www.coast.oaustech.edu.ng
COAST
JOURNAL OF THE SCHOOL OF SCIENCE
Volume 6, Issue 1. May, 2024 ISSN: 2714-3716
Coast, J. Sch. Sci. 6 (1): 981 - 988

DETECTION OF UNHEALTHY WEBSITES USING MACHINE LEARNING

[*1]Gbadamosi, O. A. and [2]Oduwale, A. M.

[1,2]Department of Computer Science, Olusegun Agagu University of Science and Technology (OAUSTECH), Okitipupa, Ondo State, Nigeria
Corresponding Author's Email: gbadamosiajoke2015@gmail.com
https://doi.org/10.61281/coastjss.v6i1.5

Abstract
In recent years, advancements in Internet and cloud technologies have led to a significant increase in electronic trading in which consumers make online purchases and transactions. Accompanying this achievement are vices like unauthorized access to users' sensitive information and damages to enterprise resources. Phishing is one of the familiar attacks that trick users to access malicious content and gain their information. This study aims to develop an efficient machine-learning program to detect phishing websites with high accuracy. Most phishing webpages look identical to the actual web pages and various strategies for detecting phishing websites, such as blacklisting, and heuristics, among others have been suggested. Existing research works showed that the performance of the phishing detection system is limited and there is a demand for intelligent techniques to protect users from cyber-attacks. A Uniform resource locator (URL) detection technique based on a supervised machine learning approach – Naïve Bayes-is employed and implemented in Python programming language. The efficacy of this approach was determined on a phishing dataset made up of 7900 malicious and 5800 legitimate sites, respectively. The results show that using the proposed methodology an accuracy of 96% can be achieved by using stacking, filtering along the Naïve Bayes and logistic regression. This study thoroughly investigates the use of machine laearning with features extracted from the URLs and was able to showcase common words for the identification of either phishing (unhealthy) or good websites and proffered a guide to end users against the recent approaches in malicious URLs detection.
Keywords: Machine learning, Phishing, Social Engineering, Model, Uniform Resource locators.

Introduction

The digital world is rapidly expanding and evolving, and likewise, some cybercriminals have relied on the illegal use of digital assets especially personal information to inflict damage on individuals. One of the most threatening crimes of all internet users is that of 'identity theft' which is defined as impersonating a person's identity to steal and use their personal information (that is, bank details, social security number, or credit card numbers, and so on) by an attacker for the individuals' gain not just for stealing money but also for committing other crimes (Ramanathan, *et al.*, 2012). Cybercriminals have also developed methods for stealing information, but social-engineering-based attacks remain their favorite approach. One of the social engineering crimes that allow the attacker to perform identity theft is called a phishing attack. Phishing has been one of the biggest

concerns as many internet users fall victim to it. It is a social engineering attack wherein a phisher attempts to lure the users to obtain their sensitive information by illegally utilizing a public or trustworthy organization in an automated pattern so that the internet user trusts the message and reveals the victim's sensitive information to the attacker. In phishing attacks, phishers use social engineering techniques to redirect users to malicious websites after receiving an email and following an embedded link. Alternatively, attackers could exploit other mediums to execute their attacks such as Voice over IP (VoIP), Short Message Service (SMS), and, Instant Messaging (IM). Phishers have also turned from sending mass email messages, which target unspecified victims, into more selective phishing by sending their emails to specific victims, a technique called spear-phishing (Gandotra & Gupta, 2021).

