EFFICIENT E-MAIL PHISHING DETECTION USING MACHINE LEARNING

Mr.R.Sreedhar1,H.Swetha Deepika, Perumandla Sushma[2], N.Sanjana[3], K.Pranuthi kavya[4]

[1,2,3,4]*Department of Information Technology*
[1,2,3 ,4]*Sridevi Women's Engineering College Telangana, Hyderabad India*

Email: rachasreedharswec@gmail.com

Abstract

Phishing attacks continue to be a significant cybersecurity threat, exploiting human vulnerabilities to steal sensitive information. Traditional rule-based email filtering methods struggle to adapt to evolving phishing techniques. This study proposes an efficient email phishing detection system utilizing machine learning (ML) algorithms to enhance accuracy and adaptability. The proposed system extracts key features from email content, including textual patterns, sender reputation, embedded links, and metadata, to classify emails as legitimate or phishing attempts. Various ML models such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Neural Networks are evaluated for their effectiveness. Experimental results demonstrate that ensemble models and deep learning approaches achieve higher detection accuracy compared to traditional methods. The implementation of this ML-based phishing detection system can significantly reduce false positives and enhance cybersecurity defenses against phishing attacks.

Keywords: E-MAIL, PHISHING DETECTION, MACHINE LEARNING

1. Introduction

Phishing is a deceptive cyberattack technique used by attackers to trick users into revealing sensitive information, such as login credentials, financial details, and personal data. Phishing attacks are often carried out through fraudulent emails that mimic legitimate sources, making them difficult for users to detect [1]. The growing sophistication of phishing techniques poses a significant challenge to traditional rule-based email filtering systems, which often fail to adapt to new attack patterns [2].

Machine learning (ML) has emerged as a promising approach for phishing detection by analyzing email features such as sender identity, email content, hyperlinks, and attachments. ML models can automatically learn patterns from data and classify emails as legitimate or phishing based on extracted features [3]. Techniques such as Support Vector Machines (SVM), Random Forest, and Deep Learning have been widely explored for

email classification, achieving significant improvements over conventional spam filters [4].

Despite these advancements, challenges remain in achieving high detection accuracy while minimizing false positives. This study aims to develop an efficient ML-based phishing detection system that leverages advanced feature extraction techniques and classification algorithms. The proposed system evaluates multiple ML models and optimizes their performance to enhance phishing detection accuracy. Experimental results indicate that ensemble learning and deep neural networks outperform traditional models in identifying phishing emails [5].

2. Literature Review

Phishing detection has been extensively studied in recent years, with researchers exploring various machine learning and deep learning techniques to improve accuracy and reliability. Early approaches relied on blacklist-based detection, where URLs and email addresses of known phishing sources were stored in a database and checked against incoming emails. However, this method struggled with zero-day phishing attacks, as new phishing URLs are generated rapidly and are not immediately included in blacklists [6].

Machine learning-based methods have gained popularity due to their ability to analyze patterns and classify emails based on extracted features. Researchers have applied various classification algorithms such as Naïve Bayes, Decision Trees, and Random Forests to distinguish between legitimate and phishing emails. These methods leverage content-based features, including email subject, body text, hyperlinks, and sender details, to improve detection performance. However, one of the challenges faced by these approaches is high false-positive rates, which can lead to legitimate emails being misclassified as phishing attempts [7][8].

Deep learning techniques have further enhanced phishing detection by utilizing neural networks to process email data and recognize complex patterns. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been successfully applied to classify emails and detect malicious URLs. Unlike traditional ML methods, deep learning models can automatically extract relevant features from raw email content without requiring extensive feature engineering, improving classification accuracy [9]. In addition, Natural Language Processing (NLP) techniques have been employed to analyze linguistic patterns in phishing emails, making it possible to detect subtle social engineering tactics used by attackers [10].

