Grenze International Journal of Engineering and Technology, June Issue

ML Shield: Detecting Malicious Websites with Machine Learning

Aditya Shankar Tripathi[1], Abhishek Maurya[2], Aniket Jain[3], Anjali Kumari[4] and Manu Singh[5]
[1-5]Department of Computer Science, ABES Engineering College, Ghaziabad
Email: ashankartri@gmail.com, {abhishek.am669, aniketjain2711, anjali.k4221, drmanuajaykumar}@gmail.com

***Abstract*— Phishing websites have become a significant cybersecurity threat, hosting malware and exploiting users by mimicking popular sites. Victims suffer financial loss, compromised privacy, and damaged reputation. Urgent solutions are needed to mitigate these threats promptly. Our goal is to develop a Chrome extension that utilizes machine learning and deep learning techniques to identify phishing websites and alert users when they visit such sites. When a user opens a website in their browser, the extension operates in the background. If the website is recognized as phishing, a pop-up box with an 'OK' button informs the user. If the website is not flagged as phishing, no action is taken. We employed machine learning algorithms (such as Random Forest) and a Single Layer Perceptron for training the model. The browser extension is written in JavaScript, and its JavaScript file extracts URL features like browser popups, use of frames, and shortening services. The algorithm then determines whether the website is legitimate or not.**

***Index Terms*— Malicious websites, machine learning methodologies, phishing attacks, feature extraction, URL analysis**

I. INTRODUCTION

In the rapidly evolving landscape of cybersecurity, the proliferation of malicious websites poses a profound threat to the digital ecosystem, jeopardizing user privacy, security, and data integrity. This burgeoning menace encompasses a diverse array of harmful online destinations, including phishing sites, drive-by download sites, malicious ads, and more. The overarching goal of these websites is to exploit unsuspecting users, coercing them into divulging sensitive information, downloading malicious software, or engaging in activities leading to unauthorized data access. The consequences of such interactions are far-reaching, ranging from financial losses to identity theft and the compromise of confidential information.

In 2021, a staggering 323,972 individuals were reported as victims of phishing attacks, indicating that half of the users affected by data breaches succumbed to phishing tactics. The height of the pandemic witnessed a substantial 220% surge in phishing incidents. Conventional methods employed to identify and combat malware websites have historically relied on blacklists, heuristics, and signature-based detection. While these approaches have offered a degree of protection, their limitations become apparent in the face of the dynamic and sophisticated tactics continually evolving in the hands of cybercriminals. As malicious actors refine their strategies, there arises an urgent need for advanced techniques capable of effectively detecting and thwarting the ever-changing landscape of malicious websites.

Embracing the Power of Machine Learning, this review paper advocates for the adoption of machine learning

Grenze ID: 01.GIJET.10.2.55

techniques as the most effective approach to detect malware websites. The primary aim is to enhance the precision and efficiency in identifying and categorizing malicious websites, thus bolstering proactive defence against cyber threats. Machine learning provides a dynamic and data-driven methodology for identifying and combating these malicious websites. Through the analysis of extensive datasets and the extraction of patterns, machine learning models autonomously classify websites as benign or malicious, continually adapting to evolving threats.

This paper's primary goal is to develop an efficient and accurate malware website detection system rooted in a comprehensive analysis of various machine learning algorithms. By exploring diverse features, ranging from domain attributes and URL structure to content analysis and SSL certificates, the research aims to uncover distinctive characteristics that differentiate malicious websites from legitimate ones. Furthermore, the study endeavours to evaluate the performance of various machine learning models, comparing them against traditional detection methods to underscore their potential advantages. This comprehensive exploration serves as a foundation for understanding and progressing the cutting-edge techniques in machine learning for the identification of malicious websites.

By bridging the gap between evolving cyber threats and advanced detection mechanisms, this paper aims to contribute to the ongoing efforts in fortifying cybersecurity, providing a valuable resource for researchers, practitioners, and cybersecurity professionals.

II. Background And Literature Review

A malicious website is a website that has been designed to harm or exploit users, typically by delivering malware, engaging in phishing attacks, or conducting other forms of cybercrime. These websites are created with the intent of stealing sensitive information, spreading malware, or carrying out various types of fraudulent activities. Malicious websites often mimic legitimate websites to deceive users into visiting them.

