International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 04 | Apr 2025 www.irjet.net p-ISSN: 2395-0072

Detecting Phishing URLs Using Machine Learning

Dr. G. L. Lakshmi [1], K Jishitha [2], N Srujana Sree [3], M Naga Poojitha [4], B Siva Naga Lakshmi Neeharika[5],

[1,2,3,4,5]*Dept of AI&DS, VVIT, Andhra Pradesh, India*

Abstract - *Phishing sites are now a serious security threat to online security. The origin of the majority of cyberattacks that compromise business and customer data confidentiality, integrity, and availability is phishing. Various automatic phishing site detection methods have been generated by numerous research studies across several decades. Emerging solutions enhance performance but require substantial manual feature engineering and fail to detect new, emerging phishing attacks well. This area requires solutions to automatically detect phishing sites at speed to combat zero-day phishing attacks since detecting such methods is an ongoing research topic. The site at the provided URL provides multiple data sources that enable the assessment of the possible maliciousness of servers. Machine learning is the perfect method for detecting phishing activity. The method does away with all the issues associated with the earlier approach.*

Keywords: **Phishing websites, cyberattacks, machine learning**

1. INTRODUCTION

Individuals and organizations around the globe are exposed to significant risks from phishing attacks in the cyber world. Increased use of banking websites, education websites, and social media websites now provides cybercriminals with new means of tricking users for personal information. Attackers construct misleading fake e-mails or messages with fictitious Trusted Source identities to redirect users to fraud sites where phishers obtain system access and financial information while adding malicious software. Cyberattacks through secure websites have become more complex, and psychological manipulation strategies have enabled attackers to avoid traditional filters and blocklists through sophisticated deceptive methods. As a result, machine learning-based phishing detection has become a necessary security measure against online scams, especially with hackers being adept at evading the old security systems. Phishing attacks are growing by leaps and bounds across every sector all over the world, impacting businesses, entrepreneurs, and even governments. Cybercriminals continuously evolve their strategies to pilfer sensitive data, exploiting mediums like emails, social media, calls, and SMS. As more and more people are getting habituated to virtual transactions and e-communication, email phishing, smishing, vishing, and social networking scams are growing in scale. Organizations and individuals must stay alert, authenticate sources of information before release, and adopt sound cybersecurity to prevent falling prey to such attacks.

Fig 1 Increase of Phishing Attacks

Note: The blue histogram represents real threat detection rates. They go up along with simulated threat report rates.

1.1 Types of Phishing Attacks

Phishing attacks are turning into a rising cybersecurity menace since the attackers use so many tricks to deceive people and companies. The following are some of the most common types:

- **Email Phishing**: Email phishing is the most common type of phishing, in which cybercriminals send fake emails that seem to be from organizations that are believed to be trustworthy. They sometimes register fake domains that are similar to the genuine ones by replacing small characters (e.g., "m" with "rn"). Sometimes, they also add the company name to the domain to make it look real.

- **Spear Phishing**: Spear phishing is a very targeted type of phishing, as opposed to mass email phishing. Attackers research specific victims thoroughly, collecting information such as their name, job title, email address, and company. This allows them to craft very realistic emails because

they know enough about the particular victim to make sure that the information is correct.

- **Whaling:** Whaling is a more evolved form of spear phishing, specifically targeting senior managers or key decision-makers within an organization. Because these individuals are more cautious, spammers will not employ glaring tricks like fictitious links. Instead, they design e-mails that mimic genuine requests by other senior officers, usually requiring immediate financial wire transfers or sensitive data.

- **Smishing & Vishing:** These scams use text messages (smishing) or phone calls (vishing) instead of email. The attackers tend to impersonate banks, government institutions, or customer care representatives to try to deceive victims into sharing personal information or sending money. For example, they may send a notification with a message stating suspicious activities in your bank account, asking you to click on a malicious link or call them directly.

- **Angler Phishing:** Phishing on social media, in which cybercriminals use imposter profiles, cloned posts, and direct messages to trick users. They may impersonate customer support teams or trick users into clicking on malicious links, causing data theft or malware installation.

