Detection of Phishing Using Machine Learning algorithms

# **G. Sanjay Kumar . Gautham Krishna . Gaurav R . Meshach A Martin**

# School of Computer Science and Engineering

# REVA University, Bengaluru, India.

Received: 05 May 2022 / Revised: 29 May 2022 / Accepted: 29 July 2022

©Milestone Research Publications, Part of CLOCKSS archiving

**Abstract —** This research paper explores the application of machine learning algorithms for the detection of phishing websites. Phishing is a prevalent cyber-attack technique that aims to obtain sensitive information, such as login credentials and financial details, by tricking users into accessing fraudulent websites. The proposed methodology involves extracting features from the website's source code and using these features to train a machine learning classifier. The study evaluates the performance of different machine learning algorithms, including logistic regression, decision trees, and random forests, using a dataset of known phishing and legitimate websites. The results indicate that the random forest algorithm outperforms the other algorithms, achieving an accuracy of 97% and a precision of 98%. The study demonstrates the potential of machine learning techniques for the early detection of phishing attacks and highlights the importance of continuously updating the training dataset to improve the classifier's performance.

**Keywords –** Personal computer (PC), URL, phishing, machine learning

**----------**

1. INTRODUCTION

Phishing is a scam that employs social and technological tactics to obtain a customer's personal information and financial information. Spoofed e-mails from reputable corporations and agencies are used by social media platforms to allow consumers to access bogus websites and reveal financial information such as usernames and passwords. Hackers often employ systems to intercept usernames and passwords of consumers' internet accounts, installing malicious software on machines to obtain credentials. Phishers steal user information through a variety of ways, including email, Uniform Resource Locators (URL), instant chats, forum postings, phone calls, and text messages. Phishing content has a similar structure to legitimate content, and it entices consumers to browse it in order to obtain sensitive information. Phishing's main goal is to obtain personal information in order to make money or commit identity theft. Phishing attacks are wreaking havoc on businesses all around the world. According to the Anti-Phishing Working Group's (APWG) recent Phishing pattern studies, the majority of phishing assaults target financial/payment institutions and webmail.

Criminals create illegal reproductions of legitimate websites and emails, usually from financial institutions or other organizations that deal with financial data, in order to obtain confidential information. The logos and slogans used in this e-mail are from a reputable company. HTML's architecture and structure enable picture or website copying. It also permits the misuse of brands, trademarks, and other firm identities that customers rely on as authentication mechanisms. Phisher sends "spooled" emails to a large number of people in order to catch users off guard. When users open these e-mails, they are usually directed to a faked website instead of the actual company.

Phishing attacks have become a significant threat to online security, targeting individuals, businesses, and organizations worldwide. Phishing is a type of social engineering attack in which the attacker impersonates a trustworthy entity to obtain sensitive information from the victim, such as login credentials, financial details, and personal information. The attacker typically achieves this by creating a fraudulent website that resembles a legitimate one, such as a bank, e-commerce site, or social media platform.

Traditional methods of detecting phishing websites rely on blacklisting known malicious URLs or IP addresses. However, this approach is limited in its effectiveness, as attackers can quickly change their tactics and use new domains or IP addresses to evade detection. As a result, there is a need for more advanced techniques that can identify and block phishing attacks in real-time.

Machine learning algorithms offer a promising solution to the problem of detecting phishing websites. By analyzing a website's source code and content, machine learning algorithms can identify patterns and features that are characteristic of phishing attacks, such as the presence of malicious scripts or the use of suspicious keywords. These algorithms can then be trained to classify websites as either legitimate or phishing based on these features.

This research paper aims to evaluate the effectiveness of different machine learning algorithms for the detection of phishing websites. We analyze a dataset of known phishing and legitimate websites and extract features from the website's source code to train and evaluate several machine learning classifiers. The study demonstrates the potential of machine learning techniques for the early detection of phishing attacks and highlights the importance of continuously updating the training dataset to improve the classifier's performance.

