iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57518

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

A Review on Online Phishing Detection Using Machine Learning

Adinath Admane[1], Abhishek Andhale[2], Aaditya Assalkar[3], Om Bastapure[4], Aparna V. Mote[5]
[1, 2, 3, 4]*Student,* [5]*Asst. Professor, Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra*

Abstract: ***The rapid growth of the internet and the increasing dependence on online services have led to a significant rise in cyber threats, with phishing attacks being a prevalent and pervasive threat. Phishing attacks often involve the use of deceptive techniques to trick users into divulging sensitive information, such as login credentials, credit card details, and personal information. Detecting and preventing phishing attacks in real-time has become a critical challenge for individuals, businesses, and organizations. This research paper presents an approach for online phishing detection using machine learning techniques. The primary objective is to develop a system that can automatically identify and classify phishing websites and emails, thereby enhancing cyber security and protecting users from falling victim to these malicious activities. Online phishing detection has gained significant importance, which will only grow with the amount of dependency on cyberspace, the proposed system provides an easy solution, during cases when the user is unsure about the authenticity of the website visited, they can try to copy the Uniform Resource Locator (URL) and paste the link into the online phishing detection system. Through the system process, it will help the user to identify whether given links were legitimate website or it is a phishing website. Therefore, the user will not be in a doubtful situation the whole day in wondering whether the information they gave in a certain website is safe or not. Providing complex decision with simplicity, the system will help the user to detect each variable of URL given accurate.***

I. INTRODUCTION

The rapid evolution of the internet has transformed the way we communicate, work, and conduct transactions. While this digital age has brought numerous benefits, it has also exposed individuals and organizations to various cyber security threats, with phishing being one of the most prevalent and damaging. Phishing attacks involve malicious actors attempting to deceive users into divulging sensitive information such as login credentials, financial data, or personal details by posing as trustworthy entities. These attacks can take the form of deceptive emails, fake websites, or social engineering tactics. As the sophistication of phishing attacks continues to grow, traditional methods of detection, such as signature-based approaches and rule-based filters, are becoming less effective. To combat this evolving threat landscape, there is a growing need for advanced, adaptive, and proactive solutions. Machine learning (ML) has emerged as a powerful tool in the fight against phishing, offering the potential to identify and thwart phishing attempts in real-time.

This paper explores the application of machine learning techniques for online phishing detection, emphasizing their role in enhancing security in today's digital ecosystem. We will delve into the key challenges posed by phishing attacks, outline the fundamentals of machine learning, and discuss how ML can be harnessed to build robust phishing detection systems. Additionally, we will examine various features and datasets commonly used for training and testing machine learning models in the context of phishing detection. Phishing detection is a critical cyber security problem that involves identifying and preventing fraudulent attempts to deceive individuals or organizations into revealing sensitive information, such as usernames, passwords, financial data, or personal information. Phishing attacks typically occur through various online channels, including email, social media, websites, and messaging platforms.

The primary goal of phishing detection is to differentiate between legitimate communications and phishing attempts. Implementing and maintaining phishing detection systems can be resource-intensive for organizations. This includes costs related to hardware, software, training, and ongoing updates to stay effective against new threats. Integrate real-time threat intelligence feeds that provide information on the latest phishing tactics and indicators of compromise. Use sandboxing techniques to analyse suspicious email attachments and URLs in a controlled environment. Provide ongoing and interactive cyber security training to users to keep them informed about the latest phishing trends and attack techniques. Conduct simulated phishing campaigns to assess user susceptibility and offer targeted training to those who fall for them

