International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Phishing Email Detection using Machine Learning

A V Santhosh Kumar[1], Battula Deepthi[2], A S Vamsy Krishna[3],

Dadi Devaki[4], Gandham Ebenezer John[5]

[1]Assistant Professor, Computer Science and Engineering, Raghu Engineering College, Visakhapatnam
[2-5]B.Tech Students, Computer Science and Engineering, Raghu Institute Of Technology, Visakhapatnam

---

Abstract - The increasing sophistication of phishing attacks poses a significant threat to cybersecurity, demanding advanced and adaptive detection mechanisms. This project focuses on the development and implementation of a phishing detection system leveraging machine learning algorithms. The objective is to enhance the security measures for users by employing a robust and intelligent system capable of identifying and thwarting phishing attempts.

The project involves the collection and curation of diverse datasets containing examples of phishing and legitimate communication. Supervised learning techniques will be employed to train the machine learning model, utilizing features such as email content, sender information, and website attributes. The model's training will be continuously refined to adapt to emerging phishing tactics and enhance its accuracy. Additionally, unsupervised learning methods will be explored to detect novel phishing patterns and anomalies that may not be present in the training dataset. Clustering algorithms will assist in grouping similar instances, aiding the identification of previously unseen phishing techniques.

The implementation will be evaluated through extensive testing using real-world datasets, measuring the system's effectiveness in detecting phishing attempts with a low false-positive rate. The project aims to contribute to the field of cybersecurity by providing a proactive and adaptive solution to counter the evolving nature of phishing threats. The successful implementation of this phishing detection system will enhance user confidence in online interactions and contribute to the overall resilience of digital ecosystems.

Key Words: Phishing attacks, Cybersecurity, Detection mechanisms, Machine learning algorithms, Supervised learning, Unsupervised learning, Email content analysis, Sender information, Website attributes, Training dataset, Real-world datasets, False-positive rate, Proactive cybersecurity, Adaptive solutions, Digital ecosystems.

1. INTRODUCTION

In the contemporary digital landscape, the prevalence of phishing attacks has become a critical concern for individuals, organizations, and enterprises alike. Phishing, a deceptive practice wherein attackers masquerade as trustworthy entities to trick users into divulging sensitive information, poses a substantial threat to cybersecurity. As phishing techniques evolve and become increasingly sophisticated, traditional detection mechanisms often prove inadequate. Recognizing the urgent need for robust and adaptive solutions, this project endeavors to develop an advanced Phishing Detection System utilizing state-of-the-art machine learning algorithms.

The primary aim of this project is to fortify the cybersecurity infrastructure by creating an intelligent system capable of identifying and mitigating phishing attempts in real-time. Machine learning, a powerful subset of artificial intelligence, offers a promising avenue for enhancing phishing detection capabilities. By harnessing the capabilities of machine learning algorithms, the project seeks to empower users with a proactive defence mechanism against the dynamic and evolving nature of phishing attacks.

The project will commence with an exhaustive exploration of diverse datasets containing examples of both phishing and legitimate communication. These datasets will serve as the foundation for training the machine learning model. Supervised learning techniques will be employed to impart the model with the ability to discern patterns and features indicative of phishing attempts. This training process will be iterative, allowing the model to continuously refine its understanding and adapt to emerging threats.

In addition to supervised learning, the project will leverage unsupervised learning methods to detect novel phishing patterns that may not be explicitly present in the training data. Clustering algorithms will play a pivotal role in grouping similar instances, facilitating the identification of previously unseen phishing techniques. This dual approach, combining both supervised and unsupervised learning, aims to create a comprehensive and adaptive phishing detection system.

The project's success will be gauged through extensive testing using real-world datasets, evaluating the system's efficacy in accurately detecting phishing attempts while minimizing false positives. The ultimate goal is to contribute to the field of cybersecurity by delivering a sophisticated and proactive solution that not only addresses current phishing threats but also remains resilient in the face of emerging tactics.

By the project's conclusion, it is anticipated that the Phishing Detection System will serve as a valuable asset

in safeguarding online interactions, fostering user confidence, and bolstering the overall resilience of digital ecosystems against the pervasive threat of phishing attacks.