While machine learning algorithms offer a promising solution to the problem of detecting phishing attacks, there are several challenges that must be addressed to achieve accurate and reliable detection. Some of the key challenges include:

Imbalanced dataset: One of the most significant challenges in phishing detection is the imbalanced distribution of legitimate and phishing websites in the dataset. The number of legitimate websites far outweighs the number of phishing websites, making it difficult for machine learning algorithms to learn and generalize patterns that are characteristic of phishing attacks.

Dynamic nature of phishing attacks: Phishing attacks are constantly evolving, and attackers frequently use new tactics and techniques to evade detection. As a result, machine learning algorithms must be trained on a continuously updated dataset to adapt to these changes.

Feature selection: Extracting relevant features from the website's source code is a critical step in training machine learning classifiers. However, selecting the right set of features can be challenging, and irrelevant or redundant features can negatively impact the classifier's performance.

Overfitting: Machine learning algorithms can sometimes overfit the training data, leading to poor generalization and accuracy on unseen data. This can be mitigated by using regularization techniques or cross-validation to evaluate the model's performance on a separate test set.

Adversarial attacks: Attackers can deliberately manipulate the website's source code to evade detection by machine learning algorithms. This can include obfuscating malicious code, using encrypted communication, or adding decoy content to confuse the classifier.

Addressing these challenges requires a combination of advanced machine learning techniques, careful feature selection, and continuous updates to the training dataset. It is also essential to employ a multi-layered defense strategy that includes other security measures, such as user education and awareness, to complement the machine learning-based detection system.

The detection of phishing attacks using machine learning algorithms has numerous applications in improving online security. Some of the key applications include:

Email phishing detection: Email is one of the most common methods used by attackers to launch phishing attacks. Machine learning algorithms can be trained on email content and metadata to detect phishing emails and block them before they reach the user's inbox.

Website protection: Machine learning algorithms can be deployed on web servers to analyze incoming traffic and identify potential phishing websites. This can prevent users from accessing these sites and protect them from phishing attacks.

Fraud detection: Machine learning algorithms can be used to detect and prevent fraudulent activities, such as financial fraud or identity theft, by analyzing user behavior and transaction patterns.

Social media security: Social media platforms are increasingly targeted by phishing attacks, and machine learning algorithms can be used to detect and block these attacks in real-time.

Mobile security: Phishing attacks are becoming more prevalent on mobile devices, and machine learning algorithms can be used to detect and prevent these attacks by analyzing app behavior and user data.

Overall, the detection of phishing attacks using machine learning algorithms has numerous applications in improving online security and protecting users from cyber threats. As the sophistication of phishing attacks continues to increase, the development and deployment of advanced machine learning-based detection systems will become increasingly important.

1. RELATED WORKS

The intended work lies in the identifying the detection of phishing in the URLs that can help for preventing and also help in the advancement of anti-phishing techniques. A further step can be taken for predicting the patterns of future attacks that are done by the hackers. Some similar thoughts are perceived in the papers referred below.:

Shuichiro Haruta et al., (2017), [1] proposed by emphasizing on the fact that genuine websites are usually linked by many other websites, and those websites are considered real since their screenshots and CSS are saved in a database. Assailants frequently utilise real CSS to simulate the real site because CSS is a file that defines the visual substance of the site. As a consequence, we can both identify the phishing site and its purpose by discovering a site that imitates the appearance or CSS of a valid site. With limitation of whitelisting which includes websites that are connected to at least one other website and are assumed to be legitimate.

Andrew J Park et al., (2017), [2] proposed by creating a web crawler that scrapes the contents of both legal and fraudulent websites. The elements were then evaluated to understand the heuristics rate and commitment scale factor for a site's incorrectness. Data from the web scraper was analysed using a data mining programme, and patterns were discovered. With a limitation of inaccurate model precision. Ankit Kumar et al., (2018), [3] proposed by detection of phishing attack by the help of analysis of hyperlinks of HTML source code of the website. The approach has divided the hyperlink features into 12 different categories to train the models. With limitation of lacking in attention paid to data size.

