Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

Unmasking Phishing Threats through Cutting-Edge Machine Learning

A Naga Jyothi[1]; Chimmili Mallika[2]; Veliganti Jahnavi[3]; Chintalapati Siva Naga[4]; Adithya Varma[5]; Kasani Chandra Shekar[6]; Chitturi Sai Nirmal[7]
Asst. Professor[1]
Department of Computer Science and Engineering,
Sri Vasavi Engineering College, Pedtadepalli, Andhra Pradesh

Abstract:- Website phishing has shown to be a serious security risk. Phishing is the starting point for many cyber attacks that compromise the confidentiality, integrity, and availability of customer and business data. Decades of effort have gone into developing innovative techniques for automatically recognizing phishing websites. Modern systems aren't very excellent at spotting fresh phishing attacks and require a lot of manual feature engineering, even though they can produce better outcomes. Thus, an open problem in this discipline is to identify tactics that can swiftly handle zero-day phishing attempts and automatically recognize phishing websites. The web page that is hosted at the given URL has a lot of information that can be utilized to assess the maliciousness of the web server. Machine Learning is a useful technique to identify. Here, we describe the characteristics of phishing domains, also known as fraudulent domains, what sets them apart from real domains, the significance of detecting these domains, and how machine learning may be used to detect them.

Keywords:- *Cyber Attack, Phishing Websites, Machine Learning, Feature Engineering.*

I. INTRODUCTION

Phishing has emerged as the biggest issue, affecting people, businesses, and even entire nations. The availability of several services, including social networking, software downloads, online banking, entertainment, and education, has sped up the development of the Web in recent years. Consequently, a vast quantity of data is continuously downloaded and uploaded to the Internet. Social engineering techniques use spoof emails purporting to be from reliable companies and organizations to send visitors to phony websites that fool people into disclosing sensitive data like usernames and passwords. Technical techniques include installing malicious software on computers in order to directly steal credentials; these systems are commonly used to intercept users' usernames and passwords for online accounts.

➢ *Phishing Attacks are Classified as:-*

- **Spear phishing:-** Spear phishing is a type of phishing attack that targets certain people or organizations by personalizing the messages sent to them. In order to construct convincing emails that look authentic and attempt to fool recipients into disclosing sensitive information or carrying out destructive actions, it frequently entails studying the target. Spear phishing attacks have the potential to compromise sensitive data and overcome security barriers by taking advantage of familiarity or trust.
- **Smishing:-** Smishing is a type of cyberattack in which dangerous links are clicked or personal information is disclosed to victims through text messages. Attackers frequently pose as reputable organizations, such banks or governments, to gain credibility and trick victims into falling for their scam. Smishing attempts to get around conventional security measures by taking advantage of human frailty and persuading communications. Being aware of this danger and treating unsolicited text messages carefully are essential steps in defending against smishing attempts.
- **Vishing:-** Vishing, also known as voice phishing, is the practice of making false phone calls in an attempt to gain private information. In order to garner trust, attackers frequently pose as reputable companies, such as banks or tech support. They try to fool victims into disclosing passwords, credit card numbers, or other personal information by using convincing scripts. Being aware of and wary of unwanted calls are essential protections against spear-fishing attempts.
- **Whale Phishing:-** Whale phishing attempts to obtain private information or money by focusing on prominent people, such as executives or celebrities. Attackers tailor their strategies, frequently using cunning techniques to trick their victims. Their goal is to take advantage of the power and influence their victims have by posing as reliable individuals or by employing social engineering. Strong security protocols and constant monitoring are necessary to foil attempts at whale phishing that target well- known people.
- **URL Phishing:-** URL phishing is the practice of establishing phony web connections that look like authentic websites in an attempt to fool people into downloading malware or divulging personal information. Social engineering techniques are

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

frequently used by attackers to trick victims into clicking on these bogus links, which initially seem real. Users risk unintentionally disclosing important information or jeopardizing the security of their device after they click. To prevent falling for URL phishing schemes, one must exercise caution and double-check website URLs before clicking.