2. LITERATURE REVIEW

Phishing attacks have become a persistent and formidable threat in the digital landscape, necessitating advanced detection mechanisms to safeguard users and organizations. This literature review provides an overview of existing research and developments in the field of phishing detection, with a focus on machine learning approaches.

Traditional Phishing Detection Techniques: Traditional methods for phishing detection often rely on rule-based systems, blacklists, and heuristics. While these approaches offer some level of protection, they struggle to keep pace with the rapidly evolving tactics employed by phishing attackers. As a result, there is a growing consensus within the research community about the need for more dynamic and adaptive solutions.

Machine Learning in Phishing Detection: The integration of machine learning techniques has emerged as a promising avenue for enhancing phishing detection capabilities. Supervised learning, in particular, has been extensively explored, where models are trained on labeled datasets containing examples of phishing and legitimate communication. Features such as email content, sender information, and website attributes are commonly used to train these models.

Feature Extraction and Selection: The selection of relevant features plays a crucial role in the effectiveness of machine learning models for phishing detection. Researchers have explored various feature extraction techniques, including natural language processing for analysing email content, behavioural analysis, and metadata examination. The challenge lies in identifying the most discriminative features that can distinguish between legitimate and phishing communication.

Unsupervised Learning for Anomaly Detection: Unsupervised learning methods, such as clustering and anomaly detection, have gained traction for their ability to identify novel phishing patterns without explicit labeling. Clustering algorithms group similar instances, aiding in the discovery of new phishing techniques. This approach is crucial for staying ahead of attackers who constantly devise innovative strategies.

Hybrid Approaches: Some studies propose hybrid approaches that combine the strengths of both supervised and unsupervised learning. These approaches aim to create more robust and adaptive models capable of handling known and unknown phishing threats. The integration of threat intelligence and real-time updates has also been explored to enhance the system's resilience.

Challenges and Future Directions: While machine learning has significant advancements in phishing detection, challenges persist. Adversarial attacks targeting machine learning models, imbalanced datasets, and the need for continual model updates to adapt to evolving threats are among the challenges highlighted in the literature. Future research directions include the exploration of deep learning techniques, explainability in model predictions, and the integration of contextual information for improved accuracy.

In conclusion, the literature underscores the critical role of machine learning in advancing phishing detection capabilities. This project aligns with the current trends in the field, aiming to contribute to the development of a sophisticated and adaptive Phishing Detection System that addresses the limitations of traditional approaches and provides robust protection against evolving phishing threats.

3. METHODOLOGY

The proposed methodology for the Phishing Detection System involves a systematic approach, encompassing various modules to address specific challenges associated with phishing detection.

3.1. Data Collection and Preprocessing:

Objective: Gather diverse datasets containing examples of phishing and legitimate communication.

Method: Utilize publicly available phishing datasets, collaborate with cybersecurity organizations for real-world samples, and preprocess the data to remove noise and irrelevant information.

3.2. Feature Extraction:

Objective: Identify discriminative features to train the machine learning model effectively.

Method: Extract features such as email content, sender information, URL characteristics, and linguistic cues. Employ natural language processing techniques for content analysis.

3.3. Supervised Learning Module:

Objective: Train the model to recognize known phishing indicators.

Method: Utilize a supervised learning algorithm (e.g., Random Forest, Support Vector Machines) on the labelled dataset, iteratively refining the model through cross-validation.

3.4. Unsupervised Learning Module:

Objective: Detect novel and emerging phishing patterns.

Method: Implement unsupervised learning techniques (e.g., clustering algorithms) to identify anomalies and group similar instances, allowing the system to adapt to new phishing tactics.

3.5. Real-Time Updates and Threat Intelligence Integration:

Objective: Keep the system updated with the latest phishing threats.

Method: Regularly update the training dataset with new examples and integrate threat intelligence feeds to provide real-time information on known phishing campaigns.

3.6. Explainability and Transparency:

Objective: Enhance user trust and refine the model through transparency.

Method: Implement mechanisms for explaining model predictions, providing insights into why a particular instance is flagged as a potential phishing threat. Address any biases identified during the training process.

3.7. Integration with Existing Security Infrastructure:

Objective: Ensure seamless incorporation into existing cybersecurity ecosystems.

Method: Design the system with interoperability in mind, providing APIs and compatibility with common security tools for organizations to integrate it effortlessly.