II. LITERATURE REVIEW

1) Bhagwat M. D., Dr. Patil P. H.; "A Methodical Overview on Detection, Identification and Proactive Prevention of Phishing Websites" [2021]. In this paper an approach to fuzziness resolution and an open and intelligent phishing website detection model will be proposed in the Phishing website assessment. This approach is based on smooth logic and machine learning algorithms that define various factors on the phishing website. A total of 30 characteristics or features and phishing website attributes can be used for phishing detection with high accuracy.
2) Srushti Patil, Sudhir Dhage; "A Methodical Overview on Phishing Detection along with an Organized Way to Construct an Anti-Phishing Framework" [2019]. This paper presents a focused literature survey of methods available to detect phishing websites. A comparative study of the in-use anti-phishing tools was accomplished and their limitations were acknowledged. We analysed the URL-based features used in the past to improve their definitions as per the current scenario which is our major contribution. Also, a step wise procedure of designing an anti- phishing model is discussed to construct an efficient framework which adds to our contribution. Observations made from this study are stated along with recommendations on existing systems.
3) Nathezhtha T., Sangeetha D., Vaidehi V; "WC-PAD: Web Crawling based Phishing Attack Detection" [2019] In this paper The phishing websites target individuals, organizations, the cloud storage hosting sites, and government websites. Currently, hardware-based approaches for anti-phishing are widely used but due to the cost and operational factors software-based approaches are preferred. The existing phishing detection approaches fails to provide solution to problem like zero-day phishing website attacks. To overcome these issues and precisely detect phishing occurrence a three-phase attack detection named as Web Crawler based Phishing.
4) Ropak's, Athira P Vijayaraghavan, Tony Thomas; "On Effectiveness of Source Code and SSL Based Features for Phishing Website Detection" [2019] In this paper we extract the relevant rules based on webpage source code and Secure Socket Layering (SSL) based features from a training dataset using Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm. Further, we check for the presence of these rules in a test dataset. Our implementation results show that the webpage source code-based rules can identify phishing websites with an accuracy of 0.92.
5) Yazan A. Al-Sariera1, Victor Elijah Adeyemo2, Abdullateef O Balogun 3; "AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites" [2019]. In this paper proposed AI-based meta-learners were fitted on phishing website datasets (currently with the newest features) and their performances were evaluated. The models achieved a detection accuracy not lower than 97% with a drastically low false-positive rate of not more 0.028. In addition, the proposed models outperform existing ML-based models in phishing attack detection. Hence, we recommend the adoption of meta-learners when building phishing attack detection models.
6) Shoma Tanaka, Takashi Matsunaka, Akira Yamada; "Phishing Site Detection Using Similarity of Website Structure" [2020]. In this paper, our method can identify phishing sites that differed in appearance but have similar website structures. Our method is particularly effective for detecting phishing sites constructed by the same phishers or using the same tools, as our method identifies structural similarity between websites. We conducted an evaluation to confirm the correctness of our assumption using phishing sites constructed using phishing kits and the Phish Tank dataset. We found many phishing sites that were structurally like phishing sites constructed using phishing kits.
7) Su Yang; "Research on Website Phishing Detection Based on LSTM RNN" [2020]. In this paper LSTM has the advantage of capturing data timing and long-term dependencies. LSTM has strong learning ability, can automatically learn data characterization without manual extraction of complex features, and has strong potential in the face of complex high-dimensional massive data.
8) A S S V Lakshmi Pooja, Sridhar M.; "Analysis of Phishing Website Detection Using CNN and Bidirectional LSTM" [2020]. In this paper Feature engineering is remaining essential for website-detection phishing solutions, although the quality of detection depends ultimately on previous knowledge of its features. Moreover, while the functionalities derived from different measurements are more precise, these characteristics take a lot of time to remove. This suggests a multidimensional approach to the detection of phishing focused on a quick detection mechanism through deep learning to overcome these limitations.
9) Athulya A. A., Praveen K.; "Towards the Detection of Phishing Attacks" [2020]. In this paper Phishing fraud might be the most popular cybercrime. Phishing is one of the risks that originated a couple of years back but still prevailing. This paper discusses various phishing attacks, some of the latest phishing evasion techniques used by attackers and anti-phishing approaches. This review raises awareness of those phishing strategies and helps the user to practice phishing prevention. Here, a hybrid approach of phishing detection also described having fast response time and high accuracy.