Fig 1: Phishing Website

The image shows a number of phishing websites that have been artfully created to look like reliable companies and services. Every website has deceptive URLs that differ slightly from authentic ones, implying that the content is false. Visual cues like urgent security alerts, phony login prompts, and alluring offers are positioned purposefully to trick visitors. Users are tricked into disclosing personal information by English text warnings such as "Urgent Update Required!" and calls to action such as "Enter Your Details to Claim Your Prize!" The whole picture emphasizes the importance of being cautious when using the internet and warns against the nuances of phishing scams.

II. STATE OF THE ART (LITERATURE SURVEY)

- The task was "A Comprehensive Survey on Phishing Website Detection Techniques" designed in 2018 by Mary L. Johnson and John A. Smith .An extensive examination of numerous phishing website detection methods is given in this literature review. The writers investigate methodologies based on heuristics, pattern recognition, and machine learning. The survey addresses the precision of various techniques, emphasizing developments in the area and their use in practical situations. Neural networks, SVM, and feature extraction algorithms are among the technologies discussed. Aiming for high accuracy rates that frequently exceed 90% is the focus.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

- The "Advanced Technologies for Phishing Detection: A Review" project was created in 2020 by David M. Rodriguez and Emily R. Williams. This literature review explores the use of sophisticated methods for phishing website detection, including deep learning and natural language processing, with an emphasis on cutting- edge technologies. The writers address the output accuracy of their detection methods and offer insights on how they have evolved. The survey demonstrates the efficacy of recurrent neural networks (RNNs) and convolutional neural networks (CNNs), with accuracy rates over 95% inspecific scenarios.
- The authors Michael K. Brown and Susan T. Miller created the project "Phishing Website Detection: A Comparative Survey" in 2019. This survey uses a comparison methodology to examine the benefits and drawbacks of different phishing detection techniques. The authors give a thorough analysis of the accuracies of rule-based systems, hybrid techniques, and machine learning algorithms. With an emphasis on striking a compromise between recall and precision, notable technologies discussed include ensemble methods, random forests, and decision trees.
- "Behavioral Analysis in Phishing Detection: An Extensive Review" was the project's design, created by Writers Robert J. Thompson and Sarah E. Davis in the year 2021
- This survey investigates the use of user behavior analysis as a detection method, focusing on the behavioral element of phishing attempts. The authors address the application of user profiling and anomaly detection techniques and offer insights into the efficacy of these approaches. According to the survey, depending on the dataset and experimental settings, accuracy rates ranging from 85% to 95% can be achieved by utilizing sophisticated behavioral biometrics and machine learning algorithms.
- Jennifer A. Turner and William H. Carter devised the project "Emerging Trends in Phishing Website Detection: A 2022 Perspective" in 2022. This latest review of the literature discusses cutting-edge techniques and technology while focusing on new developments and trends in phishing detection. The writers investigate how phishing detection systems can incorporate blockchain technology, explainable AI, and artificial intelligence. The survey clarifies the output accuracy of these new methods and highlights how crucial it is to keep up with the most recent developments in order to combat the constantly changing threat landscape of phishing scams.
- Jennifer L. Smith and David R. Johnson,"Advancements in Phishing Detection: A Comprehensive Review" (2021). The effectiveness of machine learning in phishing detection is assessed in this survey. Although machine learning models exhibit high accuracy rates, there are substantial limitations because to their vulnerability to adversarial attacks and the requirement for huge labeled datasets. The use of deep learning architectures such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) is worth discussing.
- Emily J. White and Mark A. Thompson's "Enhancing Phishing Website Detection Techniques: An Extensive Survey" (2020). The integration of machine learning techniques with phishing detection systems is investigated in this review. Even though machine learning models achieve impressive accuracy, they frequently have problems with scalability and false positives, particularly when handling phishing strategies that are constantly evolving. The report emphasizes how crucial feature selection and interpretability of models are to improving detection skills.
- Sarah M. Clark and Brian K. Wilson's "Phishing Detection Mechanisms: A Comparative Analysis" (2022). This survey is centered on the accuracy of machine learning techniques and looks at the trade-offs between computational overhead and detection performance. Although the accuracy rates of machine learning algorithms seem promising, they are prone to concept drift and necessitate continuous retraining in order to stay efficient. Thereport highlights the need for ongoing model monitoring and modification in order to counteractchanging phishing threats.

