Intelligent Phishing Detection with Attribute-Based Policy Enforcement and PhiNet Model

Vasant Lohar , Shantanu Kamble , Gourang Lawane , Shivam Kamble , Digvijay Patil

Dr. D.Y. Patil Pratishthan's College of Engineering Salokhenagar, Kolhapur Maharashtra, India.

Assistant professor : Suyog Vilas Tate-Patil

---***---

ABSTRACT

Phishing emails are a form of cyberattack that target individuals and organizations aiming to steal sensitive information such as passwords, financial data and personal details. In the past ten years, various methods to counter phishing attacks have been developed to identify and mitigate these threats. Despite these efforts, many of these approaches remain ineffective and imprecise, highlighting the urgent need for more reliable and precise detection strategies to combat such attacks[1]. In this paper we proposed an approach to detect phishing attacks that occur through misleading email communications. It classifies incoming emails as either phishing or legitimate based on various features like email content, Sender details, embedded links, and attachments. The detection process is automated using machine learning algorithms. The system includes components like dataset training, feature extraction, model prediction, and user interface for interaction. This web-based system allows users to verify whether a specific email is safe or poses a threat. The system employs a specially designed machine learning model to distinguish between phishing and genuine emails, reaching a notable accuracy of 96.10% on testing data, which surpasses the 94% accuracy achieved by the Random Forest algorithm on the same datasets[2].

Key Words: *Phishing Email, Machine learning, Cyberattack, Feature extraction, Random Forest*

1. INTRODUCTION

Phishing attacks continue to be one of the most widespread cybersecurity threats, It exploits human vulnerabilities to gain unauthorized access to sensitive data. In 2024, there is a concerning increase in phishing emails every year. These attacks mainly target individuals and organizations through deceptive emails that pose as legitimate entities, resulting in credential theft, financial fraud, and malware infections[3].

The sophistication of phishing techniques has advanced, utilizing AI-generated emails and deepfake technology to evade traditional security measures[4]. Furthermore, phishing-as-a-service (PhaaS) platforms have simplified the process for cybercriminals to execute large-scale attacks, leading to a drastic increase in credential phishing attempts over the past few years.

To address this escalating threat, researchers are investigating various detection methodologies, including machine learning classification, heuristic analysis, and blacklist strategies[1]. While traditional rule-based detection methods struggle to keep up with the evolving tactics of phishing, machine learning and deep learning models have shown promising results in accurately detecting these phishing emails[1].

This paper focuses on examining recent phishing attack patterns, reviewing existing detection methods, and introducing a tailored machine learning approach that utilizes unique features. By examining real-world phishing emails datasets and assessing the effectiveness of various classification algorithms, this study seeks to improve email security and reduce the risks associated with phishing attacks.

2. LITERATURE REVIEW

Phishing attacks remain a significant cybersecurity challenge, preying on human weaknesses to obtain sensitive information. Traditional detection methods often fall short against the ever-evolving tactics used by attackers, making it essential to adopt more advanced solutions like machine learning-based detection systems. Recent studies have demonstrated the effectiveness of supervised learning models, especially Random Forest and Decision Trees, in accurately identifying phishing emails.

Traditional Phishing Detection Techniques

Initially, phishing detection relied on rule-based systems, blacklists, and heuristic analysis[3]. While these methods were effective against known threats, they struggled with zero-day phishing attacks and adaptive techniques[3]. Blacklists approach needs frequent updates, and heuristic approaches can produce false positives, undermining their reliability.

Using ML Techniques to Identify Phishing Attempts

The use of machine learning has revolutionized phishing detection, enabling models to uncover patterns in large datasets[5]. models to identify patterns within extensive datasets. Supervised learning involves training models with labeled phishing and legitimate emails datasets, whereas unsupervised learning focuses on detecting anomalies without predefined labels[5]. Research indicates that Techniques from Natural Language Processing (NLP) have been shown to enhance phishing detection by analyzing factors such as email content, sender behavior, and embedded links analyzing email content, sender behavior, and embedded URLs[1].