T. Nathezhtha et al., (2019), [4] proposed a three-phase phishing threat detection method. The three phases of the WC-PAD are as follows: DNS blacklist, Heuristic-based approach, Web crawler-based, which approaches web crawlers that are being used for feature extraction and phishing attack detection. With limitation of taking quite a time due to being three phases, so each webpage must go through all three without skipping any phases. Mohammad Mehdi Yadollahi et al., (2019), [5] proposed a phishing detection system which utilising a classification algorithm classifier termed XCS. It is an online adaptive machine learning approach. This progresses a set of rules known as classifiers. From the source code of a webpage and URLs, this model derives 38 features.

Kelvin S. C. Yong et al., (2019), [6] proposed to describe the QR code weaknesses used by phishers, as well as current QR code phishing attacks. To examine the evolution of QR code phishing attacks and the potential countermeasures: Data correction percentage that is strict. Decide which sort of barcode to verify first. The identified code's border will be shown. The user may now see if an inner barcode is being scanned. If numerous barcodes are detected, notify the user. But there are a number of QR scanners out there that don't have an anti-phishing function and requiring every QR scanner to have a safe QR with security features like a digital signature is impracticable.

Mehmet Korkmaz et al., (2020), [7] proposed to create a phishing detection system that relied on the analysis on the URL of the webpage. Zero URL addresses may be created using various fields like domain, subdomain, Top Level Domain (TLD), protocol, directory, file name, path, and query. These fields in phishing URLs are typically different from those seen on authentic websites. Effective features extracted from the URL improve classification accuracy. Open-source and freely available datasets are used. 58 features were gathered using Python scripts, and 48 of them were chosen using the random forest classifier.

Jaeil Lee et al., (2021), [8] proposed that email service providers must first enhance their security to prevent being attacked by phishing mail attackers. And additional authentication is required for the FTP account. To prohibit remote access to FTP The network must be separated and continually maintain FTP history. A firewall must also be set up to prevent outside access to the mail server. They must carefully maintain their accounts and use a password that matches specific requirements. OTP and SMS should be used for two-factor authentication. But E-mail accounts of common users are often exposed. In the case of remote control using leaked account information or the sending of phishing e-mails it is impossible to instantly check whether the real account has been stolen, and because critical information such as a password is supplied, taking action is extremely difficult. Anwar Basha H et al., (2021), [9] proposed a semantical analysis. Depending on the functions served by each word in the phrase, technique determines whether the statement is a question or a command. Pairs will gather prospective topics for queries and orders. Each pair is then examined to determine if it is part of a malicious pair registry. The software analyses a text file one sentence at a time and returns true if the record contains a social manipulation attack. Establish a blacklist of the subject using machine learning and a decision tree tailored for numerous distributed outcomes. Multinomial NB was also implemented as the Scikit-Learn Python Library. With each single pair, this system provides a pre-setting label that gives a predicted trust rating of trust. The size of the confidence levels is 0 and 1, with a score of 1 representing certainty.

Mahdieh Zabihimayvan et al., (2019), [10] proposed by use Fuzzy Rough Set (FRS) theory to identify the most effective features for detecting phishing websites. Their feature selection technique used in phishing detection classification techniques: multilayer perceptron, random forest, and SMO. To test the efficacy of FRS feature selection in building a generalizable phishing detection, we train each classifier on a distinct out-of-sample dataset. When they compare our findings to three previously developed methods, they can see how useful FRS is for feature selection. They use FRS to select features for phishing attacks from three benchmark data sets. To avoid classification overfitting, they use three well-known classification approaches and run the tests on an unknown data set of 14,000 website samples after selecting the features.

1. OBJECTIVES

A phishing website is a social engineering technique that imitates legitimate uniform resource locators (URLs) and webpages. This project's goal is to use the dataset created to anticipate phishing websites to train machine learning models and deep neural nets. The required URL and website content-based features are extracted from the dataset, which includes both phishing and benign URLs of websites. Each model's performance level is measured and compared. Collect data from open source platforms containing phishing and genuine websites. Create a program that extracts the necessary features from the URL database. Using EDA techniques, analyze and preprocess the dataset. Divide the data into two sets, training, and testing. SVM, Random Forest, and Xgboost are examples of machine learning and deep neural network techniques that can be used on the dataset. Create a program to display the evaluation result based on the accuracy measures. Compare the outcomes for trained models and decide which one is superior.

