INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [Impact Factor: 9.47]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 12, Issue - 4, April - 2026

DOIs:10.2015/IJIRMF/202604006 --:-- Research Paper / Article / Review

Email Phishing Detection Using Artificial Intelligence and Machine Learning

[1.]Dr. Dnyanda Hire, [2.]Shivam Suryavanshi, [3.]Tejas Dhokane, [4.]Pratik Deshmukh
[1.]Professor, [2.,3,4]Student,
[1,2,3,4.]Electronics and Telecommunication Engineering, Dr. D Y Patil Institute of Engineering Management and Research, Pune, India
Email – [1]dnyanada.hire@dypiu.ac.in , [2]shivamsuryavanshi1010@gmail.com ,
[3] tejasdhokne2003@gmail.com, [4]pratikdeshmukh93599@gmail.com

Abstract: *Phishing attacks have now been identified as one of the biggest security threats in the recent past. Phishing attacks may involve the use of emails that may resemble emails from trusted organizations with the aim of extracting information from the victims, such as login information, financial information, etc. The traditional approach to phishing attacks that relies on the filtering and blacklisting approach has now become obsolete due to the constant changes in the approaches and formats of the emails by the attackers to evade the detection systems. Recently, Artificial Intelligence and Machine Learning approaches have been identified as effective in the automatic detection of phishing attacks. These approaches have the ability to analyse different characteristics of the emails that may contain malicious content, such as the content of the emails, sender details, links, etc. The advantages and disadvantages of different research work on the effectiveness of Machine Learning and Deep Learning approaches in the detection of phishing attacks and malicious emails have been discussed in this study.*

Key Words: *Phishing Attacks, Cybersecurity, Email Security, Malicious Emails, Artificial Intelligence (AI), Machine Learning (ML), Deep Learning, Email Analysis, Spam Detection.*

1. INTRODUCTION:

With the increase in the popularity of internet services and communication, email communication is one of the most frequently used communication platforms by both individuals and organizations. Although there are a number of advantages of email communication, this type of communication is mostly targeted by cybercriminals. Among various types of cyber-attacks, one of the most frequently occurring cyber-attacks during email communication is phishing, which is a type of social attack. They act as genuine entities and try to obtain sensitive information such as login information, financial information, or personal information from their victims by sending malicious emails with fake links or attachments. The phishing attacks have also become sophisticated in recent times, and it is difficult to identify them due to the changing strategies of the attackers. The conventional methods used in the system are not enough to identify the phishing attacks designed by the attackers using the blacklist techniques or rule-based systems. To overcome the limitations of the conventional methods, the researchers have started to use Artificial Intelligence and Machine Learning techniques to identify the phishing attacks. The techniques use different features like email content, sender, hyperlink, and URL features to identify the phishing attacks. The usage of AI techniques in the phishing detection system has improved the accuracy of the system, and the phishing detection domain has becoa major domain in the field of cybersecurity.

2. LITERATURE REVIEW:

Phishing detection has been a major research focus in the past decade or so, with different machine learning and deep learning algorithms being proposed to enhance the accuracy of the detection process with minimal false positives. Initially, different rule-based filtering and blacklisting approaches were the most popular approaches for

phishing detection. These approaches were not effective for zero-day attacks and dynamic phishing attacks. Thus, the data-driven Artificial Intelligence approaches have been widely accepted.

A research paper by R. Alotaibi et al. proposed a CNN-based approach for phishing email classification. In this research paper, the author proposed an automated approach for phishing email classification by extracting the text features from the email content and achieving high values for the performance parameters such as accuracy, precision, recall, and F-measure. The CNN-based approach was highly successful for the detection of structural and semantic patterns in emails. However, the author recommended that the hyperparameters could be tuned along with the exploration of the data with different other deep learning approaches that could potentially produce better results Another major contribution was achieved by Y. Fang et al. in their paper on an advanced Recurrent Convolutional Neural Network model incorporating an attention mechanism. In this paper, the model was able to analyse the header as well as the body of the email through the application of a multi-level vector representation approach. Through the application of the attention mechanism, the model was able to focus on the key parts of the email, hence eliminating the unwanted information. This improved the accuracy of the classification model. Even though the model demonstrated impressive accuracy in the detection of phishing emails, it was computationally expensive and complex to implement. Another approach that has shown impressive accuracy in the detection of phishing emails is the application of the hybrid ensemble approach for the detection of phishing emails.

