

Ensemble Model for Detecting Phishing and Trojan using Latest Machine Learning Technique

Vaibhav Bhamare¹, Krishna Deore², Anand Sonawane³, Dhikale Shubham⁴
Prof. Vidya Kale⁵

Department of Information Technology^{1,2,3,4,5}
Matoshri Aasarabai Polytechnic, Eklahare, Nashik, Maharashtra, India

Abstract: *Phishing is an online threat where an attacker impersonates an authentic and trustworthy organization to obtain sensitive information from a victim. One example of such is trolling, which has long been considered a problem. However, recent advances in phishing detection, such as machine learning-based methods, have assisted in combatting these attacks. Therefore, this paper develops and compares four models for investigating the efficiency of using machine learning to detect phishing domains. It also compares the most accurate model of the four with existing solutions in the literature. These models were developed using artificial neural networks (ANNs), support vector machines (SVMs), decision trees (DTs), and random forest (RF) techniques. Moreover, the uniform resource locator's (URL's) UCI phishing domains dataset is used as a benchmark to evaluate the models. Our findings show that the model based on the random forest technique is the most accurate of the other four techniques and outperforms other solutions in the literature.*

Keywords: phishing detection; machine learning; phishing domains; artificial neural networks; support vector machine; decision tree; random forest

REFERENCES

- [1]. Cabaj, K.; Domingos, D.; Kotulski, Z.; Respcio, A. Cybersecurity Education: Evolution of the Discipline and Analysis of Master Programs. *Comput. Secur.* 2018, 75, 24–35. [CrossRef]
- [2]. Iwendi, C.; Jalil, Z.; Javed, A.R.; Reddy, G.T.; Kaluri, R.; Srivastava, G.; Jo, O. KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against CyberAttacks. *IEEE Access* 2020, 8, 72650–72660. [CrossRef]
- [3]. RehmanJaved, A.; Jalil, Z.; AtifMoqurrab, S.; Abbas, S.; Liu, X. Ensemble AdaboostClassifier for Accurate and Fast Detection of Botnet Attacks in Connected Vehicles. *Trans. Emerg. Telecommun. Technol.* 2020, 33, e4088. [CrossRef]
- [4]. Conklin, W.A.; Cline, R.E.; Roosa, T. Re-Engineering Cybersecurity Education in the US: An Analysis of the Critical Factors. In *Proceedings of the 2014 47th Hawaii International Conference on System Sciences*, IEEE, Waikoloa, HI, USA, 6–9 January 2014; pp. 2006–2014.
- [5]. Javed, A.R.; Usman, M.; Rehman, S.U.; Khan, M.U.; Haghghi, M.S. Anomaly Detection in Automated Vehicles Using Multistage Attention-Based Convolutional Neural Network. *IEEE Trans. Intell. Transp. Syst.* 2021, 22, 4291–4300. [CrossRef]
- [6]. Mittal, M.; Iwendi, C.; Khan, S.; RehmanJaved, A. Analysis of Security and Energy Efficiency for Shortest Route Discovery in Low-energy Adaptive Clustering Hierarchy Protocol Using Levenberg-Marquardt Neural Network and Gated Recurrent Unit for Intrusion Detection System. *Trans. Emerg. Telecommun. Technol.* 2020, 32, e3997. [CrossRef]
- [7]. Bleau, H.; Global Fraud and Cybercrime Forecast. Retrieved RSA 2017. Available online: <https://www.rsa.com/en-us/resources/2017-global-fraud> (accessed on 19 November 2021).
- [8]. Computer Fraud & Security. APWG: Phishing Activity Trends Report Q4 2018. *Comput. Fraud Secur.* 2019, 2019, 4. [CrossRef]

- [9]. Hulten, G.J.; Rehfuss, P.S.; Rounthwaite, R.; Goodman, J.T.; Seshadrinathan, G.; Penta, A.P.; Mishra, M.; Deyo, R.C.; Haber, E.J.; Snelling, D.A.W. Finding Phishing Sites; Google Patents: Microsoft Corporation, Redmond, WA, USA, 2014.