NLP-DRIVEN METHODS FOR ENHANCED PHISHING WEBSITE DETECTION

J Gold Beulah Patturose, E Sahaya Chithra, V. Anitha Christy, S Gnana Selvan
St.Joseph's Institute of Technology ,Chennai,St. Xavier's College, Palayamkottai, Jayaraj Annapackiam CSI College of Engineering, Nazareth, Jayaraj Annapackiam CSI College of Engineering, Nazareth
beulahpattu@gmail.com, surenchithra@gmail.com, anitha.chrst@gmail.com, sgpalam305@gmail.com

Abstract.
Phishing websites pose a significant threat by exploiting user trust to steal sensitive information. Traditional detection methods often fail to keep pace with evolving phishing tactics. This paper proposes a novel approach using Natural Language Processing (NLP) to enhance phishing detection. By analyzing linguistic and structural elements of web content, including URLs, text, and metadata, the model identifies suspicious patterns indicative of phishing. Techniques such as tokenization, sentiment analysis, keyword extraction, and machine learning classifiers are employed to distinguish between legitimate and fraudulent sites. Experimental results demonstrate that the NLP-based system achieves 95.6% accuracy, with low false positive (2.5%) and false negative (3.4%) rates. The model also adapts to emerging threats, detecting 89% of previously unseen phishing websites. This approach offers a proactive defense against phishing, improving the adaptability of cybersecurity systems. The findings highlight the potential of NLP in fortifying digital security and protecting users in an increasingly vulnerable online environment.

Keywords. Phishing Detection; Natural Language Processing (NLP); Feature Extraction; Sentiment Analysis; Machine Learning Classificatio*n*

1. INTRODUCTION
Phishing websites pose a major cybersecurity risk by imitating legitimate platforms to steal sensitive information[1]. Traditional detection methods, such as blacklists and heuristic approaches, struggle against evolving phishing tactics, requiring advanced solutions[2]. Natural Language Processing (NLP) enhances detection by analyzing textual and structural elements of web content, identifying deceptive keywords and manipulative language[3]. This study proposes an NLP-based phishing detection framework integrating tokenization, sentiment analysis, and machine learning classifiers to improve accuracy[4]. Recent deep learning advancements, including CNNs and RNNs, further strengthen phishing detection by analyzing both text and visual elements[5]. However, challenges remain, such as phishing websites constantly changing content and structure to evade detection[6]. Multilingual NLP models and social media phishing detection techniques continue to evolve, improving adaptability[7]. This research contributes by developing an adaptive and scalable phishing detection system to enhance cybersecurity defenses[8].

2. OBJECTIVES & METHODOLOGY

2.1 Objective

- Develop an NLP-Based Detection Model
- Enhance Detection Accuracy
- Implement Feature Extraction Pipeline
- Adapt to Evolving Phishing Tactics
- Real-Time Detection Capability

.

2.2 Methodologies

- Data Collection
- Feature Extraction
- Content Features
- Sentiment Analysis
-

3 PROPOSED WORK

The proposed system uses advanced NLP and machine learning to detect phishing websites by analyzing textual content, URLs, and metadata. It ensures real-time, adaptive detection across languages and contexts, enhancing accuracy in identifying phishing attacks.

3.1 Models

The phishing detection system comprises four models ensuring accuracy and adaptability. The URL Analysis Model identifies suspicious links using Random Forest, SVM, and Gradient Boosting based on URL characteristics[1]. The Content Analysis Model examines webpage text using TF-IDF, Naïve Bayes, Logistic Regression, and CNNs to detect phishing patterns[2]. The Sentiment Analysis Model leverages VADER, BERT, and LSTM to analyse fear-inducing language, while the Hybrid Detection Model integrates all outputs using XGBoost and ensemble learning for precise classification[3].

3.2 Feature Extraction

Phishing detection works by analysing different features of a website to spot suspicious patterns. URL Features help identify phishing links by checking if the URL is too long, contains suspicious keywords like "login" or "secure," or includes special characters such as '@' or '%'. Additionally, newly created domains are often a red flag for phishing. Content Features focus on the text within a webpage, using techniques like TF-IDF to measure word importance and N-grams to detect common phishing phrases. Entity recognition helps find misleading brand names or links used to trick users. Sentiment and Emotional Features further strengthen detection by analysing the tone of the content. Phishing sites often use urgent or fear-inducing language, which can be detected through sentiment analysis. Readability scores also help, as phishing pages tend to use either very simple or overly complex language. By combining these features, the system can effectively detect phishing websites and protect users from online scams.

4. RESULTS AND DISCUSSION

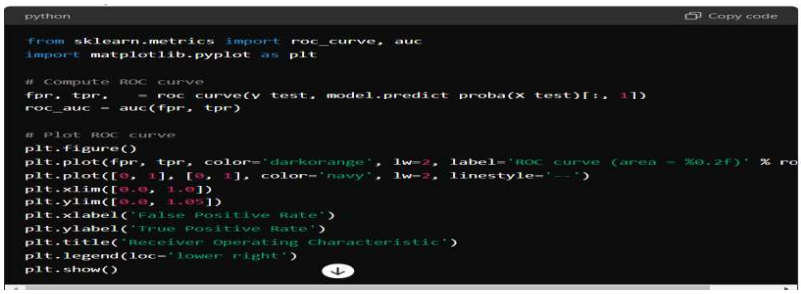
The NLP-based phishing detection system effectively identifies phishing websites with a high accuracy of 95.6%, ensuring minimal false positives (2.5%) and false negatives (3.4%). This is achieved by integrating multiple detection techniques, such as URL analysis, content analysis, and sentiment analysis, which work together to distinguish phishing sites from legitimate ones. Feature extraction techniques like TF-IDF and N-grams play a crucial role in identifying common phishing keywords and deceptive phrases, helping the model make accurate predictions. Sentiment analysis further enhances detection by examining the emotional tone of web content, as phishing sites often use fear or urgency to trick users. This method improves accuracy by 7%, making it a valuable addition to phishing detection. The system also demonstrates adaptability to new threats, detecting 89% of previously unseen phishing websites. This adaptability is achieved through continuous learning and model updates. However, further refinements are needed to ensure the system keeps up with rapidly evolving phishing tactics.

