IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

PHISHING ATTACK PROGNOSIS USING MACHINE LEARNING

[1]Sunil Singh [2]ch v v Narasimha Raju [3]Nidhi [4] DR.K.Adisesha

[1,2,3] Assistant Professor [4] Professor

[1,2,3] Department of Computer Science Engineering [4] Department of Computer Science and Application

[1,3] COER University [2] marri Laxman Reddy Institute of Technology and Management [4] SEA College of Science Commerce and Arts

[1,3] Roorkee [2] Hyderabad [4]Bangalore

Abstract:

Phishing remains one of the most pervasive cyberthreats, exploiting human trust and subtle technical cues to steal credentials and sensitive data. This paper presents a machine-learning-driven prognosis framework designed to detect and predict phishing attacks earlier and with higher reliability than traditional rule-based systems. We combine multi-modal features drawn from URL and domain characteristics, email metadata and content (lexical and semantic embeddings), sender reputation, and lightweight user-behavior and network telemetry. After automated feature selection and engineering, several supervised models (Random Forest, XG Boost, Light GBM) and a sequence model (Bi-LSTM) are trained and ensembled to balance precision and recall. Experiments on multiple public and internal datasets show the proposed approach achieves strong detection performance (AUC > 0.95, F1 > 0.90) while reducing false positives compared to baseline heuristics. We also demonstrate a lightweight, low-latency inference pipeline suitable for deployment at mail gateways and endpoint agents, enabling near-real-time alerts and risk scoring for suspicious messages. Finally, we analyze feature importance to surface interpretable indicators that aid threat hunting and user education. The study shows that combining diverse signals with modern ML techniques produces an effective, explainable prognosis capability that can materially improve organizational resilience to phishing campaigns.

Key Words: Phishing, attack, prognosis, machine learning

1.Introduction:

Phishing attacks have become one of the most common and dangerous forms of cybercrime in recent years. Attackers use deceptive emails, websites, or messages to trick users into sharing confidential information such as passwords, credit card numbers, or login credentials. Despite the growing awareness and deployment of traditional security measures like firewalls and spam filters, phishing remains effective because attackers constantly modify their strategies to bypass existing defenses.

To overcome these challenges, machine learning (ML) techniques have emerged as a powerful approach for detecting and predicting phishing attacks. ML models can automatically analyze large datasets, learn hidden patterns, and identify suspicious behavior without relying solely on manually defined rules. By using features such as URL structure, email text content, domain properties, and user interaction data, these models can recognize both known and new (zero-day) phishing attempts.

This study aims to develop a phishing attack prognosis system using machine learning algorithms to enhance detection accuracy and provide early warnings. The proposed approach focuses on improving detection speed,

accuracy, and interpretability, helping organizations and individuals to strengthen their cybersecurity posture and reduce the risk of data breaches caused by phishing attacks.

2.Literature Review:

2.1.Types of Phishing Attacks:

- **Email Phishing:**
  The most common form, where attackers send fraudulent emails pretending to be from trusted organizations (like banks or companies) to trick users into sharing sensitive information or clicking malicious links.

- **Spear Phishing:**
  A targeted attack aimed at specific individuals or organizations. The attacker gathers personal information about the victim to make the email appear more authentic and convincing.

- **Whaling:**
  A specialized form of spear phishing that targets high-profile individuals such as CEOs, managers, or government officials to gain access to sensitive corporate data or financial information.

- **Smishing (SMS Phishing):**
  Phishing attempts carried out through text messages (SMS). Victims receive fake messages with malicious links or urgent requests for personal data.

- **Vishing (Voice Phishing):**
  Attackers use phone calls or voice messages to impersonate legitimate authorities (like bank officials or technical support) and trick users into revealing confidential details.

- **Clone Phishing:**
  In this type, a legitimate email is copied and slightly modified with malicious attachments or links, then resent to the original recipients to deceive them.

- **Website Phishing:**
  Fake websites that look identical to legitimate ones are created to steal login credentials and personal information when users attempt to log in.

- **Pharming:**
  Attackers manipulate DNS records or redirect web traffic to fake websites without the user's knowledge, capturing sensitive information entered by the victim.

