Detection Of E-mail Phishing Attacks Using Machine Learning And Deep Learning: A Review

[1]Kanchan Dhondiba Waghmare, [2]Prof.S.Khan

[1]Student, [2]Assistant Professor
[1]Computer Science and Engineering(Cyber Security),
[1]P.E.S College of Engineering, Chh.Sambhajinagar, India
[1]kanchanwaghmare2002@gmail.com, [2]khan.sanazarrin@gmail.com

Abstract— **Email phishing remains one of the most persistent cybersecurity challenges, contributing to significant data breaches and financial damage across the globe. This dissertation presents an in-depth exploration of developing a machine-learning-driven system for identifying phishing emails. The primary objective is to apply ML techniques to improve both the precision and reliability of phishing detection. Several classification algorithms—including Support Vector Machine (SVM), Decision Tree, Naive Bayes, Random Forest, and Logistic Regression—are implemented to distinguish malicious emails from legitimate ones. To maximize performance, hyperparameters are carefully optimized, and regularization strategies are applied to reduce the risk of overfitting. Model effectiveness is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The results indicate that the Random Forest model, when properly tuned, delivers the highest detection accuracy and notably surpasses conventional approaches. This research demonstrates the strong potential of machine learning in strengthening email security and provides a comprehensive foundation for future work in phishing detection. Overall, the findings underscore the ongoing need for continuous improvements in ML techniques to defend against ever-advancing cyber threats. [2]**

Index Terms— **Email Phishing, Phishing, Machine Learning, Deep Learning.**

I. INTRODUCTION

Phishing emails represent a distinct type of spam in which an attacker sends fraudulent messages that appear to originate from legitimate organizations. These messages often contain harmful attachments or malicious URLs designed to install malware on the victim's device. Once executed, such malware can compromise critical system components, leading to operational failures or unauthorized access. In many cases, phishing emails direct users to counterfeit websites engineered to steal sensitive information, including banking details, login credentials, and credit card numbers. The primary objective behind these attacks is financial gain, making phishing a persistent and dangerous threat to both individuals and organizations.[1]

Electronic mail, or email, has revolutionized communication for both personal and professional purposes, offering a fast, convenient, and cost-effective means of exchanging information across the globe. However, the widespread adoption of email has also led to a substantial increase in unwanted and potentially harmful messages, commonly referred to as spam. Beyond being a mere annoyance, spam emails represent a significant security threat, often serving as vectors for phishing attacks, malware distribution, scams, and other cyber threats. Current cybersecurity statistics suggest that more than 45% of global email traffic consists of spam, underscoring the critical need for robust and intelligent spam detection mechanisms. [4]

Email allows users to send and receive messages to and from anyone with a valid email address anywhere in the world. These messages are transmitted using protocols within the TCP/IP suite. When setting up an email account, users typically need to provide their email address, password, and the servers responsible for sending and receiving emails. Most web-based email services automatically configure these settings, requiring users only to enter their credentials. However, when using desktop email clients such as Microsoft Outlook or Apple Mail, manual configuration may be necessary. In such cases, users must input not only their email address and password but also specify the incoming and outgoing mail servers along with the correct port numbers for each server.[5]

Email continues to be one of the most widely used and critical channels of digital communication for both individuals and organizations. Yet, the rapid growth in email volume, coupled with the rising threats of spam, phishing, and information overload, has created significant challenges for efficient and secure email management. Conventional email servers provide limited automation and often rely heavily on manual effort to sort, filter, and respond to messages. To overcome these challenges, this project proposes an AI-enhanced email server that delivers intelligent, adaptive, and secure management of email communications. By integrating machine learning and natural language processing (NLP), the system can automate tasks such as email classification, prioritization, and generation of context-aware responses. Furthermore, it incorporates robust security measures, including advanced spam detection, phishing mitigation, and end-to-end encryption, ensuring comprehensive protection against evolving cyber threats.[8]

In this study, we explore the use of bulk email and messaging systems for distributing large volumes of messages within organizations, primarily for informational and alert purposes. According to Text Local, bulk messaging platforms are particularly effective in large-scale commercial settings, enabling organizations to communicate efficiently with a wide audience. Services like Logon Utility provide integrated bulk email and SMS gateways, allowing messages to be delivered directly to mobile devices. Essentially, these tools offer a streamlined solution for sending mass emails and SMS notifications over the internet to users' handheld devices, improving organizational communication and outreach.[11]