10) Wieheng Bai; "Phishing Website Detection Based on Machine Learning Algorithm" [2020]. In this paper phishing websites impersonate legitimate entities to steal sensitive information such as usernames, passwords, and financial details from unsuspecting users. Traditional methods of detecting phishing websites primarily rely on static rules and blacklists, which often struggle to keep pace with the dynamic nature of phishing attacks. In this context, machine learning algorithms have emerged as a promising solution for enhancing the accuracy and efficiency of phishing website detection.

III. OBJECTIVES

1) To continuously monitor online channels for potential phishing threats.
2) To phishing attacks can evolve quickly, so real-time monitoring is crucial to stay ahead of cybercriminals.
3) To detect phishing attempts across various communication channels, including email, SMS, social media, and instant messaging platforms.
4) To identify suspicious patterns. This includes monitoring for unusual login activities, unexpected changes in account settings and a typical browsing behavior.

IV. LIMITATIONS

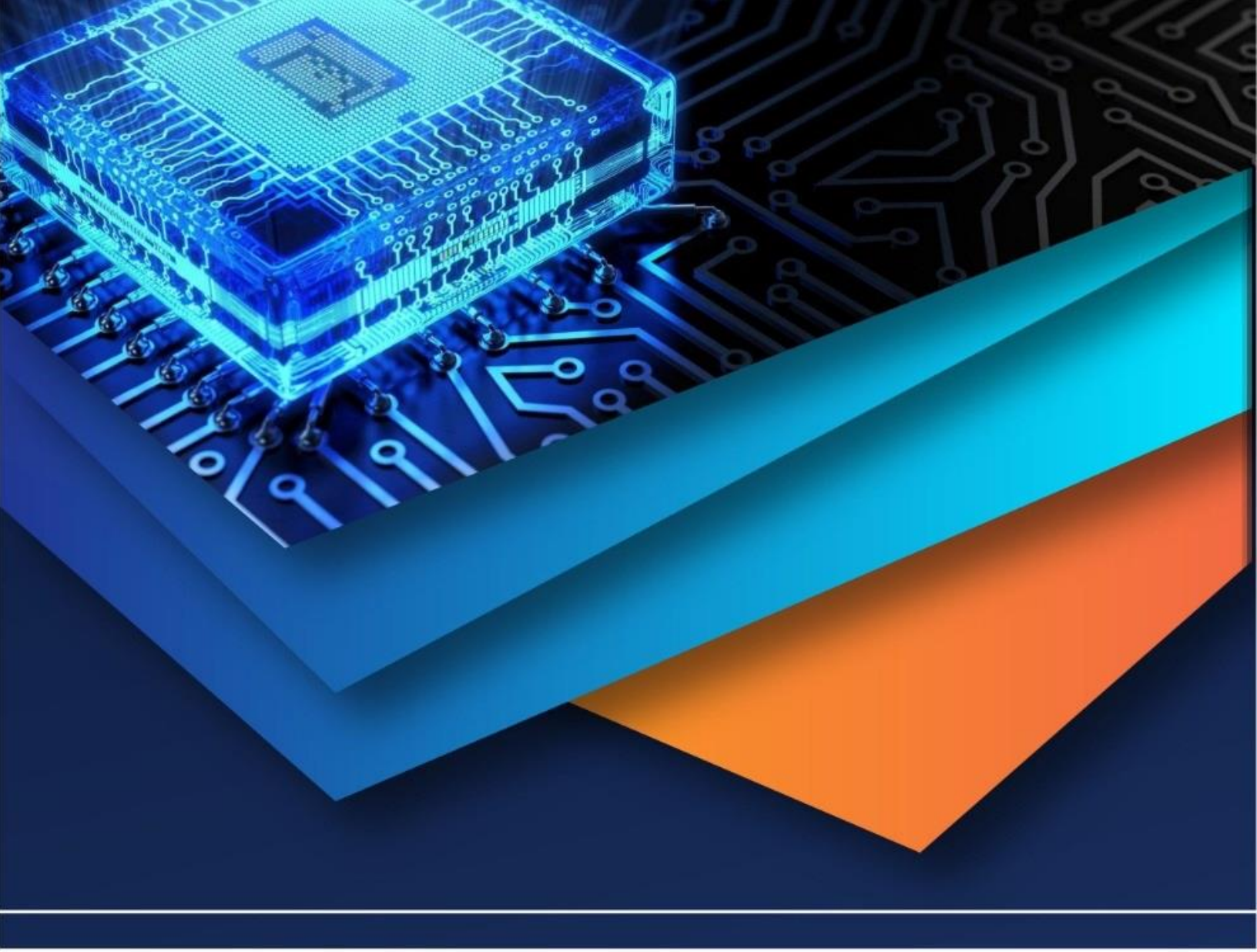
1) *Dependency on Technology:* The system relies on stable internet connectivity and modern smartphones. Users without access to these resources may face challenges in using the app.
2) Evolving Tactics: Phishing attackers continually adapt and develop new tactics to evade detection. ML models may struggle to keep up with the rapidly changing landscape, leading to a potential lag in identifying novel phishing techniques.
3) Privacy Concerns: Some ML-based phishing detection methods involve scanning user communications, raising privacy concerns. Users may resist the idea of their emails being analyzed, even if it is for security purposes, leading to potential adoption challenges.
4) Human Factor: ML models may not fully account for the human factor in phishing attacks. Social engineering tactics, such as exploiting human emotions or creating urgent scenarios, can be challenging for ML models to detect effectively.
5) *Maintenance and Updates:* Ongoing maintenance and updates are necessary to address bugs, improve features, and stay current with changes in technology and user preferences.

