

Phishing Website and Spam Content Detection using Machine Learning Algorithms

Tanishka Narang¹, Snehal Kamble², Pallavi Sadavarte³,

Shivani Paygude⁴, Prof. Anjali P. Kadam⁵

Students, Dept. of Computer Engineering^{1,2,3,4}

Guide, Dept. of Computer Engineering⁵

Bharati Vidyapeeth's College of Engineering for Women, Pune, India

Abstract: *Phishing attacks continue to pose a major threat for computer system defenders, often forming the first step in a multi-stage attack. There have been great strides made in phishing detection; however, some phishing emails appear to pass through filters by making simple structural and semantic changes to the messages. In this paper, a Phishing and Spam Content Detection System is proposed that deals with the data uncertainty to which we are applying the SVM and NLP algorithm. There have been great strides made in phishing detection; however, some phishing URLs appear to pass through filters by making simple structural and semantic changes to the spellings. The phishing problem is big and there does not exist only one solution to reduce all susceptibilities effectively, thus multiple techniques are implemented. We can reduce the threat of phishing by analyzing various features of URL, then by checking the legitimacy of the website by knowing where the website is being hosted and who is managing it, another approach is to check visual appearance to analyze the genuineness of the website. The next step is to make sure the content on the analyzed website is spam or not. By using Natural Language Processing we process the content present on the website and determine whether it is spam or not. We make use of Machine Learning techniques and algorithms for the evaluation of these different features of URLs and websites. Using different approaches can improve the accuracy and enhance the system, thus helping better detect and prevent these threats.*

Keywords: Detection, Phishing, Spam, Support Vector Machine, Natural Language Processing

REFERENCES

- [1]. Vaibhav Patil, Pritesh Thakkar, Chirag Shah, Tushar Bhat, and S. P. Godse, "Detection and Prevention of Phishing Websites using Machine Learning Approach", ICCUBEA 2018.
- [2]. Amani Alswailem, Bashayr Alabdullah, Norah Alrumayh, and Aram Alsedrani, "Detecting Phishing Websites Using Machine Learning", ICCAIS 2019.
- [3]. Awishkar Ghimire, Avinash Kumar Jha, Surendrabikram Thapa, Sushruti Mishra, and Aryan Mani Jha, "Machine Learning Approach Based on Hybrid Features for Detection of Phishing URLs", ICoCCDSE 2021
- [4]. Pradeepthi. K V and Kannan. A "Performance Study of Classification Techniques for Phishing URL Detection", ICoAC 2014
- [5]. Weiheng Bai, "Phishing Website Detection Based on Machine Learning Algorithm", CDS 2020
- [6]. Mahajan Mayuri Vilas, Kakade Prachi Ghansham, Sawant Purva Jaypralash, and Pawar Shila, "Detection of Phishing Website Using Machine Learning Approach", ICEECCOT 2019