International Journal of Advance and Innovative Research
Volume 12, Issue 2 (XVII): April - June 2025
ISSN 2394 - 7780

PHISHING DETECTION AND PREVENTION USING MACHINE LEARNING ALGORITHMS

[1]Kavya Chouhan, [2]Ms. Rohana Deshpande and [3]Ms. Fatima Shaikh
[1]Bachelor of Science and [2, 3]Assistant Professor, Information Technology, Jai Hind College
Churchgate, Mumbai, India

ABSTRACT
The Internet has become an integral part of our lives. With the increased use of the internet, fraudsters have become more active and advanced. Phishing attacks are one of the most serious security risks on the Internet today. Due to technological advancements, detecting phishing attempts has become more challenging. Understanding if a web page is legitimate or not is a difficult task because of its semantics-based attack strategy, which primarily exploits computer users' vulnerabilities. The goal of this research is to create a machine learning model that can enhance accuracy and reduce false positives when detecting phishing emails and URLs.

There are several methods for determining whether a website is legitimate or not. Many software companies are launching anti-phishing technologies that use approaches such as signatures, blacklists, heuristics, and visuals. However, each has its advantages and disadvantages. Therefore, there is a need for real-time machine learning methods. Using algorithms such as random forests, xgboost and neural networks for the classification of phishing webpages, we can achieve our goal. To increase detection accuracy, we use advanced feature extraction techniques such as URL Length, Web Traffic, Page Rank, and HTTPS token. Data will be collected from publicly available Kaggle databases. Model performance will be assessed using accuracy, precision, and recall.

This study not only explores how we may use several classification models to improve the accuracy of detecting phishing websites but also addresses the practical implementation challenges associated with integrating these models into existing security infrastructures.

Keywords- *Phishing detection, machine learning, cybersecurity, data analysis, classification*

I. INTRODUCTION
One of the most serious problems with using the web is phishing attacks, which tricks users to provide sensitive information including login credentials, financial details, and personal data by impersonating them as legitimate websites [5]. This type of social engineering takes advantage of users' trust in a recognizable digital environment, which can have financial and data security risks affecting not only people but also companies and industries worldwide [1]. This is because traditional security solutions, such as blacklist-based filtering and signature-based detection alone, are not sufficient because in today's world, attackers continuously improve their tactics to compromise systems. These methods are unable to keep up with high-volume phishing attempts, and there is generally a delay in detection, which creates more risk [2].

The application of machine learning algorithms has demonstrated considerable success in overcoming these difficulties as it improves the detection accuracy and adaptability of phishing protection techniques. ML-based methods perform feature extraction from web pages, such as URLs, HTML content, and hyperlink structures of web pages, to classify websites based on learned patterns between legitimate or malicious sites. This is particularly beneficial for “zero-hour” detection, where a new phishing site can be indicated as fake because of its feature properties and does not need to appear in previous blacklists [4].

Machine learning models, such as decision trees and support vector machines, to more effective ensemble methods, such as Random Forest for checking the domain name Structure or Link Patterns or Embedded Scripts in phishing sites, are another frontier domain where machine learning has been able to display high classification accuracy. These advances highlight the importance of machine learning in phishing detection, not only to provide more reliable defenses, but also scalable solutions that can be applied at both the enterprise and individual levels.

ISSN 2394 - 7780

Fig. 1 Lifecycle of a Phishing Attack

II. CHALLENGES IN INTEGRATION

A. Existing Security Infrastructure

Protecting our digital realm from unauthorized access and criminal activity is not a new concept; firewalls, intrusion detection systems (IDS), and anti-virus software are time-tested measures. These technologies, however highly capable they may be in their own niches of defense, are by no means adaptable or fast enough to adapt quickly to new types of threats like zero-days and sophisticated phishing campaigns. However, including machine learning models inside such security systems can enhance their detection accuracy and also help in speed via pattern recognition and predictive analysis.

Today, machine learning poses as the next generation solution to improve traditional security systems by providing a range of benefits. Machine learning algorithms like Random Forest, Support Vector Machines (SVM) etc have been successful in enhancing IDS and can accurately categorize legitimate or malicious traffic with high accuracy. Moreover, machine learning models can learn with new data and enhance systems to function better in unknown threats detection than existing static rule-based firewalls & antivirus software [2].