Feature Extraction and Selection

Effective phishing detection depends heavily on identifying and utilizing the most informative features. The process of selecting these features is essential, as it directly impacts the precision and efficiency of the machine learning models used. Frequently considered features include characteristics from email headers, content-based indicators, URL patterns, and metadata, all of which help differentiate phishing emails from legitimate ones:

1) Email Content Analysis: Keywords and urgency indicators
2) Sender Information: Domain reputation and email headers
3) URL Characteristics: Length and obfuscation techniques
4) Metadata Examination: Time of sending and presence of attachments[1]

Employing feature selection techniques like TF-IDF, word embeddings, and statistical analysis enhances model performance by minimizing noise and emphasizing relevant attributes.

Random Forest Classification

Random Forest is a one of the popular Supervised Machine Learning option for phishing detection because of its robustness and capacity to manage complex datasets[6]. It works by building multiple decision trees and combining their predictions, which helps reduce overfitting and enhance accuracy[6]. Research shows that Random Forest classifiers can achieve over 90% accuracy in detecting phishing emails by utilizing a wide range of feature sets[6].

Decision Trees in Phishing Detection

By linking email properties to classification results, decision trees provide interpretable models for phishing detection[5]. Even though they are effective, ensemble approaches like Random Forest are often used because they are less prone to overfitting[5]. However, decision trees are still frequently used in hybrid techniques that combine machine learning models with rule-based detection to improve security.

Future research and challenges

Future work should concentrate on utilizing advanced deep learning strategies, such as transformer architectures and recurrent neural networks, to address current limitations and enhance detection precision[1]. To increase transparency and user trust, integrating explainable AI is essential, as it helps clarify how phishing classification decisions are made[1]. Additionally, for timely defence against evolving cyber threats, phishing detection systems should be connected with threat intelligence solutions.

To conclude, this research emphasizes the use of machine learning techniques, especially Random Forest and Decision Tree models, to enhance the recognition of phishing emails. Through the use of effective feature extraction techniques and supervised algorithms, the system achieves better accuracy in distinguishing between phishing and genuine emails. Ultimately, this project contributes to improved cybersecurity by raising awareness, boosting email protection measures, and reducing the risks associated with malicious email attacks.

3. METHODOLOGY

The phishing detection strategy designed in this project leverages machine learning and follows a structured process. This process is divided into multiple stages that assist in determining whether an email should be classified as phishing or legitimate.

Data Collection and Preprocessing:

The dataset used is collected from publicly available sources which contains emails with Legitimate and phishing email labels[5]. The dataset may contain some errors and missing Values which we need to clean and process first to make our dataset ready for training model.

Feature Extraction:

To build an effective supervised learning model, it is important to identify and select key features that can distinguish phishing emails from legitimate ones. These features may include aspects such as the content of the email, sender metadata, URL patterns, and language-related indicators[1]. Techniques from natural language processing (NLP) are applied to analyze the email text and extract relevant information for further processing[1].

Machine Learning Module:

Create a supervised machine learning system to detect phishing emails[5]. Employ supervised learning algorithms, such as Random Forest and Decision Trees, utilizing a labeled dataset[6]. Regularly enhance the system's accuracy using cross-validation method[5], with consideration for continual learning to address performance challenges as phishing techniques evolve[7]. Subsequently, construct a custom feature-based machine learning model. Evaluate both models to determine which achieves superior accuracy on the dataset. Ultimately, select and utilize the model that demonstrates higher accuracy and robust performance.

Web app for User interaction:

Create a web application using FLASK which will serve as the front-end interface for phishing email detection, integrating a trained custom machine model and its associated feature engineering component. The application will allow users to submit emails for analysis. The application generates a dynamic response, displaying the classification results on a user-friendly web page. By leveraging Flask, the application provides a seamless and lightweight framework for real-time phishing detection, offering an intuitive interface for general users.