Cybercriminals usually exploit users with a lack of digital/cyber ethics or who are poorly trained in addition to technical vulnerabilities to reach their goals. Susceptibility to phishing varies between individuals according to their attributes and awareness level, therefore, in most attacks, phishers exploit human nature for hacking, instead of utilizing sophisticated technologies. Even though the weakness in the information security chain is attributed to humans more than technology, there is a lack of understanding about which ring in this chain is first penetrated. Studies found that certain personal characteristics make some persons more receptive to various lures. For example, individuals who usually obey authorities more than others are more likely to fall victim to a Business Email Compromise (BEC) that pretends to be from a financial institution and requests immediate action by seeing it as a legitimate email (Zainab, *et al.,* 2021). Various channels are used by the attacker to lure the victim through a scam or through an indirect manner to deliver a payload for gaining sensitive and personal information from the victim. However, phishing attacks have already led to damaging losses and could affect the victim not only through a financial context but could also have other serious consequences such as loss of reputation, or compromise of national security. Cybercrime damages have been expected to cost the world $6 trillion annually by 2021, up from $3 trillion in 2015 according to Cybersecurity Ventures (Hung *et al.,* 2017). Phishing attacks are the most common type of cybersecurity breach as stated by the official statistics from the cybersecurity breaches survey 2020 in the United Kingdom. Although these attacks affect organizations and individuals alike, the loss for the organizations is significant, which includes the cost of recovery, the loss of reputation, fines from information laws/regulations, and reduced productivity. There is a significant chance of exploitation of user information. For these reasons, phishing in modern society is highly urgent, challenging, and overly critical. There have been several recent studies against phishing based on the characteristics of a domain, such as website URLs (Uniform Resources Locators), website content, incorporating both the website URLs and content, the source code of the website, and the screenshot of the website (Hodžić *et al.,* 2016). However, there is a lack of useful anti-phishing tools to detect malicious URLs in an organization to protect its users. In the event of malicious code being implanted on the website, hackers may steal user information and install malware, which poses a serious risk to cybersecurity and user privacy (Anuraag, 2021). Malicious URLs on the internet can be easily identified by analyzing them through the Machine Learning (ML) technique using the Bayes

classifier (Pujara *et al.,* 2018).

Related works

An effective phishing domain name detection approach based on Heterogenous information networks (HIN) named HinPhish. HinPhish extracts link relationships from web pages and constructs a HIN model of domains and resource objects. HinPhish leverages the characteristics of different relations to calculate the phish score of each node object effectively. Moreover, HinPhish not only improves the accuracy of detection but also can increase the phishing cost for attackers. Extensive experimental results demonstrate that HinPhish can achieve an accuracy of 0.9856 and an F1-score of 0.9858 (Bingyang*et al.,* 2021). A Comparative Analysis of Machine Learning Algorithms for Phishing Website Detection was employed using machine learning techniques to train models that can detect phishing websites that involve a dataset and some programmed code that performs computations allowing the code to analyze a portion of the data and observe relationships between the features and the classification of the data, and its performance was measured and scored (Sarma *et al.,* 2021; Lord, 2018). Another study was surveyed by (Yuan *et al.,* 2018), (Anuja & Gayatri, 2022), on the detection of phishing websites using machine learning and was deduced as one of the best machine learning method based on accuracy, false-positive rate, and false-negative rate, phishing, feature classification, random forest classifier, and other terms were used in this study, and was similarly carried out by (Zhang *et al.,* 2021; Sneha & Thosar, 2018).

Materials and Methods

This study focused on detecting phishing websites with Uniform Resource Locators (URLs). Analysis of detecting phishing websites based on the content/link analysis was carried out. Various classification methods have been investigated in the sense of filtering phishing links. A supervised machine learning technique was used consisting of logistics regression and Naïve Bayes implemented in a Python environment. The proposed system's architecture is shown in Figure 1, and each of these components is briefly discussed.

