LITERATURE SURVEY

Weiheng Bai in [1] analysed the structural features of the URL of the phishing website, extracts 12 kinds of features, and used four machine learning algorithms for training. Then, the best performing algorithm is used as our model to identify unknown URLs. After the recognition is completed, a snapshot of the web page is extracted and compared with the regular web page snapshot to implement the recommendation of the original regular web page of the phishing web page.

Lizhen Tang et.al in [2] implemented the framework as a browser plug-in capable of determining whether there is a phishing risk in real-time when the user visits a web page and gives a warning message. The real-time prediction service combines multiple strategies to improve accuracy, reduce false alarm rates, and reduce calculation time, including whitelist filtering, blacklist interception, and machine learning (ML) prediction.

Farashazillah Yahya et.al in [3] compared the three models such as Decision Tree, K-Nearest Neighbour (KNN), and Random Forest. However, KNN was the prime candidate for the best model considering that it has the highest accuracy. However, Random Forest is deemed more suitable for the dataset even though the accuracy is lesser because of the lowest false-negative value than the other models. The experiments can be further investigated with different datasets and models for comparative analysis.

Qasem Abu Al-Haij et.al in [4] presented an efficient phishing websites detection system that analyzes the phishing websites URL addresses to learn data patterns that can identify authentic and phishing websites. Our system employs machine learning techniques such as neural networks and decision trees to learn

data patterns in websites URLs. The result shows that decision trees models provide 97.40% classification accuracy on the almost balanced-class dataset.

Rizka Widyarini Purwanto et.al in [5] proposed a feature-free method for detecting phishing websites using the Normalized Compression Distance (NCD), a parameter-free similarity measure which computes the similarity of two websites by compressing them, thus eliminating the need to perform any feature extraction. They used the Furthest Point First algorithm to perform phishing prototype extractions, in order to select instances that are representative of a cluster of phishing webpages.

Nguyet Quang Do et.al in [6] proposed a taxonomy of deep learning algorithm for phishing detection by examining 81 selected papers using a systematic literature review approach. The paper first introduces the concept of phishing and deep learning in the context of cybersecurity. Then, taxonomies of phishing detection and deep learning algorithm are provided to classify the existing literature into various categories. Next, taking the proposed taxonomy as a baseline, this study comprehensively reviews the state-of-the-art deep learning techniques and analyzes their advantages as well as disadvantages.

Manuel Sánchez-paniagua et.al in [7] compared the machine learning and deep learning techniques to present a method capable of detecting phishing

websites through URL analysis. In most current state-of-the-art solutions dealing with phishing detection, the legitimate class is made up of homepages without including login forms. On the contrary, we use URLs from the login page in both classes because we consider it is much more representative of a real case scenario and we demonstrate that existing techniques obtain a high false-positive rate when tested with URLs from legitimate login pages.

Smriti Dangwal et.al in [8] combined two datasets with 30 and 48 features respectively, to identify 18 common features. Moreover, feature selection was conducted to identify 13 optimal features for a more robust model. A comparison with prior research works on the same datasets showed that the best models built on all features using the random forest algorithm scored lower on the 30 feature dataset, and achieved better performance on the 48 features dataset. The best model on the 13 features achieved an accuracy of 0.937.

Athulya A.A. et.al in [9] discussed various phishing attacks, some of the latest phishing evasion techniques used by attackers and anti-phishing approaches. This review raises awareness of those phishing strategies and helps the user to practice phishing prevention. Here, a hybrid approach of phishing detection also described having fast response time and high accuracy.

Lakhita et.al in [10] lexically analysed the URLs can enhance the performance and help to differentiate between the original email and the phishing URL. As assessed in this study, in addition to textual analysis of phishing URL, email classification is successful and results in a highly precise anti phishing.

Malak Aljabri et.al in [11] used intelligent techniques mainly Machine Learning (ML) and Deep Learning (D L) are increasingly applied in the field of cybersecurity due to their ability to learn from available data in order to extract useful insight and predict future events. The effectiveness of applying such intelligent approaches in detecting phishing web sites is investigated in this paper. A set of ML models were then applied, and a comparative performance evaluation was conducted. Results proved the importance of features selection in improving the models' performance. Furthermore, the results also aimed to identify the best features that influence the model in identifying phishing websites.

Aaisha Makkar et.al in [12] developed a safe data sharing architecture for various IIoT devices using federated learning (FL). The proposed architecture incorporates FL into the edge computing consensus process, allowing the consensus computing activity to be used for federated training as well. The proposed framework achieves high efficiency, and better security, according to numerical findings generated by experimenting deep learning models. More

precisely, the proposed framework named as SecureIIoT, is able to achieve 99.79% accuracy by detecting attacks as a binary classification problem.

