indications of deception or fraudulent activity.

Typically, content analysis involves the utilization of automated tools or algorithms that analyze the textual content of a website's pages, including keywords, phrases, and patterns that are commonly associated with phishing attacks. Additionally, the content analysis may encompass the examination of images and links within the website, the number of words, number of characters as these elements can also provide clues to the authenticity of a website [10]. Through content analysis, cybersecurity experts can identify and flag websites that exhibit suspicious characteristics, such as the presence of phishing-related keywords in the content of the website or other indications of potentially fraudulent activity. This analysis can aid in the early detection of potential phishing attacks, enabling the prompt implementation of mitigation measures to safeguard users from falling prey to such attacks.

Security Indicator Analysis: this is a pivotal facet of content analysis when it comes to the detection of phishing websites. It encompasses a meticulous examination of security indicators, such as SSL certificates, which are instrumental in establishing secure communication between a website and its users. SSL certificates serve as an indication of a website's possession of a valid encryption certificate, which ensures a secure connection for transmitting sensitive information. In the context of phishing attacks, malicious websites may lack SSL certificates altogether or may utilize invalid or expired certificates. Such instances can serve as red flags, suggesting potential malicious intentions, as legitimate websites typically uphold up-to-date SSL certificates to ensure secure communication with their users. In instances where criminals engage in phishing attacks, they employ analytical techniques on specific components of the feature set content, such as obfuscating certain strings and manipulating address numbers. In such scenarios, the preservation of a backup record of the target source, character string, and

SSL certificate number located in the address bar of the legitimate website serves as a crucial point of reference during the detection process[11].

1.2 - Structure of A URL

Figure 1 illustrates the structure of a URL.

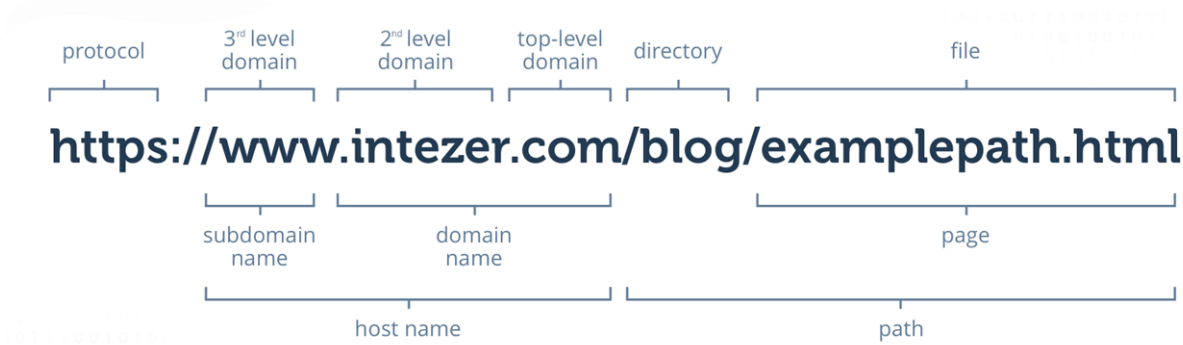

Fig .1: Structure of A URL

3 The Challenges of Phishing Website Detection

The process of detecting phishing websites can be a demanding and intricate task, primarily due to the following factors:

Sophisticated Techniques: The prevalence of phishing attacks is on the rise, and they are becoming increasingly advanced in their methods, which can include the utilization of various techniques such as social engineering tactics to deceive users into disclosing confidential information. There is a possibility that malicious actors may possess the necessary expertise and motivation to bypass URL classification algorithms by creating instances that can evade detection by such algorithms[12].

1. Evolving Tactics: Adversaries often modify their methods and approaches to avoid detection, which poses a significant challenge for security tools to keep up with. They employ sophisticated tools and techniques to infiltrate computer networks and systems. These attacks are capable of circumventing firewalls and antivirus programs to illicitly obtain confidential information[13].

2. Use Of Subdomains: Adversaries can employ subdomains to develop deceptive phishing websites that mimic authentic ones, thereby complicating the task for users to differentiate between them [14].

