Identifying Phishing Websites with Machine Learning

1st Harsh Chauhan
Student, Faculty of IT and Computer Science
Parul University
Vadodara, India
200511202008@paruluniversity.ac.in

2nd Prof. Dharmendrasinh Rathod
Assistant Professor, Faculty of IT and Computer Science
Parul University
Vadodara, India
dharmendrasinh.rathod@paruluniversity.ac.in

***Abstract*—The Internet has turned into a piece of our life as everything is becoming conceivable by the click of our finger. Nevertheless, it has similarly given opportunities to perform cybercrimes and harmful activities like Phishing. In the Phishing, attackers try to deceive their victims to steal information by social engineering or making counterfeit sites to take basic data like record ID, username, passwords from people and associations which would brings about extreme monetary misfortune, loss of reputation and client's trust. Despite the fact that numerous techniques have been proposed to identify phishing sites, One of the best strategies to distinguish such malevolent exercises is Machine Learning. This is because most phishing sites have some common sorts of features which can be recognized by AI strategies. The objective of this research is to foster the new system to guard and utilize various ways to categorize websites. In this paper, an outline of the diverse AI approaches is introduced just when contrasted with observe which Machine Learning algorithm served the best in identifying those fake sites.**

***Index Terms*—Phishing, Anti-Phishing, Phishing Detection, Machine Learning, Cybercrime**

I. Introduction

Phishing attacks are the act of sending false communication that which feels like it seems to come from a legitimate source. It is generally done through email [1]. The objective includes the robbery of the sensitive information like debit card or credit card and login data, or to introduce malware on the victim's machine. Phishing incorporates different kinds of methods like link manipulation, filter evasion, website fabrication and social engineering. Most normal methodology of phishing attack is to set up a spoofing site page which emulates the original page. Phishing is like fishing in the water, however rather than attempting to get a fish, meanwhile, attackers attempt to take client's personal data. At the point when a user opens a phony webpage and enters some sensitive information, for example, username and password the phisher acquires the data of that user which can be utilized for malevolent exercises.

Phishing sites look basically the same in appearance to genuine sites to attract sizable number of users. A phishing attack is the point at which an attacker sends an email or the URL claiming to be a person or thing that he/she isn't, to get vital data from the person. The victim as to his/her interest or a feeling of frenzy or urgency, they enter the details, similar to a username, password, or debit/credit card number, they are probably going to fall in such trap without figuring it out. The recent example is a Gmail phishing trick which focused on around 1 billion Gmail clients around the world. Phishing procedures utilized against victims by hackers or attackers to fool the user into entering their sensitive credentials, for example, usernames, passwords, and credit card details into an illegitimate entity as a website. In this kind of attack, unauthentic entities camouflage themselves as genuine and reliable sources. Along these lines, users get deceived by the look and feel of the phony site which is practically indistinguishable from the real site. For the most part, attackers use banking and installment sites, social media websites and E-Commerce websites to draw their expected victims. Thus, phishing tops the way to deal with convey ransomware and other malware. On the off chance that any association succumbed to such an attacks regularly supports extreme monetary misfortunes, declining market shares, reputation, and consumer trust. Depending upon scope, a phishing attempt may give alarming opportunity to any business assuming that they are not cautious with regards to where precisely their information is going through.

A. Classification of Phishing Attacks

There are numerous methods of Phishing. The attackers are always on the way to find new alternatives as well as the techniques to steal information and their credentials. There are various ways that a phishing attack can be executed. Such characterization of Phishing attacks is given beneath [2]:

Technical Subterfuge: Technical Subterfuge incorporates planting crimeware onto PCs to steal credentials directly, often using systems to intercept consumers' online account information like user identification and passwords– and to corrupt local navigational infrastructures to misdirect consumers to counterfeit websites [3].

Keyloggers: A keylogger (short form for keystroke logger) is programmed tool which logs and tracks the keys on the keyboard when struck on during typing, consistently in a secret way so users don't realize that their activities are being monitored. It incorporates noxious intent to assemble user's account data, credit/debit card numbers, usernames, passwords, and other private information[4].

Malware: Malicious Software or Malware, is a term for viruses, worms, trojans, and other dangerous computer pro-

grams hackers use to wreak computer systems and gain access to sensitive information [5].

DNS Poisoning: Domain Name System (DNS) poisoning and spoofing of cyberattacks that exploit the weaknesses of the DNS Server to redirect traffic from authentic servers towards malevolent ones [6].

Social Engineering: Social Engineering is the term which incorporates leading malicious activities achieved through human cooperations. They control the user mentally into committing a type of security errors or giving away sensitive information. In Phishing, the attacks are made through Websites, Emails and Mobile applications. [7].

Spear-Phishing: Spear phishing is the act of sending an email to explicit and well-researched targets while implying to be a trusted sender. The point is to either taint device with malware or persuade the victim to hand over information or money [8].

Whaling: A whaling attack, otherwise called whaling phishing or a whaling phishing attack, is a particular sort of phishingattack that targets representative who holds prominent in an organization, like the CEO, to steal sensitive information froma company [9].