1.2 Phishing Architecture

A phishing attack usually follows a systematic procedure. The attacker first sends a false email to the victim, making it appear to be from a reputable organization such as a bank or service provider. The email is usually filled with lots of urgent details, pressuring the victim into responding promptly by clicking on a link that has been provided. After the victim clicks on the link, the victim is taken to a phishing website that has been created to resemble an original one. They unwittingly input sensitive login information like usernames, passwords, or monetary details. The attacker intercepts these login credentials, storing them for malicious intent.

Finally, the attacker exploits the hijacked credentials to access unauthorized legitimate accounts, which may result in identity theft, financial fraud, or other forms of cybercrime. This step emphasizes the significance of understanding phishing and cybersecurity to avoid becoming a victim of such attacks.

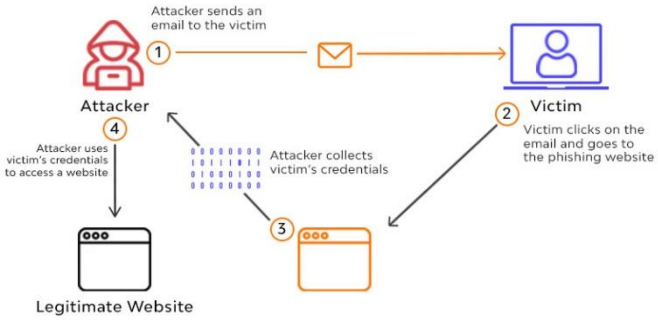

Fig 1.1.1 Architecture

2. Related Work

The Detection of phishing has been a research-oriented field in the domain of cybersecurity, and several methods have been proposed to reduce the threat. The most widely discussed methods in the literature are blacklist-based detection, heuristic-based detection, machine learning-based detection, and deep learning-based approaches.

- **Blacklist-Based Detection:** The traditional detection systems for phishing are based on maintaining blacklists of known phishing URLs, like Google Safe Browsing and PhishTank databases. When a user attempts to log into a website, the URL is matched with these blacklists. If it is present in the database, then the user is warned. However, blacklist-based methods do not identify zero-day phishing attacks that well because criminals continue to register new domains and URLs in order to go unnoticed. Ma et al. (2009) stated that blacklists are very reactive and can't react to new phishing methods in real time.

- **Heuristic-Based Detection:** Heuristic-based methods are attempting to overcome the weakness on which blacklists suffer by conducting a scan of web characteristics such as URL generation, domain name characteristics, SSL certificates, and page content. Zhang et al. (2013) have suggested a heuristic system taking into account URL length, use of special characters, and sites and subdomains used in phishing for detecting phishing. Heuristics, though capable of detecting new phishing, suffer from high false positives and variable accuracy on different datasets.

- **Machine Learning-Based Detection:** Machine learning was an efficient method of phishing attack detection based on pattern learning of URL features. A supervised model of learning suggested by Mohammad et al. (2014) extracted lexical, host-based, and network-based attributes in an attempt to train classifiers such as Decision Trees, Random

Forest, and Support Vector Machines (SVM). Verma and Dyer (2015) also established research in a setting where ensemble learning techniques could be applied towards enhanced phishing detection accuracy using more than one classifier. These approaches are more precise than blacklists and heuristics but are dependent on feature engineering quality.

- **Deep Learning Approach:** Recent advancements have explored deep learning models such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Transformer-based models for phishing detection. Bahnsen et al. (2017) proposed a Recurrent Neural Network (RNN)-based model to analyse sequential patterns in URLs with higher classification accuracy. Deep learning models, nonetheless, require vast quantities of label training data and vast quantities of computational horsepower, rendering them unsuitable for real-time applications.

- **Hybrid and Stacking Ensemble Models:** In order to leverage the potential of different classifiers, researchers applied stack ensemble methods, where multiple machine learning algorithms are invoked together and multiple machine learning algorithms are listed together to provide better performance. Random Forest, XGBoost, and Multilayer Perceptron (MLP) have been employed here as the base classifiers and Logistic Regression as the meta-classifier. Similar ensemble strategies have also been investigated by Sahingoz et al. (2019), where stacked classifiers were discovered to be better compared to one model in the context that they decrease false positives but improve detection performance. Past studies revealed the limitations of traditional approaches like blacklists and heuristics, while machine learning and deep learning greatly improved phishing detection. Even with this, there is a problem of detecting new phishing methods and keeping the models updated in real-time defence. This paper continues the existing research with a hybrid machine learning model for enhancing accuracy and immunity to phishing attacks.