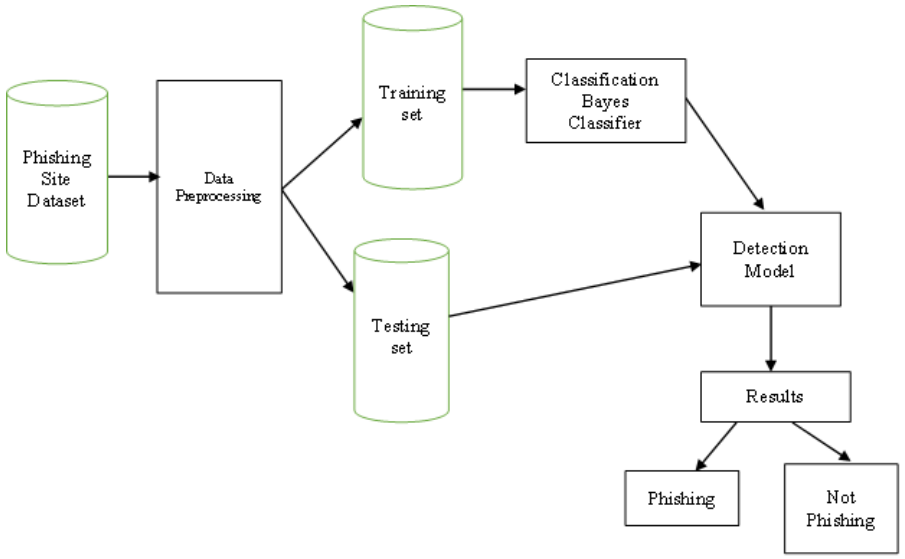

Figure 1: Phishing Website System Architecture

Components of System Architecture

Phishing Site Dataset; the phishing dataset is made up of 7900 malicious and 5800 legitimate sites, respectively. Data Preprocessing; the process of generating raw data for machine learning models. This is the first step in creating a machine-learning model. This is needed in machine learning algorithms to reduce its complexities. Training set; trains the machine learning model, allowing it to learn the patterns and relationships within the data. Testing set; used after the model has been trained and validated, to provide an unbiased evaluation of the model performance on completely new, unseen data. Classification Method; Naive Bayes Classifier is a probabilistic machine learning model based on Bayes' theorem. It assumes independence between features and calculates the probability of a given input belonging to a particular class. Results; Phishing or Not Phishing

Phishing Site Dataset

The dataset in this study was obtained from the public online repository Mendeley. The dataset is made up of 7900 malicious and 5800 legitimate sites, respectively. The final output showed that the proposed method outperformed the recent approaches in malicious URL detection providing maximum efficiency and an improved feature extraction technique was employed for this dataset by using the browser automation framework. The class label indicated two outcomes where 0 was a phishing website, and 1 was a real website. When dealing with classification tasks, supervised learning algorithms are implemented by the training data set to learn and determine the optimal combinations of variables that would develop a predictive model. The main aim is to produce a well-trained model. The trained model is evaluated using “new” examples from the test datasets to estimate the model's accuracy in classifying new data.

Data Preprocessing

It is important to preprocess data to convert raw data into an understandable format to allow the model to train in the best way possible. The dataset was split into two parts. 70% of the dataset was used for training purposes and the rest of the 30% for testing/selecion purposes and this is to boost the algorithm classifier by applying some simple techniques to the dataset, like dta preprocessing and features selection.. Training dataset provided features and labels together to learn the relationship between them so that the model could later on test its knowledge against the test set, where it was only provided with the features and was set to generate labels for each set of features and check how many of its predictions were done correctly. Handling of of missing data is a major decision during data preprocessing for most model and to detect and eliminate the predominant features, the use of logistics regression and network X library in Python was adopted.

Classification Models

Given one or more inputs a classification model will try to predict the value of one or more outcomes. Outcomes are labels that can be applied to a dataset, for this study when filtering phishing website URLs we classify them as “Bad” or “Good”. Conclusions are obtained from a certain set of observed values. A simple supervised machine learning technique for constructing classifiers, thereby building models that give labels to inputs (problem instances) called the Naïve Bayes Technique was deployed.

This technique utilizes the Bayes' theorem whereby:

i. The probability of a link is computed to be phishing, given a certain word(s) is contained in the link
ii. The probability of a link is computed

to be phishing (not safe), considering all words contained in the link

iii. Considering new words (not found in the dataset) or words that rarely appear.

Using Naïve Bayes Technique to compute if a given word is a phishing link.

$$\Pr(S \mid W) = \Pr(W \mid S).\frac{\Pr(S)}{\Pr.(W \mid S).\Pr(S) + \Pr(W \mid H).\Pr(H)}$$

Where:

Pr(S) represents the overall probability that any given word is a phishing site or link.

Pr (W|S) represents the probability that a random link as "Kazan-pacir.rs" appears on phishing websites/links.

Pr(H) represents the overall probability that any given link is not spam.

Pr(W|H) represents the probability that a random website "google.com" appears as a valid phishing site/link.

Based on statistics, it is shown that the current probability of a website being a phishing website or link is 80% minimum. That is the factor to determine if a website is a phishing website or not based on the presence of a static random word "dfghjkd.com" is error-prone, and the reason this program considered several words contained in the dataset, combines the phishing probabilities and determines a link's overall probability of being a phishing link. However, accuracy can be a misleading metric to determine the quality of a model especially when the class inbalance is high,due to this attribute of accuracy, it is wise to check out measures when the output is a numerical score. F-measure (F1 score) is used to evaluate the the classifiers in which the Recall and Precision would have been defined.