Panagiotis Bountakas and Christos Xenakis proposed the hybrid ensemble learning phishing email detection model, known as HELPHED. The author was able to apply the hybrid ensemble approach by combining different machine learning models such as the Decision Trees and K-Nearest Neighbours models by applying the stacking and soft voting approach. The hybrid ensemble approach was able to make the model more robust and thus improve the accuracy of the model by achieving a high F1-score with an increase in the instances of the abuse of QR codes in phishing, the researchers have moved on to improve the detection capabilities beyond the conventional content of the email. A real-time AI system, which could be utilized in the detection of malicious QR code link content, was proposed by Mohammed S. Al-Zahrani et al. This proposed system was directed towards improving the extraction of lexical features from the URLs of the QR codes, which was successful in achieving satisfactory accuracy in the detection of malicious content. Though the proposed system was quite effective in the detection of malicious content, the accuracy results provided by the system showed that there was a need to improve the extraction of features to counter the changing nature of the attacks Recently, the idea of utilizing the deep learning-based methods involving the application of Variational Auto Encoders (VAE) was considered for enhancing the generalization abilities of the system. This is to be done to leverage the benefits of utilizing the latent variables associated with both malicious as well as benign URLs for enhancing the detection abilities of the system against the increasing number of phishing attacks. The challenges associated with the system are the increasing rates of false positive rates and concerns related to adversarial attacks. In summary, as per the literature available, it has been recognized that individual machine learning as well as deep learning-based methods have excellent detection abilities. However, none of the methods are completely secure for detecting phishing attacks. The methods are mainly reliant on the analysis of emails, URLs, and even QR codes. A need to develop a hybrid system with data analysis methods, semantic analysis methods, and even image processing methods exists. The objective of the proposed research is to develop a hybrid system with ML and DL methods to detect phishing attacks with the potential to identify all sorts of attacks.

3. OBJECTIVES:

The main objective of this research is to analyse and develop an intelligent phishing detection system using Artificial Intelligence techniques. The objective of this research is to improve the accuracy and reliability of phishing detection by analysing the malicious emails, URLs, and QR code-based phishing attacks. The objective of this research is to enhance the capability of cybersecurity systems for detecting phishing attacks using various Machine Learning and Deep Learning techniques.

The specific objectives of this research are as follows:

- To study various phishing detection methods using Machine Learning and Deep Learning methods.
- To study various data sets used for phishing detection research.
- To develop a hybrid phishing detection system using Machine Learning and Deep Learning methods.
- To study various URLs and QR code-based phishing attacks.
- To minimize false positive and false negative rates in phishing detection systems.
- To develop a reliable phishing detection system for various cybersecurity applications.

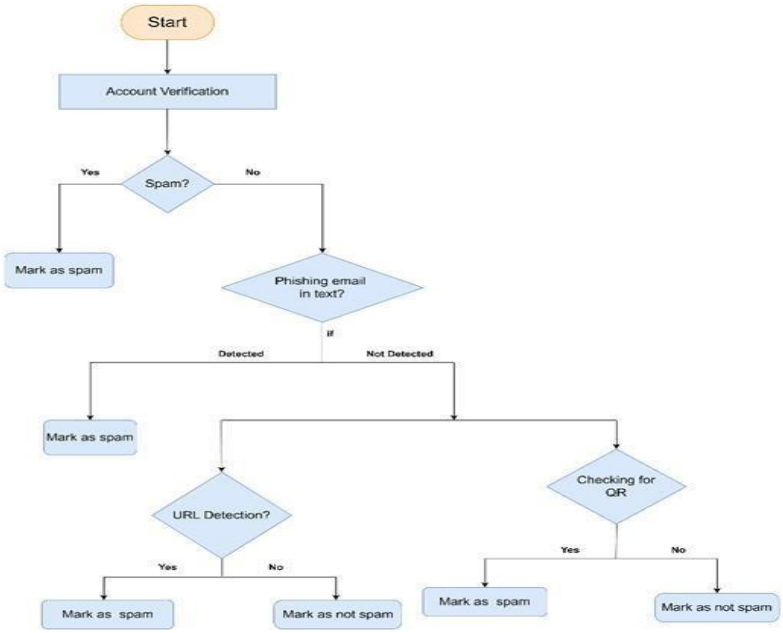
4. METHODOLOGY:

The research methodology adopted for this research is the development of an intelligent phishing detection framework leveraging the power of Artificial Intelligence techniques through a systematic approach.
The methodology adopted for this research is as follows:

Firstly, the data sets containing phishing emails, legitimate emails, malicious URLs, and QR code-based links are collected from publicly available sources. Then, the data sets collected are pre-processed, i.e., unwanted data, redundant attributes, and format issues are removed from the data sets. Next, the features extracted from the data sets collected are analysed, which will help in identifying the patterns that are different from phishing content and legitimate content. Some of the major features that are extracted from the data sets collected include email headers, email senders, URL hyperlinks, email text, lexical attributes of URLs, etc. Moreover, Natural Language Processing techniques are also used for identifying semantic patterns within the data sets collected, i.e., emails. Then, the machine learning algorithms are applied for processing the data sets, which include metadata and attribute data of URLs. The deep learning algorithms are used to detect complex patterns in the data sets, where the data sets include text and images. The ensemble learning algorithms are used to classify the data sets accurately by combining the predictions of the models. Finally, the performance of the detection system is evaluated by using the accuracy, precision, recall, and F1-score metrics to assess the efficiency of the proposed method.

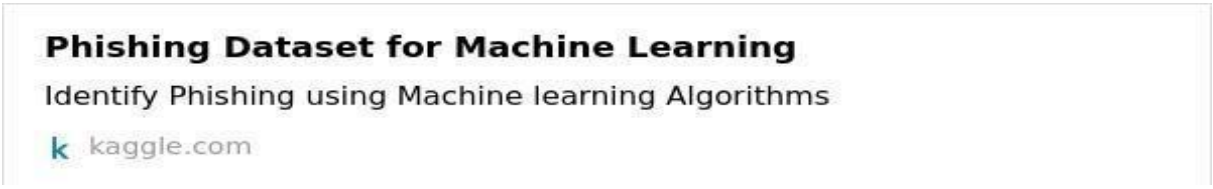
4.1 Data Collection and Preprocessing:

A comprehensive set of phishing emails, legitimate emails, malicious URLs, benign URLs, and QR code embedded links needs to be collected from publicly available sources of cybersecurity information and benchmark data. The data needs to be pre-processed in a manner such that there exists no form of bias in the training of the model. Preprocessing of data occurs by removing noise from the data, tokenizing the emails, normalizing URLs, decoding QR codes, and eliminating redundant data. Feature extraction occurs at different levels: email analysis, URL analysis, and QR code analysis. In email analysis, structured feature extraction occurs, followed by Natural Language Processing principles. For URL analysis, feature extraction of the lexical features takes place, followed by analysis of the QR code, which is then decoded to extract the URLs, which are then considered as inputs to be analysed. The proposed system will include Machine Learning as well as Deep Learning approaches. The proposed system will employ traditional Machine Learning methods for analyzing the numerical features, and Deep Learning techniques such as Convolution Neural Networks for identifying the hidden patterns and relationships involving the text and URL features. The proposed system will employ the Supervised Learning approach for training the system with the labeled data divided into training and testing sets. Hyperparameter optimization will be performed to optimize the performance of the model. Various metrics like accuracy, precision, recall, F1 measure, false positive rate, etc., will be used to measure the efficacy of the proposed system. The comparative analysis will be performed with the existing approaches to validate the improvement achieved with the proposed system. The proposed system will ensure that it is not only adaptive but also able to identify not only known but also unknown phishing attacks.