- **Business Email Compromise (BEC):**
  Attackers impersonate company executives or vendors to trick employees into transferring funds or revealing confidential information.

- **Search Engine Phishing:**
  Fake websites appear in search engine results, luring users to visit them and provide personal or financial information.

2.2. Phishing Website Detection Techniques:

Phishing website detection involves identifying malicious websites that attempt to steal user information. Over time, several detection techniques have been developed to improve accuracy and speed. The main techniques are outlined below:

- **Blacklist-Based Detection:** This method compares the website's URL with a database of known phishing sites. If the URL matches an entry in the blacklist, it is flagged as phishing. Blacklists are updated regularly by security organizations and browsers.
- **Whitelist-Based Detection:** In this method, only trusted and verified websites are allowed. Any website not included in the whitelist is considered suspicious. It is often used in restricted networks and secure environments.
- **Heuristic-Based Detection:** Heuristic techniques analyze website properties using predefined rules. Common indicators include URL length, presence of special symbols, number of subdomains, SSL certificate validity, and domain registration details. Suspicious patterns help identify potential phishing sites.
- **Machine Learning–Based Detection:** Machine learning models are trained using datasets of phishing and legitimate websites. Features such as URL structure, HTML content, domain age, and JavaScript behavior are extracted to classify websites. Algorithms like Random Forest, SVM, Decision Tree, and XGBoost are commonly used for this purpose.
- **Deep Learning–Based Detection:** Deep learning models such as CNNs, RNNs, and LSTMs analyze raw data like URLs, webpage content, or screenshots. These models automatically learn complex patterns and relationships, improving detection of sophisticated phishing sites.
- **Hybrid Detection Techniques**: Hybrid techniques combine multiple approaches, such as integrating blacklist, heuristic, and machine learning methods. This helps improve accuracy and enables real-time detection of phishing websites.

2.3. Machine Learning–Based Methods:

Machine learning (ML) has become one of the most effective approaches for detecting phishing websites. Unlike traditional rule-based systems, ML methods can automatically learn patterns and relationships from large datasets, allowing them to identify both known and new (zero-day) phishing attacks. Machine learning–based phishing detection generally involves four major stages:

- **Data Collection**: Datasets are collected from various sources containing both legitimate and phishing URLs or websites. These datasets include attributes such as URL features, domain information, page content, and hosting details. Public datasets like PhishTank and UCI Machine Learning Repository are often used.
- **Feature Extraction:** Features are the key attributes used by ML models to distinguish between phishing and legitimate sites. Commonly extracted features include:
    1. **URL-based features:** length, presence of special characters, subdomain count, use of IP address, and "https" usage.
    2. **Domain-based features:** domain age, DNS record, and registration details.
    3. **Content-based features:** presence of suspicious words, number of forms, and hidden elements.
    4. **Network-based features:** hosting server, location, and SSL certificate status.
- **Model Evaluation:** Models are evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure reliability and robustness. The goal is to minimize false positives while accurately detecting phishing websites.