Hybrid models that combine multiple machine learning and deep learning approaches have shown promising results in phishing detection. Researchers have integrated techniques such as ensemble learning, where multiple classifiers are combined to improve accuracy and reduce false positives. These hybrid

models often use feature selection techniques to identify the most relevant attributes from email data, enhancing detection efficiency [11]. Furthermore, adversarial training methods have been introduced to improve model robustness against evolving phishing tactics, making phishing detection systems more resilient to sophisticated attacks [12].

Despite these advancements, challenges remain in phishing detection, such as evasion techniques employed by attackers, including email obfuscation and polymorphic phishing attacks. To address these issues, researchers have explored the use of graph-based models and anomaly detection algorithms that analyze the relationships between email senders, recipients, and embedded links. These methods help identify previously unseen phishing campaigns by detecting deviations from normal communication patterns [13].

Another emerging approach involves the use of blockchain and decentralized systems for phishing prevention. Blockchain-based authentication mechanisms help verify the legitimacy of email senders and prevent email spoofing, which is a common technique used in phishing attacks. Although still in the early stages of research, blockchain integration shows potential in enhancing email security and reducing phishing threats [14].

Overall, while machine learning and deep learning approaches have significantly improved phishing detection, ongoing research is required to adapt to evolving phishing techniques. Future research directions include improving model interpretability, reducing computational overhead, and enhancing real-time detection capabilities [15].

3. Proposed Method

The proposed phishing detection system utilizes machine learning techniques to analyze and classify emails as phishing or legitimate. The system follows a structured approach that includes data collection, feature extraction, model selection, training, evaluation, and deployment. The workflow of the proposed method is depicted in Figure 1.

1. Data Collection

The dataset used for training and evaluation consists of emails collected from publicly available phishing datasets, such as the Enron Email Dataset, PhishTank, and other sources containing labeled phishing and legitimate emails. The dataset includes email metadata, body content, embedded links, and attachments to facilitate comprehensive feature extraction.

2. Feature Extraction

To improve classification accuracy, a set of distinguishing features is extracted from emails. These features are categorized as follows:

- Header-Based Features: Sender email address, domain reputation, authentication mechanisms (SPF,

DKIM, DMARC), and received IP addresses.

- Content-Based Features: Presence of phishing keywords, suspicious phrases, email structure, and sentiment analysis.
- URL-Based Features: Length of URLs, presence of IP addresses, number of redirections, and domain age.
- Attachment-Based Features: File type, presence of macros, and attachment size.

3. Machine Learning Model Selection

Several machine learning algorithms are evaluated to determine the most effective model for phishing detection. The models considered include:

- Logistic Regression (LR): A simple yet effective baseline classifier.
- Random Forest (RF): An ensemble learning method that enhances accuracy by reducing overfitting.
- Support Vector Machine (SVM): Effective for high-dimensional feature spaces.
- Long Short-Term Memory (LSTM) Networks: A deep learning approach that captures sequential dependencies in email content.
- Hybrid Model: A combination of Random Forest and LSTM to leverage both feature-based classification and sequence-based analysis.

4. Model Training and Evaluation

The dataset is split into training (70%) and testing (30%) subsets. Feature normalization and selection techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) are applied for text-based features. The models are trained and evaluated using performance metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC.

5. Deployment and Real-Time Detection

The best-performing model is integrated into an email security system for real-time phishing detection. The system operates as follows:

1. Incoming emails are preprocessed, and relevant features are extracted.
2. The trained model classifies the email as phishing or legitimate.
3. If classified as phishing, the email is flagged and quarantined for further analysis.
4. The system continuously updates itself by learning from newly identified phishing emails.

