

Machine Learning-Based Phishing Website Detection

Dr. Pawar Santosh Barku¹, Prof. Jadhav Ankita Ganpat², Vaibhav Sanjay Khalkar³

Vidya Niketan Institute of Management & Computer Application, Bota, Maharashtra, India

Abstract: *Phishing websites are a major cybersecurity threat because they attempt to deceive users into sharing sensitive information, including usernames, passwords, banking details, and personal data. Traditional detection methods, such as blacklists and rule-based techniques, may fail to identify newly created or modified phishing websites. Machine learning offers an effective approach to analyzing website features and classifying websites as legitimate or phishing. This review examines machine learning techniques used for phishing website detection, including Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, Logistic Regression, and Deep Learning. It discusses website features, dataset quality, feature selection, and evaluation measures such as accuracy, precision, recall, and F1-score. The review also highlights challenges involving changing phishing strategies, classification errors, generalization, computational requirements, and privacy. Future research directions include real-time detection, improved feature extraction, explainable models, and updated datasets. Overall, machine learning shows promise, but reliable detection requires careful evaluation, continuous improvement, and complementary cybersecurity measures.*

Keywords: Phishing Websites, Machine Learning, Cybersecurity, Website Classification, Phishing Detection, Feature Selection, Classification Algorithms

I. INTRODUCTION

The rapid growth of internet usage, online banking, e-commerce and digital communication has increased the importance of cybersecurity. People regularly use websites to access email accounts, make online payments, share personal information and perform various online activities. However, this growing dependence on the internet has also created opportunities for cybercriminals to conduct phishing attacks.

Phishing is a type of cyberattack in which attackers create fraudulent websites or messages that appear to be genuine. These websites attempt to trick users into entering sensitive information such as usernames, passwords, banking details and other personal data. The stolen information may be misused for financial fraud, identity theft and unauthorized access to online accounts.

Traditional phishing detection techniques often use blacklists, predefined rules and website reputation information. Although these methods can identify many known phishing websites, they may not always recognize newly created websites or websites that frequently change their characteristics. Therefore, researchers have explored machine learning techniques to improve the identification of suspicious websites.

Machine learning is a branch of artificial intelligence that enables computer systems to learn patterns from data and use those patterns to make predictions or classifications. In phishing website detection, a model can be trained using website-related features and examples of legitimate and phishing websites. Depending on the dataset and selected features, algorithms such as Decision Tree, Random Forest, Support Vector Machine and Naive Bayes can be studied for classification.

This research review examines existing approaches to machine learning-based phishing website detection. It focuses on the techniques discussed in previous research, the features used to identify phishing websites, the evaluation of detection models and the challenges that affect their performance. The purpose is to understand how machine learning can support phishing detection and identify areas that require further investigation.



II. BACKGROUND OF THE STUDY

The development of the internet has transformed the way people communicate, access information and conduct financial transactions. Online banking, e-commerce, social networking and digital services have become an important part of everyday life. As more users depend on online platforms, protecting personal and financial information has become a major cybersecurity concern.

Phishing is one of the common threats associated with internet usage. In a phishing attack, an attacker attempts to deceive users by creating a fake website that resembles a trusted website. The main purpose is to collect confidential information such as login credentials, credit card details, banking information and other sensitive data. These attacks may cause financial losses, identity theft and privacy violations.

Initially, phishing detection systems commonly relied on blacklists and rule-based techniques. Blacklists contain the addresses of websites already identified as malicious, while rule-based systems examine predefined characteristics. However, attackers can create new phishing websites and modify existing ones, making it difficult for traditional approaches to identify every threat.

Machine learning has emerged as a potential approach for improving phishing website detection. It enables a computer system to learn patterns from previously collected website data. Features such as URL length, the presence of suspicious symbols, domain-related information, HTTPS usage and other website characteristics may be examined during classification, depending on the available dataset.

Researchers have investigated different machine learning algorithms to distinguish legitimate websites from phishing websites. Decision Tree, Random Forest, Support Vector Machine and Naive Bayes are examples of algorithms that can be evaluated for this task. Their performance may vary according to the quality of the dataset, the features selected, the training method and the evaluation procedure.

The background of this study is based on the need to investigate how machine learning techniques can contribute to phishing website detection. Reviewing existing research helps identify commonly used approaches, their advantages, their limitations and the challenges that remain in detecting changing phishing threats.