Mustafa Al Fayoumi et.al in [13] treated the detection of a phished email as a classification problem and this paper shows how machine learning methods are used to categorize emails as phished or not. SVM classifier attains a maximum accuracy of 0.998 percent in email classification.

Parv Rastogi et.al in [14] proposed to ascertain malicious URLs, which is formulated on random forest, support vector machine (SVM), deep neural network (DNN), convolutional neural network (CNN). The several datasets are considered containing malicious and benign URLs to train the model to detect URL behaviour and attributes. The empirical results show that the suggested method can detect malicious URLs efficiently, based on URL behaviour and attributes. Thus, the solution may be advised as an efficient and reliable solution for the problem of malicious URL detection.

Areti Nagendra Soma Charan et.al in [15] used URLs as a dataset to detect phishing websites. The dataset contains 6000 URLs, from which ten features were extracted and utilized to determine if the website was phishing or not. Eight machine learning algorithms were designed for this research. The performance

analysis results show that the Multilayer perceptron algorithm got the highest accuracy of 85.41% and an F1 score of 85.17% compared with other algorithms.

REFERENCES

1. Weiheng Bai “Phishing Website Detection Based on Machine Learning Algorithm” in 2020 International Conference on Computing and Data Science (CDS) 10.1109/CDS49703.2020.00064
2. Lizhen Tang and Qusay H. Mahmoud “ Deep Learning-Based Framework for Phishing Website Detection”
3. Farashazillah Yahya, Ryan Isaac W Mahibol, Chong Kim Ying, Magnus Bin Anai, Sidney Allister Frankie, Eric Ling Nin Wei, Rio Guntur Utomo “Detection of Phising Websites using Machine Learning Approaches” in 2021 International Conference on Data Science and Its Applications (ICoDSA) 10.1109/ICoDSA53588.2021.9617482
4. Qasem Abu Al-Haij, Ahmad Al Badawi “URL-based Phishing Websites Detection via Machine Learning” in 2021 International Conference on Data Analytics for Business and Industry (ICDABI) 10.1109/ICDABI53623.2021.9655851

5. Rizka Widyarini Purwanto, Arindam Pal, Alan Blair, and Sanjay Jha "PHISHSIM: Aiding Phishing Website Detection With a Feature-Free Tool " in IEEE Transactions On Information Forensics and Security, Vol. 17, 2022

6. Nguyet Quang Do, Ali Selamat, Ondrej Krejcar ,Enrique Herrera-viedma , and Hamido Fujita " Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions" Volume 10, 2022.

7. Manuel Sánchez-paniagua , Eduardo Fidalgo Fernández , Enrique Alegre , Wesam Al-nabki , and Víctor González-castro " Phishing URL Detection: A Real-Case Scenario Through Login URLs " Volume 10, 2022 .

8. Smriti Dangwal, Arghir-Nicolae Moldovan " Feature Selection for Machine Learning-based Phishing Websites Detection " in 2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA) 10.1109/CyberSA52016.2021.9478242

9. Athulya A.A., Praveen K. " Towards the Detection of Phishing Attacks " in 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) 10.1109/ICOEI48184.2020.9142967

10. Lakhita, Surendra Yadav, Brahmdutt Bohra, Pooja " A review on recent phishing attacks in Internet" in 2015 International Conference on Green

Computing and Internet of Things (ICGCIoT)

10.1109/ICGCIoT.2015.7380669

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

12. A. Makkar, T. W. Kim, A. K. Singh, J. Kang and J. H. Park, "SecureIIoT Environment: Federated Learning Empowered Approach for Securing IIoT From Data Breach," in IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 6406-6414, Sept. 2022, doi: 10.1109/TII.2022.3149902.

13. M. A. Fayoumi, A. Odeh, I. Keshta, A. Aboshgifa, T. AlHajahjeh and R. Abdulraheem, "Email phishing detection based on naïve Bayes, Random Forests, and SVM classifications: A comparative study," 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022, pp. 0007-0011, doi: 10.1109/CCWC54503.2022.9720757.

14. P. Rastogi, E. Singh, V. Malik, A. Gupta and S. Vijh, "Detection of Malicious Cyber Fraud using Machine Learning Techniques," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022, pp. 520-524, doi: 10.1109/Confluence52989.2022.9734181.

15. A. N. S. Charan, Y. -H. Chen and J. -L. Chen, "Phishing Websites Detection using Machine Learning with URL Analysis," 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), 2022, pp. 808-812, doi: 10.1109/AIC55036.2022.9848895.