Evaluation and improvement:

Thorough testing should be carried out using authentic datasets to assess the system's performance. This includes analyzing the frequency of false positives and false negatives. Enhancements should be introduced based on these findings, and the model should be continuously refined to adapt to fresh data and newly evolving threats.

4. IMPLEMENTATION

The email detection framework was developed using Python, leveraging essential libraries like Pandas and NumPy for handling data, Scikit-learn to execute machine learning processes, and Flask to develop the web-based server side. The dataset contained pre-labelled email instances, with the text from each message treated as an input attribute.

Before feeding the text into the learning models, it was transformed into numerical form using the Term Frequency–Inverse Document Frequency (TF-IDF) technique[8]. This conversion allowed the system to efficiently extract important textual features from the raw email content[8].

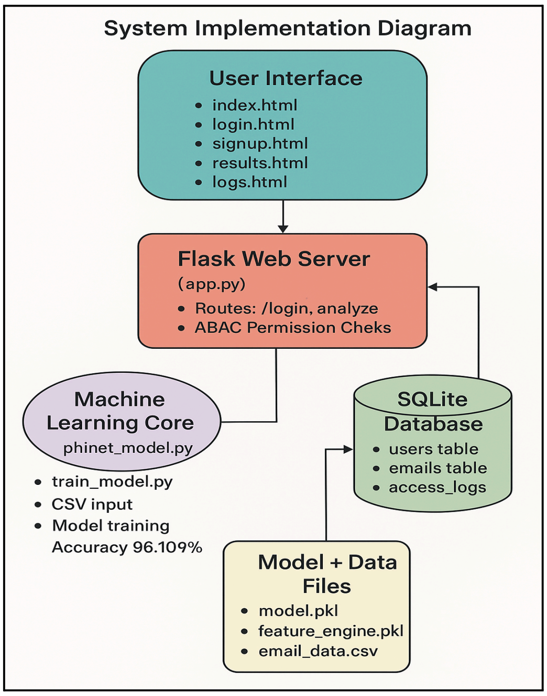
Two classification models were developed for evaluation: a Random Forest model and a tailored decision tree ensemble named the PhiNet Classifier.

The PhiNet Classifier serves as the main machine learning engine in the system[5]. It is an ensemble model constructed using multiple decision trees sourced from the Scikit-learn library[5]. The ensemble works in a style similar to boosting, where each tree in the sequence is trained on the errors made by its predecessor. The final classification result is obtained by aggregating the predictions through a majority voting strategy.

This approach offers a balance between speed and accuracy, avoiding the complexity of deep learning methods while maintaining interpretability. Upon testing, the PhiNet Classifier attained an accuracy of **96.10%**, outperforming the Random Forest model, which reached **94% accuracy** on the same dataset, showing PhiNet's effectiveness in accurately separating phishing emails from legitimate ones.

The system also implements a responsive user interface created with HTML and styled using Tailwind CSS. Email content, along with optional URLs and attachments, can be submitted via a form, and the system provides immediate phishing detection results. To ensure security and control access, the Web app implements Attribute-Based Access Control (ABAC), allowing only authorized users to access features like model retraining, log viewing, and CSV downloads[4]. These permissions are managed based on user roles, departments, verification status, and clearance levels.

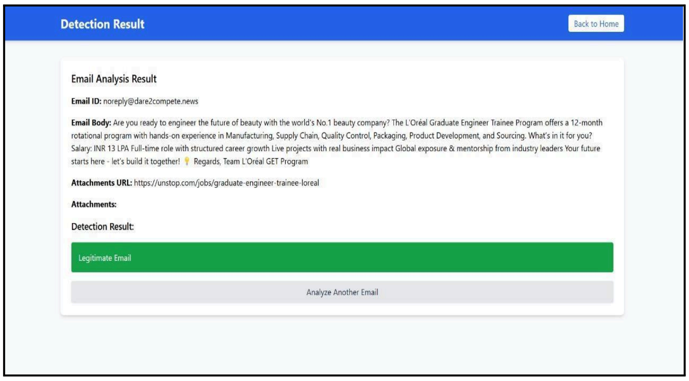
The System successfully implementation combination of machine learning, secure access control, and web technologies into a cohesive tool for real-time phishing email detection