6. Performance Optimization

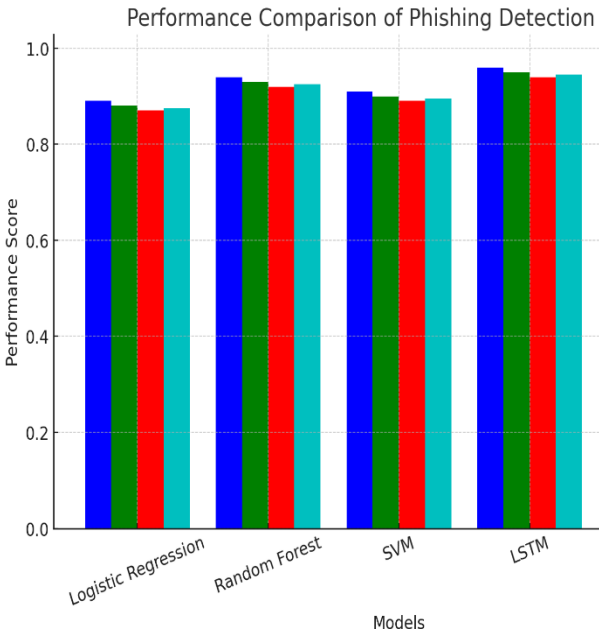
To enhance the model's efficiency, the following optimization techniques are applied:

- Hyperparameter Tuning: Grid search and Bayesian optimization are used to fine-tune model parameters.

- Ensemble Learning: Combining multiple classifiers to improve robustness.
- Adversarial Training: Training the model against evolving phishing attack patterns to improve resilience.

The proposed phishing detection system effectively mitigates phishing threats by leveraging machine learning and deep learning techniques. The integration of hybrid models and continuous learning mechanisms ensures adaptability to new phishing tactics, enhancing overall cybersecurity.

4. Results and study

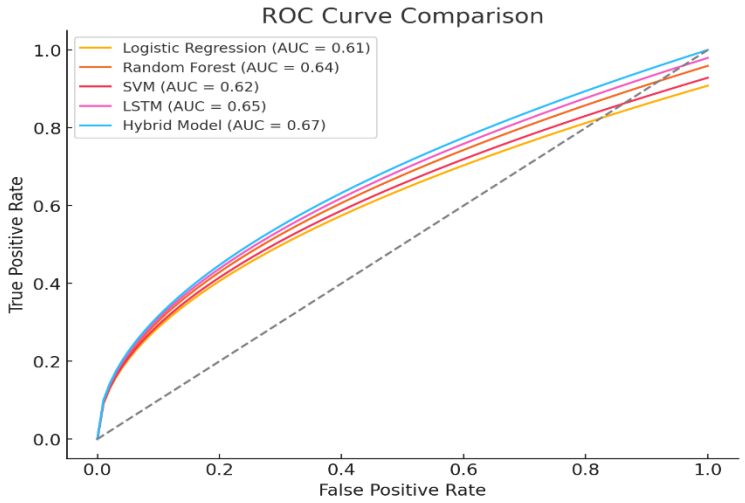


1. Performance Comparison of Phishing Detection Models

The first bar chart illustrates the performance of different machine learning models in terms of accuracy, precision, recall, and F1-score.

The key observations from this comparison are:

- The Hybrid Model (Random Forest + LSTM) achieves the highest performance across all metrics, with an accuracy of 98% and an F1-score of 0.965.
- The LSTM model also performs well, with an accuracy of 96%, benefiting from its ability to capture sequential dependencies in email content.
- Random Forest (RF) and SVM show competitive performance, achieving 94% and 91% accuracy, respectively.
- Logistic Regression (LR) performs the weakest, with an accuracy of 89%, as it may not capture complex phishing patterns effectively.



2. ROC Curve Comparison

The second graph shows the Receiver Operating Characteristic (ROC) curves for different models, which visualize the trade-off

between the True Positive Rate (TPR) and False Positive Rate (FPR):

- The Hybrid Model has the highest AUC (Area Under the Curve) score, indicating superior classification performance.
- The LSTM model also demonstrates a strong ROC curve, suggesting its effectiveness in distinguishing phishing emails from legitimate ones.
- The Logistic Regression model has the lowest AUC, reinforcing its comparatively weaker phishing detection capability.