III. REVIEW OF LITERATURE

3.1 Mohammad, Thabtah and McCluskey (2012)

Mohammad, Thabtah and McCluskey studied features related to phishing websites using an automated approach. Their research focused on identifying website characteristics that can help distinguish phishing websites from legitimate websites. The study is relevant to machine learning-based phishing detection because the selection of suitable features is an important step in building a classification model. The findings provide a foundation for understanding how website features can contribute to phishing identification.

3.2 Mohammad, Thabtah and McCluskey (2014)

Mohammad, Thabtah and McCluskey investigated intelligent rule-based classification of phishing websites. Their work examined website features and classification techniques for identifying fraudulent websites. The research highlights that phishing websites may contain several indicators that can be examined during detection. It also emphasizes that no single detection approach can completely eliminate the phishing threat. This study is relevant to the present research because it provides insight into feature-based website classification.

3.3 Tang and Mahmoud (2021)

Tang and Mahmoud presented a survey of machine learning-based solutions for phishing website detection. Their review examined different machine learning approaches and the methods used to identify phishing websites. The study discusses the importance of selecting suitable features and classification techniques for improving detection. It provides a useful overview of existing research and helps identify challenges that researchers must consider when developing phishing detection models.



3.4 Odeh, Keshta and Abdelfattah (2021)

Odeh, Keshta and Abdelfattah reviewed machine learning techniques for website phishing detection and discussed their promises and challenges. Their work considered conventional machine learning algorithms, including Random Forest, Support Vector Machine and Naive Bayes, as well as deep learning approaches. The review identifies challenges such as overfitting and insufficient training data. This research is relevant to the present study because it highlights the need to select appropriate algorithms and evaluate their performance carefully.

3.5 Hasane Ahammad et al. (2022)

Hasane Ahammad and co-authors studied phishing URL detection using machine learning methods. Their research considered algorithms such as Random Forest, Decision Tree, LightGBM, Logistic Regression and Support Vector Machine. The approach uses URL-related features to classify URLs as malicious or legitimate. The study demonstrates the importance of feature extraction and model selection in automated phishing detection. It is directly relevant to the present research because it focuses on identifying malicious URLs through machine learning techniques.

3.6 Safi and Singh (2023)

Safi and Singh conducted a systematic literature review of phishing website detection techniques. Their review examined multiple approaches, including blacklist-based, visual similarity, heuristic, machine learning and deep learning techniques. The study reviewed 80 research publications and discussed the datasets and algorithms used in phishing detection. It reported that machine learning was widely used in the reviewed studies and that Random Forest was a frequently applied classifier. This review helps establish the importance of comparing detection techniques and considering dataset selection when developing phishing detection systems.

3.7 Overall Review of Literature

The reviewed studies indicate that machine learning is an important approach for phishing website detection. Researchers have investigated website features, URL characteristics, rule-based classification, conventional machine learning algorithms and deep learning techniques. The literature also identifies challenges related to dataset quality, feature selection, overfitting and the changing nature of phishing attacks. These studies provide a foundation for examining machine learning algorithms and evaluating their effectiveness in detecting phishing websites. The present research focuses on understanding these approaches and identifying suitable techniques for improving phishing website detection.

IV. EDUCATIONAL ROLE OF DIGITAL PLATFORMS AND SOCIAL MEDIA**4.1 Decision Tree**

Decision Tree is a supervised machine learning algorithm that classifies data using a sequence of decision rules. In phishing website detection, it can analyze features such as URL characteristics, domain information and other website-related attributes. The model follows a tree structure to classify a website as legitimate or phishing. Its main advantage is that the classification process can be relatively easy to understand. However, a Decision Tree may overfit the training data if it is not properly controlled.

4.2 Random Forest

Random Forest is an ensemble learning algorithm that combines predictions from multiple decision trees. It can analyze several website features and use the combined predictions to classify a website. This approach can reduce some of the limitations of an individual decision tree and is commonly considered for phishing detection tasks. Its performance depends on the quality of the dataset, feature selection and model configuration.

4.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm that identifies a decision boundary between different classes. In phishing website detection, it can be trained using features extracted from legitimate and phishing websites. Its effectiveness depends on the nature of the data, feature scaling and the selected kernel. For large datasets, training and parameter selection may require additional computational resources.



4.4 Naive Bayes

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It estimates the probability that a website belongs to a particular class using the available features. It can provide a relatively simple and computationally efficient approach to phishing detection. However, its assumption that features are conditionally independent may not always reflect the relationships among website characteristics.

4.5 Logistic Regression

Logistic Regression is a supervised classification algorithm that estimates the probability of a website belonging to a particular class. It can be used to classify websites as legitimate or phishing based on selected features. Its simplicity and interpretability make it a useful baseline for comparison with other algorithms. However, its performance may be limited when the relationship between website features and phishing behaviour is highly complex.

4.6 Deep Learning Techniques

Deep learning uses neural networks to learn complex patterns from data. In phishing detection, deep learning approaches may analyze URL sequences, website content or other representations of websites, depending on the model design and available data. These techniques can learn useful patterns but may require larger datasets and greater computational resources. Their effectiveness must be established through appropriate testing and comparison.

4.7 Comparison and Evaluation of Algorithms

Different machine learning algorithms may perform differently on the same phishing detection dataset. Their results depend on the number and quality of features, class distribution, data preparation and training procedure. Common evaluation measures include accuracy, precision, recall and F1-score. Recall is particularly important because it measures how many actual phishing websites are correctly identified. Comparing several measures provides a more complete assessment of a model's performance.

4.8 Importance of Feature Selection

Feature selection involves identifying the website characteristics that are most useful for classification. Examples may include URL length, suspicious symbols, domain-related information, HTTPS usage and webpage characteristics. Appropriate feature selection can reduce unnecessary information and may improve model efficiency. However, features that are useful in one dataset may not be equally effective in another, so their contribution should be evaluated experimentally.

V. CHALLENGES AND LIMITATIONS IN EDUCATIONAL USE

Machine learning-based phishing website detection has shown potential for identifying fraudulent websites. However, several challenges can affect the accuracy, reliability and practical use of these systems. The performance of a detection model depends on the quality of the dataset, the selection of website features, the classification algorithm and the changing nature of phishing attacks.

5.1 Dataset Quality and Availability

The quality of the dataset plays an important role in training a machine learning model. A dataset should contain sufficient examples of both legitimate and phishing websites. If the dataset is outdated, incomplete or unbalanced, the model may not learn the patterns required for reliable classification. Incorrect labels can also affect the performance of the trained model.

5.2 Changing Phishing Techniques

Cybercriminals continuously change their methods to avoid detection. They may create new domain names, modify URLs, copy the appearance of legitimate websites or use compromised websites for phishing activities. A model trained on older examples may fail to identify phishing websites that use new techniques. Therefore, detection systems may require regular evaluation and updated training data.

5.3 Feature Selection and Extraction

Machine learning models use website-related features to distinguish legitimate websites from phishing websites. These features may include URL length, suspicious symbols, domain information, HTTPS usage and webpage characteristics.



Selecting relevant features is important because irrelevant or insufficient features can reduce classification performance. Some features may also become less useful as phishing techniques change.

5.4 False Positives and False Negatives

A false positive occurs when a legitimate website is incorrectly classified as a phishing website. A false negative occurs when a phishing website is incorrectly classified as legitimate. Both errors create difficulties for users and security systems. False positives may block genuine websites, while false negatives may allow users to access dangerous websites. Therefore, evaluation should consider precision, recall and F1-score in addition to accuracy.

5.5 Generalization to New Websites

A machine learning model should perform well on websites that were not included in its training dataset. However, a model may learn patterns that are specific to the training data and perform poorly on new examples. This problem can occur when the training and testing datasets are not sufficiently representative of real-world websites. Appropriate data splitting and testing on independent datasets can help assess generalization.

5.6 Computational Requirements and Scalability

Some machine learning methods require additional processing time, memory and computational resources, particularly when working with large datasets or complex models. A practical phishing detection system should provide results within a reasonable time and handle a large number of website requests. The choice of algorithm should therefore consider both detection performance and computational cost.

5.7 Privacy and Data Security

Some detection approaches examine webpage content, domain information or other website-related data. The collection and processing of such information should be handled carefully, especially when personal or sensitive information may be involved. A detection system should follow appropriate privacy and data-security practices while collecting, storing and processing data.

5.8 Need for Continuous Evaluation

Phishing websites and attack strategies change over time. A model that performs well on one dataset may not maintain the same performance in a different environment. Continuous evaluation, updated datasets and testing against emerging threats can help identify weaknesses in the detection process. Machine learning should be considered one component of a broader cybersecurity strategy rather than a complete guarantee of protection.