Results and Discussion

In developing the program that classifies website link datasets to be good or bad, IDE (Integrated Development Environment) was used. The IDE (Integrated Development Environment) used was Spyder in the Python language for detection as shown in Figure 2 and represented in Table 1 considering the various metrix of classifier.

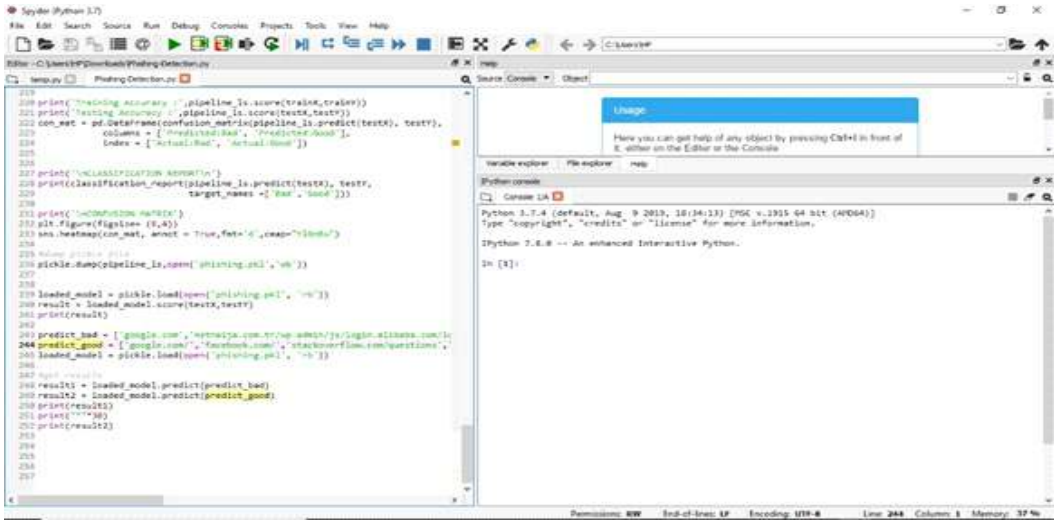

Figure 2: Phishing detection Code snippet in IDE

Figure 2 shows the common words detected in the good URLs, likewise, Figure 3 shows the common words detected in the bad URLs that were placed in the code after execution for ease of detection while working on the internet.

Table 1: Classification report on detecting a Phishing Website

CLASSIFICATION REPORT

	precision	recall	f1-score	support
Bad	0.91	0.97	0.94	36951
Good	0.99	0.97	0.98	100386
accuracy			0.97	137337
macro avg	0.95	0.97	0.96	137337
weighted avg	0.97	0.97	0.97	137337

CONFUSION MATRIX
0.967008162403431
['good' 'bad' 'bad' 'bad' 'bad']
****
['good' 'good' 'good' 'good' 'good']