Despite these advantages, blending machine learning with traditional security systems also brings complications, especially with outdated infrastructure. Most of the conventional IDS and firewall systems are rule-based which will make it difficult to fit in a machine learning model that is driven by data results. The resource constraints can also happen in the case of traditional systems with limited processing power and lack of real-time processing capabilities which are necessary for running machine learning algorithms. Traditional systems may also resist due to protocol incompatibility or security policies that are not compatible with newer, AI-driven capabilities [4].

As such, the integration of machine learning into existing security frameworks has great potential to improve defenses in general, but its utility will hinge on the cross-compatibility between themselves, as well as power-wise support and adaptability with legacy systems, so it is possible for these benefits without disturbing traditional norms in terms of securities.

B. Real-Time Application Issues

Deploying machine learning models for security in real-time comes with various difficulties, mainly surrounding latency and computational efficiency. Machine learning based phishing and intrusion detection systems rely upon rapid threat response times to detect the threats early, so as not impact users or systems. However, the higher latencies of complex models (such as deep learning architectures) make them unsuitable for real-time applications. When working with large data streams, even feature-rich models can be slow to compute; thus delaying detection response.

It is a fundamental trade-off between model complexity and detection speed among real-time cybersecurity applications. However, complex models such as Random Forest or deep neural networks, provide great accuracy in phishing detection but may be inadequate for high-speed situations without proper optimization due to processing overhead [3]. On the other hand, simpler models like Logistic regression or Decision Trees could

ISSN 2394 - 7780

provide slightly worse detection but it will be faster as compared to complex learning schemes and may fail to detect advanced phishing/intrusion patterns.

These trade-offs emphasize the need of how models should be selected and optimized according to requirements from real-time systems. Achieving a fine balance between accuracy and fast detection usually necessitates ensemble or hybrid methods that combine efficient techniques, complementing each other to retain high performance while also ensuring computational efficiency, which are key for effective real-time security solutions.

C. Limitations During Deployment

There are many difficulties associated with deploying machine learning models in cyber security, such as data quality or model drift. In this case, phishing detection and other security models require a list of high-quality labeled datasets to achieve an accurate evaluation. However, real-world data are often noisy or incomplete; therefore, model predictions are inaccurate and include false positives or negatives.

Approving usage is another obstacle, because customers and administrators might be hesitant to authorize fully automated machine learning-based systems. Complex models may lack interpretability and reliability, which leads to potential distrust of the system's outputs. In high-stake organizations, the problem worsens because false positives slow down operations and put security at risk by producing unfiltered data. This means that building trust in machine learning for cybersecurity involves increasing transparency and delivering users with concrete actionable behaviors [6].

In summary, efficient deployment requires careful attention to data quality and modeling updates as well as an overall approach to improve user trust in hands-off systems while also achieving the right balance between advanced threat detection functionality and system usability and reliability.

III. METHODOLOGY

A. Data Collection

The dataset used in this study was obtained from a publicly available Kaggle-phishing data collection that contains legitimate and phishing samples retrieved from active phishing databases [8]. Using open-access data such as this is crucial for training machine learning models because it captures a wide range of phishing attack patterns and characteristics observed in real-world situations. Current phishing data allow models to be trained on the most recent malicious practices and real user behaviors, which helps them percolate accurate results and be adept through live detection environments. Additionally, these datasets have pre-processed and normalized data that help the model generalize better across multiple types of phishing cases [2].

The dataset has 10,000 rows by 50 columns in which 5000 phishing webpages and 5000 legitimate webpages. The columns, or features, represent several attributes of the URLs: structural properties of resources and pages; behavioral measures capturing browsing habits or interaction data; and metadata that are already known about Web Resources from other repositories.