Smishing: Smishing is a cyberattack that uses misleading text messages to deceive victims. Their aim is to trick the user into believing that a message has arrived from a trusted person or organization, and then convincing the user to take action that leads to give them information (like bank account login credentials, for example) or access to your mobile device [10].

Vishing: Vishing or voice phishing, is a type of attack or scam in which the fraudsters try to convince their victims to give away the valuable information over the phone call [11].

Mobile Applications: In mobile apps, the attacker can try to steal the information through SMS, MMS, camera, through any social media, or even by installing an app from an untrusted source. Apps that are from untrusted sources can be leaking away from information like phone number, online activity, device information, etc.

B. Application Areas

The utilization of Internet applications has expanded enormously as of recent years. This has prompted to new wave of phishing and targeting the users of these applications. These phony sites show appealing offers on social networking applications to draw end-users. A portion of the applications that the phishing destinations depended on are:

Social Media: These days getting to social media has become more straightforward with the increment of smart devices. Also, there are such various social media applications on the Internet presently, and consistently new users are joining these social media applications. This expansion in users has opened more ways for attackers as there will be more potential victims. Consequently, attackers continue to make phishing sites on social networking brands and claims to offer types of services.

Dating/Matrimonial websites: Fake offers are made on dating or matrimonial sites by luring the victim to enter their login credentials to proceed further to chat with the opposite gender.

Blogs: Many attractive blogs are used in login as unaware users might enter sensitive information on a phishing website.

Gaming: Some games provide in-game currency which can be bought by trading real-time money. The phishing website tricks players by offering free in-game currencies for that particular game.

Banking: The attacker uses social engineering skills to make the victim panic and fall into a phishing scam by taking them to a phishing website and asking them to give away their sensitive banking information.

Job Recruitment: A fraudulent page deceives the job seekers to a job portal where he/she is promised for their job placement and in the process, information is exchanged which might include the job seeker's identity data or bank details.

E-commerce: The attacker creates a fake e-commerce website and the victim feels like a good or affordable price on that online shopping website and pays for the product.

C. Challenges

- As Attacker's techniques keep evolving and their phishing attacks are getting more sophisticated, their systematic attack strategies are becoming harder for even security professionals to keep up.
- Using database phishing prevention techniques such as blacklists and whitelists have a huge limitation due to their requirement to update the databases sometimes taking several days, whereas phishing campaigns normally take significantly lower times (a few hours) in their attacks.
- People with less or no knowledge about the internet are most likely to become victims of phishing attacks.
- Phishing attacks affect people throughout the globe, which means that the attack could be conducted from an international source. This makes it difficult to file lawsuits against them.

II. Literature Review

In [12], they have proposed a three-pronged approach to deal with phishing. The approaches consist of preventing phishing using blacklist and filters, detecting phishing using the indicators which are based in the browsers and different detection tools like Anti-Phishing Toolbars and stakeholder training which would feel like a game-based AI training. they have introduced how various kinds of phishing attacks are utilized by the hackers, the problems and challenges due to phishing and the solution approach that ought to be taken to counter such phishing attacks.

In [13], First, they analyzed different features of the URL. Second, they checked authenticity of the site by knowing where the site was being hosted and third, they utilized the visual appearance-based analysis for checking how much the website is real. The proposed framework methodology monitors the traffic on the user's system and afterwards the

URL is compared with the whitelist of the genuine domain websites.

In [2], they have collected different types of machine learning and deep learning approaches of phishing detection techniques used to counter phishing attacks and explained the content of the different previous researches conducted under the anti-phishing techniques using the machine learning approach. They have classified various anti-phishing methods and their machine learning approaches with their strengths and limitations from the literature survey.

In [14], they utilized three techniques which would assist with being prepared when the phishing attack will be executed. they combined several weak learners into a stronger one, this is perhaps the primary reason why ensemble-based learning is used in practice for most of the classification.

In [15], They have developed a system that uses machine learning techniques to classify websites based on their URL. The pruned decision tree was able to generate better accuracy with least false positive rate.

III. Proposed Work

There are different strategies which are proposed to keep away from phishing attacks by investigating some of website's behavior. Indeed, even a client can likewise foresee some of such attacks via training and knowledge about phishing sites. However, the approach might not be always keep working as we as the internet users would visit hundreds of websites in a day and predicting every website visited through training and knowledge is practically not possible. One more alternative to distinguish phishing site is by utilizing a software whose fundamental task is to monitor every single site visiting and recognizing the dubious one before user continues to enter it any further. The software should be proficient to investigate the content from other websites, emails, social media, and numerous other ways of getting URL link to a website.

A software with machine learning approach have proved to be the best and a powerful tool which helps to classify phony websites. These methods require training data and for that, there are many samples of websites which would help to train the machine learning model. Multiple features are extracted in the dataset from the websites which shows the originality of the website. So, multiple machine learning algorithms have been used to detect phishing websites like K-Nearest Neighbors, Logistic Regression, XGBoost, Random Forest and Decision Tree. These machine learning models provide the accuracy performance with the training data and testing data and shows the result of their accuracy.

