IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

Detection of Phishing Websites using Machine Learning

Pratik Pramod Sarda and Dr. Pushpalata Aher
Sandip University, Nashik, India

Abstract: *Phishing attacks are a rapidly expanding threat in the cyber world, costing internet users billions of dollars each year. It is a criminal crime that involves the use of a variety of social engineering tactics to obtain sensitive information from users. Phishing techniques can be detected using a variety of types of communication, including email, instant chats, pop-up messages, and web pages. This study develops and creates a model that can predict whether a URL link is legitimate or phishing. The data set used for the classification was sourced from an open source service called 'Phish Tank' which contain phishing URLs in multiple formats such as CSV, JSON, etc. and also from the University of New Brunswick dataset bank which has a collection of benign, spam, phishing, malware & defacement URLs. Over six (6) machine learning models and deep neural network algorithms all together are used to detect phishing URLs. This study aims to develop a web application software that detects phishing URLs from the collection of over 5,000 URLs which are randomly picked respectively and are fragmented into 80,000 training samples & 20,000 testing samples, which are equally divided between phishing and legitimate URLs. The URL dataset is trained and tested base on some feature selection such as address bar-based features, domain-based features, and HTML & JavaScript-based features to identify legitimate and phishing URLs. In conclusion, the study provided a model for URL classification into phishing and legitimate URLs. This would be very valuable in assisting individuals and companies in identifying phishing attacks by authenticating any link supplied to them to prove its validity..*

Keywords: Phishing attacks

I. INTRODUCTION

1.1 Background Information

The Internet has become an important part of our lives for gathering and disseminating information, particularly through social media. According to Pamela (2021), the Internet is a network of computers containing valuable data, so there are many security mechanisms in place to protect that data, but there is a weak link: the human. When a user freely gives away their data or access to their computer, security mechanisms have a much more difficult time protecting their data and devices. Therefore, Imperva (2021) defines social engineering (a type of attack used to steal user data, including login credentials and credit card numbers) as a type of attack that is one of the most common social engineering attacks. The attack happens when an attacker fools a victim into opening an email, instant message, or text message as if it were from a trusted source. Upon clicking the link, the recipient is fooled into believing that they've received a gift and unsuspectingly clicks a malicious link, resulting in the installation of malware, the freezing of the system as part of a ransomware attack, or the disclosure of sensitive information. Computer security threats have increased substantially in recent years, owing to the rapid adoption of technology improvements, while simultaneously increasing the vulnerability of human exploitation. Users should know how the phishers do it, and they should also be aware of techniques to help protect themselves from becoming phished.

The strategies employed by cybercriminals are becoming more complex as technology advances. Other than phishing, there are a variety of methods for obtaining personal information from users. KnowBe4 (2021) stated the following techniques: a) Vishing (Voice Phishing): This kind of phishing includes the phisher calling the victim to get personal information about the bank account. The most common method of phone phishing is to use a phony caller ID. b) Smishing (SMS Phishing): Phishing via Short Message Service (SMS) is known as Smishing. It is a method of luring a

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

target through the SMS text message service by sending a link to a phishing website. c) Ransomware: A ransomware attack is a type of attack that prevents users from accessing a device or data unless they pay up. d) Malvertising: Malvertising is malicious advertising that uses active scripts to download malware or push undesirable information onto your computer. The most prevalent techniques used in malvertisements are exploits in Adobe PDF and Flash. Hence, this is a rapidly evolving threat to individuals as well as big and small corporations. Criminals now have access to industrial-strength services on the dark web, resulting in an increase in the amount of these phishing links and emails, and, more frighteningly, they are increasing in 'quality,' making them tougher to detect

1.2 Statement of Problem

Phishing attacks have gotten increasingly complex, it is very difficult for an average person to determine if an email message link or website is legitimate. Cyber-attacks by criminals that employ phishing schemes are so prevalent and successful nowadays. Hence, this project seeks to address fake URLs and domain names by identifying phishing website links. Therefore, having a web application that provides the user an interface to check if a URL is Phishing or legitimate will help decrease security risks to individuals and organizations.

1.3 Aim of Study

This project aims to detect phishing websites using machine learning and deep neural networks by developing a web application that allows users to check if a URL is phishing or legitimate and have access to resources to help tackle phishing attacks. 1.4 Objectives of the Study To accomplish the project's purpose, the following particular objectives have been established: i. dataset collection and pre-processing; ii. machine- learning model selection and development ; iii. development of a web-based application for detection;
iv. Integration of the developed model to a web application.

II. METHODOLOGY

An extensive review was done on related topics and existing documented materials such as journals, e- books, and websites containing related information gathered which was examined and reviewed to retrieve essential data to better understand and know how to help improve the system. The methodology used to achieve the earlier stated objectives is explained below. The dataset collection consists of phishing and legitimate URLs which were obtained from open-source platforms. The dataset was then pre-processed that is cleaned up from any abnormality such as missing data to avoid data imbalance. Afterward, expository data analysis was done on the dataset to explore and summarize the dataset. Once the dataset was free from all anomalies, website content-based features were extracted from the dataset to get accurate features to train and test the model. An extensive review was done on existing works of literature and machine learning models on detecting phishing websites to best decide the classification models to solve the problem of detecting phishing websites. Hence, Series of these machine learning classification models such as Decision Tree, Support Vector Machine, XGBooster, Multilayer perceptions, Auto encoder Neural Network and Random Forest was deployed on the dataset to distinguish between phishing and legitimate URLs. The best model with high training accuracy out of all the deployed models was selected then integrated into a developed web application. Thus, a user can enter a URL link on the web application to predict if it is phishing or legitimate.

Significance of the Project

According to Abdelhamid, Thabtah, and Abdel-jaber (2017), various fraudulent websites have been built on the World Wide Web in the previous decade to resemble reputable websites and steal financial assets from users and organizations. This type of online scam is known as phishing, and it has cost the internet community and other stakeholders hundreds of millions of dollars. As a result, robust countermeasures that can identify phishing are required. These are the challenges to be addressed in this project:
a. Reduce the rate of financial theft from users and organizations online.
b. Educate Internet Users on the deception of phishers.
c. Educate

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

Internet users on the countermeasures of a phishing attack.

Scope of the Study

This study explores data science and machine learning models that use datasets gotten from open-source platforms to analyze website links and distinguish between phishing and legitimate URL links. The model will be integrated into a web application, allowing a user to predict if a URL link is legitimate or phishing. This online application is compatible with a variety of browsers

Project Report Arrangement This report is structured into five chapters. The rest of the chapters are stated as follows: Chapter two describes the summary of works of literature review on the research and related works which is divided into five areas which are: Introduction, Theoretical Review, Conceptual Review, and Empirical Review and Summary of the chapter. Chapter three describes extensively the methodology, system analysis, and design of the system model. Chapter four describes the implementation of the machine learning model in the methodology and discussion on the result. Chapter five discusses the Summary, Conclusion, Recommendation, Limitation of Study, and Contribution to knowledge.

III. LITERATURE REVIEW

Overview of the Study This chapter offers an insight into various important studies conducted by excellent scholars from articles, books, and other sources relevant to the detection of phishing websites. It also provides the project with a theoretical review, conceptual review, and empirical review to demonstrate understanding of the project. 2.2 Theoretical Review Ankit and Gupta (2017) mentioned that Statistics show that according to Internet world stats ("Internet world stats usage and population statistics", 2014), the total numbers of Internet users worldwide are 2.97 billion in 2014; that is, more than 38% of the world population uses the Internet. Hackers take advantage of the insecure Internet system and can fool unaware users to fall for phishing scams. A phishing e-mail is used to defraud both individuals and financial organizations on the Internet. (“RSA Anti-Fraud Command Center”, n.d.) Said the Anti-Phishing Working Group (APWG) is an international consortium that is dedicated to promoting research, education, and law enforcement to eliminate online fraud and cyber- crime. In 2012, total phishing attacks increased by 160% over 2011, signifying a record year in phishing volumes. The total phishing attacks detected in 2013 were approximately 450,000 and led to financial losses of more than 5.9 billion dollars (“RSA Anti-Fraud Command Center”, n.d.). Total attack increases by 1% in 2013 as compared to 2012. The total number of phishing attacks noticed in Q1 (first quarter) of 2014 was 125,215, a 10.7 percent increase over Q4 (fourth quarter) of 2013. More than 55% of phishing websites contain the name of the target site in some form to fool users and 99.4% of phishing websites use port 80 ("Anti-Phishing Working Group (APWG)Phishing activity trends report first quarter", 8 2014). According to the APWG report in the first quarter of 2014, the second-highest number of phishing attacks ever recorded was between January and March 2014 ("AntiPhishing Working Group (APWG) Phishing activity trends report first quarter", 2014) and payment services are the most targeted industry. During the second half of 2014, 123,972 unique phishing attacks were observed ("APWG report", 2014). In the year 2011, total financial losses were 1.2 billion, and they rose to 5.9 billion dollars in 2013. The financial losses due to phishing attacks in 2014 and 2015 were 4.5 and 4.6, respectively, as shown in Figure 2.1 ("The RSA Current State of Cybercrime", n.d.). The growth of phishing attacks from 2005 to 2015 is shown.

Feature selection

Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data. It is also the process of automatically choosing relevant features for your machinelearning model based on the type of problem you are trying to solve. We do this by including or excluding important features without changing them. It helps in cutting down the noise in our data and reducing the size of our input data. Figure 2.5 shows the feature selection process. Feature selection models are of two types:

i. Supervised Models: Supervised feature selection refers to the method which uses the output label class for feature selection. They use the target variables to identify the variables which can increase the efficiency of the model.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

ii. Unsupervised Models: Unsupervised Feature selection refers to the method which does not need the output label class for feature selection. We use them for unlabeled data. Figure 2.6 shows the flow of the feature selection model

Feature extraction

Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process. Feature extraction is the name for methods that select and or combine variables into features, effectively reducing the amount of data that must be processed, while still accurately and completely describing the original data set (deepAI, n.d.). Why is Feature Extraction Useful? The process of feature extraction is useful when you need to reduce the number of resources needed for processing without losing important or relevant information. Feature extraction can also reduce the amount of redundant data for a given analysis. Also, the reduction of the data and the machine's efforts in building variable combinations (features) facilitate the speed of learning and generalization steps in the machine learning process. (DeepAI, n.d.) According to (Rami, Fadi & Lee, 2015), they have compounded important features that have proved to be sound and effective in predicting phishing websites. In addition, they have proposed some new features, experimentally assign new rules to some well-known features and update some other features. These feature selections include:

i. Address Bar based Features
ii. Abnormal Based Features
iii. HTML and JavaScript-based
iv. Domain-based Features

All the listed feature selection above consists of feature extraction which are guided by rules. The feature extraction is as follows: 2.2.3.1 address bar-based features (1) Using the IP Address If an IP address is used as an alternative to the domain name in the URL, such as "http://125.98.3.123/fake.html", users can be sure that someone is trying to steal their personal information. Sometimes, the IP address is even transformed into hexadecimal code as shown in the following link "http://0x58.0xCC.0xCA.0x62/2/paypal.ca/index.html"

Rule: IF { If The Domain Part has an IP Address → Phishing Otherwise → Legitimate

Algorithm and model evaluation (Performance Metrics) Algorithms or Models are used when it comes to machine learning includes the most important topics in Artificial Intelligence. It is further divided into Supervised and Unsupervised learning which can be related to labeled and unlabeled data analysis or data prediction. In Supervised Learning, we have two more types of business problems called Regression and Classification. Classification is a machine learning algorithm where we get the labeled data as input and we need to predict the output into a class. If there are two classes, then it is called Binary Classification. If there are more than two classes, then it is called Multi-Class Classification. There are so many classification algorithms available but for this project, let focus on the commonly used algorithms use by researchers to predict phishing websites. I. According to Rishikesh and Irfan (2018), three machine learning classification models such as Decision Tree, Random forest, and Support vector machine was selected to detect phishing websites. i. Decision Tree Algorithm (Rahul, 2017) One of the most widely used algorithms in machine learning technology. The decision tree algorithm is easy to understand and also easy to implement. The decision tree begins its work by choosing the best splitter from the available attributes for classification which is considered as a root of the tree. The algorithm continues to build a tree until it finds the leaf node. Decision tree creates training model which is used to predict target value or class in tree representation each internal node of the tree belongs to attribute and each leaf node of the tree belongs to the class label. In the decision tree algorithm, Gini index, and information gain methods are used to calculate these nodes. ii. Random Forest Algorithm (Saimadhu, 2017) Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on the concept of decision tree algorithm. Random forest algorithm creates the forest with several decision trees. A high number of a tree gives high detection accuracy. The creation of trees is based on the bootstrap method. In the bootstrap method features and samples of a dataset are randomly selected with replacement to construct a single tree. Among randomly selected features, a random forest algorithm will choose the best splitter for the classification, and as the decision tree algorithm; the Random forest algorithm also uses Gini index and information gain methods to find the best splitter. This

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

process will get continue until a random forest creates N number of trees. Each tree in the forest predicts the target value and then the algorithm will calculate the votes for each predicted target. Finally, the random forest algorithm considers the high-voted predicted

iii. Support Vector Machine Algorithm (Noel, 2016) Support vector machine is another powerful algorithm in machine learning technology. In the support vector machine algorithm, each data item is plotted as a point in n-dimensional space, and the support vector machine algorithm constructs separating lines for the classification of two classes, this separating line is well known as a hyperplane. 38 Support vector machine seeks for the closest points called support vectors and once it finds the closest point it draws a line connecting to them. Support vector machine then constructs separating line which bisects and perpendicular to the connecting line. To classify data perfectly the margin should be maximum. Here the margin is a distance between hyperplane and support vectors. In a real scenario, it is not possible to separate complex and nonlinear data, to solve this problem support vector machine uses kernel trick which transforms lower dimensional space to higher-dimensional space.

Kondeti, Konka, and Kavishree (2021) argued that phishing website detection is best done using Machine learning supervised classification models. Models such as Decision Tree Classifier, Random Forest Model, Multilayer Perceptron's, XGBoost Classifier, Auto Encoder, and Support Vector Machines are trained on a categorical input URL such as phishing (1) and legitimate (0). Since Rishikesh et al. (2018) has stated the significance of Decision tree classifier, Random Forest, and Support Vector Machines models on detecting phishing websites, let consider other classification models stated by Kondeti et al. (2021) for detecting phishing websites which are as follows: i. XGBooster is best used because it is not any different for classification or regression processes, it is meant for speed and performance. It will add gradient boosting to decision trees. ii. Auto Encoder Neural Network It is like a neural network that has the same no of input neurons as that of output neurons. It has fewer neurons in the hidden layers of the network that are called predictors. The input neurons pass information to the predictors and process the output. iii. Multilayer Perceptrons: Multilayer perceptrons are known as feed-forward neural networks. They are used to process multiple stages simultaneously and result in an optimal decision for the processed stage.

In 2016, the paper titled A Literature Review on Phishing Crime, Prevention Review and Investigation of Gaps by Anjum N. Shaikh, Antesar M. Shabat, and M.A. Hossain produced comprehensive research in which different reviews of previous pieces of literature are presented. The paper proposes an innovative approach of CRI - Crime, prevention Review and Investigation of research gap to combat the phishing in exploring the phishing problem. The strengths of the paper state the theory of phishing crime, a review of the antiphishing techniques offered by different research, and an investigation of the research gaps. This research aims to develop a practice to understand the growing threats of phishing under a theoretical model of CRI and discover new literature. The weakness of the paper is that there is need for a thorough research in terms of evaluating the effectiveness of countermeasures for phishing and developing a holistic technique to generating blacklists that are free from the demerits mentioned in the paper. Also, the paper didn't implement the research into a phishing detection system particularly against phishing emails since it is considered the most common way of attack. To address this weakness, this project will ensure proper and accurate evaluation from different resources on Blacklisting and Countermeasures to a Phishing attack. Also, this project will develop a phishing detection system to validate phishing emails

Implementation and result

(Rishikesh & Irfan, 2018) stated the implementation and result for detecting phishing websites. Let consider and evaluate the models used by Rishikesh et al. (2018) to carry out phishing websites detection. Rishikesh et al. (2018) said that the scikit-learn tool is best used to import Machine learning algorithms for phishing detection. A Dataset is divided into a training set and testing set in 50:50, 70:30, and 90:10 ratios respectively. Each classifier is trained using the training set and a testing set is used to evaluate the performance of classifiers. The performance of classifiers has been evaluated by calculating the classifier's accuracy score, false-negative rate, and false-positive rate from Table 2.2. The result shows that the Random forest algorithm gives better detection accuracy which is 97.14% with the lowest false- negative rate than decision tree and support vector machine algorithms. Result also shows that the detection accuracy of phishing websites increases as more datasets are used as the training dataset. All classifiers perform well

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

when 90% of data is used as a training dataset. Fig. 2.7 shows the detection accuracy of all classifiers when 50%, 70%, and 90% of data is used as a training dataset, and the graph clearly shows that detection accuracy increases when 90% of data is used as a training dataset and random forest detection accuracy is maximum than other two classifiers.

Conceptual Review

Phishing mechanism Figure 2.8 below shows a fake website is the clone of a targeted genuine website, and it always contains some input fields (e.g., textbox). When the user submits his/her details, the information is transferred to the attacker. An attacker steals the credential of the innocent user by performing the following steps: i. Construction of Phishing Site. In the first step, the attacker identifies the target as a well- known organization. Afterward, the attacker collects detailed information about the organization by visiting their website. The attacker then uses this information to construct the fake website. ii. URL Sending. In this step, the attacker composes a bogus e-mail and sends it to thousands of users. The attacker attached the URL of the fake website in the bogus e-mail. In the case of a spear-phishing attack, an attacker sends the e-mail to selected users. An attacker can also spread the link of phishing websites with the help of blogs, forums, and so forth (Kruegel, Kirda, Mutz, Robertson, & Vigna, 2005). iii. Stealing of the Credentials: when the user clicks on the attached URL, consequently, fake site is opened in the web browser. The fake website contains a fake login form that is used to take the credential of an innocent user. Furthermore, the attacker can access the information filled by the user. iv. Identity Theft: attacker uses this credential for malicious purposes. For example, an attacker purchases something by using the credit card details of the user. (Ankit & Gupta, 2017)

Taxonomy Of Phishing Attack

An Attacker performs a phishing attack by utilizing technical subterfuge and social engineering techniques [("RSA Anti-Fraud Command Center", n.d.), (Almomani, Gupta, Atawneh, Meulenberg, & Almomani, 2013)]. In social engineering techniques, attackers carry out this attack by sending bogus e-mail. Attackers often convince recipients to respond using the names of banks, credit card companies, e-retailers, and so forth (Tewari, Jain, & Gupta, 2016). Technical subterfuge strategies install malware into the user's system to steal credentials directly using Trojan and key logger spyware (Gupta, Tewari, Jain, & Agrawal, 2016). The malware also misaddresses users to fake websites or proxy servers. Attackers attached malware or embedded malicious links in the fraudulent emails and when the user opens the fraud hyperlink, malicious software is installed on the user's system, which collected the confidential information from the system and sent it to the attacker (e.g., key logger software sends the details of every key hit by the user). Attackers may also get remote access to the victim's computer and collect data whenever attackers want. In this paper, we focus on social engineering schemes, as it is the most popular way to steal victim's information by phishing. Classification of various phishing attacks is shown in Figure 2.9. (Ankit & Gupta, 2017).

Antiphishing technique

The method to a phishing scam starts with spreading bogus e-mail. To prevent a phishing attack, after receiving such e-mail, antiphishing techniques need to be activated, either by redirecting the phishing mail in the spam folder or by showing a warning when an online user clicks on the link of the phishing URL. The lifecycle of phishing attacks is shown in Figure 2.10. The following steps are involved in the phishing lifecycle: i. Step 1. The attacker creates the fake copy of a popular organization and sends the URL of the fake website to a large number of Internet users using e-mail, blogs, social networking sites, and so forth. ii. Step 2. In the case of a fake e-mail, every e-mail is first to pass through the DNS-based blacklist filters. If the domain is found in the blacklist, then email is blocked before it reached the SMTP mail server. There are also various solutions available which block fake e-mail based on structural features of mail (Almomani, Gupta, Atawneh, Meulenberg, & Almomani, 2013). iii. Step 3. If a fake e- mail bypasses the blacklist and features-based solutions and if the user opens the attached link in the email then some browser-based blacklist techniques block the site on the client-side. iv. Step 4. Some other solutions like the heuristic and visual similarities-based approaches also blocked the webpage only when the browser requests for any suspicious webpage

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

API (Application Programming Interface)

The work of an API here is that it serves as an intermediary between the web server and web application. A python framework called Django was used on the model prediction on the web application. These two ends communicate using a JSON (Javascript Object Notation) to send a request and receive a response. Figure 4.19 shows how the API was created by using Django python code for communication between the JSON dataset and feature extraction formula for detection phishing URL

Summary

Phishing attacks are a rapidly expanding threat in the cyber world, costing internet users billions of dollars each year. It involves the use of a variety of social engineering tactics to obtain sensitive information from users. Hence, Phishing techniques can be detected using a variety of types of communication, including email, instant chats, pop-up messages, and web pages. This project was able to categorize and recognize how phishers carry out phishing attacks and the different ways in which researchers have helped to solve phishing detection. Hence, the proposed system of this project worked with different feature selection and machine learning and deep neural networks such as Decision Tree, Support Vector Machine, XGBooster, Multilayer Perceptions, Auto Encoder Neural Network, and Random Forest to identify patterns in which URL links can be detected easily. The Model with the highest accuracy based on the feature extraction algorithm used to identify phishing URL from legitimate URL links was integrated to a web application where users can input website URL links to detect if it is legitimate or phishing.

Contribution to Knowledge

This project provides a new and faster way to help users detect if a URL link is phishing or legitimate and also provide them access to educational resources about phishing attacks.

Conclusion

The system developed detects if a URL link is phishing or legitimate by using machine learning models and deep neural network algorithms. The feature extraction and the models used on the dataset helped to uniquely identify phishing URLs and also the performance accuracy of the models used. It is also surprisingly accurate at detecting the genuineness of a URL link. 5.4 Recommendation

Through this project, one can know a lot about phishing attacks and how to prevent them. This project can be taken further by creating a browser extension that can be installed on any web browser to detect phishing URL Links.

REFERENCES

[1]. Abdelhamid, N., Thabtah F., & Abdel-Jaber, H. Phishing detection: A recent intelligent machine learning comparison based on models' content and features," 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, 2017, pp. 72-77, DOI: 10.1109/ISI.2017.8004877.

[2]. Anjum N. S., Antesar M. S., & Hossain M.A. (2016). A Literature Review on Phishing Crime, Prevention Review and Investigation of Gaps. Proceedings of the 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), Chengdu, China, 2016, pp. 9-15, DOI: 10.1109/SKIMA.2016.7916190.

[3]. Almomani, A., Gupta, B. B., Atawneh, S., Meulenberg, A., & Almomani, E. (2013). A survey of phishing email filtering techniques, Proceedings of IEEE Communications Surveys and Tutorials, vol. 15, no. 4, pp. 2070–2090.

[4]. Ashritha, J. R., Chaithra, K., Mangala, K., & Deekshitha, S. (2019). A Review Paper on Detection of Phishing Websites using Machine Learning.Proceedings of International Journal of Engineering Research & Technology (IJERT), 7, 2. Retrieved from www.ijert.org. Anti-Phishing Working Group (APWG) Phishing activity trends report the first quarter. (2014) Retrieved from http://docs.apwg.org/reports/apwg trends report q1 2014.pdf APWG report. (2014). Retrieved from http://apwg.org/download/document/245/APWG Global Phishing Report 2H

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

[5]. Ayush, P. (2019). Workflow of a Machine Learning project. Retrieved from https://towardsdatascience.com/workflow-of-a-machine-learning-projectec1dba419b94

[6]. Camp W. (2001). Formulating and evaluating theoretical frameworks for career and technical education research. Journal of Vocational Education Research, 26(1), 4- 25.

[7]. DeepAI (n.d.). About clinical psychology. Retrieved from https://deepai.org/machinelearning-glossary- and-terms/feature-extraction

[8]. Engine K., & Christopher K. (2005). Protecting Users Against Phishing Attacks. Proceedings of the Oxford University Press on behalf of The British Computer Society, Oxford University, 0, 2005, Retrieved from: https://sites.cs.ucsb.edu/~chris/research/doc/cj06_ph ish.pdf

[9]. Gandhi, V. (2017). A Theoretical Study on Different ways to identify the Phishing URL and Its Prevention Approaches: presented at International Conference on Cyber Criminology, Digital Forensics and Information

[10]. Security at DRBCCC Hindu College, Chennai. Retrieved from https://www.researchgate.net/publication/319006943

[11]. _A_Theoretical_Study_on_Different_ways_to_Identify_the_Phishing_URL_and_It s_Prevention_Approache s

[12]. Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2016). Fighting against phishing attacks: state of the art and future challenges, Neural Computing and Applications. Internet world stats usage and population statistics. (2014). Retrieved from http://www.internetworldstats.com/stats.htm.

[13]. Imperva. (2021). Phishing attacks. Retrieved from https://www.imperva.com/learn/application-security/phishing-attack-scam/

[14]. Kiruthiga, R., Akila, D. (2019, September). Phishing Websites Detection Using Machine Learning. Retrieved from

[15]. https://www.researchgate.net/publication/337049054 Phishing Websites Detection Using_Machine_Learning.