3. Time-Sensitive Attacks: Malicious actors often execute time-sensitive attacks that remain active for brief durations, posing a challenge for security professionals to detect and dismantle their associated websites or platforms. A prime illustration of such attacks is the Watering Hole Attack[15].

4. Limited Data Availability: Limited or incomplete data may be available to identify a phishing website, which can impede the detection process[16].

5. False Positives: Authentic websites could occasionally activate phishing alerts due to multiple factors, such as obsolete databases, erroneous algorithms, or analogous domain names. The resulting false positives may cause superfluous warnings to users, generating feelings of exasperation and diminishing reliance on the anti-phishing mechanism. [17].

6. Resource-Intensive: The application of online learning and semi-supervised learning in a real-time anti-phishing mechanism underscores the necessity of significant computing resources and intricate algorithms to scrutinize voluminous data streams in real-time [18]. The process of scrutinizing web content, URLs, and other pertinent features to recognize plausible phishing websites is an arduous computational endeavor, thereby emphasizing the need for effective algorithms to achieve optimal detection precision while minimizing the false positive rate.

In summary, detecting phishing websites necessitates a blend of technical know-how, advanced tools and techniques, and a comprehensive grasp of current phishing attack trends to effectively identify and mitigate these threats.

4. Literature Review

In this paper, we will discuss previous studies in terms of their methods, datasets, contributions, and results.

S. Arvind Anwekar, V. Agrawal [19]: In this study, the authors focused on extracting features from URLs, in addition to other features such as the age of the SSL certificate and the universal resource locator of the anchor, IFRAME, and website rank. They collected URLs of phishing websites from PhishTank and URLs of benign websites from the Alexa website. Using a combination of the random forest (RF), decision tree

(DT), and support vector machine (SVM), contributed to improving the detection mechanism for phishing websites and achieved a high noticeable detection accuracy of 97.14%, with a low rate of false positives at 3.14%. The results also showed that the classifier's performance improves with more training data.

N. Choudhary b, K. Jain, S. Jain [20]: This study emphasizes the significance of only using attributes from the URL. Both the Kaggle and Phishtank websites make it easy to get the dataset used in this study. The researchers used a hybrid approach that combined Principal Component Analysis (PCA) with Support Vector Machine (SVM) and Random Forest algorithms to reduce the dataset's dimensionality while keeping all important data, and it produced a higher accuracy rate of 96.8% compared to other techniques investigated.

A. Lakshmanarao, P. Surya, M Bala Krishna [21]: This thesis collected a dataset of phishing websites from the UCI repository and used various Machine learning techniques, including decision trees, AdaBoost, support vector machines (SVM), and random forests, to analyze selected features (such as web traffic, port, URL length, IP address, and URL_of_Anchor). The most effective model for detecting phishing websites was chosen, and two priority-based algorithms (PA1 and PA2) were proposed. The team utilized a new fusion classifier in conjunction with these algorithms and attained an accuracy rate of 97%. when compared to previous works in phishing website detection

L. Tang, Q. Mahmoud [22]: The proposed approach in the current study uses URLs collected from a variety of platforms, including Kaggle, Phish Storm, Phish Tank, and ISCX-UR, to identify phishing websites. The researchers made a big contribution since they created a browser plug-in that can quickly recognize phishing risks and offer warnings. Various datasets and machine learning techniques were investigated, and the proposed RNN-GRU model outperformed SVM, Random Forest (RF), and Logistic Regression with a maximum accuracy rate of 99.18%. On the other hand, the suggested method is not always accurate in identifying if short links are phishing risks.

A. Kulkarni & L. Brown[23]: A machine learning system was created to categorize websites based on URLs from the University of California, Irvine Machine Learning Repository. Four classifiers were used: SVM, decision tree, Naive Bayesian, and neural network. The outcome of experiments utilizing the model developed with the support of a training set of data demonstrates that the classifiers were able to successfully differentiate authentic websites from fake ones with an accuracy rate of over 90%. Limitations include a small dataset and all features being discrete, which may not be suitable for some classifiers.