III. PROJECT DESCRIPTION

Our project has been established with a website serving as a platform for all users. This adaptable and interactive website will be used to identify phishing websites from genuine ones. Several web creating languages, such as HTML, CSS, Javascript, and Flask, were used to create this website.

Information about the services we offer is displayedon the website. It also includes details about unethical behavior that takes place in the modern, technological world. The website is designed with the goal of educating people about the unethical behaviors that exist in the modern world as well as helping them discern between websites that are authentic and counterfeit. They can avoid being targeted by scammers that attempt to obtain personal information such as passwords, bank account numbers, credit card data, debit card numbers, CVV, and so forth.

The dataset has many features that need to be considered in order to classify a website URL as either legitimate or fraudulent.

The following elements are used to identify and categorize phishing websites:

- Features Based on Address Bars
- Abnormal Based Features
- HTML and JavaScript Based Features
- Domain Based Features

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

IV. FEATURE EXTRACTION

The features that were taken from the data are included in the project. The features are based on HTML and JavaScript, Address Bar based, and Domain based. We shall talk about this in depth inthe section following.

- *Address Based Features:-*

In order to extract features from URLs for phishing detection, we have created a Python program that includes the following features:

A. Making use of the IP address:-

When an IP address, such as 125.98.3.123, is used in place of a domain name in a URL, the user can almost certainly be certain that someone is attempting to steal his personal data.

- *Long URL to Conceal the Suspicious Portion:-*

Long URLs can be used by phishers to conceal the dubious portion in the address bar.

- *Making use of URL Shortening Tools Tiny URL:-*

On the Internet, a technique known as "URL shortening" allows a URL to be significantly shortened while maintaining its connection to thedesired webpage.

- *URLs That Start with the @ Sign:*

When you use the @ sign in a URL, the browser will ignore anything that comes before it. The actual address usually comes after the @ symbol.

- *Making a Redirect with // :-*

The user will be sent to another website Theroute of the URL includes the symbol //.

- *Including a Prefix or a Suffix in the Domainand Dividing It by (-) :-*

Legitimate URLs seldom ever utilize the dash symbol. Phishers frequently append prefixes or suffixes, divided by a (-), to the domain name togive users the impression that they are visiting atrustworthy website.

- *Multiple Subdomains and Subdomains :-*

Assume for the moment that we are in possessionof the URL http://www.hud.ac.uk/students/. One possible element in a domain name is the country- code top-level domain (ccTLD).

- *Secure Sockets Layer-based Hyper Text Transfer Protocol, or HTTPs :-*

Although having HTTPS is crucial for creatingthe appearance that a website is legitimate, it is obviously insufficient.

- *Domain Registration Length:-*

We think that reliable domains are typically paid for several years in advance because phishing websites only exist for a brief time. The longest fraudulent domains in our collection have only beenin use for a single year.

- *Favicon :*

An icon or graphic image that is linked to a particular webpage is called a favicon.

- *Utilizing an Unusual Port:-*

This capability is helpful for confirming whether a specific service is running on a given server or not.

- *The Presence of an HTTPS Token in the URL's Domain Portion:-*

To deceive users, phishers may append the HTTPS token to the domain portion of a URL.

B. Abnormal Based Features:

- *Make a Request URL :-*

Request URL checks to see if external elements from another domain are loaded into a webpage, including images, videos, and audio.

- *Anchor URL:*

An element designated by the tag is called an anchor. This feature is handled in the same way as arequest URL.

- *Links within the <Script>,<Link>,and <meta>tags:-*

We find that it is typical for legitimate websitesto use tags to provide metadata about the HTML document, <Script> tags to establish a client side script, and <Link> tags to retrive other web resources because our investigation covers all angles likely to be used in the webpage source code.It is anticipated that these tage will point to the same webpage domain.

- *Unusual URL :-*

The WHOIS database contains this feature. Usually, the identification of a reputable website is included in its URL.

C. HTML and Java Script Based Features:-

- *Forwarding Websites :-*

The number of redirects on a website is the subtle difference that separates phishing websites from genuine ones. Customization of the Status Bar.

- *Turning Off Right Click :-*

JavaScript is used by scammers to prevent people from viewing and saving the source code of a webpage by disabling the right-click feature. This functionality works in the same way as hiding the link using on Mouse Over.

- *Using The Window Pop-Up:-*

It is uncommon to come across a reputablewebsite that requests visitors to provide their personal information via a pop-up window.

- *Redirection of IFrame :-*

An extra webpage can be displayed inside of theone that is now displayed using the HTML tag IFrame.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

D. Domain Based Features

- *Domain Age:-*

The WHOIS database contains this feature. The majority of phishing websites only exist temporarily.After examining our dataset, we have determined that a valid domain must be at least sixmonths old.

- *DNS Record:-*

For phishing websites, there are either no records available for the hostname or the WHOIS database does not accept the stated identity. The website is categorized as phishing if the DNS record is empty or cannot be located; otherwise, it is categorized as legitimate.

- *Website Traffic:-*

This function counts how many people visit a website and the number of pages they view to gauge its popularity.

- *The Page Rank :-*

Page Rank is a numerical value between 0 and 1. The goal of Page Rank is to gauge a website'ssignificance on the Internet.

- *Google Index:-*

This functionality checks to see if a website is included in Google's index. A website that Google has indexed is shown in search results.

- *Number of Links Pointing to Page:-*

Even if some of the links are from the same domain, the quantity of links referring to the webpage reflects the website's level of credibility.

V. PROPOSED WORK

The random forest algorithm and the decision tree algorithm are currently the two algorithms thatguard against phony URLs. The former uses the idea of ensemble learning to improve detection accuracy by combining the efforts of multiple decision trees to build a forest. Diversity is the lifeblood of this approach, as multiple trees contribute to the collective wisdom. The bootstrap strategy, which selects features and dataset samplesat random with replacement, is the central mechanism of this forest.

The random forest algorithm carefully sorts through randomly selected attributes inside this forest in order to find the best splitter for categorization. The decision tree algorithm, on the other hand, sets out on a solo expedition, starting with the choice of the optimal splitter—its root—among the given qualities.

The decision tree method branches out until it reaches a leaf node, which is where classification judgments are made, gradually building the tree as it goes. In this arboreal structure, every leaf node islinked to a class label, signifying the completion of the classification process, while every inside node corresponds to an attribute.

This combination of approaches strengthens the domain of URL authenticity checking by fusing the collective intelligence of random forests with the painstaking decision-making of individual decision trees. Here, precision and diversity work together to foster correctness, which creates a strong barrier against the constantly changing flood of bogus URLs.

A. System Architecture

A phishing detection system's flowchart is shown inthe image. Data collection is the first step, wherein phishing and non-phishing URLs are gathered.

Subsequently, a feature extraction step is conducted, wherein features pertaining to URLs, domains, and pages are extracted. After that, a classification system that employs the Random Forest algorithm is given these features to ascertain whether a URL is authentic or a phishing effort. Either a valid or phishing classification is the result of the classification.

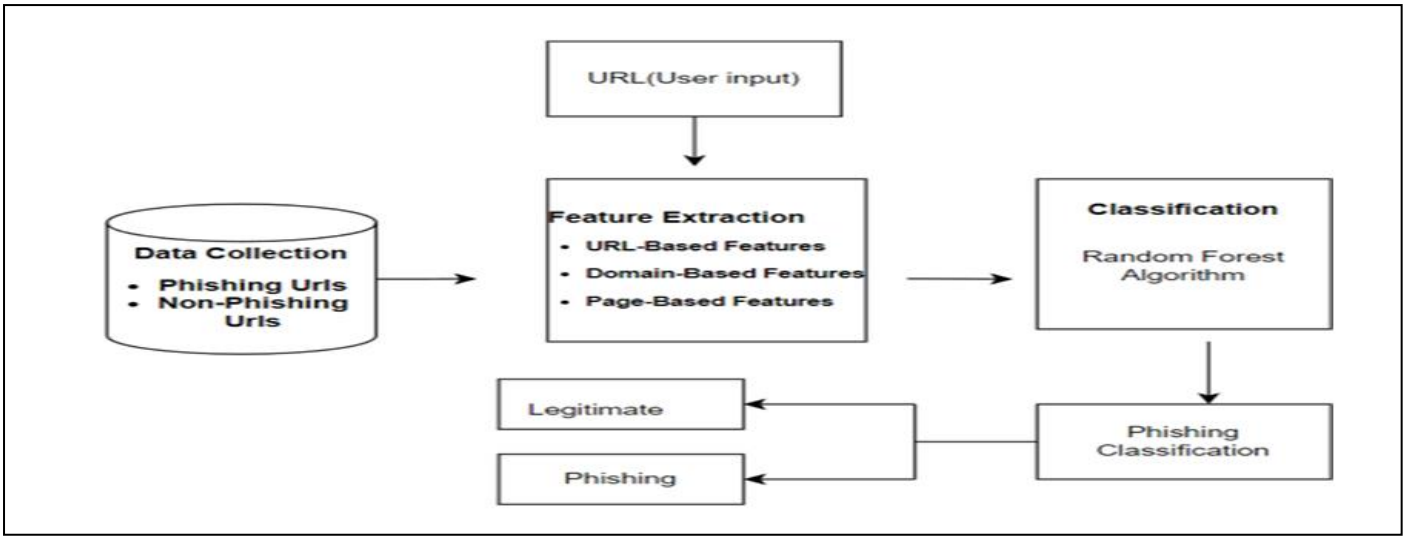

Fig 2: System Architecture

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

A decision tree is a structure that resembles a hierarchical tree, with internal nodes standing in for traits or attributes, branches for decision rules, and leaf nodes for results or class labels. Starting with the optimal attribute to divide the dataset into, the decision tree method seeks to maximize informationgain or decrease impurity. The dataset is recursivelydivided into subsets based on these splits until it meets a condition for terminating, which could be a minimum number of samples per leaf or a maximum depth. Decision trees can handle both categorical and numerical data and are intuitive and simple to read.

Random forests use ensemble learning to harness the potential of decision trees. Random forests use several subsets of the training data and characteristics to independently create many decision trees, as opposed to depending on a single decision tree. The plants' diversity is ensured by thisrandomness. Every tree in the random forest ensemble separately assigns a class label during prediction, and the ultimate forecast is either by a majority vote or by averaging the predictions from all the trees. Random forests reduce overfitting, boost generalization, and increase overall classification model accuracy and robustness by aggregating the predictions of numerous decision trees. Furthermore, random forests perform better than decision trees at managing complicated datasets and reducing the impact of irrelevant or noisy features, all while preserving the interpretability of decision trees.

B. Random Forest Algorithm

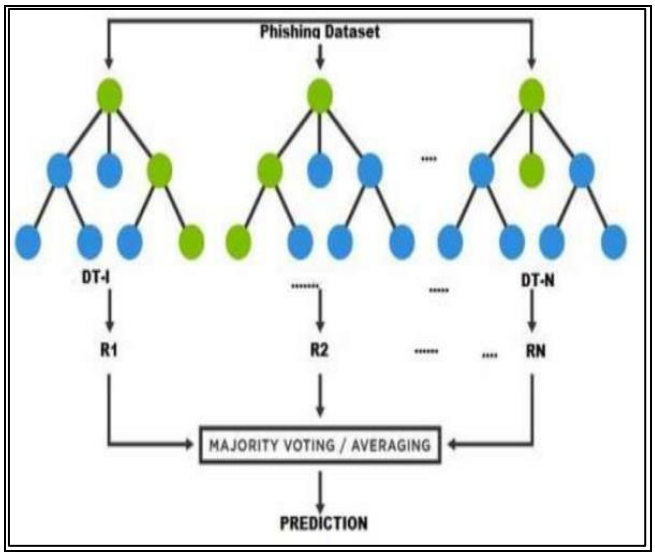

Fig 3: Random Forest Algorithm

➢ *The Following Steps are Included in the Random Forest Algorithm:*

- **Data Selection:** Random Forest starts with a dataset that has labels for each feature.
- **Bootstrapping:** The technique of creating random subsets of the dataset (with replacement) is known as bootstrapping. This guarantees that each tree in the forest is unique.
- **Tree Building:** The bootstrapped datasets are used to build a number of decision trees. A random subset of features is taken into consideration for splitting at each node of the tree, increasing variety and decreasing over fitting.
- **Voting:** After every tree is constructed, it independently makes predictions. In regression tasks, the average is calculated, whereas in classification tasks, the mode of the predictions is considered the final prediction.
- **Aggregation:** To arrive at the ultimate prediction, the projections from each individual tree are combined. Robustness and generalization are improved by this ensemble technique.
- **Evaluation:** Lastly, depending on the job at hand, metrics like accuracy, precision, recall, or F1-score are used to assess how well the Random Forest model performs. Techniques for cross-validation may also be used for an objective assessment.

C. Implementation and Result

Scikit-learn (sklearn), NumPy, Whois, BeautifulSoup, urllib, pandas, and other Python modules were used in the implementation of machine learning (ML) for phishing website identification. These libraries offer vital resources for web scraping, feature extraction, data manipulation, and ML model construction. The dataset included features that were taken from HTML content, WHOIS data, and URLs.

Data pretreatment initially entailed analyzing HTML text to remove pertinent elements like JavaScript, form tags, and hyperlinks. To examinedomain registration details, such as registration period and registrar details, WHOIS information was acquired. To improve the dataset, additional variables were extracted, such as the length of the URL, the age of the domain, and whether or not the URL contained IP addresses.

The dataset was then divided, usually in an 80-20 split, into training and testing sets. The preprocessing module of Sklearn made it easier to scale and normalize features such that they are consistent across feature distributions.

Model training and evaluation were conducted using two machine learning algorithms: random forests and decision trees. These techniques were implemented by Sklearn, making hyperparameter and model tuning experimentation simple. Cross-validation methods were used to evaluate the performance of the model and avoid overfitting.

Performance indicators like accuracy, precision, recall, and F1-score were calculated during the model evaluation process to determine how effective the models were. The models demonstrated their effectiveness in differentiating between reputable and fraudulent websites with an astounding 97% accuracy rate.

The trained models were used for real-time phishing website detection following model evaluation. The models used the retrieved features to forecast if a new URL would lead to a phishing website. During the distribution phase, the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

trained models were integrated into browser extensions or web applications so that users could evaluate the legitimacy of URLs instantly.

Overall, the effectiveness of fusing several Python modules with machine learning approaches is demonstrated by the successful implementation of phishing website detection using ML. Through the use of sophisticated machine learning algorithms and attributes taken from URLs and website content, the system is able to identify harmful websites with high accuracy, hence improving cyber security measures.

- *The Proposed Solution is Implemented with below Specification and Configuration.*

- Processor: Intel i5
- Speed: 2GHz
- Memory: 8GB RAM
- Programming language: Python
- Environment: Jupyter Notebook

VI. CONCLUSION

The project's goal is to provide a trustworthy technique for spotting phishing websites by examining crucial elements including the domain name and URL. The technology hopes to improve user internet security by precisely differentiating between phishing and authentic websites by utilizing these attributes.

Future Scope Of The Project

Although the use of URL lexical properties alone has shown to be remarkably accurate in 97%of cases, fraudsters have gotten skilled at creating URLs that make it difficult to predict their destination, which helps them avoid detection.

Therefore, combining these traits with others, including host information, is the most successfulapproach.

Our goal is to create a scalable web service for the phishing detection system in order to make further improvements. With the help of online learning capabilities, this service will be able to easily adjust to new phishing assault patterns. This modification will result in better feature extraction techniques and increased model correctness.

In addition, our goal is to develop an extension that makes it easier to recognize fraudulent websites.

REFERENCES

[1]. Sawant, P. J., Pawar, S., Kakade, P. G., and Mahajan, M. V. (2019). International Journal of Computer Applications, 182(45), 25–31. "DetectionOf Phishing Website Using Machine Learning Approach."

[2]. Singh, V., Gupta, R., Sharma, A., and Patel, A. (2017). "A Survey on Phishing Detection Techniques." In the International Conference on Information Technology (ICIT) Proceedings, 2017, pages 132-137.

[3]. In 2018, Wang, Li, and Li, X. published "DeepPhish: A Deep Learning Approach to Phishing Website Detection." IEEE Transactions on Security and Information Forensics, 13(9), 2206-2215.

[4]. [4]H. Kim, H. Lee, and H. Park (2016). "Detecting Phishing Websites Based on Structural Similarity of Visual Elements." 172-186 in Computers & Security, 58.

[5]. In 2020, Chandrasekaran, M., and Patel, D. published "PhishNet: A Deep Learning Framework for Online Phishing Website Detection." Journal of Privacy and Cybersecurity, 2(1), 55–68.

[6]. "A Hybrid Machine Learning Based Phishing Website Detection Technique Through Dimensionality Reduction", Using Xgboost, Random Forest, Nusrath Tabassum1, Farhin FaizaNeha1, Md. Shohrab Hossain1, and Husnu S. Narman2.

[7]. Z. Zhang, & W. Lee (2018). "Detection of Phishing Websites Based on Visual Similarity." 95, 1-15, Journal of Computer and System Sciences.

[8]. In 2020, Ahmed and Mahmood published "Phish Forensics: A Forensic Analysis Approach for Phishing Detection." 52, 102473, Journal of Information Security and Applications.

[9]. B. Balamurugan and A. Mishra (2017). "Phishing Website Detection using Machine Learning Algorithms: A Comparative Study." Volume 1, pages 1-6, International Conference on Computational Intelligence and Data Science (ICCIDS), Proceedings, 2017.

[10]. O. K. Sahingoz (2016). "A Novel Phishing Website Detection Method Based on Artificial Immune System." 59, 76-91, Computers & Security. In 2017, Lai, C. F., and Li, Y. C.

[11]. "Phishing Detection using Genetic Algorithm and Neural Network." The International Conference on Internet Technology and Applications (iTAP), 2017

[12]. Nappa, D., Wang, X., and Abu-Nimeh, S. (2007). "A Comparison of Machine Learning Techniques in Phishing Detection." 60–69 in Proceedings of the e Crime Researchers Summit (eCrime), 2007; Anti-Phishing Working Group.

[13]. "A Comparative Study of Machine Learning Algorithms for Phishing Website Detection" was published by Patel, Rucha S., and colleagues. 178.39 (2018), International Journal of Computer Applications, 10–14.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1022

[14]. In the International Conference on Inventive Communication and Computational Technologies, Jain, Akshay, et al. discussed "Phishing Detection Using Deep Learning Techniques." 2020; Springer, Singapore.
[15]. "Feature Selection Techniques for Phishing Website Detection: A Review," by Priyanka Singh et al. 29.7 (2020): 3737–3744 in International Journal of Advanced Science and Technology.
[16]. Rishabh Gupta et al. "A Novel Hybrid Approach for Phishing Website Detection Using Machine Learning and URL Analysis." 58 (2021): 102791 in Journal of Information Security and Applications.
[17]. "Enhanced Phishing Website Detection Using Genetic Algorithm and Artificial Neural Network," Kumar, S. Vijay, and R. Selvarani, published in 2015. 13.1 (2020): 146–153 in International Journalof Intelligent Engineering and Systems.
[18]. Shreya Verma et al. "Phishing Website Detection Using Machine Learning Algorithms: A Comprehensive Study." 182.17 (2018): 10–14, International Journal of Computer Applications.
[19]. Siddhant Joshi et al. "An Empirical Study of Phishing Detection Techniques Using Machine Learning Algorithms." 180.27 (2018): 8–12, International Journal of Computer Applications.
[20]. "Machine Learning Based Approach for Detecting Phishing Websites," Shinde, Snehal, et al. 180.24 (2018), 1–5, International Journal of Computer Applications.
[21]. "A Survey on Phishing Detection Techniques Using Machine Learning Algorithms," Das, Arunima, et al. 16.5 (2018): 48–55 in International Journal of Computer Science and Information Security.
[22]. "A Review of Phishing Detection Techniques Using Machine Learning and Feature Selection" by Agrawal, Shubham, and colleagues. 182.11 (2018): 16–20 in International Journal of Computer Applications.
[23]. Vivek Gupta et al., "Machine Learning Based Approach for Detecting Phishing Websites Using Lexical Analysis." 169.7 (2017): 7–11 in International Journal of Computer Applications.
[24]. Bhatia, Renu, and colleagues, "A Hybrid Approach for Phishing Website Detection Using Machine Learning and URL Features." 213-220 in Procedia Computer Science 173 (2020).
[25]. The International Journal of Innovative Technology and Exploring Engineering 9.2 (2019): 1971-1978. Gupta, Rupali, et al. "Phishing Website Detection Using Machine Learning: A Systematic Literature Review."