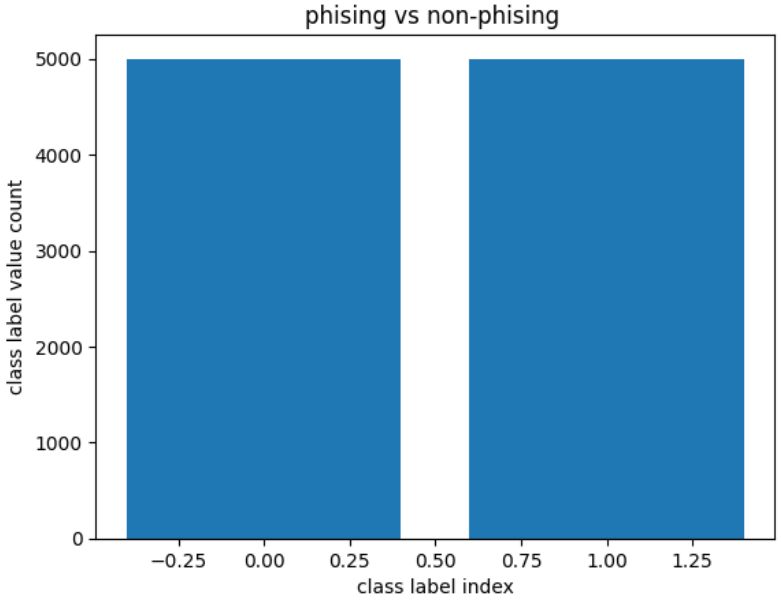

Fig. 2 Distribution of Phishing vs. Non-Phishing URLs

ISSN 2394 - 7780

B. Model Selection

In this study, an implementation was performed using three machine learning algorithms: Random Forest, XGboost and Neural Network models. Each method was chosen to use different components of machine learning for optimal detection accuracy and model robustness based on their performance, as reported in previous studies.

1) **Random Forest**: Random Forest is a powerful ensemble method that has shown promising performance and robustness, particularly when applied to phishing detection problems. This makes the model a popular choice for use in cybersecurity applications because it can deal with imbalanced datasets. Existing research also sheds light on the utility of Random Forest to identify phishing sites, exploiting characteristics in URL and web page features with a bagging wentropy within itself and radical characteristic randomness.
2) **XGBoost**: XGB is a popular choice for phishing detection because it has GPU support that boosts the efficiency and scalability of data training, making large-scale datasets processed quickly with high speed. This has been proven in the research, which made use of studies regarding phishing detection and how boosting algorithms like XGBoost are able to sequentially identify complex non-linear patterns due to the way it adjusts focus on areas where they were previously scored less [6]. This model is quite advantageous for our goals and it has a strong advantage against the simpler models when dealing with false negatives in phishing detection.
3) **Artificial Neural Network (ANN):** ANN is a fully connected network that shows an advantageous performance while capturing deep information between the data inputs; therefore, it performs well in intricate feature interactions. With phishing detection, neural networks have been exploited to recognize the complex dynamics in fishing strategies, and they can model a flexible propensity within feature learning [7]. Neural Networks: Although they are immensely powerful, NNs can be computationally intensive and may have issues with time lag, that is, latency, thereby rendering them a suitable candidate for real-time detection without proper hardware optimization [5].

Random Forest outperformed these models in terms of performance scores on the Kaggle phishing dataset, which is expected given previous works that have shown random forest to be one of the best trade-offs between accuracy and computational efficiency [8]. The combat model is a winner in terms of both toughness and instantaneous adaptability, which makes it ideal for applications such as cybersecurity, where fast identification with accuracy plays an important role.

C. Evaluation Metrics

To assess the effectiveness of phishing detection models after deployment, important evaluation measures such as **accuracy**, **precision**, and **recall** are utilized to determine how well each model makes decisions about whether a URL is a phishing or legitimate. These measures provide a comprehensive view of model reliability, notably for detecting phishing attacks.

1) **Accuracy**: This indicates the ratio of the total URLs tested, which has been correctly classified as both phishing and legitimate. For instance, in this study, the Random Forest model achieved an accuracy of 99%, XGBoost reached approximately 97%, and the Neural Network had an approximate score of 96%. High accuracy indicates that a model generally performs well, although the above evolution process may not correctly capture the performance of imbalanced datasets, especially where phishing instances are rare compared with legitimate ones.
2) **Precision**: Precision measures the accuracy of a classifier when creating phished or legit urls. The necessity for high precision is especially acute in phishing detection; low false positives are essential for avoiding interference with legitimate sites. This study has achieved approximately 99% precision from Random Classifier, 97 % from XGBoost and a Neural Network Model of approximately 95%.
3) **Recall**: Recall, also known as sensitivity, shows how well a model can find phishing URLs in all real-life instances. The recall scores in this research were approximately 99% for Random Forest, 97% for XGBoost and 96% for The Neural Network model.