In their review, Kiren et al. [1] comprehensively explored various types of phishing attacks and corresponding detection methods. Additionally, the authors introduced several mitigation techniques for phishing. The paper suggested that employing a machine learning approach stands out as the most effective among all anti-phishing strategies, positing the potential for achieving 100% accuracy in phishing detection.

Rana et al. [2] conducted a thorough review and analysis of contemporary phishing attack techniques, aiming to enhance reader awareness and provide education on various attack methodologies. The authors articulated the purpose of the paper as not only informing readers about phishing techniques but also advocating for the adoption of anti-phishing methods.

Felegyhazi et al. [3] discover new phishing URLs by using the domain name and server information from blacklisted URLs. The method is comparing newly registered phishing URLs' Domain Name System (DNS) zone data and registration details with the data obtained from banned URLs.

Athulya et al. [4] provided an in-depth exploration of various phishing attacks, the latest techniques employed by malicious actors, and underscored several anti-phishing methodologies. The paper aims to heighten awareness regarding phishing attacks and strategies, encouraging readers to adopt anti-phishing practices. Additionally, the authors introduced a novel phishing detection approach designed to efficiently identify phishing websites.

Shahrivari et al. [5] proposed an approach that used feature engineering to build a dataset and extract 30 features from the URL, the content of the web page, and the host data. They then used twelve machine learning approaches, such as decision trees and random forests, to identify phishing websites.

Xin Mei Choo et al.[6] focuses on a tool that detects phishing websites by analyzing log information from legitimate websites. The technique identifies phishing sites by examining reports registered on official website servers when users enter phishing webpages, which may refer to legitimate webpages by demanding services.

D.Sahoo et al.[7] explores several machine learning methods, including SVM, Logistic Regression, Naive Bayes, Decision Trees, Ensembles, Online Learning, etc., for the purpose of identifying dangerous URLs. But the application of the Random Forest (RF) and Support Vector Machine (SVM) algorithms is the particular emphasis of this paper. The outcomes of the experiment will demonstrate how accurate these two algorithms are with varying parameter setups

Aaron Blum et al.[8] investigated using confidence-weighted classification and content-based phishing URL detection to create a dynamic system. This system aims to detect both current and emerging phishing domains, offering proactive defence against new threats, unlike reactive methods used in traditional blacklisting techniques.

Anand et al. [9] employ diverse features for testing on an oversampled dataset, which comprises both genuine phishing URLs and artificially generated ones. Highlighted features indicating a phishing URL include URL length, absence of domain name/numerous digits, multiple subdomains, URL separated by "-", and the use of "@"

symbol. The utilization of an oversampled dataset significantly enhances the detection performance of phishing URLs.

Youness Mourtaji et al.[10] explored evaluated a number of techniques, such as lexical, content, black-list, and security/identity-based methodologies, to create a model utilizing machine learning classifiers for efficient phishing website identification.

Ramesh et al. [11] introduced a phishing site detection method that examines web pages to identify both direct and indirect links.Despite achieving high accuracy in detection, the method is time-consuming as it relies on search engines and third-party services like DNS queries.

Mogimi et al. [12] introduced a phish detector employing the support vector machine (SVM) algorithm for initial phishing model training, followed by the decision tree (DT) algorithm to uncover hidden phishing elements. The method achieves a true positive rate of 0.99 and a false negative rate of 0.001 in a large dataset. However, its assumption that phishing web pages solely utilize benign content is not practical.

Kazemian el al. [13], implemented three supervised machine learning methods: Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). Furthermore, unsupervised methods such as K-Means and Affinity Propagation were used to identify dangerous websites. Based on experimental data, the proposed technique obtained 98% accuracy for supervised learning and 96% accuracy for unsupervised learning.

Yang et al. [14] proposed a rapid deep learning-based approach for phishing detection. They extracted character sequence features from URLs in the initial stage, enabling swift categorization through deep learning. The second stage integrated various features, including statistical information about the URL, webpage code characteristics, and textual content. The model achieved an impressive 98.99% accuracy with a low false positive rate of 0.59%

Blum et al. [15] introduced a model suggesting the possibility of detecting malicious URLs without inspecting web page content. The model relies on URL-based characteristics such as the domain name, presence of special characters, etc. A machine learning model incorporating these features was developed for the detection process.