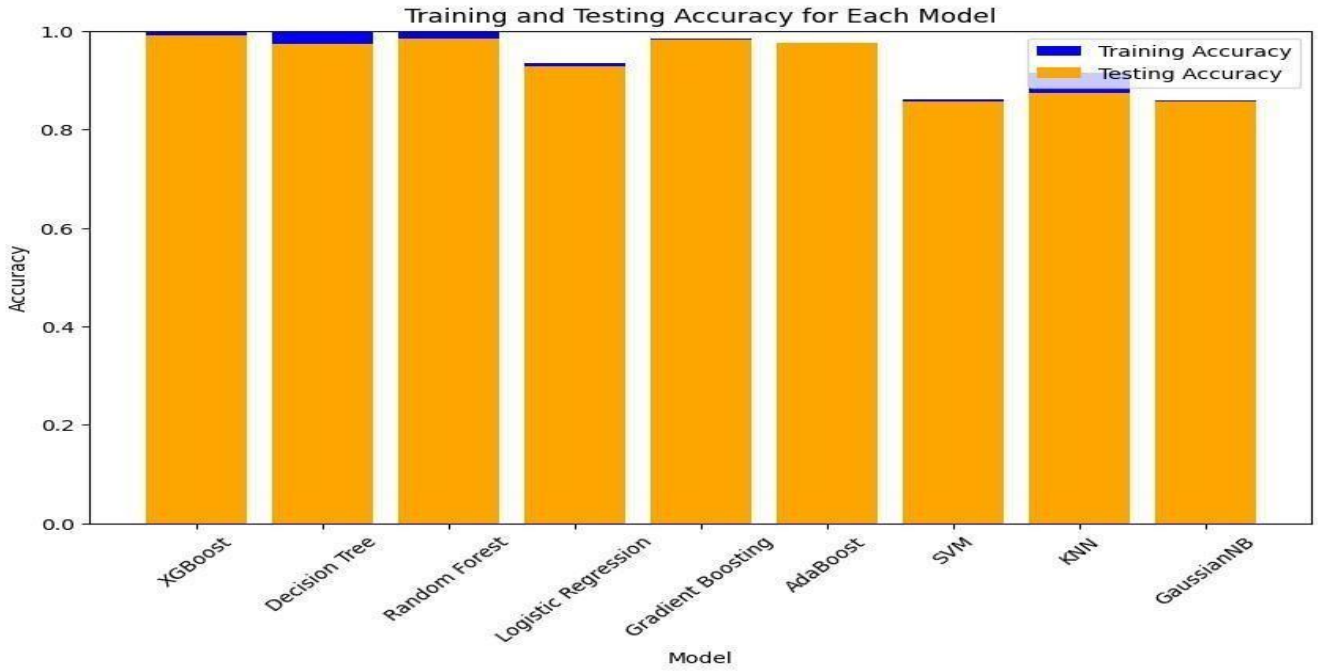


5. DATASET DETAILS:

The datasets that are used for the present research are collected from various sources that are publicly available, such as cybersecurity datasets like Kaggle. The email phishing dataset consists of thousands of emails that contain both phishing and legitimate emails and is used for training the classification models. The URL dataset consists of hundreds of thousands of URLs that are classified into different types such as benign URLs, phishing URLs, malware URLs, and defacement URLs and is used for analyzing the patterns of suspicious URLs. The QR code dataset is also taken into account for the detection of the phishing links that are present in the form of QR codes.The datasets contain various information that is useful for the detection of the phishing attacks that are taking place in different communication channels.



6. PERFORMANCE TABLE:



7. FINDINGS:

From the experimental results, there is a potential for enhancing the efficiency of phishing detection systems using the combination of Machine Learning and Deep Learning techniques. The proposed system uses a multi-level analysis method, where different attributes are analysed, including email content, email metadata, URLs, and QR codes. The proposed system is also capable of detecting not only conventional phishing attacks but also sophisticated phishing attacks that are being carried out by utilizing different techniques for evading conventional detection mechanisms. The efficiency of machine learning techniques is also evident from the processing of structured attributes, such as the sender, headers, and lexical information in URLs. At the same time, it should be noted that deep learning methods are efficient in identifying semantic patterns in emails. The usage of ensemble learning methods is efficient in improving the performance of the proposed system. This also improves the performance of the detection system. The observations of experiments show that the proposed hybrid method improves the precision and recollects, as well as decreases false positives and false negatives. This shows that the proposed hybrid method is efficient in detecting modern phishing attacks.

8. DISCUSSION:

The experimental result shows that the proposed hybrid model of Machine Learning and Deep Learning is a more comprehensive solution for phishing detection. The traditional machine learning model is more efficient in handling structured attributes such as headers, senders, and lexical features of the URL. However, it is not necessarily efficient in processing complex relationships related to the content of emails. In contrast, deep learning models like CNN and NLP are more efficient in detecting hidden semantic patterns related to phishing attacks and deceptive language structures used in phishing emails. The addition of multi-level analysis of emails, like content, URLs, and QR codes, also enhances the robustness of the system in detecting phishing attacks. Most of the existing phishing detection systems are more focused on analysing the content of the emails, which is in text format. This, therefore, makes them vulnerable to attacks like URL obfuscation and to attacks like QR code phishing attacks. The addition of URL and QR code analysis modules in the proposed system eliminates this problem, thereby ensuring that the system is effective in detecting and preventing phishing attacks.

The ensemble learning technique is advantageous in the sense that it enhances the stability of the classifier, thereby ensuring that it can generalize when new phishing attacks are encountered, which have not been previously seen. In addition, hyperparameter optimization is important in ensuring that false positives are reduced, thereby ensuring that users trust the system. However, despite the improvement in performance, there are some issues that exist. These issues include the existence of adversarial manipulation techniques and the sophisticated phishing attack patterns that may influence the effectiveness of the long-term detection approach. The second issue is the complexity

that may be involved in the implementation of the deep learning model in a real-time environment. Thus, the need to optimize the resources and implement the adaptive learning strategies arises.