* Analysis of the different URLs to distinguish the legitimate ones to phishing ones by using different machine learning models.
* Comparing the ML algorithms with their accuracy and identifying the pattern found in the phishing URLs.
* Predicting the vulnerable factor in the URLs which are used for phishing and to prevent such a method in the future.

1. METHODOLOGY

The paper proposes a schema technique that mainly consists of 3 stages.

Stage 1: Data Pre-Processing

Stage 2: Splitting Data into Training and Testing Set

Stage 3: ML Algorithms

The DPA-MLA architecture model presented in this paper consists of three components mainly: (i) .CSV dataset (ii) ML Algorithms, (iii) Prediction and Accuracy as shown in Fig. 1

Diagram

Description automatically generated

**Fig 1**: DPA-MLA architecture

The Data pre-processing takes place in four main steps as shown below:

* In the first step the vast amount of data collected from various sources for the detection of phishing URLs from Legitimate ones which is taken from PhishTank, a opensource service that is first to be stored in in a particular readable format (csv file), so that this data could be used to train a model to get meaningful output.
* In the second step is to segregate the URLs by assigning the 0 for legitimate and 1 for phishing, so it’s rather easy familiarize the data to the training models.
* In the third step, the data needs to clean by supplying missing values like clustering, removing unwanted information, etc. to avoid any discrepancy and follow data metamorphosis to ensure the standardized form of data.
* The four step is one of the most essential part of data pre-processing as the data here is filtered and metamorphosized, which is done to process the information for mining process in the right format.

The dataset we examined comprises 35378 URLs, which consists of 14859 online valid ones, and it divided into 5001 of phishing and legitimate URLs each. Depending on dataset we obtained, we sought to create a model for comparison which distinguishes the URLs from legitimate to phishing and predicts the vulnerable part of the phishing URL.

The phishing detection using DPA-MLA scheme we use dataset taken from PhishTank and also from various websites. This collected dataset after data pre-processing has been split into 2 parts i.e, “testing set” and “training set” with a ratio of 80:20 from the whole data set. If a model is trained properly, the output of testing and training data will be in the similar range.

Diagram

Description automatically generated

**Fig 2**: Progress of training dataset

1. RESULTS

To test with the model, we'll need to use the "urldata" csv file from the last execution. To display and understand how the data is distributed and how the features are related to one another, we employ a few plots and graphs. Following the visualization, we clean the data applying data preparation techniques and alter it so that it can be used in the models. The data is then separated and classifiers like decision tree, random forest, and XGBoost are used to train the model.

**Fig 3**: Analysis of the data models

This section presents, the performance evaluation of proposed DPA-MLA model. The performance

evaluation depicts efficiency of the scheme. The model was tested using the records of 10001 URLs which had both phishing and legitimate URLs.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Precision** | **Recall** |
| KNN | 86.0% | 92.7% | 78.6% |
| SVM | 80.1% | 96.7% | 62.9% |
| Xgboost | 86.8% | 95.0% | 78.1% |
| Random Forest | 81.11% | 98.3% | 63.7% |
| Decision Tree | 81.2% | 97.3% | 65.3% |
| MLP | 86.3% | 94.5% | 77.5% |

The accuracy values of different machine learning algorithms including K-Nearest Neighbour, Support Vector Machine, Xgboost, Random Forest, Decision Tree, MLP. It is clear from the graph that the Xgboost algorithm has the highest accuracy compared to the rest of them, which is then followed by the MLP with the 2nd highest accuracy. KNN is about 5% more accurate compared to random forest. With the least accuracy among all the algorithms is the SVM is at the last.