3. Methodology

3.1 User Interface

The User Interface (UI) serves as the primary interaction point for users, presenting diverse career options and key platform features in an intuitive and visually engaging manner. Developed using HTML, CSS, and JavaScript, the interface ensures seamless navigation through carousels, search bars, and interactive elements. The design focuses on enhancing usability and user satisfaction by providing easy access to personalized career recommendations and real-time insights.

Fig 3.1.1 Input Slide 1

The phishing detector interface possesses a friendly, intuitive interface that has been made easy and handy. As indicated in the above figure, the interface has a middle input field where users can input a URL to scan for possible phishing attacks. A "Scan URL" button is provided to begin the scan. One user should be able to scan any number of URLs.

Fig 3.1.2 Input Slide 2

In addition, the system incorporates an awareness section that educates users on phishing prevention. The system recognizes key security habits such as being cautious of suspicious emails, using strong passwords, enabling Two-Factor Authentication (2FA), verifying URLs, and staying up-to-date on phishing scams. This blend of real-time URL scanning and educational support ensures security while enhancing user awareness. It equips users with information, making them more alert and helping toward overall cybersecurity knowledge and security.

3.2 Feature Extraction

Feature extraction is a key operation in phishing detection since it facilitates structural and lexical analysis

of URLs to classify them as malicious or benign. The feature extraction module provided in this work verifies if URLs are well-formed, validated, and processed before extracting the meaningful traits.

The function then verifies if the given URL contains a valid scheme (i.e., http:// or https://) and appends "http://" if not.

This is performed since URLs can be passed without the scheme by some users. It subsequently uses the 'urlparse' function of Python's 'urllib.parse' module to parse different parts of the URL. If the parsed URL does not have a valid domain, the URL is considered invalid and is skipped.

Once validated, the function retrieves some of the lexical and structural features essential in identifying phishing:

- Hostname Length & URL Length: Longer because the attacker inserts deceptive words into it to mimic reputable domains.

- First Directory Length: Passes over the first directory's length in the URL path since phishing URLs contain deceptive directory names.

-Special Character Count: It counts the occurrence of special characters like '@', '-', '_', and '.' since the phishing links are often made of such characters with the intention to hide their character.

-Digit and Letter Count: It pulls out digits and letters that occur in the URL because the phishing links contain an unnecessary number of numbers.

-Directories Count: Examines the structure of the URL by identifying the number of directories, as there might be unnecessary directories in phishing URLs in a bid to look legitimate.

- IP Address Detection: Checks whether the URL contains an IP address instead of a domain name since phishing sites use naked IPs in an effort to evade detection.

The features extracted are passed back as a list to be analysed further. When there are errors in processing a URL, they are handled elegantly, and the function continues beyond the faulty URL. These features are crucial inputs to machine learning algorithms that enable URLs to be labeled as phishing or benign based on patterns learned.

3.3 Machine Learning Models for Phishing Detection

Phishing detection has evolved beyond traditional methods with the emergence of machine learning models, enabling a proactive approach to identifying malicious URLs. Unlike conventional techniques, such as blacklist-based detection, which depend on pre-compiled lists of known phishing sites, machine learning models analyze URL features and effectively detect previously unknown phishing threats. This study employs the stacking ensemble technique where Random Forest, XGBoost, and Multilayer Perceptron (MLP) serve as base classifiers and Logistic Regression as the meta-classifier to produce the output prediction.

• Random Forest for Phishing Detection

Random Forest is a type of ensemble learning where a large number of decision trees are built, and the majority votes for most of their predictions to classify better. Each decision tree in the forest is trained on a random subset of the data, and the prediction of the forest is made by majority vote. Random Forest is extremely useful when applied to phishing detection because:

1.It minimizes overfitting, generalizing better.

2.It can handle continuous and categorical features, hence appropriate for URL structure examination.

3.It incorporates feature importance analysis, which proves useful when performing the most meaningful phishing indicator discovery, e.g., URL length, special characters, and usage of HTTPS.

• XGBoost for Greater Efficiency:

XGBoost (Extreme Gradient Boosting) is a boosting algorithm that follows the idea of sequentially building weak learners in which each new tree learns from the errors of the past ones. XGBoost is used extensively in phishing detection due to the following reasons:

1.It is more efficient and quicker than regular boosting algorithms.