9. CONCLUSION:

In the dynamic and constantly changing digital world, phishing is an important and constantly changing threat to computer security. It is no longer possible to counter the sophisticated methods of hackers by using traditional methods of detecting and filtering out such threats. This study has tried to analyse various Artificial Intelligence-based methods of detecting and filtering out such threats and their efficacy in filtering out such threats from emails, URLs, and QR codes.

It is evident from this study that there is a potential for improving the efficacy of such filtering and detection mechanisms by using Machine Learning and Deep Learning algorithms. The proposed framework for filtering out and detecting such threats is based on various methods of analysis carried out simultaneously to improve efficacy in a hybrid manner. Such a mechanism is effective in filtering out traditional and sophisticated forms of phishing threats by simultaneously analysing emails, their metadata, and links embedded in such emails. Future research directions: Improving diversity, simplicity, and flexibility of such models to detect new forms of phishing threats that are constantly emerging in this field.

10. LIMITATIONS:

- The efficiency of the system largely depends on the quality of the training data.
- The representation of the newly emerging phishing schemes may not adequately contribute to the generalization.
- The implementation of the deep learning model is computationally expensive.
- The execution of the model may cause delays.
- The system may be vulnerable to attacks that try to evade AI.
- The complexity of the URL may make the analysis of the lexical features difficult.
- The detection of the QR code depends on the analysis of the decoded URL, which may not reveal the hidden redirects.
- The system may need to be continuously updated for long-term efficiency.
- The addition of ensemble models may make the system complex.
- 10. False positives, though minimized, may affect the user experience.

11. RECOMMENDATIONS:

- The system should be updated with new datasets of phishing, which have been collected to enable its adaptability.
- The retraining of both machine learning and deep learning models is recommended to improve generalization.
- The implementation of advanced feature extraction is recommended to improve system detection capabilities.
- The implementation of adversarial training is recommended to improve system robustness against evasion and manipulation attacks.
- The application of lightweight optimization is recommended to improve system simplicity, especially in a real-time scenario.
- The implementation of various mechanisms of explainable AI is recommended to improve user trust in system decision-making processes.
- The implementation of various verification mechanisms is recommended to improve system security features.
- The periodic evaluation of system reliability is recommended, especially using updated datasets.
- In addition, user awareness and cybersecurity training programs must be implemented to reduce human susceptibility to phishing attacks.
- Future research should focus on exploring multimodal learning strategies, such as integrating both textual and behavioural features in achieving accurate phishing detection.

REFERENCES:

1. R. Alotaibi, I. Al-Turaiki, and F. Alakeel, "Mitigating Email Phishing Attacks Using Convolutional Neural Networks," in *Proceedings of the 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, 2020, pp. 1–6, doi: 10.1109/ICCAIS48893.2020.9096821.
2. Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang, "Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism," *IEEE Access*, vol. 7, pp. 56329–56340, 2019, doi:10.1109/ACCESS.2019.2913705.
3. P. Bountakas and C. Xenakis, "HELPHED: Hybrid Ensemble Learning Phishing Email Detection," SSRN Electronic Journal, 2022. [Online]. Available: https://ssrn.com/abstract=4147334
4. M. S. Al-Zahrani, H. A. M. Wahsheh, and F. W. Alsaade, "Secure Real-Time Artificial Intelligence System Against Malicious QR Code Links," *Journal of Cybersecurity Research*, 2022.
5. A. Thompson, J. Adams, and C. Garcia, "Deep Learning Techniques for Phishing Email Detection: A Comparative Study," *Journal of AI Cybersecurity*, 2023.
6. R. Miller, B. Wilson, and S. Evans, "NLP Approaches for Phishing Email Detection: A Review," in *Proceedings of the International Symposium on Cybersecurity (ISC)*, 2020.
7. C. Harris, J. Clark, and D. Taylor, "Adversarial Deep Learning for Phishing Email Detection: Challenges and Opportunities," *Journal of Cyber Defense Studies*, 2024.
8. S. Ismail and M. H. Alkawaz, "Quick Response Code Validation and Phishing Detection Tool," *International Journal of Information Security Research*, 2022.
9. "An Enhanced Deep Learning-Based Phishing Detection Mechanism to Effectively Identify Malicious URLs Using Variational Autoencoders," *International Journal of Advanced Computer Science*, 2023.