Tyagi; J. Shad; S. Sharma; S. Gaur Gagandeep Kaur [24]: The research taken into account focuses on the use of various machine learning algorithms to identify if a website is legitimate or a phishing site based on a URL. This study's most important contribution is the creation of the Generalized Linear Model (GLM), a brand-new model. This model combines the results of two various methods. With a 98.4% accuracy rate,

the Random Forest and GLM combination produced the best results for detecting phishing websites.

M. Karabatak and T. Mustafa [25]: The objective of this research is to assess the effectiveness of classification algorithms on a condensed dataset of phishing websites obtained from the UCI Machine Learning Repository. The paper investigates how data mining and feature selection algorithms affect reduced datasets through experiments and analysis, finally selecting the methods that perform the best in terms of classification. According to the results, some classification strategies improve performance while others have the opposite impact. Ineffective classifiers for condensed phishing datasets included Lazy, BayesNet, SGD Multilayer Perceptron, PART, JRip, J48, RandomTree, and RandomForest. However, it was discovered that KStar, LMT, ID3, and R.F.Classifier were efficient. Lazy produced the highest classification accuracy rate of 97.58% on the compressed 27-feature dataset, whereas KStar performed at its best on the same dataset.

X. Zhang, Y. Zeng, X. Jin, Z. Yan, and G. Geng [26]: A phishing detection model that applies Bagging, AdaBoost, SMO, and Random Forest algorithms to learn and test phishing detection strategies is offered as a contribution to this work. The model is based on features from URLs and extracts multi-level statistical characteristics, semantic features of word embedding, and semantic features from Chinese web content. Legal URLs from DirectIndustry online instructions and phishing data from the Anti-Phishing Alliance of China (APAC) are included in the dataset used to test the algorithm. The study's findings suggest that a fusion model that primarily employed semantic data to identify phishing sites with high detection efficiency had the best performance, leading to a new contribution with an F-measure of 0.99%. Keep in mind that this approach is specific to Chinese websites and is language-dependent.

W. Fadheel, M. Abusharkh, and I. Abdel-Qader [27]: The present study utilized datasets from the UCI machine learning repository, including Domain, HTML, Address Bar, and URLs, the main contribution was conducting a comparative analysis of the impact of feature selection on detecting phishing websites. The KMO test was applied in the study to evaluate the dataset using (LR) and (SVM) classification algorithms. The test was conducted based on a correlation matrix to analyze the performance. Results showed that LR with the KMO test achieved an accuracy of 91.68%, while SVM with the KMO test yielded an accuracy of 93.59%

A. Ahmed and N. A. Abdullah [28]: The research team developed a software program known as Phish Checker, which is designed to distinguish between legitimate and phishing websites. The proposed approach focuses on identifying phishing attacks by analyzing the URLs and domain names of suspected phishing websites to determine their authenticity. Data was collected from the Yahoo and PhishTank directories and the results indicate that PhishChecker has an accuracy rate of 96% for identifying phishing websites. However, it should be noted that this method is based on heuristics and its effectiveness is reliant on the availability of certain discriminative elements that aid

in identifying the type of website. Additionally, the study only examines the validity of URLs in determining website authenticity.

Ankit Kumar Jain & B. B. Gupta [29]: The proposed strategy utilizes an Innovative methodology for defending counteract phishing attempts by incorporating a URL and DNS matching module with a white list of trusted websites that are automatically updated based on each user's browsing history. This method offers quick retrieval speeds, high rates of detection, and alerts users to not disclose personal information when attempting to access a website, not on the white list. Additionally, hyperlink properties are utilized to verify the validity of a website by retrieving hyperlinks from the source code and applying them to the phishing detection method. The performance of this strategy was evaluated using data from reputable sources such as Stuffgate, Alexa, and PhishTank and achieved an accuracy rate of 89.38 %