II. Literature Survey

Varshney conducts a comprehensive review and critical assessment of the most important and recent models for detecting fraudulent websites. Similarly, Vijayalakshmi M. examines historical phishing patterns, proposes a taxonomy, and summarizes state-of-the-art techniques across the literature. The researchers group defenses against sensitive-data theft by their input features—for example, web-based approaches are split into list-based, heuristic-rule, and learning-driven strategies, while content-focused methods are classified into rule-based and machine-learning-based solutions. They evaluate these approaches according to segment-specific performance, shortcomings, reliance on third-party services, and the ability to detect zero-hour attacks. Their analysis suggests that hybrid systems tend to offer superior accuracy and are more suitable for real-time deployment. Ultimately, they predict that deep learning methods will play a pivotal role in guiding future developments.[1]

The study investigates the use of several machine learning techniques to classify emails as phishing or legitimate. The authors outline a workflow that includes key stages such as data preprocessing, feature extraction, and the application of multiple ML models. In the preprocessing stage, tasks like tokenizing text, removing stop words, and addressing missing data are carried out to prepare the dataset for analysis. The algorithms tested in the study include Naive Bayes (NB), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). Results show that SVM and Naive Bayes consistently deliver the strongest performance in identifying phishing emails. The authors also note that fine-tuning the hyperparameters of these models can lead to even better classification accuracy. For future work, they recommend integrating optimization strategies with Naive Bayes and incorporating additional feature sets to further enhance detection effectiveness.[2]

The growing impact of phishing on user security has brought significant attention to the problem of detecting malicious emails. In response, a wide range of techniques has emerged, spanning communication-based solutions—such as authentication protocols, blacklists, and whitelists—as well as content-driven approaches. While blacklisting and whitelisting have been applied in practice, their effectiveness remains limited across diverse domains, reducing their overall adoption. In contrast, content-based phishing filters have become widely used due to their strong performance in identifying fraudulent messages. Consequently, recent research has increasingly focused on machine learning and data-mining methods that analyze email components such as the header, body text, and other content features to enhance detection accuracy. [3]

Spam detection has become an important area of study within Natural Language Processing (NLP) and Machine Learning (ML). The rapid expansion of digital communication—especially email—has led to a significant increase in unsolicited and potentially harmful messages, making effective spam filtering a critical component of modern email platforms. Over time, researchers have experimented with numerous email classification techniques, evolving from basic rule-based approaches to sophisticated deep learning models. Early spam filters typically relied on keyword heuristics, blacklists, and pattern-matching strategies; however, these methods lacked flexibility and were unable to keep pace with the constantly changing tactics used by spammers. This limitation paved the way for probabilistic models such as the Naive Bayes (NB) classifier. By assuming independence among words, NB offers a simple yet powerful approach to text classification, and multiple foundational works have recognized it as a reliable baseline for spam detection due to its efficiency, low computational cost, and ease of implementation.[4]

Phishing attacks have become a significant security concern for both individuals and organizations. They typically involve deceiving users into accessing fraudulent websites in order to steal personal or confidential information. Among the machine learning techniques used for detecting phishing websites, the Backpropagation (BP) neural network remains one of the most widely applied due to its strong learning capability and effective classification performance on many datasets. Despite its advantages, the BP neural network often suffers from limitations such as sensitivity to the initial parameters—particularly the starting weights and thresholds—which can cause the model to converge slowly or become trapped in local minima. To address these challenges, this paper introduces a DF-GWO–Backpropagation Neural Network (BPNN) model for phishing website detection. The proposed approach integrates an enhanced BP neural network with a dual-feature evaluation mechanism. The dual-component assessment process significantly improves the accuracy of phishing site identification. The proposed framework is compared with several existing phishing detection models, and experimental results demonstrate that our method offers higher precision, better robustness, and improved adaptability.[5]

Automation of E-Mail Handling and Management using Robotic Process Automation examines the challenges associated with managing large volumes of email and explores how Robotic Process Automation (RPA) is reshaping organizational workflows. To address the issue of email overload, the authors propose the use of Case-Based Reasoning (CBR), presenting various techniques for effective email classification. The study further investigates how RPA can streamline email processing, reduce manual workload, and lower human resource costs. Additionally, the paper highlights the research aspects of RPA, outlining its benefits and potential applications in data-intensive operations and scientific computing environments.[6]

Bulk email and SMS communication involve sending short, concise messages to a large number of recipients—often ranging from hundreds to thousands. The same content is distributed to everyone on the mailing list, making this approach ideal for marketing campaigns, product promotions, and general announcements. Organizations commonly use bulk messaging to notify audiences about upcoming events, such as concerts, product launches, or special offers. Importantly, these messages are typically directed at individuals who have willingly subscribed to receive updates. Sending bulk messages to people without their consent can frustrate recipients, damage an organization's reputation, and may even be viewed as a form of harassment.[11]

