Discover Phish: A Survey

Chetan Bharambe, Satyam Shinde, Ruturaj Bendale. , Mahesh Kathale,

Prof.SairaBanu Pansare, Prof.Tejaswini Zope

Assistant Professor, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Talegaon Dabhade, Pune, MH, India

Student, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Talegaon Dabhade, Pune, MH, India

ABSTRACT

Phishing is a growing cyber crime that involves posing as a trustworthy entity in order to steal sensitive information. Because of the rise in phishing attacks, there is a greater need for effective solutions to detect and prevent these attacks. Machine learning algorithms have been proposed as a high-accuracy solution for detecting phishing attacks. This paper provides an overview of recent research studies on machine learning-based phishing detection. The survey summarises the methods employed, the evaluation metrics, and the outcomes obtained. In this field, various machine learning algorithms are being utilised, including decision trees, random forests, and support vector machines.The models are developed using a large - scale dataset of phishing and anti-phishing URLs before being applied to new URLs should be classified as phishing or non-phishing..The URL frame, website content, and website reputation can all be used for training. Machine learning offers a promising approach for detecting phishing attacks and has the potential to significantly improve internet users' security. However, more research is needed to improve these systems' performance and robustness against evolving phishing threats.

Keywords:- Phishing attack, Machine learning, Phishing detection, Algorithm.

INTRODUCTION

Phishing is a malicious technique used by cybercriminals to steal sensitive information from unsuspecting victims, such as login credentials and financial information. The threat of phishing attacks has grown in prominence as the internet has become more widely used and online transactions have increased. As a result, there is an increasing demand for phishing detection methods. Traditional anti-phishing solutions, such as URL blacklists and anti-virus software, are limited in their ability to detect new and unknown phishing attacks. Machine learning algorithms have been proposed as a solution for detecting phishing attacks with high accuracy in response to this challenge.This study sought to summarise the current situation is the field by providing an overview of recent research studies on machine learning-based phishing detection..This survey will provide a thorough overview of the field and will be useful for cyber security researchers, practitioners, and decision-makers.

LITERATURE SURVEY:-

The paper "Phishing Website Detection Using Machine Learning Algorithms" published in 2018 investigates the use of Machine learning to detect phishing websites. The authors tested several algorithms, including Random Forest, and found it to be the most effective, with a 94.6% accuracy rate. According to the findings, machine learning can be a useful tool in detecting phishing websites. [1]

The paper "Detecting Phishing Websites Using Machine Learning" investigates how phishing websites can be detected using machine learning. With a 96.2% accuracy rate, the Random Forest algorithm has been discovered to be the most effective. According to the findings,

machine learning can be a valuable tool for improving online security by detecting phishing websites. [2]

Waiheng Bai's paper, "Phishing Website Detection Using Machine Learning Algorithms," will be presented in December 2020 at theComputer Apps and Infosec World Conference. It investigates the use of machine learning algorithms to detect phishing websites. Phishing is a major online security threat because attackers create fake websites to obtain classified info from users. The author compares several machine learning algorithms to classify phishing websites, including Naïve Bayes classifier, Decision Tree, Random Forest, and Support Vector Machine[3]

"Feature Selection for Phishing Website Classification," published in 2020, focuses on using feature selection to enhance the accuracy of machine learning algorithms through detecting phishing websites.The researchers observed that utilising feature selection approaches such as the ReliefF algorithm and the Naive Bayes classifier may enhance the accuracy of phishing website identification greatly. Feature selection, according to the authors, is an excellent approach for improving phishing website detection systems. [4]

Selvakumari M et al presented their paper "Phishing Website Detection Using Machine Learning and Deep Learning Techniques" at the 2021 Global Forum on Computing, Interaction, and Emission Systems (ICCBS).Phishing attacks are a major risk to internet security. Using these strategies boosted the accuracy of detecting phishing websites, according to the study. According to the authors, this method has the potential to enhance phishing website detection. [6]

The 2022 article "Phishing Attack Detection Using Machine Learning" looks into the usage of machine learning techniques for identifying phishing attacks. In the investigation, the Random Forest algorithm was determined to be the most successful at identifying phishing assaults. The authors conclude that machine learning techniques can aid in the detection of phishing attempts and enhance internet security. [7]

The Journal of Autonomous Systems Research and Potential Applications published a study titled "Phishing website analysis and identification using Machine Learning" in 2022. It looked into how machine learning algorithms may be used to detect phishing websites. The Support Vector Machine method was shown to be the most efficient in detecting phishing websites in the study, and the researchers concluded that machine learning algorithms can be a strong tool for increasing online security by detecting phishing websites. [8]

CONCLUSION

These studies on Phishing websites are identified using machine learning. have concluded that machine learning algorithms can be an effective tool for detecting phishing websites and improving online security. These studies compared various algorithms and discovered that the Support Vector Machine, Random Forest, and Naive Bayes algorithms performed well in detecting phishing websites. Some studies have also looked into feature selection procedures to improve the identification of phishing websites. The findings of these studies highlight the potential for machine learning to help with online security by detecting phishing websites. Furthermore, some studies have assessed the effectiveness of machine learning algorithms to other traditional techniques, such as signature-based detection, and discovered that machine learning algorithms outperform other traditional techniques in detecting phishing websites. Machine learning can identify phishing websites in real-world settings, reducing the likelihood of online fraud.

REFERENCES

[1] Rishikesh Mahajan.,et al "Phishing Website Detection using Machine Learning Algorithms", International Journal of Computer Applications, October 2018.

[2] Amani Alswailem.,et al "Detecting Phishing Websites Using Machine Learning", International Conference on Computer Applications & Information Security (ICCAIS), IEEE Xplore: 25 July 2019.

[3] Waiheng Bai., "Phishing Website Detection Based on Machine Learning Algorithms", International Conference on Computer Applications & Information Security (ICCAIS), IEEE Xplore December 2020.

[4] Shafaizal Shabudin., et al "Feature Selection for Phishing Website Classification", International Journal of Advanced Computer Science and Applications, Vol. 11, No. 4, 2020

[5] Naresh Kumar D., et al "Detection of Phishing Websites using an Efficient Machine Learning Framework", International Journal of Engineering Research & Technology (IJERT),Vol. 9 Issue 05, May-2020.

[6] Selvakumari M., et al "Phishing website detection using machine learning and deep learning techniques", ICCCEBS 2021

[7] Mafaz Alanezi., "Phishing Detection Methods", Technium Vol. 3, Issue 9 pp.19-35 (2021)

[8] Sundara Pandiyan S., et al "Phishing attack detection using Machine Learning", Measurement: Sensors Volume 24, December 2022.

[9] R. Sakunthala Jenni and S. Shankar., "Semantic Based Greedy Levy Gradient Boosting Algorithm for Phishing Detection", Computer Systems Science & Engineering DOI:10.32604/csse.2022.

[10]Ameya Chawla., "Phishing website analysis and detection using Machine Learning", International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, VOL. 10 NO. 1 (2022)

[11]http://dataaspirant.com/2017/01/30/how-decision-treealgorithm-works/

[12]http://dataaspirant.com/2017/05/22/random-forestalgorithm-machine-learing/