V. CONCLUSION

Machine learning has made significant strides in improving online phishing detection, providing a valuable layer of defense against cyber threats. However, it is important to recognize that no single solution is foolproof, and a holistic approach to cyber security that includes user education, regular model updates, and continuous monitoring is essential to combat the ever-changing landscape of online phishing attacks

REFERENCES

[1] Bhagwat M. D., Dr. Patil P. H.; "Fuzzy Logic and Machine Learning in Phishing Website Detection" [2021]. Proposes an approach based on smooth logic and machine learning for phishing website assessment, using 30 characteristics to achieve high accuracy.
[2] Srushti Patil, Sudhir Dhage; "Phishing Detection Methods and Framework Construction" [2019]. Presents a literature survey, comparative study of anti-phishing tools, and an efficient anti-phishing framework design procedure.
[3] Nathezhtha T., Sangeetha D., Vaidehi V; "Web Crawler-Based Phishing Attack Detection" [2019]. Introduces a three-phase attack detection method named "Web Crawler based Phishing" to address evolving phishing attacks.
[4] Ropak's, Athira P Vijayaraghavan, Tony Thomas; "Source Code and SSL Features for Phishing Detection" [2019]. Extracts rules from webpage source code and SSL features to identify phishing websites with 0.92 accuracy.
[5] Yazan A. Al-Sariera1, Victor Elijah Adeyemo2, Abdullateef O Balogun 3; "AI Meta-Learners for Phishing Detection" [2019]. Evaluates AI-based meta-learners with detection accuracy not lower than 97% and minimal false positives.
[6] Shoma Tanaka, Takashi Matsunaka, Akira Yamada; "Phishing Site Detection Using Website Structure" [2020]. Identifies phishing sites with similar structures, effective for sites constructed using phishing kits.
[7] Su Yang; "Phishing Detection with LSTM RNN" [2020]. Highlights LSTM's advantage in capturing timing and long-term dependencies in phishing detection.
[8] A S S V Lakshmi Pooja, Sridhar M.; "Phishing Detection Using CNN and Bidirectional LSTM" [2020]. Advocates a deep learning approach to overcome feature extraction limitations in phishing detection.
[9] Athulya A. A., Praveen K.; "Phishing Attack Detection" [2020]. Discusses various phishing attacks, evasion techniques, and a hybrid approach for quick and accurate phishing detection.
[10] Wieheng Bai; "Machine Learning in Phishing Website Detection" [2020]. Emphasizes the use of machine learning algorithms for accurate and efficient phishing website detection.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY