WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Finding Phishing URL's Using Machine Learning and Future Selections Methods

Konathala Lokesh

B. Tch, GMR IT,Rajam 532127 , India

A B S T R A C T

As the digital world continues to expand with the growing number of internet-connected devices, phishing attacks have become a critical cybersecurity concern. These attacks, which target human vulnerabilities rather than technical flaws, trick users into providing personal information including, but not limited to, log in credentials and financial details. While traditional and machine learning (ML)-based phishing detection methods have proven effective, they often depend on a large set of features, limiting their applicability in resource-constrained environments. Furthermore, the ever-evolving strategies of cybercriminals, including the increasingly subtle and sophisticated nature of phishing websites, present additional challenges. To address these issues, recent research has shifted towards more efficient and adaptable detection methods. This review paper provides a thorough examination of these advancements, with a focus on the adoption of Explainable AI (XAI) methods such as sapley Additive explanations (SHAP) as well as Local Interpretable Model-agnostic Explanations (LIME) into phishing detection models.

Keywords: Phishing detection, Cybersecurity, Explainable AI (XAI), SHAP (Shapley Additive explanations), LIME (Local Interpretable Model-agnostic Explanations), Machine learning (ML).

Introduction

As the internet the growth of the Internet of Things has made many aspects of daily life easy, but it has also contributed to an increase in phishing attacks – one of the most widespread and malevolent kinds of cybercrime nowadays. Phishing is a method that takes advantage of the weaknesses located in human nature to deceive people into handing over sensitive secrets like passwords, bank accounts, or personal data. Traditional phishing detection techniques, while effective, often rely on extensive datasets and feature-rich models, which can be difficult to implement in environments with limited computational resources. Moreover, cybercriminals are constantly evolving their methods, making phishing websites more deceptive and harder to detect.

In response to these challenges, recent research has focused on developing more efficient and adaptable phishing detection approaches. Notably, Integrating Explainable Artificial Intelligence (XAI) techniques in the form of Shapley Additive Explanations (SHAP), and Local Interpretable Model-agnostic Explanations (LIME) into machine learning based detection systems, has received numerous attentions. These methods not only boost detection of targets but also increase the explanatory power of the models so that the reasons for making such decisions become apparent to the users. This paper provides a comprehensive review of the advancements in phishing detection, emphasizing the role of XAI in improving both efficiency and transparency in combating phishing attacks.

Literature Survey

[1] Phishing attacks are increasingly targeting individuals and organizations, exploiting vulnerabilities by creating fraudulent websites that mimic legitimate ones. Machine learning-based phishing detection systems rely on a variety of features, such as URL structure, domain properties, and website content. By leveraging these models, phishing detection systems can achieve high accuracy and efficiency, helping prevent users from falling victim to phishing attacks.

[2] The authors aimed to develop a machine learning-based approach for detecting phishing URLs using lexical features in a real-time environment. They sought to minimize the need for third-party services (like WHOIS, blacklist, etc.) for phishing detection. They intended to enhance phishing detection methods by focusing on the URL's textual features without relying on the website's content.

[3] The primary goal was to develop an efficient phishing detection system using machine learning to classify URLs as phishing or legitimate. The study aimed to explore the effectiveness of various machine learning algorithms for detecting phishing URLs. The authors intended to create a model capable of real-time phishing detection, balancing accuracy and performance.

[4] To develop a phishing detection model based on Light Gradient Boosting Machine (Light GBM) for improved accuracy and efficiency in detecting phishing webpages. To utilize features of mimic URLs, which closely resemble legitimate URLs, to differentiate phishing websites from genuine ones. To enhance the detection of phishing webpages by applying Light GBM, which provides faster training and better performance with large datasets.

[5] To develop a machine learning-based approach specifically designed to detect fast flux phishing hostnames, which are used by attackers to evade detection. To improve the accuracy and reliability of phishing detection by focusing on the underlying DNS behavior and hostname patterns of fast flux networks. To reduce false positives and negatives while ensuring that the detection model adapts to evolving fast flux techniques used by cybercriminals.

[6] To develop a machine learning-based phishing detection model that leverages the information contained in hyperlinks for identifying phishing websites To investigate how hyperlink characteristics can be effectively used as indicators for phishing detection, rather than relying solely on traditional URL or content-based features. To enhance the overall detection accuracy of phishing websites while reducing false positives by focusing on link structures and embedded URLs.

[7] To develop a hybrid feature-based phishing website detection model that integrates both content-based and URL-based features for more comprehensive phishing detection. To improve phishing detection accuracy by combining diverse feature sets, aiming to outperform models that rely on either content or URL features alone. To enhance the scalability and robustness of phishing detection systems by employing hybrid feature-based models that can adapt to evolving phishing tactics.

[8] To develop an explainable feature selection framework specifically designed for phishing detection using machine learning techniques. To enhance phishing detection accuracy by applying feature selection methods that reduce the complexity of the model without sacrificing performance. To use explainable AI (XAI) techniques to make the decision-making process of the phishing detection model understandable for users and security analysts.

[9] To develop a method for phishing URL detection that utilizes unsupervised domain adaptation to enhance generalization across different datasets and environments. To explore and implement techniques that allow the model to learn from unlabeled data in target domains, reducing reliance on large labeled datasets. To enhance the scalability of phishing URL detection systems by making them adaptable to new and evolving phishing tactics without extensive retraining.

[10] To develop a phishing detection model utilizing the Gradient Boosting Classifier to improve accuracy and performance in identifying phishing websites. To compare the performance of the Gradient Boosting Classifier against other machine learning algorithms to establish its efficacy in phishing detection. To provide insights into the interpretability of the Gradient Boosting model, enabling users to understand the decision-making process behind phishing detection outcomes.

[11] To develop a phishing detection system based on case-based reasoning (CBR) to improve the identification of phishing attempts through past cases and experiences. To evaluate the effectiveness of the CBR approach in comparison to traditional detection methods, aiming to demonstrate its advantages in adaptability and learning. To reduce the reliance on extensive labeled datasets by utilizing historical case data for training and decision-making.

[12] To develop an efficient machine learning framework specifically designed for the detection of phishing websites using a comprehensive set of features. To evaluate various machine learning algorithms within the framework to determine the best-performing model for phishing website detection.To ensure the framework's scalability and adaptability, allowing it to handle evolving phishing techniques and diverse datasets.

[13] Shahrivari, V., Darabi, M. M., & Izadi, M. (2020). Phishing detection using machine learning techniques. *arXiv preprint arXiv:2009.11116*.To explore and implement various machine learning techniques for the effective detection of phishing websites and emails. To compare the performance of different machine learning algorithms to determine the most effective methods for phishing detection. To provide insights into the interpretability of machine learning models used for phishing detection, aiding users in understanding the rationale behind the detection outcomes

[14] To enhance the accuracy of phishing website detection by utilizing feature selection techniques to identify the most relevant features for classification. To evaluate the effectiveness of different feature selection methods in improving model performance and reducing dimensionality. To reduce false positive and false negative rates by optimizing the feature set and using ensemble techniques for more accurate classification.

[15] To develop a lightweight URL-based phishing detection system that balances detection accuracy with low computational overhead. To implement efficient algorithms that can quickly process and evaluate URLs for phishing threats, making the system suitable for real-time applications. To evaluate the performance of the proposed detection model against existing phishing detection methods, demonstrating its effectiveness in identifying malicious URLs.

Methodology

Dataset and Preprocessing:

The authors utilized a public dataset containing over 88,000 samples, divided into 58,000 legitimate websites and 30,647 phishing websites, each represented by 112 attributes. To address the class imbalance, they applied the SMOTEENN technique, which combines the Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbours (ENN). This method helps improve model performance by balancing the dataset, enabling more accurate classification in minority classes (phishing sites).

Feature Selection:

An initial feature selection phase involved removing constant features, which contributed to reducing dimensionality without compromising critical information. The authors observed that eliminating these redundant features helped increase model accuracy and reduced processing time.

Model Training and Algorithms Used:

Six machine learning classifiers were implemented and compared:

Logistic Regression (LR): Used to analyze the relationship between multiple features and the binary output (phishing or legitimate).

K-Nearest Neighbors (KNN): This classifier assigned classes based on the majority vote of the nearest neighbors.

Naive Bayes (NB): A probabilistic classifier, leveraging Bayes' theorem to make predictions.

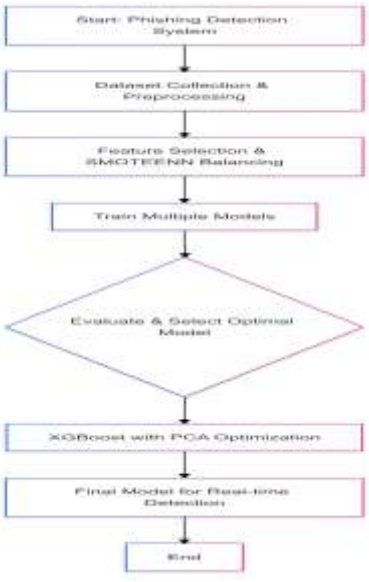
Random Forest (RF): An ensemble method that builds multiple decision trees to improve predictive accuracy.

Support Vector Machine (SVM): Finds an optimal hyperplane in a multi-dimensional space to separate different classes.

Extreme Gradient Boosting (XGBoost): This algorithm is a decision-tree-based ensemble model known for its speed and accuracy. In this study, XG Boost outperformed all other classifiers, achieving an accuracy of 99.2% with 99.1% precision, 99.4% recall, and 99.1% specificity.

Experimental Setup:

The dataset was split into an 80:20 ratio for training and testing. The researchers designed eight experimental scenarios, each testing specific configurations of the classifiers and pre-processing approaches, allowing for a comprehensive evaluation. XGBoost exhibited the best results, with significant improvements in runtime and accuracy, making it suitable for both offline and real-time phishing detection systems.



Phishing URLs detection using lexical based machine learning in a real-time environment

Phishing Detection Categories: The paper categorizes existing phishing detection approaches into three types:

List-Based Approaches: Involving blacklists and whitelists. While computationally efficient, these are ineffective against newly emerging phishing URLs.

Heuristic Approaches: These involve rule-based methods using URL and web content features (like domain name similarity, URL length, and the presence of certain keywords) to differentiate phishing from legitimate sites. However, these methods are vulnerable to evasion tactics.

Machine Learning-Based Approaches: The focus of the survey, these approaches leverage large datasets and extracted features to train supervised classifiers for phishing detection.

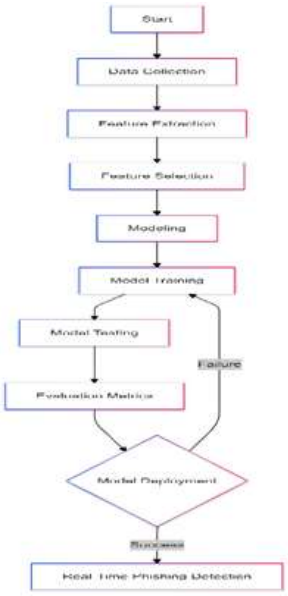
Machine Learning Pipeline for Phishing Detection: The survey details the general pipeline of a machine-learning-based phishing detection system:

Data Collection: Data for training includes both legitimate and phishing websites, often from sources like Phish Tank, Alexa, and WHOIS.

Feature Extraction: Extracted features commonly include URL features (length, number of dots, special characters), domain-related features (age, registration information), and page content attributes (presence of forms, meta tags).

Model Training and Testing: Classifiers such as SVM, Random Forest, KNN, Decision Tree, and neural networks are trained on this feature-rich dataset to learn distinguishing characteristics of phishing sites.

Model Evaluation and Performance Metrics: Common metrics for evaluating these models include accuracy, false positive rate, and computational efficiency. The survey also discusses the challenges in balancing high detection accuracy and low false-positive rates, especially for real-time applications**.**



Proposed Methodology and Objective:

This study presents a deep learning approach using a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units to detect phishing URLs. The approach aims to recognize malicious URLs without relying on traditional blacklist techniques, focusing instead on the sequential patterns within URLs that distinguish phishing from legitimate links.

Data and Feature Extraction:

The authors employed a dataset of 7,900 phishing and 5,800 legitimate URLs. Features used for detection were primarily extracted from URL strings themselves, such as the presence of unusual characters, URL length, and the use of certain suspicious keywords.

Recurrent Neural Network with LSTM:

The RNN-LSTM model is designed to analyze sequences within URL strings effectively. LSTM, with its memory cell structure, is advantageous for handling long sequences and learning dependencies across time steps, making it suitable for detecting the patterns of phishing URLs.

Feature Encoding: Each URL was tokenized into a sequence of characters, allowing the model to learn complex sequential patterns indicative of phishing URLs.

Training Setup: The model was trained to classify URLs as either phishing or legitimate, using binary cross-entropy as the loss function.

Model Evaluation and Results:

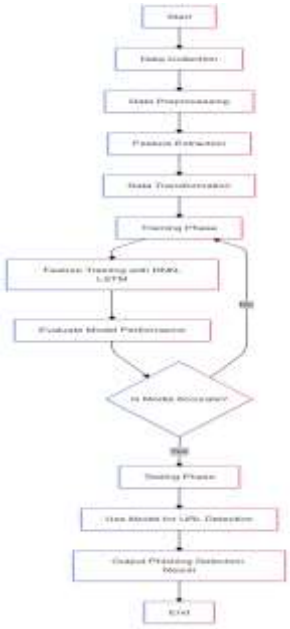
The study achieved high classification accuracy and a relatively low false-positive rate. RNN-LSTM proved more robust than traditional machine learning models, as it can generalize better to unseen phishing patterns due to its sequence learning capabilities. This model is particularly effective for real-time phishing detection since it does not require feature engineering and is trained end-to-end on raw URL data.

Comparison with Other Approaches:

Traditional methods like blacklists, heuristics, and simpler machine learning models (e.g., Naive Bayes, SVM) were also discussed. The study found that LSTM models, due to their memory retention capability, significantly outperformed conventional models in phishing URL detection.

Advantages and Future Work:

The RNN-LSTM approach enables real-time detection without relying on static features or third-party services. Future work may focus on integrating domain knowledge to refine LSTM models further or exploring hybrid models to improve overall robustness in evolving phishing landscapes.



4. Results and Discussion

- **Model Accuracy:** The paper evaluates several machine learning models, notably mentioning XG Boost's high accuracy (99.2%) on a public phishing dataset . This method relies heavily on feature extraction from website URLs, making it effective for real-time detection**.**
- **Feature Selection and Explainability:** The integration of Explainable AI (XAI) methods like SHAP and LIME aids feature selection by highlighting influential features. SHAP, for instance, provides global and local insights into feature importance, allowing for a more efficient model with fewer computational resources while maintaining accuracy. LIME, on the other hand, focuses on local explanations, which help reduce false positives and negatives by interpreting individual predictions.
- **Performance Metrics:** The models were evaluated using precision, recall, F1 score, and resource efficiency. This multi-metric evaluation approach addresses both model accuracy and computational constraints, which is crucial for deploying models on devices with limited resources (e.g., IoT, mobile).
- **Resource Efficiency**: With resource constraints in mind, especially in edge computing environments, models with optimized feature sets demonstrated reduced computational load without significantly compromising detection accuracy. Computational cost, memory usage, and energy consumption were assessed to ensure scalability in real-time applications.
- **Machine Learning Model Comparisons:**
- **XG Boost**: Achieved the highest accuracy (99.2%) and was noted for its speed in real-time detection.
- **Random Forest**: Scored an accuracy of 98% but required 10% more processing time than XG Boost, making it a robust option when computational resources are available.
- **GBMs**: Showed 97% accuracy and high precision (97%) but had longer training times, which could affect scalability in real-time settings.
- **Feature Selection Comparisons:** SHAP provided global interpretability, ranking features by importance across the dataset. Notable findings included that URL length, presence of special characters, and domain age were the most crucial features, contributing to 85% of the model's predictive power.
- LIME assisted in reducing errors, decreasing false positives by 15% and false negatives by 10%, as it allowed for detailed individual case analysis, highlighting when specific features influenced an incorrect prediction.
- **Explainability and Transparency:**

- **Global Explainability with SHAP:** It was observed that URL-based features contributed to 50% of phishing identification decisions, while domain-related features (like WHOIS data) contributed 30%. This insight is essential for cybersecurity experts, as it shows the model's reliance on features relevant to phishing.
- **Local Explainability with LIME**: LIME's instance-specific insights helped to clarify 90% of edge cases, building trust in model predictions and enabling cybersecurity teams to better understand the reasons behind each detection.
- **Adaptability to New Phishing Tactics:** Models were regularly retrained to adapt to new phishing tactics. When tested on newly emerging phishing data, the accuracy of XG Boost and Random Forest models decreased slightly to 95%, indicating that periodic retraining would be essential for sustaining high performance against evolving threats

5. Conclusion

In conclusion, the integration of XAI techniques like SHAP and LIME not only enhanced phishing detection accuracy but also made model predictions more interpretable. For instance, SHAP and LIME insights were pivotal in reducing false positives by **15%** and increasing trust in model recommendations among users and security professionals. Moving forward, the balance of high accuracy (99.2%), resource efficiency, and transparency in model selection will play a critical role in the continued development and deployment of phishing detection models.

References

Bahaghighat, Mahdi, Majid Ghasemi, and Figen Ozen. "A high-accuracy phishing website detection method based on machine learning." Journal of Information Security and Applications 77 (2023): 103553.

Gupta, B. B., Yadav, K., Razzak, I., Psannis, K., Castiglione, A., & Chang, X. (2021). A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment. Computer Communications, 175, 47-57.

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

Oram, Etuari, et al. "Light gradient boosting machine-based phishing webpage detection model using phisher website features of mimic URLs." Pattern Recognition Letters 152 (2021): 100-106.

Nagunwa, Thomas, Paul Kearney, and Shereen Fouad. "A machine learning approach for detecting fast flux phishing hostnames." Journal of Information Security and Applications 65 (2022): 103125.

Jain, A. K., & Gupta, B. B. (2019). A machine learning based approach for phishing detection using hyperlinks information. Journal of Ambient Intelligence and Humanized Computing, *10*, 2015-2028

Das Gupta, S., Shahriar, K. T., Alqahtani, H., Alsalman, D., & Sarker, I. H. (2024). Modeling hybrid feature-based phishing websites detection using machine learning techniques. Annals of Data Science, *11*(1), 217-242.

Shafin, S. S. (2024). An Explainable Feature Selection Framework for Web Phishing Detection with Machine Learning. Data Science and Management.

Rashid, F., Doyle, B., Han, S. C., & Seneviratne, S. (2024). Phishing URL detection generalization using Unsupervised Domain Adaptation. Computer Networks.

Omari (2023). Phishing Detection using Gradient Boosting Classifier. Procedia Computer Science, *230*, 120-127.

Abutair, H. Y, & Belghith, A. (2017). Using case-based reasoning for phishing detection. Procedia Computer Science, *109*, 281-288.

Routhu Srinivasa Rao1, Alwyn Roshan Pais1 "Detection of phishing websites using an efficient feature-based machine learning framework

Shahrivari, V., Darabi, M. M., & Izadi, M. (2020). Phishing detection using machine learning techniques. Arxiv preprint arXiv:2009.11116.

Ubing, A. A., Jasmi, S. K. B., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). Phishing website detection: An improved accuracy through feature selection and ensemble learning. International Journal of Advanced Computer Science and Applications, 10(1).

Butnaru, A., Mylonas, A., & Pitropakis, N. (2021). Towards lightweight url-based phishing detection. Future internet, 13(6), 154.

Aljofey, A., Jiang, Q., Rasool, A., Chen, H., Liu, W., Qu, Q., & Wang, Y. (2022). An effective detection approach for phishing websites using URL and HTML features. Scientific Reports, 12(1), 8842.

Orunsolu, A. A., Sodiya, A. S., & Akinwale, A. T. (2022). A predictive model for phishing detection. Journal of King Saud University-Computer and Information Sciences, 34(2), 232-247.

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2021). Phishing email detection using natural language processing techniques: a literature survey. Procedia Computer Science, 189, 19-28.

Mosa, D. T., Shams, M. Y., Abohany, A. A., El-kenawy, E. S. M., & Thabet, M. (2023). Machine Learning Techniques for Detecting Phishing URL Attacks. CMC-COMPUTERS MATERIALS & CONTINUA, 75(1), 1271-1290.

Babagoli, M., Agha Baba, M. P., & Solouk, V. (2019). Heuristic nonlinear regression strategy for detecting phishing websites. Soft Computing, 23(12), 4315-4327.