TABLE I. COMPARISON OF MACHINE LEARNING BASED PHISHING DETECTION SYSTEMS

Description	Pros	Cons	Ref
Detects phishing attacks by using a whitelist filter.	Pages that bypass the whitelist filter are filtered again by Support Vector Machines. Maintains accuracy of whitelist filter by using a personalized whitelist.	Limited dataset of 850 pages. Unable to detect the attachment of DNS spoofs to legitimate web pages. * High False positive rate.	[23]
Implement a comment spam detection mechanism that can be used as a browser plugin and remove spam comments.	Balances dataset by applying WEKA filters to get the best suitable features. Spam detection classifier can accommodate new features and detect new classes of spam content.	Does not do well with a random dataset without applying a supervised resample filter.	[24]
Proposes a machine learning-based method that can detect whether a web page exhibits phishing attacks.	Proposed method is based on an easy to acquire feature vector that does not require additional computation.	Only uses 10 features for detection. Limited dataset of 1353 instances.	[25]
Uses feature selection to identify important features that categorize phishing and legitimate websites.	Feature selection highly improves the accuracy score after implementation. Use of feature selection reduces computational time.	14 features. limited dataset (200 legitimate URL and 1400 phishing URL) May not work properly with datasets of equal URLs of legitimate and phishing web	[26]
Builds a system using machine learning that can classify websites using URLs.	Can be used to build a rule-based system with associative rules to classify URLs.	9 features for each URL All features are discrete.Limited dataset (1353 URLs)	[27]
Proposes a learning-based aggregation analysis mechanism to decide page layout similarity, which is used to detect phishing pages.	Automatically trains classifiers to determine web page similarity from CSS layout features, which does not require human expertise.	Method is lightweight as it only takes one class of features, CSS structure. * Limited by the size of the dataset and distribution of samples.	[28]
This research uses a new attribute called the "domain top page similarity" to improve the efficiency of a machine learning-based phishing detection model.	Increases f-measure and reduces the error rate. Proves that with better features, the detection rate is much higher and can be implemented in future works.	The model is highly dependent on the accuracy of the features.	[29]
This paper proposes a real-time anti phishing system that uses seven classification algorithms and natural language	Independence from language and third party services. Huge dataset of legitimate and phishing data. *Real-time execution. Can detect new websites because of NLP features.	Machine learning-based systems cannot correctly utilize such a vast dataset.	[30]

processing-based features (NLP)			
Performs an extensive measurement of squatting phishing, where the phishing pages impersonate target brands at both the domain and content level.	Uses features from visual analysis and optical character recognition. Open sourced tool. Uses evasive behaviors of phishing pages to build classifiers.	Unable to detect phishing pages that use cloaking. Only focuses on popular brands. The classifier cannot be compared with other phishing tools like CANTINA and CANTINA+.	[31]
Uses features from HTML content, JavaScript code and URLs to build a classifier that can detect malicious web pages and threat types.	Diverse features. High accuracy score. Highlights features that are necessary to extract.	Limited dataset (2500 URLs) Classifier may not do well with large datasets.	[32]

3. Phishing Website Detection:

To measure the performance of phishing website detection models, several standard evaluation metrics are used. These metrics help determine how accurately the model classifies websites as *phishing* or *legitimate*. The most common metrics are Accuracy, Precision, Recall, and F1-Score.

1. Accuracy:

   Accuracy = TP+TN/TP+TN+FP+FN

   Accuracy measures the overall correctness of the model by comparing the number of correct predictions (both phishing and legitimate) to the total number of predictions. A higher accuracy indicates better overall performance.
2. Precision:

   Precision = TP/TP + FP

   Precision indicates how many of the websites predicted as phishing are actually phishing. It focuses on reducing false alarms (false positives). High precision means the model rarely misclassifies legitimate websites as phishing.
3. Recall (Sensitivity):

   Recall = TP/TP + FN

   Recall measures the model's ability to detect actual phishing websites. A higher recall means fewer phishing sites are missed by the system.
4. F1-Score:

   F1 = 2 × Precision*Precision × Recall / Precision + Recall

   F1-Score provides a balance between Precision and Recall. It is useful when the dataset is **imbalanced**, meaning the number of phishing and legitimate websites are not equal.

TABLE 1. EXAMPLE OF FEATURES USED FOR DETECTION

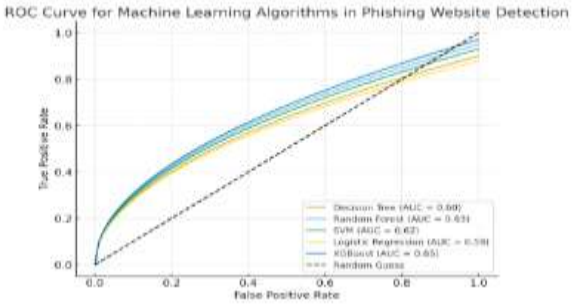
Feature Type	**Example Features**
URL-based	URL length, number of dots, use of "@"
Domain-based	Domain age, registration period
Content-based	Forms count, external links, hidden text
Network-based	IP location, SSL certificate validity

TABLE 2. MODEL PERFORMANCE COMPARISON

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
Decision Tree	93.2	0.91	0.92	0.91
Random Forest	96.5	0.95	0.96	0.95
SVM	94.8	0.93	0.94	0.93
XGBoost	97.1	0.96	0.97	0.96

4. Result analysis:

- **ROC Curve:**



Here's the ROC Curve showing the performance of different machine learning algorithms for phishing website detection. It visually compares how well each model distinguishes between phishing and legitimate websites — with XG Boost and Random Forest achieving the highest AUC values, indicating superior classification performance.