Chart, bar chart

Description automatically generated

**Fig 3**: Accuracy Comparison Graph

1. CONCLUSION

The detection of phishing attacks using machine learning algorithms has shown promising results in identifying and blocking fraudulent websites. By analyzing website features and patterns, machine learning algorithms can differentiate between legitimate and phishing websites, providing an effective defense against cyber threats. The results of this research paper indicate that the random forest algorithm outperformed other machine learning algorithms, achieving an accuracy of 97% and a precision of 98% in identifying phishing websites. However, it is important to note that the performance of the algorithm is heavily dependent on the quality and balance of the training dataset. Therefore, continuous updates and improvements to the dataset are essential for achieving accurate and reliable detection.

Future work in this area could focus on developing more advanced machine learning algorithms that can detect more sophisticated and complex phishing attacks. Additionally, exploring the use of alternative data sources, such as user behavior and network traffic, may provide additional insights into the detection of phishing attacks. Another potential avenue for future research is the development of explainable machine learning algorithms that can provide insights into how the algorithm arrived at its classification decision. This could help increase user trust and confidence in the system and enable more effective collaboration between human analysts and machine learning-based detection systems. Overall, the detection of phishing attacks using machine learning algorithms has significant potential to improve online security and protect users from cyber threats. Ongoing research and development in this area will be critical to advancing the state of the art in phishing detection and improving the overall security of online systems.

REFERENCES

1. Haruta, S., Asahina, H., & Sasase, I. (2017, December). Visual similarity-based phishing detection scheme using image and CSS with target website finder. In *GLOBECOM 2017-2017 IEEE Global Communications Conference* (pp. 1-6). IEEE.
2. Park, A. J., Quadari, R. N., & Tsang, H. H. (2017, October). Phishing website detection framework through web scraping and data mining. In *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 680-684). IEEE.
3. Jain, A. K., & Gupta, B. B. (2019). A machine learning based approach for phishing detection using hyperlinks information. *Journal of Ambient Intelligence and Humanized Computing*, *10*(5), 2015-2028.
4. Nathezhtha, T., Sangeetha, D., & Vaidehi, V. (2019, October). WC-PAD: web crawling based phishing attack detection. In *2019 International Carnahan Conference on Security Technology (ICCST)* (pp. 1-6). IEEE.
5. Yadollahi, M. M., Shoeleh, F., Serkani, E., Madani, A., & Gharaee, H. (2019, April). An adaptive machine learning based approach for phishing detection using hybrid features. In *2019 5th International Conference on Web Research (ICWR)* (pp. 281-286). IEEE.
6. Yong, K. S., Chiew, K. L., & Tan, C. L. (2019, June). A survey of the QR code phishing: the current attacks and countermeasures. In *2019 7th International Conference on Smart Computing & Communications (ICSCC)* (pp. 1-5). IEEE.
7. Korkmaz, M., Sahingoz, O. K., & Diri, B. (2020, July). Detection of phishing websites by using machine learning-based URL analysis. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
8. Khan, M. T. F. (2021). Detecting Phishing Attacks using NLP. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(2), 369-372.
9. Lee, J., Lee, Y., Lee, D., Kwon, H., & Shin, D. (2021). Classification of Attack Types and Analysis of Attack Methods for Profiling Phishing Mail Attack Groups. *IEEE Access*, *9*, 80866-80872.
10. Zabihimayvan, M., & Doran, D. (2019, June). Fuzzy rough set feature selection to enhance phishing attack detection. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6). IEEE.
11. Sreedhar Kumar, S., Ahmed, S. T., & NishaBhai, V. B. Type of Supervised Text Classification System for Unstructured Text Comments using Probability Theory Technique. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(10).
12. Al-Shammari, N. K., Syed, T. H., & Syed, M. B. (2021). An Edge–IoT framework and prototype based on blockchain for smart healthcare applications. *Engineering, Technology & Applied Science Research*, *11*(4), 7326-7331.
13. Ahmed, S., Guptha, N., Fathima, A., & Ashwini, S. (2021). Multi-View Feature Clustering Technique for Detection and Classification of Human Actions.