Professional Readiness for
Innovation, Employability and Entrepreneurship

Web Phishing Detection

Literature Survey

[1] Zichen Fan (2021), 3rd International Conference on Applied Machine Learning (ICAML). This paper thoroughly depicts a method of detecting phishing websites with joint features. Original data was crawled from PhishTank. SVM and bayes methods are mixed in training classifiers. Wireshark is used in the process of packet flow detection. After going through training, the classifier can detect 1000 websites per second. A data set of 21615 phishing and legitimate websites is used in the comparative study. In addition, 51 features are used to train and test the classifiers.

[2] Shweta Singh, M.P. Singh and Ramprakash Pandey (2020), 5th International Conference on Computing, Communication and Security (ICCCS). In this paper, a phishing detection system is implemented using deep learning techniques to prevent phishing attacks. The system works on URLs by applying a convolutional neural network (CNN) to detect the phishing webpage. In this paper, the proposed system achieved accuracy of 98.00% which is better than the earlier model. This system doesn't require any feature engineering as the CNN extracts features from the URLs automatically through its hidden layers. This is another advantage of the proposed system over the earlier model(s).

[3] Yasin Sönmez, Türker Tuncer, Hüseyin Gökal and Engin Avcı (2018), 6th International Symposium on Digital Forensic and Security (ISDFS). This study performs Extreme Learning Machine (ELM) based classification for 30 features including Phishing Websites Data in UC Irvine Machine Learning Repository database. For results assessment, ELM was compared with other machine learning

methods such as Support Vector Machine (SVM), Naïve Bayes (NB) and detected to have the highest accuracy of 95.34%.

[4] Suhas R. Sharma, Rahul Parthasarathy and Prasad B. Honnavalli (2020), IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). This paper talks about the detection of Web Phishing attacks using Machine Learning. A comparative study is made between different Machine Learning Algorithms (that can be used for binary classification) and Feature Selection Techniques when applied to Phishing datasets. The goal of this experiment is to obtain similar/comparable accuracies while achieving a significant reduction in the number of features. It uses the F1_Score and Time of execution as metrics to evaluate improvements in the overall system. Results are then rendered in the form of tables and graphs that demonstrate the same.

[5] Junaid Rashid, Toqeer Mahmood, Muhammad Wasif Nisar and Tahira Nazir (2020), First International Conference of Smart Systems and Emerging Technologies (SMARTTECH). This paper proposed an efficient machine learning based phishing detection technique. Overall, experimental results show that the proposed technique, when integrated with the Support vector machine classifier, has the best performance of accurately distinguishing 95.66% of phishing and appropriate websites using only 22.5% of the innovative functionality. The proposed technique exhibits optimistic results when benchmarking with a range of standard phishing datasets of the “University of California Irvine (UCI)” archive.

[6] Rahul Patil, Bhushan Dasharath Dhamdhere, Kaushal Sudhakar Dhonde, Rohit Gopal Chinchwade and Swapnil Balasaheb Mehetre (2014), International Conference for Convergence for Technology. This paper proposes a model to determine the phishing sites to safeguard the web users from phishers. The features of URL along with the features of Web Page in HTML tags are considered to determine the attack. Here Clustering of Database is done through K-Means Clustering and Naive Bayes Classifier prediction technique is applied to determine the probability of the web site as Valid Phish or Invalid Phish. K-Means Clustering is applied on initial URL features and Validity is checked if still we are not able to determine the Validity of Web Site then Naive Bayes Classifier is applied onto

URL as well as HTML tag features of Site and probability is evaluated based on training model.

[7] Vaibhav Patil, Pritesh Thakkar, Chirag Shah, Tushar Bhat and S. P. Godse (2018), Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). This paper discusses three approaches for detecting phishing websites. First is by analyzing various features of the URL, second is by checking legitimacy of the website by knowing where the website is being hosted and who is managing it, the third approach uses visual appearance based analysis for checking genuineness of the website. It makes use of Machine Learning techniques and algorithms for evaluation of these different features of URL and websites.