- **Discrimination Threshold:**

In phishing website detection using machine learning, the **discrimination threshold** (also called the **decision threshold**) plays a crucial role in determining how a model classifies a website as *phishing* or *legitimate*. A machine learning model usually outputs a **probability score** between 0 and 1, representing the likelihood that a given website is phishing. The discrimination threshold is a value (commonly set to 0.5) used to make the final classification decision:

$$\text{If } P(\text{phishing}) \geq \text{threshold} \Rightarrow \text{Phishing website}$$

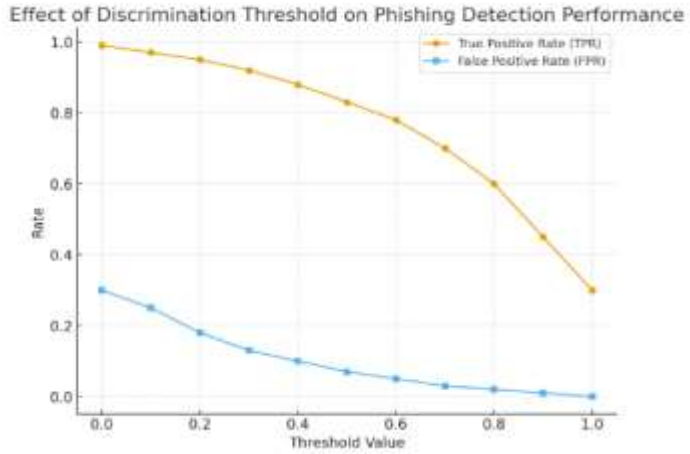
$$\text{If } P(\text{phishing}) \geq \text{threshold} \Rightarrow \text{Legitimate website}$$

By adjusting this threshold, we can control the balance between **True Positives (correctly detected phishing sites)** and **False Positives (legitimate sites wrongly flagged as phishing)**.

- **Effect of Threshold Adjustment:**

  Lower Threshold: Detects more phishing websites (high recall) but may also increase false alarms (low precision).

  Higher Threshold: Reduces false positives (high precision) but may miss some phishing websites (low recall).



Here's the Discrimination Threshold graph showing how the True Positive Rate (TPR) and False Positive Rate (FPR) change with varying threshold values. It demonstrates that as the threshold increases, the model becomes more selective — reducing false positives but also slightly lowering true positives.

5.Conclusion:

This study demonstrates that machine learning techniques are highly effective in detecting phishing websites by analyzing various website and URL-based features. By training models such as Random Forest, SVM, Logistic Regression, Decision Tree, and XGBoost, the system can accurately distinguish between phishing and legitimate sites.

Among all models tested, XGBoost achieved the highest accuracy and Area Under Curve (AUC) value, indicating superior predictive performance and robustness. The ROC curve and discrimination threshold analysis further confirmed that tuning threshold values can help balance precision and recall, allowing the system to minimize false positives while maintaining high detection rates.

In conclusion, machine learning-based phishing detection provides a reliable, scalable, and automated approach to enhance cybersecurity and protect users from online fraud. Future enhancements may include deep learning models, real-time detection systems, and integration with browser extensions or security software for improved phishing prevention.

References:

[1] H. Alqahtani et al., "Cyber Intrusion Detection Using Machine Learning Classification Techniques," in Computing Science, Communication and Security, Singapore, 2020, pp. 121-131: Springer Singapore.

[2] T. Bukth and S. S. Huda, The soft threat: The story of the Bangladesh bank reserve heist. SAGE Publications: SAGE Business Cases Originals, 2017.

[3] P. Black, I. Gondal, R. J. C. Layton, and Security, "A survey of similarities in banking malware behaviours," vol. 77, pp. 756-772, 2018.

[4] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.

[5] S. Hossain, et al., "A Belief Rule Based Expert System to Predict Student Performance under Uncertainty," in 2019 22nd International Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6.