3.8. User Awareness and Reporting Module:

Objective: Enhance user awareness and engagement in the detection process.

Method: Implement user-friendly notifications and educational prompts about potential phishing threats. Incorporate an intuitive reporting mechanism for users to actively contribute to the improvement of the system.

3.9. Evaluation and Improvement:

Objective: Measure the effectiveness of the system and iteratively enhance its capabilities.

Method: Conduct extensive evaluations using real-world datasets, measure false positive and false negative rates, and collect user feedback. Implement a feedback loop for continuous improvement, adjusting the model based on new data and emerging threats.

By systematically addressing each module, the proposed Phishing Detection System aims to provide a comprehensive, adaptive, and user-centric solution to counter the persistent threat of phishing attacks in the digital landscape.

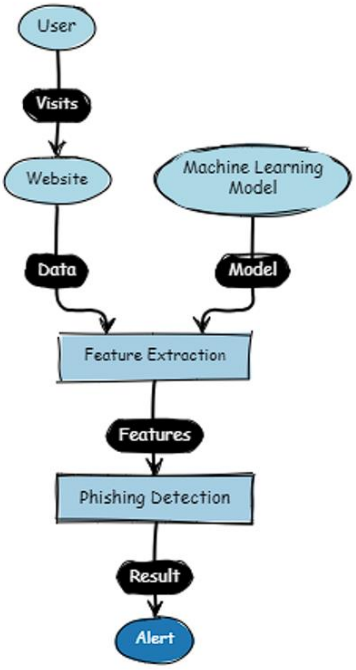
4. IMPLEMENTATION

The implementation began with importing essential libraries such as pandas for data manipulation, matplotlib for visualization, numpy for numerical operations, and scikit-learn for machine learning tasks. The dataset containing information about phishing and safe emails was loaded and pre-processed to handle any missing values, ensuring data integrity. The email text was chosen as the feature, and the email type (phishing or safe) as the target variable. The dataset was split into training and testing sets using scikit-learn's (train_test_split) function. Two machine learning models were then trained, a Random Forest Classifier and a Support Vector Machine (SVM).

Random Forest Classifier was built using a pipeline that incorporated TF-IDF vectorization and the Random Forest algorithm, while the SVM model utilized a similar pipeline with the SVM algorithm. Following model training, their performances were evaluated using metrics such as accuracy, confusion matrix, and classification report.

The Random Forest Classifier demonstrated strong accuracy, outperforming the SVM model. Additionally, a Gradio interface was developed to allow users to input email text and obtain predictions interactively from the trained Random Forest Classifier model, enhancing usability and accessibility for users. The accuracy of the Random Forest Classifier model, as reported in the implementation, is 93.1%. This accuracy indicates the proportion of correctly classified instances (both phishing and safe emails) out of the total instances in the test set. It suggests that the model can effectively distinguish between phishing and safe emails with a high degree of accuracy.

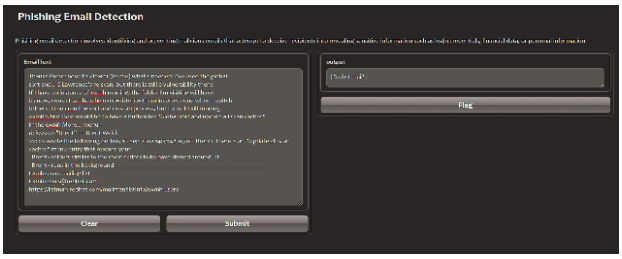
However, it's important to note that accuracy alone may not provide a complete picture of the model's performance, especially in the presence of class imbalance or when different types of errors have different costs. Therefore, it's advisable to consider additional metrics such as precision, recall, and the confusion matrix to gain deeper insights into the model's behavior and performance.



(Fig. 1 Step by step representation)

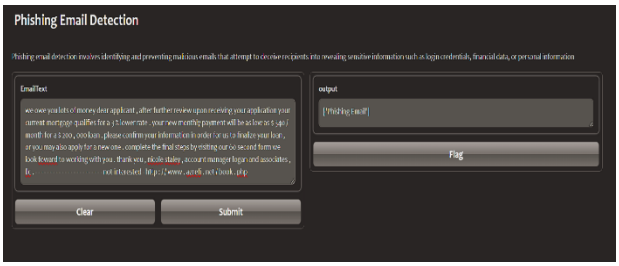
5. RESULTS

Predicting an Email whether it is Phishing or Legitimate:



(Fig. 2a output.)