The phishing email detection techniques reviewed in the selected studies were analyzed and grouped according to the specific components of the email they targeted. A majority of the researchers—18 out of 33—focused primarily on extracting and selecting features from the email body text. Others incorporated additional elements such as the header, subject line, embedded

URLs, attachments, and overall email structure to train their deep learning (DL) models. To evaluate model performance, the studies employed a range of assessment metrics, including accuracy, precision, recall, F1 score, false positive rate (FPR), receiver operating characteristic (ROC) curves, and the area under the curve (AUC). Because different datasets and evaluation criteria were used across studies, a direct comparison of their effectiveness is not feasible. Nevertheless, this review reports the accuracy values—given their ubiquity in prior research—along with F1 scores to provide insight into each model's balanced performance.[17]

Recent developments in artificial intelligence, particularly in the field of natural language processing (NLP), have profoundly impacted digital communication. Large Language Models (LLMs) such as GPT, BERT, and Gemini API exhibit remarkable abilities to generate coherent, human-like text and understand contextual nuances, making them highly suitable for automating tasks like email drafting and reply suggestions. Research on intelligent email systems underscores the increasing need for tools that not only automate responses but also preserve clarity, tone, and contextual relevance. Studies further highlight the advantages of integrating AI directly into browser-based platforms, enabling streamlined workflows without requiring users to switch between multiple applications. Chrome extensions, in particular, have proven effective for embedding intelligent functionalities into daily user tasks due to their lightweight and accessible nature. On the backend, frameworks like Spring Boot are commonly employed in enterprise environments to build scalable and maintainable APIs capable of supporting real-time data interactions. Spring AI complements this by providing a structured approach for connecting Java applications with advanced AI models, simplifying the integration process. While commercial offerings such as Grammarly and Google Smart Compose provide automated assistance, they often lack customization and deep integration with user-specific workflows. This project addresses this gap by combining the robustness of Spring Boot with the capabilities of modern LLMs to deliver a browser-native, AI-powered email assistant designed to enhance productivity, efficiency, and personalization in email communication.[10]

III. METHODOLOGY

1) Data Collection and Preprocessing:

The effectiveness of any machine learning model largely depends on the quality of the data used. In this study, the dataset was sourced from Kaggle and consists of email files in .mbox format. The dataset is divided into two categories: phishing emails and legitimate (legal) emails. Prior to model training, several preprocessing steps were applied, including email parsing, feature extraction, handling missing or inconsistent values, and data normalization. These steps ensured that the dataset was clean, structured, and suitable for machine learning analysis.

2) Feature Extraction:

Feature extraction is a critical step in phishing email detection, as it determines how effectively the model can distinguish between legitimate and malicious emails. This research focuses on extracting commonly used and informative features such as sender details, URL-related characteristics, email content attributes, and header information. Feature engineering plays a vital role in enhancing model performance by capturing meaningful patterns that differentiate phishing emails from genuine ones.

3) Classification Algorithms

This study employs three widely used machine learning algorithms for phishing email classification:

- **Support Vector Machines (SVM):**
  Support Vector Machines are supervised learning algorithms commonly used for classification and regression tasks. SVMs are particularly effective in binary classification problems, where they aim to find the optimal decision boundary that separates data points into two distinct classes. Additionally, SVMs can be extended to handle multi-class classification scenarios and are well-suited for high-dimensional datasets.
- **Random Forest:**
  Random Forest is an ensemble-based learning technique that combines multiple decision trees to improve classification accuracy and robustness. During training, several decision trees are constructed, and predictions are made based on majority voting among the trees. This approach helps reduce overfitting and enhances overall model stability, especially when dealing with large and complex datasets.
- **Decision Tree :**
  Decision Trees are intuitive and interpretable machine learning models used for both classification and regression. They work by recursively partitioning the dataset based on feature values, forming a tree-like structure of decision rules. Each internal node represents a feature-based decision, while each leaf node corresponds to a predicted class. Decision Trees are capable of capturing non-linear relationships within the data and are easy to visualize and understand. Each of these algorithms offers distinct advantages. SVMs excel in handling high-dimensional feature spaces, Random Forests provide strong performance through ensemble learning, and Decision Trees offer transparency and interpretability. The selection of an algorithm depends on factors such as dataset characteristics, problem complexity, interpretability requirements, and computational efficiency.