2.It uses regularization methods (L1 and L2) to avoid overfitting.

3.It performs quite well with imbalanced data, one of the common issues for phishing detection

• Multilayer Perceptron (MLP) for Non-Linear Patterns:

Multilayer Perceptron (MLP) is a feedforward artificial neural network (ANN) capable of detecting subtle, non-linear relationships between features at the URL level. MLP is formed with multiple layers, including:

1. An input layer, receiving the extracted feature parameters like length of the URL, hostname construction, and occurrences of digits.

2. Latent layers, on which neurons impose activation functions like ReLU to acquire a more sophisticated representation of phishing properties.

3. An output layer, which shall determine if the URL is malicious or not.

4. MLP is useful in phishing classification because it is capable of identifying patterns that are undetected in conventional models, e.g., fine text adjustment in domain names.

• Stacking Ensemble using Logistic Regression:

The output of Random Forest, XGBoost, and MLP is merged into a stacking ensemble model that uses Logistic Regression as the meta-classifier for increased detection precision. This hybrid model:

1. It harnesses the power of multiple classifiers to minimize false positives and false negatives.

2. Combines decision-tree-based (Random Forest, XGBoost) and deep learning (MLP) for enhanced phishing classification.

3. Uses Logistic Regression to generate advanced predictions and make the system more adaptive and trustworthy.

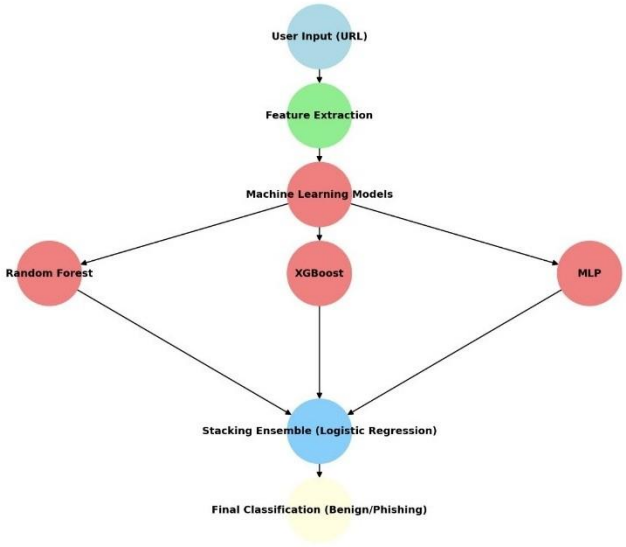

Fig 3.3.1 Model Diagram

3.4 Model Performance

The stacking ensemble method analyzed three base learners (Random Forest, XGBoost, and MLP) using Logistic Regression as the meta-classifier to achieve performance evaluation. The detection system shows clear evidence of reliable performance capabilities based on its tested results.

The accuracy of the model reaches 99.71% because it successfully classifies 99.71% of URLs as phishing or benign. This percentage represents both correct phishing and benign identifications. The model shows excellent precision regarding its ability to distinguish phishing sites from normal sites. The F1-score achieves 0.9939 because it calculates precision and recall as measures in harmonic mean to maintain balanced detection of false positives and false negatives.

When the measure approaches the value 1, it indicates low classification errors with both high precision and recall. The outstanding performance of this model proves its ability to resist phishing attacks, which makes it work as a dependable real-time detection system for phishing connections.

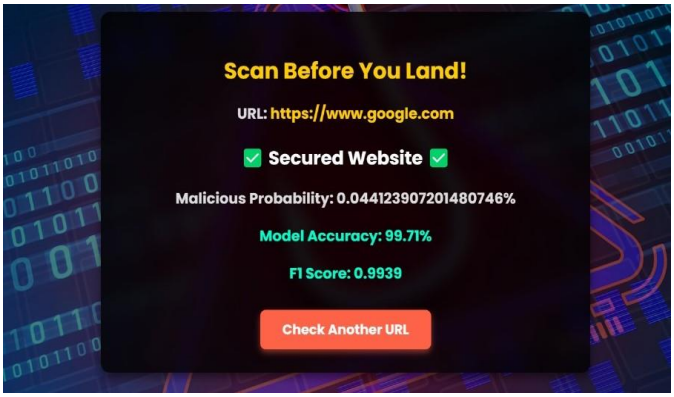

Fig 3.4.1 Output of secured website

The model successfully identifies a secure (benign) website, as shown in above Figure , based on its predicted probability.