A. Challenges

Detecting malicious websites through machine learning is a complex task, presenting its own set of challenges. Some of the key challenges include:

- *Dynamic Nature of Threats*: Malicious actors continuously evolve their tactics, making it challenging for static machine learning models to adapt quickly. This requires constant model updates and retraining.
- *Evasion Techniqu*es: Malicious actors actively try to evade detection by manipulating features or injecting noise into the data. Adversarial attacks pose a challenge to the robustness of machine learning models, requiring the development of models resistant to such manipulations.
- *Generalization Across Threat Types*: Malicious websites encompass a wide range of threats, from phishing to drive-by downloads. Creating a model that generalizes well across different types of threats requires a nuanced understanding of each threat category.

B. Opportunities

Detecting malicious websites using machine learning presents several opportunities that can significantly enhance cybersecurity measures. Some of the key opportunities include:

- *Improved Accuracy*: Leveraging machine learning algorithms can enhance the accuracy by analyzing a wide range of features and patterns. These models can better discriminate between benign and malicious websites, reducing false positives and negatives.
- *Enhanced User Experience*: Accurate detection of malicious websites contributes to a safer online experience for users. By preventing access to harmful content, machine learning-based solutions can build trust among users and foster confidence in the security of online interactions.
- *Cost-Efficiency*: Once trained, machine learning models can automate the detection process, reducing the need for extensive manual intervention. This automation can lead to cost savings and allow cybersecurity professionals to focus on more complex aspects of threat mitigation.

III. RESEARCH DESIGN AND METHODOLOGY

The development of a model for identifying and preventing phishing websites through machine learning demands a well-defined methodology. This methodology delineates the step-by-step process, starting from data collection and extending to model deployment. The following section elaborates on the methodology applied in crafting our model, as depicted in Figure 1.

A. Data Collections

Initially, data collection involves gathering various websites, comprising both genuine and deceptive ones. The dataset's size should be sufficient to capture diverse types of phishing attacks. Different methods, such as web scraping or utilizing existing datasets, can be employed to acquire this data.

B. Data Preprocessing

Upon the completion of data collection, the next step is to make it ready through pre-processing. This means cleaning and adjusting the data so it's all set for analysis. This step is crucial because it guarantees the data is top-notch and ready for machine learning. Pre-processing involves tasks like getting rid of duplicates, handling missing info, and formatting the data in a way that works well for machine learning programs. The above dataset was checked for missing values, null entries, duplicates, and noise, utilizing tools like box plots and other methods. Features that did not contribute significantly to phishing detection were then eliminated from the dataset.

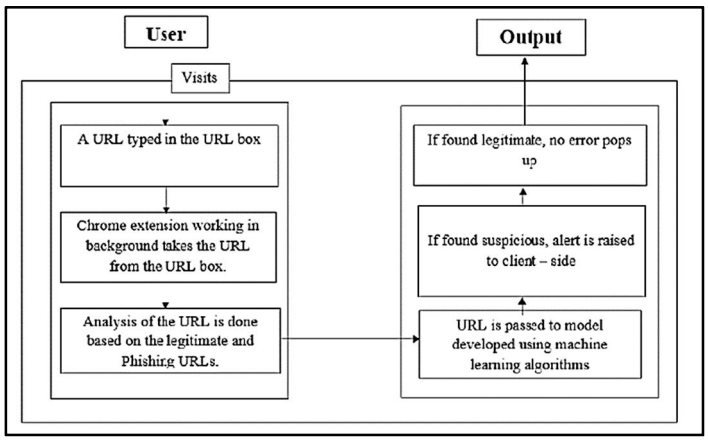

Figure 1. Design Methodology

C. Model Selection

The subsequent stage is the selection of a machine learning model. Choosing the most suitable algorithm is crucial for accurately classifying legitimate and fraudulent websites. Various machine learning algorithms are available, including Random Forest, Decision Tree, Support Vector Machines, Gradient Boosting Classifier, K-Nearest Neighbours, AdaBoost Classifier, Logistic Regression, and others. For our specific design, we opted for the Random Forest algorithm.