4) Model Training and Evaluation :

This section outlines the procedures followed for training and evaluating the phishing email detection models. The study utilizes Support Vector Machines, Random Forest, and Decision Tree algorithms. The dataset consists of labeled phishing and legitimate emails and was divided into two subsets: 80% for training and 20% for testing.

4.1 Model Training :

- **Support Vector Machines (SVM):** The SVM classifier was implemented using the scikit-learn library in Python with a radial basis function (RBF) kernel. Hyperparameters such as the regularization parameter (C) and kernel coefficient (gamma) were optimized using cross-validation techniques. The final SVM model was trained on the complete training dataset using the selected optimal parameters.
- **Random Forest:** A Random Forest classifier was trained using the scikit-learn framework. Various configurations were tested by adjusting the number of trees (n_estimators) and the maximum tree depth (max_depth). The model with the best-performing parameter combination was selected for final training.

- **Decision Tree:** A Decision Tree classifier was also implemented using scikit-learn. Hyperparameters such as maximum tree depth (max_depth) and the minimum number of samples required to split a node (min_samples_split) were explored. The final model was trained using the optimal hyperparameter settings identified during experimentation.

4.2 Model Testing and Evaluation :

After training, all three models were evaluated on the testing dataset to measure their effectiveness in accurately identifying phishing emails. The evaluation was conducted using standard performance metrics, including accuracy, precision, recall, and F1-score, providing a comprehensive assessment of each model's classification capability.

5) Objective

- To evaluate experimental evidence on the effectiveness of deep learning (DL) techniques for phishing email detection.
- To analyze the strengths and limitations of existing deep learning–based approaches used to identify phishing emails.
- To highlight research gaps and suggest potential areas for enhancement in current phishing email detection studies.

IV. Conclusion

In this study, we developed a phishing email detection system using Machine Learning methodologies. The primary objective was to design and evaluate models capable of accurately distinguishing legitimate emails from phishing attempts, thereby enhancing online safety and protecting users from cyber threats. To accomplish this, we implemented three well-established algorithms: Support Vector Machines (SVM), Random Forest, and Decision Tree. The workflow began with data preprocessing, which included thoroughly cleaning the dataset, extracting relevant features from the .mbox files, and applying appropriate scaling techniques. These foundational steps ensured high-quality input data and supported more effective model training. The selected algorithms—SVM, Random Forest, and Decision Tree—were chosen for their strong performance in binary classification tasks and their ability to capture complex relationships between features. The results indicated that the Random Forest model achieved the best balance between precision and recall, making it a dependable option for phishing detection. Nonetheless, the ideal model choice may vary depending on specific operational needs. Future research could incorporate additional features or explore advanced feature engineering strategies to further improve the models' classification accuracy and overall robustness.[3]

V. Future Scope

The findings indicate a clear need for continued research into the application of advanced deep learning techniques for phishing email detection. Approaches such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Reinforcement Learning models show strong potential for improving detection performance but remain underexplored in this domain. In recent years, the volume and sophistication of phishing emails have surged to unprecedented levels, while existing countermeasures—despite ongoing updates—have struggled to keep pace with these evolving threats. Addressing the growing risk posed by phishing requires more robust and intelligent detection mechanisms capable of adapting to novel attack patterns. Current tools, datasets, and evaluation resources remain limited, hindering progress in this research area. Consequently, there is an urgent need for further empirical studies and innovation to thoroughly evaluate and refine deep learning–based solutions for effective phishing email detection.