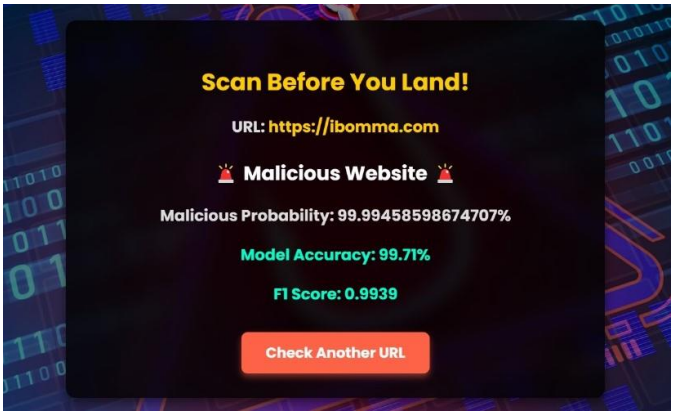

Fig 3.4.2 Output of Malicious Website

The model successfully identifies a Malicious website, as shown in the Figure, based on its predicted probability.

4. Conclusion

Attackers target victims through deceptive websites that imitate genuine services to conduct phishing attacks as a leading cybersecurity threat. The blacklisting detection technique, among other traditional methods, cannot detect newly surfacing phishing URLs, making more intelligent and flexible solutions necessary.

The solution proposed constructs a machine learning-based phishing detection system based on Random Forest and XGBoost as base classifiers, along with a Multilayer Perceptron (MLP) with incorporated Logistic Regression. The system developed utilizes protocol types, special characters, domain attributes, and redirections to differentiate phishing sites from actual sites.

The model developed shows a high accuracy rate of 99.71% with an F1-Score value of 0.9939, which is more efficient than the traditional methods of detecting phishing. The performance rate confirms that stacked ensemble methods make the system highly efficient in avoiding false detection errors. This work introduces an independent system providing real-time phishing protection capabilities and flexibility in operations. The system is constantly tested in trying to identify advanced phishing attacks because they have a very similar appearance to genuine websites.

Advances in research toward real-time deployment seek to integrate deep learning techniques like LSTMs and Transformers and adaptive techniques to counter new phishing methods that are emerging. This type of model, when deployed in security software or browser add-ons, provides improved protection against phishing using powerful and accurate detection mechanisms.

5. References

1. Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs. Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 1245-1254.

2. Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). A Self-Structuring Neural Network technology serves as the foundation of Phishing Website prediction systems. Neural Computing and Applications, 25(2), 443-458.

3. Verma, R., & Dyer, K. (2015). On the Character of Phishing URLs: Accurate and Robust Statistical Learning Classifiers. ACM Transactions on Information and System Security (TISSEC), 17(3), 1-32.

4. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine Learning-Based Phishing Detection from URLs. Expert Systems with Applications, 117, 345-357.

5. Bahnsen, A. C., Torroledo, D., Camacho, D., & Villegas, S. (2017). A study by Classifying Phishing URLs Using Recurrent Neural Networks was presented at the 13th IEEE International Conference on Machine Learning and Applications (ICMLA) during 2019. It appeared on pages 928 to 934.

6. Zhang, J., Yuan, W., & Zhang, X. (2013). An innovative Heuristic-Based Phishing Detection Approach analyses URL features as its basis for strategy development. International Conference on Computational Science and Engineering (CSE), 763-769.

7. Chiew, K. L., Chang, E. H., Sze, S. N., & Tiong, W. K. (2019). A machine learning analysis of URL-based features serves to detect phishing websites. Computers & Security, 85, 256-267.

8. Marchal, S., Saari, K., Singh, N., & Asokan, N. (2016). Know Your Phish: Novel Machine Learning-Based Phishing Detection is presented as a publication at the International Conference on Security and Privacy in Communication Systems (SecureComm) during 293-311.

9. Liu, W., Zhang, Y., Deng, X., & Lee, W. (2018). A method for phishing website detection that uses URL characteristics as examination criteria. Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM), 1059-1068.

10. Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., & Nunamaker, J. F. (2010). This paper examines how statistical learning theory assists in detecting false websites. MIS Quarterly, 34(3), 435-461.