Combined, these metrics enable an understanding of the true strengths in each model relative to a good tradeoff between accuracy and send sensitivity. (reducing overfitting).

ISSN 2394 - 7780

	Accuracy	Precision	Recall
Random Forest	0.99	0.99	0.99
XGBoost	0.97	0.97	0.96
Neural Network	0.96	0.95	0.96

Fig. 3 Model Evaluation Metrics

IV. CASE STUDY

Reference [9] provided a practical example of how machine learning models were effectively integrated into a browser plugin to detect phishing sites in real time [9]. The project involved creating PhishNet, a Google Chrome extension that detects phishing attempts on user-visited URLs using rules created by a random-forest model. PhishNet was trained on a set of URLs, and the model was developed with 14 essential parameters: domain discrepancies, IP address indicators, SSL presence, and URL length. With an accuracy of 98.35% and a True Positive Rate of 100%, the Random Forest model outperformed the other models.

When PhishNet was used in a browser, it seamlessly integrated into users' surfing experiences and immediately alerted when phishing-related website features were detected. This strategy has proven to be quite effective; end users are protected against phishing attacks without having to rely on third parties, which can create latency and security issues.

So here are some major takeaways from the case study: In a real-time, dynamic-user-facing context, machine learning models like Random Forest can be extremely effective if rules can be generated and processed as efficiently as feasible. However, it reveals some limitations with regard to dependency, because without features from the initial feature set, any phishing efforts may have a negative impact on the model's accuracy. Future applications should therefore improve feature diversity in order to respond gradually to emerging phishing strategies.

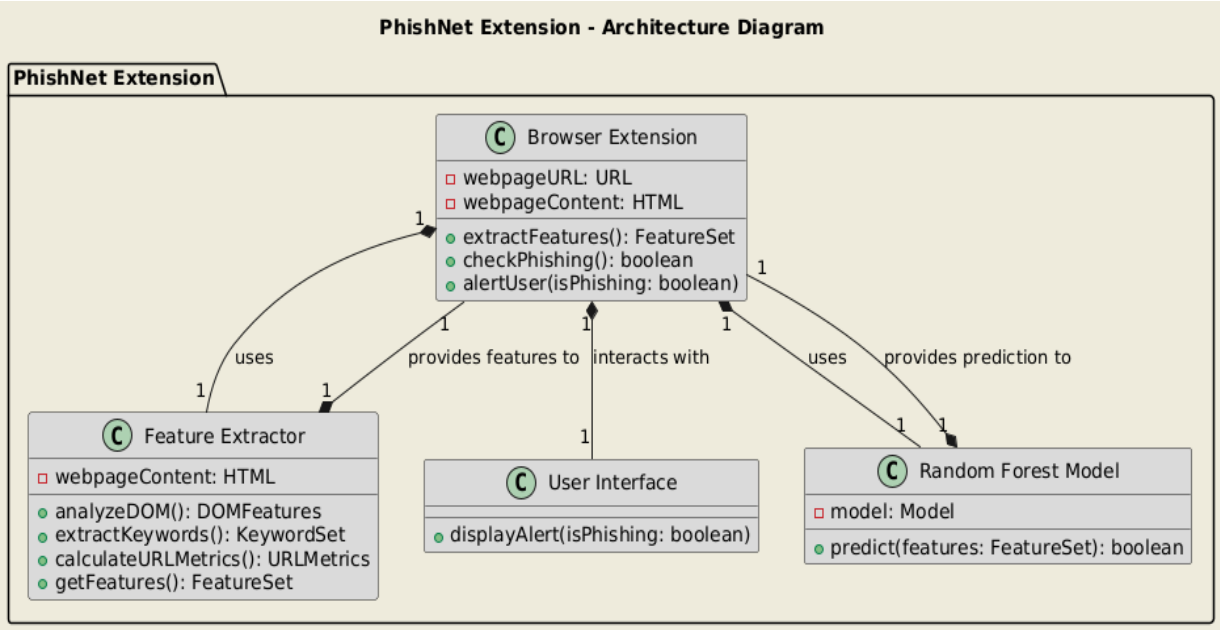

Fig. 4 Architecture Diagram of PhishNet

V. FUTURE DIRECTIONS

An adaptive learning model is the best option for handling the evolving nature of phishing attacks, and it can significantly boost the phishing detection in future. It can apply adaptive learning with mechanisms such as reinforcement learning to update and improve the detection criteria according to the latest types of phishing trends.

Additionally, hybrid model approaches that merge traditional machine learning techniques such as Random Forest with deep learning models will likely result in more robust and flexible detection systems. Hybrid models detect URL- and content-based phishing by combining the interpretability of classical models with the accuracy of deep learning.

Future research and implementation can focus on leveraging natural language processing (NLP) to improve detection of phishing attacks through email, employing more complex models that identify key aspects of the email content including the intent and sentiment of emails.

ISSN 2394 - 7780

As these technologies improve, the combination of adaptive, deep learning, and hybrid approaches will undoubtedly result in stronger and high-accuracy phishing detection frameworks that can identify and react to new types of attacks in real time.

VI. CONCLUSION

Machine learning is one such technique that can automate processes and has been proven to be effective over the traditional process. Random Forest always has the best trade-off when it comes to speed and accuracy as compared to all algorithms, so it is a suitable solution for real-time scenarios. More complicated models, such as the Neural Network, which was an effective tool for analysis, faced deployment issues because it requires significant computational capacity, particularly when operating in latency-sensitive situations.

Additionally, deploying these models into existing security infrastructures causes compatibility issues, particularly with older technologies not designed for machine-learning frameworks.

Further research may focus on applying reinforcement learning methods that enable models to adapt dynamically to changes in phishing technology. In addition, deep learning-based phishing detection also enables more content-based detection, e.g., on phishing emails rather than URL only. Making use of hybrid models that use classical algorithms alongside deep learning may prove more robust, interpretable, and adaptive to real-world cybersecurity applications.

REFERENCES

[1] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (n.d.). *Machine learning based phishing detection from URLs*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417418306067

[2] Gandotra, E., & Gupta, D. (n.d.). *An efficient approach for phishing detection using machine learning*. In *Smart Innovations in Communication and Computational Sciences* (pp. xxx–xxx). Springer. Retrieved from https://link.springer.com/chapter/10.1007/978-981-15-8711-5_12

[3] Jain, A. K., & Gupta, B. B. (n.d.). *A machine learning-based approach for phishing detection using hyperlink information*. *Journal of Ambient Intelligence and Humanized Computing*. Retrieved from https://link.springer.com/article/10.1007/s12652-018-0798-z

[4] Abdelhamid, N., Thabtah, F., & Abdel-jaber, H. (n.d.). *Phishing detection: A recent intelligent machine learning comparison based on models content and features*. Retrieved from https://ieeexplore.ieee.org/document/8004877

[5] Hossain, S., Sarma, D., & Chakma, R. J. (n.d.). *Machine learning-based phishing attack detection*. Retrieved from https://www.academia.edu/download/101989368/Paper_45-Machine_Learning_Based_Phishing_Attack.pdf

[6] Martínez Torres, J., Iglesias Comesaña, C., & García-Nieto, P. J. (n.d.). *Machine learning techniques applied to cybersecurity*. *International Journal of Machine Learning and Cybernetics*. Retrieved from https://link.springer.com/article/10.1007/s13042-018-00906-1

[7] Handa, A., Sharma, A., & Shukla, S. K. (n.d.). *Machine learning in cybersecurity: A review*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Retrieved from https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1306

[8] Tiwari, S. (n.d.). *Phishing Dataset for Machine Learning*. Kaggle. Retrieved from https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning

[9] Ojewumi, T. O., Ogunleye, G. O., Oguntunde, B. O., Folorunsho, O., Fashoto, S. G., & Ogbu, N. (n.d.). *Performance evaluation of machine learning tools for detection of phishing attacks on web pages. ICT Express.* Retrieved from https://www.sciencedirect.com/science/article/pii/S2468227622000746