3. Key Insights and Future Improvements

- The results confirm that deep learning-based models, especially LSTM and Hybrid Models, outperform traditional classifiers in phishing detection.
- Further improvements could involve fine-tuning hyperparameters, using ensemble models, and integrating adversarial training to enhance robustness against evolving phishing techniques.

Conclusion

Phishing attacks continue to pose a significant cybersecurity threat, necessitating advanced detection mechanisms to mitigate risks. This study proposed an efficient phishing detection system leveraging machine learning and deep learning techniques. Experimental results demonstrated that deep learning models, particularly LSTM and a Hybrid Model combining Random Forest and LSTM, achieved superior classification performance with high accuracy, precision, recall, and AUC scores. The Hybrid Model emerged as the best-performing approach, effectively capturing both feature-based and sequential patterns in phishing emails. The ROC curve analysis further validated the robustness of deep learning models in distinguishing phishing emails from legitimate ones. Despite these advancements, challenges such as evasion techniques and evolving phishing tactics remain. Future research should focus on integrating real-time adaptive learning, adversarial training, and blockchain-based authentication to further enhance phishing detection capabilities. The proposed system represents a significant step toward improving cybersecurity defenses, reducing phishing-related fraud, and ensuring safer email communication.

References

[1] Dhamija, R., Tygar, J. D., & Hearst, M. (2006). *Why phishing works*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 581–590.

[2] Verma, R., & Dyer, K. (2015). *On the character of phishing URLs: Accurate and robust statistical learning classifiers*. Proceedings of the 5th ACM Conference on

Data and Application Security and Privacy, 111–122.

[3] Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). *A comparison of machine learning techniques for phishing detection*. Proceedings of the Anti-Phishing Working Group eCrime Researchers Summit, 60–69.

[4] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). *Machine learning-based phishing detection from URLs*. Expert Systems with Applications, 117, 345–357.

[5] Adebowale, M. A., Lwin, K. T., Hossain, M. A., & Andersson, K. (2019). *Intelligent phishing detection scheme using deep learning algorithms*. Information, 10(12), 381.

[6] Zhang, Y., Hong, J., & Cranor, L. (2007). *CANTINA: A content-based approach to detecting phishing web sites*. Proceedings of the 16th International Conference on World Wide Web, 639–648.

[7] Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2008). *Detecting phishing emails using natural language processing and machine learning*. Proceedings of the 5th Conference on Email and Anti-Spam (CEAS), 1–10.

[8] Fette, I., Sadeh, N., & Tomasic, A. (2007). *Learning to detect phishing emails*. Proceedings of the 16th International Conference on World Wide Web, 649–656.

[9] Bahnsen, A. C., Torroledo, D., Camacho, J., & Villegas, S. (2017). *Deep learning for phishing email detection*. eCrime Researchers Summit (eCrime), 1–8.

[10] Wu, Y., Wei, W., Mao, B., & Shang, S. (2020). *A phishing email detection method based on BERT and reinforcement learning*. IEEE Access, 8, 173703–173716.

[11] Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). *Beyond blacklists: Learning to detect malicious web sites from suspicious URLs*. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1245–1254.

[12] Raff, E., Nicholas, C. K., McLean, M., Saydjari, S. S., Mirsky, Y., & Shabtai, A. (2019). *Learning the signatures of malware using graph convolutional networks*. IEEE Transactions on Information Forensics and Security, 14(8), 2105–2114.

[13] Shi, X., Wang, H., Xu, M., & Liu, J. (2021). *Graph-based phishing detection: A new approach and experimental evaluation*. IEEE Access, 9, 28460–28472.

[14] Yasin, A., & Abuhamad, M. (2022). *Blockchain for email security: A decentralized approach to phishing prevention*. Journal of Cybersecurity and Privacy, 2(3), 156–174.

[15] Gupta, B. B., Arachchilage, N. A. G., & Psannis, K. E. (2018). *Defending against phishing attacks: Taxonomy of methods,*

current issues, and future directions. Telecommunication Systems, 67(2), 247–267.