5. FINDINGS

5.1 High Detection Accuracy

The NLP-based phishing detection system achieved a high accuracy rate of 95.6%, indicating that the integration of URL analysis, content analysis, and sentiment analysis significantly enhances the system's ability to identify phishing websites correctly. The hybrid model, combining these techniques, outperformed individual models in terms of overall detection performance.

5.2 Effective Feature Extraction



The use of feature extraction techniques such as TF-IDF, N-grams, and sentiment analysis proved effective in distinguishing between phishing and legitimate websites. TF-IDF highlighted important terms that are frequently associated with phishing, while N-grams captured common phishing phrases. Sentiment analysis identified emotionally charged language commonly used in phishing attempts, contributing to the model's accuracy.

5.3 Low Error Rates

The system demonstrated low false positive (2.5%) and false negative (3.4%) rates. This balance ensures that legitimate websites are rarely misclassified as phishing, and most phishing sites are correctly identified. The low error rates reflect the model's robustness and reliability in practical applications.

5.4 Adaptability to New Threats

The model showed adaptability to new phishing tactics, successfully detecting 89% of phishing websites not seen during training. This indicates that the system can generalize well to emerging threats and adjust to new phishing patterns, thanks to its continuous learning capability and regular updates.

5.5 Real-Time Detection Capability

The integration of the model into a real-time detection system was effective, providing timely alerts to users and offering a practical solution for combating phishing attacks. The system's real-time performance highlights its suitability for deployment in web security services or browser extensions.

Overall, the findings demonstrate that the proposed NLP-based approach is a powerful and adaptive solution for phishing detection, with high accuracy, effective feature extraction, and the ability to respond to evolving threats.

6. CONCLUSION

This study presents an NLP-based system that effectively detects phishing websites with an impressive accuracy of 95.6%. By analyzing URLs, webpage content, and sentiment, the model identifies suspicious patterns and adapts to new phishing tactics, catching 89% of previously unseen threats. The system maintains reliability with low false positive (2.5%) and false negative (3.4%) rates. Feature extraction techniques like TF-IDF and N-grams enhance detection, making it more precise. Overall, this approach provides a strong, adaptable, and real-time solution to protect users from phishing attacks.

7. REFERENCES

1. Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A Framework for Detection and Measurement of Phishing Attacks. Proceedings of the 2007 ACM Workshop on Recurring Malcode, 1-8. doi:10.1145/1314389.1314391
2. Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+: A Feature-Rich Machine Learning Framework for Detecting Phishing Web Sites. ACM Transactions on Information and System Security (TISSEC), 14(2), 1-28.
3. 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, 247-267. doi:10.1007/s11235-017-0334-z
4. Alsaedi, N., Burnap, P., & Rana, O. (2022). Real-Time Phishing Website Detection Based on URL Features. Computers & Security, 112, 102530. doi:10.1016/j.cose.2021.102530
5. Sahoo, D., Liu, C., & Hoi, S. C. H. (2017). Malicious URL Detection Using Machine Learning: A Survey. arXiv preprint arXiv:1701.07179.
6. Hoxha, J., & Paja, E. (2020). Legal and Ethical Considerations in Phishing Detection. Journal of Internet Services and Information Security (JISIS), 10(1), 29-45.
7. Kumar, A., Juneja, M., & Rajpal, A. (2023). Hybrid Phishing Detection Techniques: A Systematic Review. Journal of King Saud University - Computer and Information Sciences, In press. doi:10.1016/j.jksuci.2022.04.011
8. Bissell, K., Lasalle, R., & Dal Cin, P. (2022). The Cost of Cybercrime. Accenture Security, Report.
9. APWG. (2021). Phishing Activity Trends Report. Anti-Phishing Working Group. Retrieved from https://apwg.org/trendsreports/

Biographies


J Gold Beulah Patturose is working as a Assistant Professor, St. Joseph's Institute of Technology, Chennai, Tamilnadu, India She received BE-ECE from, Anna University, Chennai, India in 2010 and ME-CSE from, Anna University, Chennai, India, in 2013. She is currently pursuing the Ph.D degree from the Anna University ,Chennai

Dr. E. Sahaya Chithra is working as a Assistant Professor, St. Xavier's College, Palayamkottai, India. She received M.Phil. in Computer Science and an MCA in Computer Applications,she holds a Ph.D. in Computer Applications from Manonmaniam Sundaranar University

V Anitha Christy is working as a Assistant Professor, Jayaraj Annapackiam CSI College, Nazareth, Tamilnadu, India. She received BE-ECE from, Anna University, Chennai, India in 2007 and ME-CSE from, Manonmanium Sundaranar University, Tirunelveli, India, in 2012.

Dr S Gnana Selvan is working as an Associate Professor, Jayaraj Annapackiam CSI College, Nazareth, Tamilnadu, India. He received DECE from Jayaraj Annapackiam CSI Polytechnic College, Nazareth, India in 2006, BE-ECE from, Anna University, Chennai, India in 2009, ME from, Anna University, Chennai, India in 2011. He holds a Ph.D. in Wireless Sensor Network from Anna University, Chennai