References

[1] Dhruv Rathee, Suman Mann,“ Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning” International Journal of Computer Applications (0975 – 8887) Volume 183 – No. 47.
[2] Iqra Faya, Zahoor Ahmad Nagar, “Email Phishing Detection Using Machine Learning”, ISSN-2349-5162.
[3] B Leela Venkata Sai Ram, A Mary Sowjanya, “Phishing Mail Detection Using Machine Learning”, ISSN: 2320-2882.
[4] Shantanu Kumar Pandey, Astha Yadav, Nitin Kumar, Abhineet Suman, “Email Spam Detection (Using Machine Learning)”, ISSN: 2320-2882.
[5] Umer Ahmed Butt, Rashid Amin, Hamza Aldabbas, Senthilkumar Mohan ,Bader Alouffi, Ali Ahmadian, “Cloud-based email phishing attack using machine and deep learning Algorithm”, Complex & Intelligent Systems (2023) 9:3043–3070.
[6] Dr.N.Mohanapriya, K.Sivapriya, “Robotic Process Automation Driven – Email Automation and Spam Detection”, SSN: 2278-0181 Vol. 13 Issue 10.
[7] Sandeep S S, Dr.Shyam R, “A Survey of Email Phishing Attack Detection Methods: State-of-the-Art and Future Directions” ISSN-2349-5162.
[8] Navneet Yadav, Saniya Singh , Ms. Shagufta Siddiqui, “AI Powered Mail Server For Smart And Secure Email Management”, ISSN-2349-5162)
[9] Nikita Sunita Mokal, Saee Suresh Dalvi,rof. Sumedh Pundkar, “ Email Spam Detection Using Machine Learning Algorithms” ISSN-2349-5162.
[10] Mr.A.Mohanasundaram, Arulraj S, Abin V, Kathirkumar S, Kavin Bala R.K.T, “AI-Powered Workflow Automation And Email Communication System”, IJSART - Volume 11 Issue 4 – APRIL 2025 ISSN [ONLINE]: 2395-1052.
[11] Aryan Dame, Drishti Lalwani, Anirudha Khode, Aditya Trivedi, Anurag Chouhan, “Bulk SMS and Email Delivery System”, ISSN 2582-7421 International Journal of Research Publication and Reviews, Vol 4, no 2, pp 1307-1311, February 2023
[12] Pritesh A. Patil, Prayag P. Bhosale, “Literature Survey on Spam Email Detection”, ISSN 2582-7421 International Journal of Research Publication and Reviews, Vol 3, no 11, pp 2688-2694 November 2022
[13] Amit Kumar Bachcha Jha, “AI for Cybersecurity: Phishing Email Detection Use Case”, Volume 11 Issue 9 ISSN: 2349-6002
[14] Naresh Vinod Wankhade, Dr.Ranjit. R. Keole, Prof.Tushar. R. Mahore, “Paper on Spam Email Detection with Classification Using Machine Learning”, Volume 9 Issue 2 ISSN: 2349-6002
[15] Ajay Reddy Yeruva, Deepika Kamboj, Dr. Poorna Shankar, Upendra Singh Aswal, Dr. A Kakoli Rao, Somu C S, “E-mail Spam Detection Using Machine Learning – KNN”, 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) 979-8-3503-9826-7/22/$31.00 ©2022 IEEE DOI: 10.1109/IC3I56241.2022.1007262.
[16] Thashina Sultana, K A Sapnaz, Fathima Sana, Mrs. Jamedar Najath, “Email based Spam Detection”, nternational Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 9 Issue 06, June-2020.

[17] Phyo Htet Kyaw , Jairo Gutierrez , Akbar Ghobakhlou, “A Systematic Review of Deep Learning Techniques for Phishing Email Detection”, Electronics 2024, 13, 3823. 10.3390/electronics13193823
[18] Kutub Thakur, Md Liakat Ali, Muath A. Obaidat, Abu Kamruzzaman, “A Systematic Review on Deep-Learning-Based Phishing Email Detection”, Electronics 2023, 12, 4545. electronics12214545
[19] Hany F. Atlam, Olayonu Oluwatimilehin, “Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review”, Electronics 2023, 12, 42. electronics12010042.
[20] Karakadzai Sithole, “Building a Cloud-Based Machine Learning System for Automated Phishing Detection and Email Security in Cloud-Based Email Platforms.”, E-ISSN: 2582-2160.
[21] Dr. Manju Arora, Chetan Jangid, Abhishek Rai, “Email Spam Detection in the Age of Cyber Threats: A Comprehensive Approach Integrating Cleaning, Analysis, and Modeling”, E-ISSN: 2582-2160.
[22] Maharsh Patel, Aastha Shukla, Raunaq Porwal, Radhika Kotecha, “Customized Automated Email Response Bot using Machine Learning and Robotic Process Automation” , 2nd International Conference on Advances in Science & Technology (ICAST-2019).
[23] Abhaykumar Dalsaniya, “AI-Based Phishing Detection Systems: Real-Time Email and URL Classification”, TIJER ISSN 2349-9249 © November 2023 Volume 10, Issue 11.
[24] Hajar Fares, Jihad Kilani, FatimaEzzahra Fagroud, Hicham Toumi, Fatima Lakrami, Youssef Baddi, Noura Aknin, “Machine Learning Approach for Email Phishing Detection”, Hajar Fares et al. / Procedia Computer Science 251 (2024) 746–751.
[25] Said Salloum, Tarek Gaber, Sunil Vadera, Khaled Shaalan, “Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey”, Said Salloum et al. / Procedia Computer Science 189 (2021) 19–28.
[26] Peter Letmathe, Elisabeth Noll, “Analysis of email management strategies and their effects on email management performance”, Omega 124 (2024) 103002.