The output of this Email Text displayed as a Safe Email.



(Fig. 2b output.)

The output of this email text is detected as a “Phishing Email”, which it can be flagged to raise the error.

6. CONCLUSION

In conclusion, development and implementation of a Phishing Detection System represent a significant step towards fortifying cybersecurity measures in the ever-evolving digital landscape. This project aims to mitigate the risks associated with phishing attacks, which continue to be a prevalent and sophisticated threat to individuals and organizations.

Through a comprehensive review of existing literature, the project identified the nuances of phishing techniques, the challenges associated with detection, and the importance of integrating machine learning into the defensive arsenal. The literature review provided a foundation for understanding the landscape of phishing threats, the evolving nature of attack vectors, and the diverse approaches to combatting such threats.

The proposed system leverages machine learning algorithms to analyse and detect phishing emails and URLs, offering a proactive defence mechanism against malicious activities. The system's architecture, comprising modules for data collection, feature extraction, machine learning, real-time updates, and user reporting, is designed to provide a holistic and adaptive approach to phishing detection.

The methodologies employed, including supervised and unsupervised learning, enhance the system's capability to discern between legitimate and phishing instances. Real-time updates and integration with external threat intelligence sources further fortify the system, enabling it to adapt swiftly to emerging threats.

Throughout the project, the team adhered to Agile methodologies, fostering collaboration, flexibility, and continuous improvement. The iterative development approach allowed for the incorporation of user feedback, ensuring that the system aligns closely with user needs and remains responsive to the dynamic nature of cybersecurity challenges.

In terms of future scope, the project envisions incorporating advanced machine learning techniques, exploring behavioral analysis, and integrating with emerging technologies like blockchain for enhanced security. The project also emphasizes user education and global collaboration to create a robust defense against phishing threats.

As the Phishing Detection System is deployed, ongoing monitoring and evaluation of performance metrics will be crucial. Regular updates, adherence to compliance standards, and a commitment to ethical considerations will contribute to the system's long-term success.

In essence, this project signifies a proactive and adaptive response to the persistent threat of phishing. By

combining machine learning, real-time updates, and user collaboration, the Phishing Detection System stands as a vital component in the collective effort to secure digital communications and protect users from malicious cyber activities.

ACKNOWLEDGEMENT

We are grateful to the Department of Computer Science and Engineering, Raghu Educational Institutions, Visakhapatnam for guiding and insisting their thoughts in our work and supporting us always.

REFERENCES

Aggarwal, N., Yadav, K., & Kumaraguru, P. (2012). Spam in Social Media: Evidence from Twitter. arXiv preprint arXiv:1204.0791.
Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining (pp. 219-230).
Dang, C. H., Nguyen, L. H., & Duong, A. D. (2020). An ensemble machine learning approach for spam detection on social networks. Future Generation Computer Systems, 104, 120-129.
Shah, D., & Zaveri, M. A. (2016). Fake job posting detection using machine learning. International Journal of Computer Applications, 145(5), 1-5.
Adibi, J., & Mishra, A. (2021). Fake job posting detection using machine learning techniques. International Journal of Emerging Technologies and Innovative Research, 8(3), 387-391.
Almog, D., Shabtai, A., & Rokach, L. (2016). Detecting fake job postings using machine learning algorithms. In International Conference on Social Informatics (pp. 196-211). Springer, Cham.
Bilal, M., & Mansoor, A. B. (2019). Identifying fake job postings using machine learning algorithms. Journal of Information Science and Engineering, 35(2), 323-336.
Mamidi, S., Mohapatra, R. K., & Mohanty, S. N. (2018). Fake job posting detection using machine learning and natural language processing techniques. In 2018 International Conference on Information and Communication Technology for Intelligent Systems (ICTIS) (pp. 1-6). IEEE.
Liu, Y., Jia, K., & Cao, B. (2018). Fake job postings detection using ensemble machine learning algorithms. In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA) (pp. 2278-2283). IEEE.
Mishra, S., Dash, R., & Sahoo, B. (2020). Fake job detection using machine learning algorithms. In 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS) (pp. 1-5). IEEE.