(Fig. 4.1 Implementation of system modules)

5. RESULTS

Determining if an Email is Phishing or Legitimate:



Detection Result

Email Analysis Result

Email ID: noreply@dare2compete.news

Email Body: Are you ready to engineer the future of beauty with the world's No.1 beauty company? The L'Oréal Graduate Engineer Trainee Program offers a 12-month rotational program with hands-on experience in Manufacturing, Supply Chain, Quality Control, Packaging, Product Development, and Sourcing. What's in it for you? Salary: INR 13 LPA Full-time role with structured career growth Live projects with real business impact Global exposure & mentorship from industry leaders Your future starts here - let's build it together! 💡 Regards, Team L'Oréal GET Program

Attachments URL: https://unstop.com/jobs/graduate-engineer-trainee-loreal

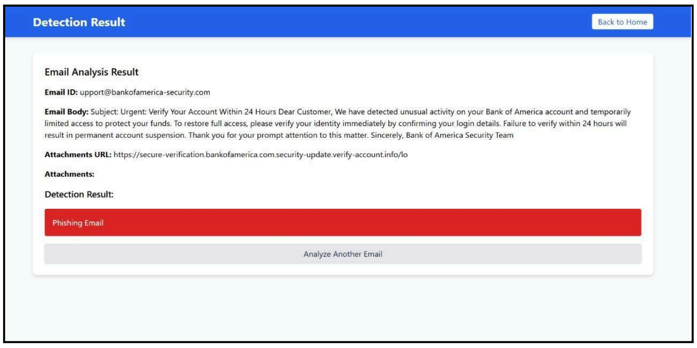
Attachments:

Detection Result:

Legitimate Email

(Fig. 5.1)

The output of this Email detection displayed as Safe Email



Detection Result

Email Analysis Result

Email ID: upport@bankofamerica-security.com

Email Body: Subject: Urgent: Verify Your Account Within 24 Hours Dear Customer, We have detected unusual activity on your Bank of America account and temporarily limited access to protect your funds. To restore full access, please verify your identity immediately by confirming your login details. Failure to verify within 24 hours will result in permanent account suspension. Thank you for your prompt attention to this matter. Sincerely, Bank of America Security Team

Attachments URL: https://secure-verification.bankofamerica.com.security-update.verify-account.info/lo

Attachments:

Detection Result:

Phishing Email

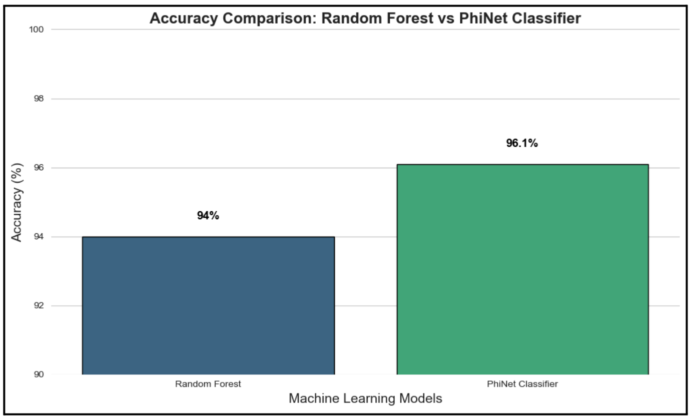
(Fig. 5.2)

The result produced by the email detection system indicated that the message was identified as a phishing attempt

6. COMPARISON OF MODELS

PhiNet Classifier outperforms Random Forest with a 2.1% higher accuracy, achieving 96.1% in detecting phishing emails[9]. The custom PhiNet model utilizes an ensemble of Decision Tree classifiers in a boosting-like approach, improving accuracy while maintaining interpretability[5]. While Random Forest being robust, does not incorporate the same level of iterative refinement as PhiNet, leading to slightly lower accuracy[6].

