Phishing Detection Using Genetic Algorithm-Based Feature Selection and Boosting Ensemble Learning

Sulaiman Mohd Yusof*, Nazhatul Hafizah Kamarudin*, Abdul Ghafar Jaafar**

*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

**Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

sulaiman.my@gmail.com, nazhatulhafizah@ukm.edu.my, abdulghafar@utm.my

***Abstract*—In line with the increasing frequency of complex cybersecurity threats such as phishing attacks, the need for accurate phishing detection has become more important. This study aims to identify the key factors thats influencing the effectiveness of phishing detection, develop an enhanced phishing detection model based on these findings, and evaluate the model in terms of accuracy, precision, sensitivity, and F1-score. To build a strong foundation, the PRISMA methodology was applied to conduct a comprehensive literature review. Using KNIME and WEKA, we created a phishing detection model that blends feature selection techniques with a stacked boosting classification approach. A Genetic Algorithm (GA) was used to pinpoint the most relevant features, while Random Forest ensured their accuracy. The classification process utilizes XGBoost, AdaBoost, and RealAdaBoost, with their results integrated through majority voting. The model was tested in two phases using datasets D1 and D2 which obtained from previous studies via Mendeley website. The outcomes were very impressive where for dataset D1, the model achieved 100\% marks across all metrics in both testing phases. For dataset D2, it scored 100\% during validation and delivered near perfect results which is 100\% precision and 99.99\% for accuracy, sensitivity, and F1-score during testing. Even more remarkable, these achievements were accomplished using just 27 out of 48 features for D1 and 21 out of 54 features for D2. This study highlights how combining GA-based feature selection with a stacked boosting based model can fortify cybersecurity measures against increasingly tricky phishing threats.**

***Keyword*— Genetic Algorithm, Machine Learning, Network Security, Phishing Detection, Phishing Websites**

Sulaiman Mohd Yusof received the Bachelor's degree in computer science, majoring in computer system, from Universiti Teknologi Malaysia, Skudai, Johor, Malaysia, in 2005. He obtained the Master's degree in cyber security from Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia, in 2025. His major field of study is computer systems and cybersecurity. He is currently serving as a Eksekutif Hasil Kanan II at the Seksyen Keselamatan Siber, Jabatan Digital, Inland Revenue Board of Malaysia. In this role, he is responsible for enterprise cybersecurity operations, including network security management and Security Operations Center (SOC) oversight. His scope of work includes the administration and optimization of SIEM, IPS, and NDR platforms to enhance threat visibility, incident response readiness, and overall organizational cyber resilience. His professional interests focus on cybersecurity operations, security monitoring, and incident handling. Mr. Sulaiman is a Professional Technologist (Ts.) registered with the Malaysia Board of Technologists. He also holds the GIAC Certified Incident Handler (GCIH) certification, issued by Global Information Assurance Certification.

Nazhatul Hafizah Kamarudin is currently a Senior Lecturer at the Centre for Cybersecurity, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM). She received her Bachelor of Engineering in Electrical Engineering and Master of Engineering in Wireless Security from Stevens Institute of Technology, New Jersey, USA. In 2019, she was awarded the Ph.D. degree in Electrical Engineering, specializing in Security and Cryptography, from Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. Her research interests include authentication, network security, the Internet of Things (IoT), and network intrusion detection. Prior to joining UKM, she served as an Assistant Professor at UCSI University and as a Lecturer at Infrastructure University Kuala Lumpur (IUKL). She is a Certified Ethical Hacker (CEH) accredited by EC-Council and a Chartered Engineer (CEng) registered with the Institution of Engineering and Technology (IET), UK.

Abdul Ghafar Jaafar obtained his PhD in Cybersecurity from Universiti Teknologi Malaysia (UTM) in 2020. He has over 10 years of industry experience, which supports his academic and research activities. He is currently serving as a Senior Lecturer at the Faculty of Artificial Intelligence (FAI), Universiti Teknologi Malaysia. His research spans a broad spectrum of cybersecurity domains, with particular emphasis on Ethical Hacking, Network and Application Security, Intrusion Detection and Prevention Systems (IDPS), Endpoint Detection and Response (EDR), Cyber Threat Intelligence (CTI), Malware Analysis and Mitigation, Cryptographic Systems, Digital Forensics, and the application of Artificial Intelligence in Cybersecurity. He holds professional certifications as a CompTIA PenTest+ Certified Professional and a Certified Ethical Hacker (CEH). He is a Professional Technologist (Ts.) registered with the Malaysia Board of Technologists (MBOT) with a specialization in Cybersecurity.