In summary, machine learning-based phishing website detection faces challenges related to dataset quality, changing attack techniques, feature selection, classification errors, generalization, computational requirements and privacy. Addressing these limitations is important for developing more reliable detection systems. Combining suitable machine learning techniques with regular evaluation and other security measures may improve the practical effectiveness of phishing website detection.

VI. RESEARCH METHODOLOGY AND EVIDENCE BASE

The research methodology adopted for this review paper is based on studying existing literature related to machine learning-based phishing website detection. The purpose of this review is to understand the methods used by researchers, identify commonly considered website features, examine the evaluation approaches and discuss the challenges associated with phishing website classification.

6.1 Research Approach

This paper follows a literature review approach. Relevant research studies, review papers and technical publications related to phishing website detection and machine learning techniques are considered to understand the development of existing approaches. The review focuses on classification algorithms, website features, model evaluation and the limitations reported in previous studies.

6.2 Literature Identification

Literature relevant to the topic can be identified through academic search platforms and digital libraries, including Google Scholar, IEEE Xplore, ScienceDirect and SpringerLink. Search terms may include “phishing website



detection,” “machine learning for phishing detection,” “phishing URL classification,” “website classification algorithms” and “machine learning-based cybersecurity.”

6.3 Selection of Relevant Studies

Studies are considered relevant when they discuss phishing website or URL detection, machine learning-based classification, features used for identifying suspicious websites, or methods for evaluating detection performance. Publications that do not directly contribute to the objectives of the review may be excluded. Priority should be given to research papers that clearly describe their methodology, datasets, algorithms and evaluation measures.

6.4 Evidence and Information Extraction

Information from selected studies can be organized according to the detection technique, machine learning algorithm, features examined, dataset characteristics and evaluation metrics. This approach helps compare the methods discussed in the literature and identify common advantages and limitations. Any reported performance values should be attributed to the original study and interpreted in the context of its dataset and experimental setup.

6.5 Evaluation Measures Discussed in the Literature

Phishing website detection models are commonly evaluated using measures such as accuracy, precision, recall and F1-score. Accuracy indicates the proportion of correct predictions, while precision measures how many websites predicted as phishing are actually phishing websites. Recall measures how many phishing websites are correctly identified. The F1-score combines precision and recall into a single measure. Considering multiple metrics provides a more informative assessment than relying on accuracy alone.

6.6 Evidence-Based Comparison

The review compares the approaches described in the selected literature rather than assuming that one algorithm is always superior. Decision Tree, Random Forest, Support Vector Machine, Naive Bayes and other methods may behave differently depending on the dataset, selected features and experimental conditions. Therefore, performance claims should be supported by the original publications, and results obtained from different datasets should not be compared without considering these differences.

6.7 Scope and Limitations of the Methodology

This review is intended to summarize and discuss existing research rather than claim new experimental findings. Its conclusions depend on the availability, relevance and quality of the literature examined. Differences in datasets, feature extraction techniques, testing procedures and evaluation measures may limit direct comparisons between studies. A systematic search and transparent selection process can improve the reliability and reproducibility of the review.

In summary, the methodology provides a structured way to examine existing research on machine learning-based phishing website detection. By organizing evidence according to algorithms, features, evaluation measures and reported limitations, the review aims to develop a clearer understanding of current approaches and identify areas that may benefit from further research.

VII. FUTURE TRENDS AND RESEARCH DIRECTIONS

The continuous development of phishing techniques has created a need for more reliable and adaptable website detection systems. Machine learning provides useful methods for identifying patterns in website-related data, but further research is required to improve detection performance, adaptability and practical usability. Future research may focus on advanced learning techniques, improved feature extraction, real-time detection and the responsible use of website data.

7.1 Integration of Advanced Machine Learning Techniques

Future studies may investigate advanced classification algorithms and ensemble learning techniques to improve phishing website detection. Combining multiple models may help capture different patterns in website data. However, the effectiveness of such approaches should be established through appropriate testing and comparison with existing methods.



7.2 Use of Deep Learning

Deep learning techniques may help identify complex patterns in URLs, webpage content and other website-related information. Neural networks can learn representations from data, potentially reducing the need to define every feature manually. Future research should examine their performance against traditional machine learning algorithms while considering training requirements, computational cost and model interpretability.

7.3 Real-Time Phishing Website Detection

Real-time detection is an important direction for future research. A practical system should be capable of analyzing a website and providing a warning within a reasonable time. Researchers may explore efficient feature extraction, lightweight classification models and integration with browser-based security tools. Such systems should also be tested against newly appearing phishing websites.