M. Aydin and N. Baykal [30]: Throughout this experiment, phishing websites were detected using subset-based feature selection methods based on URL attributes. A dataset comprising both legitimate and phishing URLs was obtained from Google and PhishTank, and multiple features were retrieved from URLs. The usefulness of two classification algorithms—Naive Bayes and Sequential Minimal Optimization (SMO)—for identifying phishing websites was investigated in this study. The results showed that SMO performed better than Naive Bayes for phishing detection, with an accuracy rate of 95.39%. The SMO algorithm also had another benefit in that it made use of more chosen features overall. The accuracy rate of the Naive Bayes method, in contrast, was 88.17% while using the same quantity of chosen features.

S. Smadi∗, N. Aslam, Li Zhang, R. Alasem, and M A Hossain [31]: The intelligent model in this study was built to be capable of distinguishing between legitimate emails and phishing emails by utilizing attributes extracted from both the email header and body. using ten data mining techniques, and it was discovered that the RF, J48, and PART algorithms had the best precision levels, obtaining 98.87%, 98.11%, and 98.10%, respectively. The legal email dataset was taken from the Spam Assassin project, while the phishing email dataset was sourced via Nazario. The study found that the outcomes of the classification model were considerably influenced by the features extracted during the preprocessing stage. Notably, when compared to comparable models at the time of publication, the model described in this study had the best accuracy and the lowest false positive rate.

L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen [32]: Using six criteria based on URL parameters such as the subdomain, principal domain, Page rank, Alexa rank, path domain, and Alexa reputation, this article suggests a novel method for identifying phishing websites. The method focuses on evaluating how closely a phishing site's URL resembles the URL of a reliable website and also takes into account the site's ranking as a crucial component in determining its validity. The approach was tested using data from PhishTank and DMOZ, and the authors showed that it could identify over 97% of phishing sites.

Weibo Chu; Bin B. Zhu; Feng Xue; Xiaohong Guan; Zhongmin Cai [33]: They tested the effectiveness of phishing detection methods based on machine learning utilizing a secure website as part of their contribution to this work. The authors presented and tested several useful features for incorporation into the detector based only on lexical and domain characteristics. Finding the ideal mix of attributes led to the creation of a detector with a detection rate higher than 98%. Support vector machines and Gaussian radial basis function algorithms were used in the study, and the datasets used included phishing URLs from the Taobao-phishing dataset, safe URLs from the Yahoo! directory, and well-known Chinese navigational websites.

5. Summary Of the Literature Review

Table 1: Summary of The Literature Review

Reference	Year of publication	Method/ Technique	Datasets	Result
[17]	2022	Decision tree, Random Forest, and Support vector machine (SVM)	Alexa And Phishtank	Accuracy: 0.97
[18]	2022	Random Forest and (SVM)	Kaggle and PhishTank website	Accuracy: 0.96 Precision: 0.96 F-Score: 0.97
[19]	2021	decision trees, support vector machines, random forests, and AdaBoost	UCI machine learning repository	Accuracy: 0.97
[20]	2021	SVM, Random Forest, , Logistic Regression and RNN-GRU	Phish Storm, Phish Tank, ISCX-UR, and Kaggle	Accuracy: 0.99
[21]	2019	decision tree, Naïve Bayesian classifier, support vector machine (SVM), and neural network	University of California, Irvine Machine Learning Repository	Accuracy: 0.90
[22]	2018	Decision Tree, Random Forest, and Generalized Linear Model (GLM)	N/A	Accuracy: 0.98 Precision: 0.97 Recall: 0.98
[23]	2018	azy, BayesNet, SGD Multilayer Perceptron, PART,	UCI machine learning repository	Accuracy: 0.97 with 27 reduced features

