Review of the machine learning methods in the classification of phishing attack

John Arthur Jupin[1], Tole Sutikno[2], Mohd Arfian Ismail[3], Mohd Saberi Mohamad[4], Shahreen Kasim[5], Deris Stiawan[6]

[1,3] Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia
[2] Department of Electrical and Computer Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[4] Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, City Campus, Pengkalan Chepa, 16100 Kota Bharu, Kelantan, Malaysia
[4] Faculty of Bioengineering and Technology, Universiti Malaysia Kelantan, Jeli Campus, Lock Bag 100, 17600 Jeli, Kelantan, Malaysia
[5] Faculty of Computer Science & Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia
[6] Department of Computer Engineering, Universitas Sriwijaya, Palembang, Indonesia

ABSTRACT

The development of computer networks today has increased rapidly. This can be seen based on the trend of computer users around the world, whereby they need to connect their computer to the Internet. This shows that the use of Internet networks is very important, whether for work purposes or access to social media accounts. However, in widely using this computer network, the privacy of computer users is in danger, especially for computer users who do not install security systems in their computer. This problem will allow hackers to hack and commit network attacks. This is very dangerous, especially for Internet users because hackers can steal confidential information such as bank login account or social media login account. The attacks that can be made include phishing attacks. The goal of this study is to review the types of phishing attacks and current methods used in preventing them. Based on the literature, the machine learning method is widely used to prevent phishing attacks. There are several algorithms that can be used in the machine learning method to prevent these attacks. This study focused on an algorithm that was thoroughly made and the methods in implementing this algorithm are discussed in detail.