Figure 3: Common words used inn good URLs

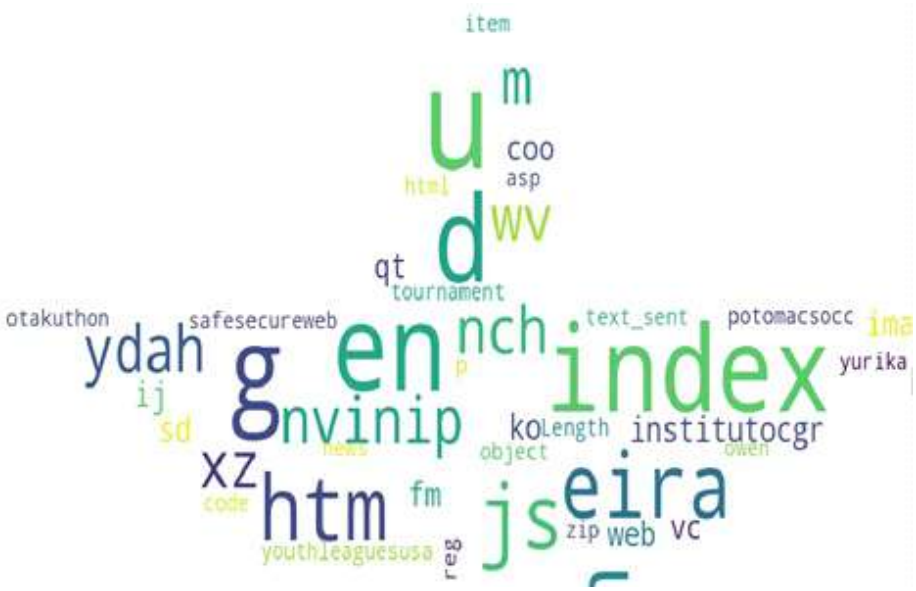

Figure 4: Common words used in bad URLs

Conclusions

The need for supervised machine learning techniques to help in the detection of phishing websites is required with the use of a classifier model to curb the increasing insecurity and the incursions of Internet marketers, and unsolicited commercial links (also known as phishing links) by users. This study emphasized the phishing technique in the context of classification, where phishing websites are considered to involve the automatic categorization of websites into a predetermined set of class values based on several features and the class variable. The machine learning-based phishing techniques are based on website functionalities to gather information that can help classify websites for detecting phishing sites. The problem of phishing cannot be eradicated, but can be reduced by combating it in two ways, improving targeted anti-phishing procedures and techniques and informing the public on how fraudulent phishing websites can be detected and identified actualising the goal of this study. To combat the ever-evolving complexity of phishing attacks and tactics, machine learning anti-phishing techniques are essential.

References

Anuja Bhosale, Gayatri Gadas, *et al* (2022). Detection of Phishing Websites using Machine Learning. International Journal of Advanced Research in Computer and Communication Engineering. 11(6). DOI: 10.17148/IJARCCE.2022.11695.

Anuraag V. 2021. Comparative study of machine learning algorithms for phishing website detection.Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 133-137.

Bingyang Guo, Yunyi Zhang, and Chengxi Xu (2021). HinPhish: An Effective Phishing Detection Approach Based on Heterogeneous Information Networks. Appl. Sci., (11): 9733. https://doi.org/10.3390/app11209733.

Gandotra E., and Gupta D. (2021). An

Efficient Approach for Phishing Detection Using Machine Learning. Algorithms for Intelligent Systems, Springer, Singapore. https://doi.org/10.1007/978-981-15-8711-5_12.

Hodžić, A., Kevrić, J., and Karadag, A. (2016). Comparison of machine learning techniques in phishing website classification. In International Conference on Economic and Social Studies. ICESoS'16: 249-256.

Hung Le, Quang Pham, Doyen Sahoo, and Steven C.H. Hoi (2017). URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection. Conference'17, Washington, DC, USA, arXiv:1802.03162.

Lord, N. (2018). What is a Phishing Attack? Defining and Identifying Different Types of Phishing Attacks. https://digitalguardian.com/blog/whatphishing-attack-defining-and-identifying-different-types phishingattacks.

Pujara, P., and Chaudhari, M. B. (2018). Phishing Website Detection Using Machine Learning: A Review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, IJSRCSEIT, **3**(7): ISSN 2456-3307.

Ramanathan, V., and Wechsler, H. (2012). PhishGILLNET-phishing detection methodology using probabilistic latent semantic analysis, AdaBoost, and co-training. EURASIP J. Info. 1–22. doi:10.1186/1687-417X.

Sarma, D., Mittra, T., and Hossain, S. (2021). Comparative Analysis of Machine Learning Algorithms for Phishing Website Detection. Springer Nature Singapore Pte Ltd. S. Smysetal. (eds.). Inventive Computation and Information Technologies, Lecture Notes in Networks and Systems 173, https://doi.org/10.1007/978-981-33-4305-4_64.

Sneha M., and Thosar, D. S. (2018). Detection of Phishing Web Sites Based on Extreme Machine Learning. **4**(6): ISSN(O)-2395-4396.

Yuan, H., Chen, X., Li, Y., Yang, Z., and Liu, W. (2018). Detecting Phishing Websites and Targets Based on URLs and Webpage Links. In 2018 24th International Conference on Pattern Recognition: 3669-3674).

Zainab A., Chaminda H., Liqaa N. and Imtiaz K. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. Front. Comput. Sci., 09 March 2021 Sec. Computer Security **3** https://doi.org/10.3389/fcomp.2021.563060.

Zhang, Y., Egelman, S., Cranor, L., and Hong, J. (2007). Phindin Phish: Evaluating Anti-Phishing Tools. In Processing of the 14th Annual Network and Distributed System Security Symposium.