		JRip, J48, RandomTree, RandomForest, KStar, LMT, and ID3		
[24]	2017	AdaBoost, Bagging, Random Forest, and SMO	DirectIndustry web guides & Anti-Phishing Alliance of China	F-Score: 0.99
[25]	2017	Logistic Regression (LR) and Support Vector Machine (SVM)	UCI machine learning repository	KMO test with LR Accuracy: 0.91 KMO test with SVM Accuracy: 0.93
[26]	2016	PhishChecker application	PhishTank and Yahoo directory datasets	Accuracy:0.96
[27]	2016	hyperlink information, and white-list	PhishTank, Alexa, Stuffgate, and Online payment service provider	Accuracy: 0.89
[28]	2015	Naive Bayes, and Sequential Minimal Optimization (SMO)	PhishTank and legitimate URLs from Google	Naive Bayes Accuracy: 0.88 Optimization (SMO) Accuracy: 0.95
[29]	2015	Random Forest (RF), J48, and PART	Nazario and SpamAssassin project	Accuracy RF: 0.98 Accuracy J48: 0.98 Accuracy PART: 0.98
[30]	2014	heuristic features detection method	PhishTank and DMOZ	Accuracy: 0.97
[31]	2013	Support Vector Machine (SVM), and Gaussian Radial Basis Function (RBF)	Taobao-phishing dataset, Yahoo!, And popular Chinese navigational websites	Accuracy: 0.98

6. Conclusion

Previous studies have shown that machine learning algorithms effectively detect phishing websites. Many studies in recent years have employed hybrid algorithms to achieve high accuracy, and a system utilizing the Random Forest algorithm as one of the hybrid algorithms can achieve an accuracy more than of 99%. However, it is important to note that limitations exist in previous studies and that a single method may not be effective in all cases due to the constantly evolving tactics used by phishers. One of the suggestions for the future studies is to explore the use of deep learning techniques, such as neural networks, for phishing detection.

7. References

[1] A. J. Ashutosh Kumar Singh, and Keshav Singh, "A Survey on Cyber Security Awareness and Perception among University Students in India," Journal of Advances in Mathematics and Computer Science, November 2021.
[2] S. Shams Hussein, W. Hashim Abdulsalam, and W. Abed Shukur, "Covid-19 Prediction using Machine Learning Methods: An Article Review," Wasit Journal of Pure Sciences, vol. 2, no. 1, pp. 217-230, 03/26 2023, doi 10.31185/wjps.124.
[3] S. Mahdi Muhammed, G. Abdul-Majeed, and M. Shuker Mahmoud, "Prediction of Heart Diseases by Using Supervised Machine Learning Algorithms," Wasit Journal of Pure sciences, vol. 2, no. 1, pp. 231-243, 03/26 2023, doi: 10.31185/wjps.125.
[4] N. Kareem, "Afaster Training Algorithm and Genetic Algorithm to Recognize Some of Arabic Phonemes."
[5] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student performance prediction model based on supervised machine learning algorithms," in IOP Conference Series: Materials Science and Engineering, 2020, vol. 928, no. 3: IOP Publishing, p. 032019.
[6] H. H. Chinaza Uchechukwu, and Jianguo Ding, "A Survey of Machine Learning Techniques for Phishing Detection," IEEE Access, August 2020.
[7] P. Kalaharsha and B. M. Mehtre, "Detecting Phishing Sites--An Overview," arXiv preprint arXiv:2103.12739, 2021.
[8] B. Sabir, M. A. Babar, R. Gaire, and A. Abuadbba, "Reliability and Robustness analysis of Machine Learning based Phishing URL Detectors," IEEE Transactions on Dependable and Secure Computing, 2022.
[9] M. Almousa, T. Zhang, A. Sarrafzadeh, and M. Anwar, "Phishing website detection: How effective are deep learning-based models and hyperparameter optimization?," Security and Privacy, vol. 5, no. 6, p. e256, 2022.
[10] H. Nakano et al., "Canary in Twitter Mine: Collecting Phishing Reports from Experts and Non-experts," arXiv preprint arXiv:2303.15847, 2023.
[11] Q. Zhang, "Practical Thinking on Neural Network Phishing Website Detection Research Based on Decision Tree and Optimal Feature Selection," in Journal of Physics: Conference Series, 2021, vol. 2031, no. 1: IOP Publishing, p. 012062.
[12] A. AlEroud and G. Karabatis, "Bypassing detection of URL-based phishing attacks using generative adversarial deep neural networks," in Proceedings of the sixth international workshop on security and privacy analytics, 2020, pp. 53-60.
[13] M. Mijwil, O. J. Unogwu, Y. Filali, I. Bala, and H. Al-Shahwani, "Exploring the Top Five Evolving Threats in Cybersecurity: An In-Depth Overview," Mesopotamian journal of cybersecurity, vol. 2023, pp. 57-63, 2023.
[14] A. A. E. K. Yassine El Hajjaji, and Abdellah Ezzati, "Phishing Attacks and Countermeasures: A Survey," IEEE Access, 2020.
[15] P. R. Brandão and G. P. Matos, "Machine Learning and APTs."
[16] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep learning for phishing detection: Taxonomy, current challenges and future directions," IEEE Access, 2022.
[17] M. H. A. a. A. A. Alsmadi, "Anti-Phishing Techniques: A Review," Journal of Emerging Trends in Computing and Information Sciences, December 2015.
[18] S. L. Xu Chen, Wei Wang, and Xiaodan Zhang, "A Real-Time Anti-Phishing Method Based on Online Learning and Semi-Supervised Learning," Journal of Computational Science, October 2021.
[19] S. A. Anwekar and V. Agrawal, "PHISHING WEBSITE DETECTION USING MACHINE LEARNING ALGORITHMS."