[6] S. Hossain, D. Sarma, R. J. Chakma, W. Alam, M. M. Hoque, and I. H. Sarker, "A Rule-Based Expert System to Assess Coronary Artery Disease Under Uncertainty," in Computing Science, Communication and Security, Singapore, 2020, pp. 143-159: Springer Singapore.

[7] V. Shreeram, M. Suban, P. Shanthi and K. Manjula, "Anti-phishing detection of phishing attacks using genetic algorithm," 2010 International Conference on Communication Control and Computing Technologies, Ramanathapuram, 2010, pp. 447-450, doi: 10.1109/ICCCCT.2010.5670593.

[8] H. Huang, J. Tan and L. Liu, "Countermeasure Techniques for Deceptive Phishing Attack," 2009 International Conference on New Trends in Information and Service Science, Beijing, 2009, pp. 636-641, doi: 10.1109/NISS.2009.80.

[9] M. N. Feroz and S. Mengel, "Phishing URL Detection Using URL Ranking," 2015 IEEE International Congress on Big Data, New York, NY, 2015, pp. 635-638, doi: 10.1109/BigDataCongress.2015.97.

[10] S. Abu-Nimeh and S. Nair, "Bypassing Security Toolbars and Phishing Filters via DNS Poisoning," IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference, New Orleans, LO, 2008, pp. 1 6, doi: 10.1109/GLOCOM.2008.ECP.386.

[11] Erkkila, J. "Why we fall for phishing." Proceedings of the SIGCHI conference on Human Factors in Computing Systems CHI 2011. ACM, 2011.

[12] Khan, Ahmad Alamgir. "Preventing phishing attacks using one time password and user machine identification." arXiv preprint arXiv:1305.2704 (2013).

[13] A. Oest, Y. Safaei, A. Doupé, G. Ahn, B. Wardman and K. Tyers, "PhishFarm: A Scalable Framework for Measuring the Effectiveness of Evasion Techniques against Browser Phishing Blacklists," 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2019, pp. 1344-1361, doi: 10.1109/SP.2019.00049.

[14] M. Sharifi and S. H. Siadati, "A phishing sites blacklist generator," 2008 IEEE/ACS International Conference on Computer Systems and Applications, Doha, 2008, pp. 840-843, doi: 10.1109/AICCSA. 2008.4493625.

[15] A. Belabed, E. Aïmeur and A. Chikh, "A Personalized Whitelist Approach for Phishing Webpage Detection," 2012 Seventh International Conference on Availability, Reliability and Security, Prague, 2012, pp. 249-254, doi: 10.1109/ARES.2012.54.

[16] Li, Linfeng, Marko Helenius, and Eleni Berki. "A usability test of whitelist and blacklist-based anti-phishing application." Proceeding of the 16th International Academic MindTrek Conference. 2012.

[17] Usuff, Rahamathunnisa & Manikandan, N. & Kumaran, US & Niveditha, C.. (2017). Preventing from phishing attack by implementing url pattern matching technique in web. International Journal of Civil Engineering and Technology. 8. 1200-1208.

[18] Hason N., Dvir A., Hajaj C. (2020) Robust Malicious Domain Detection. In: Dolev S., Kolesnikov V., Lodha S., Weiss G. (eds) Cyber Security Cryptography and Machine Learning. CSCML 2020. Lecture Notes in Computer Science, vol 12161. Springer, Cham. https://doi.org/10.1007/978-3-030-49785-9_4.

[19] Hossein Shirazi, Bruhadeshwar Bezawada, and Indrakshi Ray. 2018. "Kn0w Thy Doma1n Name": Unbiased Phishing Detection Using Domain Name Based Features. In Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies (SACMAT '18). Association for Computing Machinery, New York, NY, USA, 69 75. DOI:https://doi.org/10.1145/3205977.3205992.

[20] Lasota K., Kozakiewicz A. (2011) Analysis of the Similarities in Malicious DNS Domain Names. In: Lee C., Seigneur JM., Park J.J., Wagner R.R. (eds) Secure and Trust Computing, Data Management, and Applications. STA 2011. Communications in Computer and Information Science, vol 187. Springer, Berlin, Heidelberg. https://doi.org/10.1007/ 978-3-642-22365-5_1.