We have trained and evaluated our model using 22 parameters dataset and various algorithms. From all the algorithms used Random Forest provided the best accuracy. For the implementation of extension, 16 parameters were used and single layer perceptron was used to detect phishing websites.

(1) Random Forest: The Random Forest algorithm is a versatile machine learning technique employed for both classification and regression tasks. It operates as a group of methods, linking multiple decision trees to make predictions. Widely applied in areas like finance, healthcare, and marketing, Random Forest creates several decision trees by randomly choosing subsets of the original dataset and features. This approach aims to enhance performance by consolidating the predictions from individual trees, reducing overfitting, and improving the model's generalization ability. The final prediction of the Random Forest algorithm is established by taking into account the majority vote across all individual trees

(2) Single Layer Perceptron: A single-layer perceptron is a type of artificial neural network featuring a single layer of nodes that takes input values and generates a set of output values. This algorithm is employed in supervised learning for classification tasks, specifically when the output involves binary classification (i.e., two classes).

Comprising one or more input nodes, a set of weights, an activation function, and an output node, the single-layer perceptron connects input nodes to the output node through a set of weights. These weights are adjusted during training to minimize output errors. The activation function utilized in a perceptron is typically a threshold function, determining whether the perceptron's output is 0 or 1 based on the input and weights.

D. Model Training

Following the completion of the machine learning algorithm, the next phase is model training. This procedure entails providing the model with extracted features along with their respective labels. Through

learning from the data, the model adapts its parameters to minimize errors. The data is divided into two sets: a training set and a validation set, to help with this procedure. The validation set is used to evaluate the model's performance after it has been trained using the training set. In this specific study, the model was trained using 80% of the dataset.

E. *Model Evaluation*

Following the model training, the next step is to estimate its performance. This involves testing the model on a distinct test set that was not part of the training data. The model's effectiveness is evaluated through metrics such as sensitivity, accuracy, and recall, offering insights into its capability to identify and prevent phishing attacks.
Building a machine learning model to detect and prevent phishing websites demands a thoughtfully constructed methodology. This process includes data collection, pre-processing, incorporating relevant features, building a machine learning model, training, evaluating its performance, and deploying the model. Following this methodology enables the development of a model capable of offering real-time protection against phishing attacks.

IV. Result And Discussion

The extension is designed to display an alert whenever a user visits a phishing website, serving to safeguard them from potential harm. For this to work the extension needs to be added in the browser by the user. The extension is developed through the utilization of a single-layer perceptron in deep learning, chosen after assessing the accuracy of various machine learning models.
The extension is to be first uploaded by going to the chrome::://extensions sections and then clicking on Load unpacked as shown in the Figure.2.
Phishing websites are identified through the analysis of diverse URL features. Upon a user's visit to a website, the extension extracts the website's URL, and subsequently, a JavaScript file (content.js) processes the URL to determine whether it is safe or potentially phishing. If the URL is recognized as a phishing threat, a popup alert is triggered for the user; otherwise, no popup is displayed. This alert mechanism is implemented through the background.js file, as illustrated in Figure 3.