[20] S. Jain, "Phishing Websites Detection Using Machine Learning," Available at SSRN 4121102.
[21] A. Lakshmanarao, P. S. P. Rao, and M. B. Krishna, "Phishing website detection using novel machine learning fusion approach," in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021: IEEE, pp. 1164-1169.
[22] L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," IEEE Access, vol. 10, pp. 1509-1521, 2021.
[23] A. D. Kulkarni and L. L. Brown III, "Phishing websites detection using machine learning," 2019.
[24] I. Tyagi, J. Shad, S. Sharma, S. Gaur, and G. Kaur, "A novel machine learning approach to detect phishing websites," in 2018 5th International conference on signal processing and integrated networks (SPIN), 2018: IEEE, pp. 425-430.
[25] M. Karabatak and T. Mustafa, "Performance comparison of classifiers on reduced phishing website dataset," in 2018 6th International Symposium on Digital Forensic and Security (ISDFS), 2018: IEEE, pp. 1-5.
[26] X. Zhang, Y. Zeng, X.-B. Jin, Z.-W. Yan, and G.-G. Geng, "Boosting the phishing detection performance by semantic analysis," in 2017 IEEE international conference on big data (big data), 2017: IEEE, pp. 1063-1070.
[27] W. Fadheel, M. Abusharkh, and I. Abdel-Qader, "On Feature selection for the prediction of phishing websites," in 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2017: IEEE, pp. 871-876.
[28] A. A. Ahmed and N. A. Abdullah, "Real time detection of phishing websites," in 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016: IEEE, pp. 1-6.
[29] A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list," EURASIP Journal on Information Security, vol. 2016, no. 1, pp. 1-11, 2016.
[30] M. Aydin and N. Baykal, "Feature extraction and classification phishing websites based on URL," in 2015 IEEE Conference on Communications and Network Security (CNS), 2015: IEEE, pp. 769-770.
[31] S. Smadi, N. Aslam, L. Zhang, R. Alasem, and M. A. Hossain, "Detection of phishing emails using data mining algorithms," in 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2015: IEEE, pp. 1-8.
[32] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen, "A novel approach for phishing detection using URL-based heuristic," in 2014 International conference on Computing, management, and telecommunications (ComManTel), 2014: IEEE, pp. 298-303.
[33] W. Chu, B. B. Zhu, F. Xue, X. Guan, and Z. Cai, "Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs," in 2013 IEEE international conference on communications (ICC), 2013: IEEE, pp. 1990-1994.

Article submitted 2 March 2023. Accepted at 17 May. Published at 30 Jun 2023.