7.4 Improved Feature Selection and Extraction

Future research may explore more effective ways to select features that distinguish legitimate websites from phishing websites. URL structure, domain information, webpage content and other relevant characteristics may provide useful signals. Researchers should also investigate how feature importance changes as attackers modify their techniques. Selecting informative features may improve model efficiency and reduce unnecessary computation.

7.5 Detection of New and Previously Unseen Threats

Machine learning models may perform poorly when they encounter phishing techniques that differ from their training data. Future studies can investigate methods that adapt to changing threats and evaluate models on newer, independent datasets. Testing on previously unseen websites is important for understanding whether a model can generalize beyond the examples used during training.

7.6 Explainable and Interpretable Detection Models

In cybersecurity, it is useful to understand why a website has been classified as suspicious. Future research may focus on explainable machine learning techniques that provide understandable reasons for a prediction, such as unusual URL characteristics or suspicious domain-related features. Clear explanations may help security professionals assess alerts and understand potential model errors.

7.7 Privacy-Preserving and Secure Learning

Future phishing detection research should consider privacy and data security during the collection, storage and processing of website information. Researchers may investigate approaches that reduce the need to share sensitive data while still supporting effective model development. The security of datasets and the detection system itself should also be considered.

7.8 Standardized Evaluation and Benchmarking

A fair comparison of detection algorithms requires clear evaluation procedures and suitable datasets. Future studies should report dataset sources, feature selection methods, training and testing procedures, and metrics such as precision, recall, F1-score and accuracy. Testing across different datasets and time periods may provide a better understanding of real-world performance and help reduce misleading comparisons.

In summary, future research on machine learning-based phishing website detection may focus on advanced algorithms, deep learning, real-time detection, improved features, adaptability to new threats, explainability, privacy and standardized evaluation. These directions may contribute to more effective detection systems. However, their benefits must be demonstrated through reliable experiments and evaluation under realistic conditions.

VIII. CONCLUSION

Phishing websites are a significant cybersecurity threat because they attempt to deceive users into sharing sensitive information, including login credentials, banking details and personal data. Traditional detection techniques, such as blacklists and predefined rules, can identify many known threats but may face difficulties when dealing with newly created or modified phishing websites. Therefore, machine learning has become an important area of research for improving phishing website detection.



This review examined the role of machine learning in identifying phishing websites and discussed different classification techniques, including Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, Logistic Regression and Deep Learning. These techniques can analyze website-related features and classify websites based on patterns learned from data. However, the effectiveness of each approach depends on factors such as dataset quality, feature selection, algorithm configuration and evaluation procedures.

The review also highlighted several challenges, including changing phishing strategies, inaccurate classifications, limited generalization to new websites, computational requirements and privacy concerns. Evaluation measures such as accuracy, precision, recall and F1-score are important for assessing model performance. In particular, examining recall alongside other metrics helps determine how effectively a model identifies phishing websites without relying only on overall accuracy.

Future research may focus on real-time detection, advanced learning techniques, improved feature extraction, explainable models and evaluation using updated datasets. Models should be tested under realistic conditions to determine whether they can identify previously unseen phishing websites. Combining machine learning with other cybersecurity measures may provide an additional layer of protection.

In conclusion, machine learning offers a promising approach to phishing website detection, but no single algorithm can guarantee the identification of every phishing website. Reliable detection requires appropriate datasets, careful model evaluation, regular assessment of emerging threats and responsible system design. Continued research in this area can contribute to the development of more effective and adaptable cybersecurity solutions.

REFERENCES

1. Mohammad, R. M., Thabtah, F., & McCluskey, L. (2012). An assessment of features related to phishing websites using an automated technique. *2012 International Conference for Internet Technology and Secured Transactions (ICITST)*, 492–497. IEEE.
2. Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Intelligent rule-based phishing websites classification. *IET Information Security*, 8(3), 153–160.
3. Mohammad, R. M., Thabtah, F., & McCluskey, L. (2015). Tutorial and critical analysis of phishing websites methods. *Computer Science Review*, 17, 1–24.
4. Tang, L., & Mahmoud, Q. H. (2021). A survey of machine learning-based solutions for phishing website detection. *Machine Learning and Knowledge Extraction*, 3(3), 672–694.
5. Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. *Journal of King Saud University – Computer and Information Sciences*, 35(2), 590–611.
6. Mohammad, R., & McCluskey, L. (2012). *Phishing Websites* [Dataset]. UCI Machine Learning Repository.