Both models demonstrate high reliability in phishing email detection, but PhiNet models with custom features provides a more optimized classification



(Fig. 6.1 Comparison of Models Accuracy)

7. CONCLUSION

This research presents an advanced machine learning approach for detecting phishing email attacks, addressing the growing cybersecurity threats faced by individuals and organizations[5]. By leveraging supervised learning techniques, particularly a custom PhiNet classifier, the system significantly improves phishing email detection accuracy, achieving an impressive 96.10% accuracy compared to 94% of Random Forest[5].

The methodology encompasses robust feature extraction, email content analysis, and sender verification to distinguish legitimate emails from malicious phishing attempts. The web-based implementation offers real-time detection, enhancing accessibility and ease of use for individuals and businesses alike.

Comparative analysis highlights that the custom PhiNet classifier outperforms traditional models, reinforcing the importance of tailored feature engineering in phishing detection. The implementation of secure access control mechanisms further strengthens the reliability of the system.

Future research could explore advanced neural network methods like RNNs and transformer models to improve performance, as recent studies have demonstrated the efficacy of deep learning models such as CNNs with Bi-GRU in enhancing phishing detection accuracy[10], as well as integrating threat intelligence systems for real-time adaptability to evolving threats[1]. Additionally, enhancing model transparency through explainable AI can improve user trust and decision-making[1].

In essence, this research provides a proactive and efficient solution to phishing email detection, mitigating risks associated with cyber fraud and contributing to a safer digital ecosystem.

8. REFERENCES

[1] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, “A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques,” *IEEE Access*, vol. 10, pp. 65703–65727, 2022, doi: 10.1109/ACCESS.2022.3183083.

[2] Assistant Professor, Computer Science and Engineering, Raghu Engineering College, Visakhapatnam and A. V. S. Kumar, “Phishing Email Detection using Machine Learning,” *INTERANTIONAL J. Sci. Res. Eng. Manag.*, vol. 08, no. 04, pp. 1–5, Apr. 2024, doi: 10.55041/IJSREM32276.

[3] R. Alabdan, “Phishing Attacks Survey: Types, Vectors, and Technical Approaches,” *Future Internet*, vol. 12, no. 10, p. 168, Sep. 2020, doi: 10.3390/fi12100168.

[4] C. S. Eze and L. Shamir, “Analysis and prevention of AI-based phishing email attacks,” May 08, 2024, *arXiv*: arXiv:2405.05435. doi: 10.48550/arXiv.2405.05435.

[5] F. Salahdine, Z. El Mrabet, and N. Kaabouch, “Phishing Attacks Detection A Machine Learning-Based Approach,” in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York, NY, USA: IEEE, Dec. 2021, pp. 0250–0255. doi: 10.1109/UEMCON53757.2021.9666627.

[6] S. Rawal, B. Rawal, A. Shaheen, and S. Malik, “Phishing Detection in E-mails using Machine Learning,” *Int. J. Appl. Inf. Syst.*, vol. 12, no. 7, pp. 21–24, Oct. 2017, doi: 10.5120/ijais2017451713.

[7] A. Ejaz, A. N. Mian, and S. Manzoor, “Life-long phishing attack detection using continual learning,” *Sci. Rep.*, vol. 13, no. 1, p. 11488, Jul. 2023, doi: 10.1038/s41598-023-37552-9.

[8] E. S. Gualberto, R. T. De Sousa, T. P. De B. Vieira, J. P. C. L. Da Costa, and C. G. Duque, “From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection,” *IEEE Access*, vol. 8, pp. 76368–76385, 2020, doi: 10.1109/ACCESS.2020.2989126.

[9] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, “A comparison of machine learning techniques for phishing detection,” in *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, Pittsburgh Pennsylvania USA: ACM, Oct. 2007, pp. 60–69. doi: 10.1145/1299015.1299021.

[10] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, “Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models,” *Sensors*, vol. 24, no. 7, p. 2077, Mar. 2024, doi: 10.3390/s24072077.