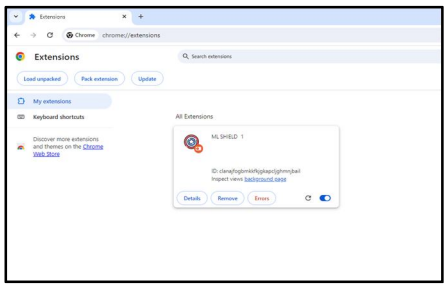

Figure 2. Loading Extension

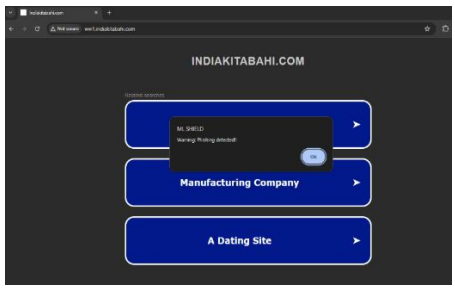

Figure 3. Phishing alert on sample website

V. Conclusions

In this study, we present a system tailored for the detection and prevention of phishing websites, with a specialized emphasis on URL classification. The dataset used for this purpose consists of 11,055 entries, each with 22 different features that aid in identifying whether a website is phishing site. The data in the dataset is represented as 1, 0, or -1, indicating legitimate, suspicious, and phishing respectively. We employed the Random Forest Algorithm for both training and testing the model. The input URL undergoes processing, calculating 21 distinct features (excluding the Result, the 22nd feature) represented as 1, 0, and -1. The model accurately classifies the input URL as Legitimate or Phishing with an impressive accuracy of 94.54%.
For possible enhancements and future integration, we have the option to inform the Google Website Blacklist database and page ranking system about recently identified phishing websites detected by our trained model. This reporting mechanism ensures timely actions, leading to the banning or appropriate processing of these sites.

References

[1] Tandale, Kiran D., and Sunil N. Pawar. "Different Types of Phishing Attacks and Detection Techniques: A Review." 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC). IEEE, 2020.

[2] R. Alabdan, “Phishing attacks survey: Types, vectors, and technical approaches,” Futur. Internet, vol. 12, no. 10, pp. 1–39, 2020, doi: 10.3390/fi12100168.
[3] K. A. Moul, “Avoid phishing traps,” Proc. ACM SIGUCCS User Serv. Conf., no. August 2017, pp. 199–208, 2019, doi: 10.1145/3347709.3347774.
[4] Athulya, A. A., and K. Praveen. "Towards the Detection of Phishing Attacks." 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184). IEEE, 2020.
[5] V. Shahrivari, M. M. Darabi, and M. Izadi, Phishing Detection Using Machine Learning Techniques, Sharif University of Technology, Tehran, Iran, 2020.
[6] Xin Mei Choo, Kang Leng Chiew, Dayang Hanani Abang Ibrahim, Nadianatra Musa, San Nah Sze, Wei King Tiong,“Feature-Based Phishing Detection Technique”,2016
[7] D. Sahoo, C. Liu, S.C.H. Hoi, “Malicious URL Detection using Machine Learning: A Survey”. CoRR, abs/1701.07179, 2017.
[8] Aaron Blum,Brad Wardman,Thamar Solorio,Gary Warner, “Lexical feature based phishing URL detection using online learning,” 3rd ACMworkshop on Artificial intelligence and security, Chicago, Illinois,USA, pp. 54-60, August 2010.
[9] Ankesh Anand, Kshitij Gorde, Joel Ruben Antony Moniz, Noseong Park, Tanmoy Chakraborty, and Bei-Tseng Chu. “Phishing URL detection with oversampling based on text generative adversarial networks”. In: 2018, pp. 1168–1177.
[10] Youness Mourtaji,Mohammed Bouhorma,Alghazzawi, “Perception of a new framework for detecting phishing web pages,” MediterraneanSymposium on Smart City Application Article No. 11, Tangier,Morocco, October 2017.
[11] Ramesh, G.; Krishnamurthi, I.; Kumar, K.S.S. An efficacious method for detecting phishing webpages through target domain identification. Decis. Support Syst. 2014, 61, 12–22.
[12] Moghimi, M.; Varjani, A.Y. New rule-based phishing detection method. Expert Syst. Appl. 2016, 53, 231–242.
[13] H. B. Kazemian and S. Ahmed, “Comparisons of machine learning techniques for detecting malicious webpages,” Expert Syst. Appl., vol. 42, no. 3, pp. 1166–1177, 2015.
[14] Peng Yang, Guangzhen Zhao, Peng Zeng. Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning. IEEE Access 2019. Digital Object Identifier 10.1109/ACCESS.2019.2892066
[15] BlumA,Wardman B, Solorio T, Warner G. Lexical feature based phishing URL detection using online learning. In: Proceedings of the 3rd ACM workshop on security and artificial intelligence, AISEC; 2010. https://doi.org/10.1145/1866423.1866434. 2010 October
[16] Fortinet: https://www.fortinet.com
[17] Coursera:https://www.coursera.org/articles/machine-learning-models
[18] https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/
[19] Detection of URL based Phishing Attacks using Machine Learning – IJERT