IV. Methodology

A. Dataset

One of the biggest challenge in this research was about availability of a precise dataset. Even though there were many researches about anti-phishing are done, there were not enough dataset that was produced for the research purpose. Another factor which was a hurdle in finding proper dataset was to have different and more features of the phishing website. So, the dataset from kaggle is used which contains sample of 11056 websites. The features in the dataset used to detect phishing websites are as follows [16]:

Using the IP Address: If an IP address is used for a replacement of the domain name in the URL, such as "http://125.45.13.133/real.html", users must make sure that someone is trying to steal their personal credential or information. Sometimes, the IP address is even converted into hexadecimal code as shown in the following link "http://0x58.0xCA.0xCC.0x62/2/paypal.in/index.html".

Long URL to hide the suspicious part: Phishers tend to use long URL to hide the doubtful part in the address bar. To ensure precision of our study, the length of URLs within the dataset is evaluated and constructed an average URL length. The results showed that if the URL's length is greater than or equal 54 characters then the URL will classify it as phishing. By reviewing the dataset, it was found that almost 1200 URLs' lengths equals to 54 or more which constitute 48.8 percent of the total dataset size.

Using URL shortening service "TinyURL": URL shortening is a method on the "World Wide Web" in which a URL is converted into considerably smaller in length and still lead to the suspicious webpage. This is done by means of an "HTTP Redirect" on a domain name, which then carries to link of the webpage that has a long URL or just an IP Address. For example, the URL "http://portal.hud.ac.ca/" becomes shortened to "bit.ly/19DYSk4". URL's having "@" symbol: Just by using "@" symbol within the URL misleads the browser to ignore everything which precedes the "@" symbol and the real address often follows the "@" symbol.

Redirecting using "//": The existence of "//" within the URL path means that the user will be redirected to another website. An example of such URL's which uses redirection is: "http://www.legitimatereal.com//http://www.phishingfake.com". they found that if the URL starts with "HTTP", that means the "//" must appear on the sixth position. However, if the URL uses "HTTPS" then the "//" will emerge on seventh position of URL.

Adding Prefix or Suffix separated by (-) to the domain: The dash symbol is rarely used in legitimate URLs. Phishers tend to attach prefixes or suffixes separating with the (-) to the domain name so that end-user notices that they are trafficking with an authorized website. For an illustration, a website might look like: http://www.confirmed-paypal.com/.

Sub domain to multi sub domains: Let us assume we have the following URL: http://www.duh.ac.us/students/. A domain name might append the country-code top-level domains (ccTLD), which in our illustration has "us". The "ac" part is shortened for word "academic", the united "ac.us" is called the second-level domain (SLD) and "duh" is the domain name of the website. To produce a rule for extracting this feature, we first have removed the "www." from the link which is actually the sub domain itself. Then, we removed the country-code top-level domains (ccTLD).

HTTP/HTTPS: HTTPS existence in a URL is very important in determining the impression of legitimacy of website.

Certificate Authorities that are regularly being listed among the top dependable names which includes: GoDaddy, VeriSign, GeoTrust, Comodo, Doster, Thawte, and Network Solutions. Furthermore, by testing out datasets, they found that the minimum age of a reputable certificate must be of two years. **Domain Registration length:** Based on the fact that a phishing website lives for a short period of time, they believe that trustworthy domains are regularly paid for several years in advance. In the dataset, we find that the longest fraudulent domains have been used for one year only.

Favicon: A favicon is a graphic image (icon) associated with a specific webpage. Many active user agents like browsers and news reader display favicon as the pictorial identity remainder of the website on the address bar. If the favicon reveals any domain other than that shown in the address bar, then the URL is possibly to be believed as a phishing setup.

Using non-standard port: This element is valuable in approval assuming some particular assistance like for instance, HTTP is up or down on a specific server. In the point of controlling interruptions, it is vastly improved to only open ports that you want. A few firewalls, Proxy and Network Address Translation (NAT) servers will, of course, block all or the greater part of the ports and just open the ones chose. On the off chance that all ports are open, phishers can run practically any help they need and subsequently, client data is undermined.

HTTPS Token: The phishers might add the ”HTTPS” token to the domain of a URL to deceive clients. For instance, http://https-www-paypal-it-webapps-mdd-home.soft-hair.com/.

Request URL: Request URL analyzes whether the outside objects held inside a site page, for example, pictures, record-ings and sounds are stacked from another area. In authentic website pages, the site page address and a large portion of articles inserted inside the site page are having a similar space.

URL of Anchor: An anchor is a component characterized by the <a> tag. This element is dealt with precisely as ”RequestURL”. In any case, for this component we look at: 1) If the anchor labels and the site have diverse domain names. This is like request URL include. 2) If the anchor doesn’t connection to any website page. Links in tags: Given that our investigation covers all angles likely to be used in the webpage source code, we find that it is common for legitimate websites to use tags to offer metadata about the HTML document.

Server form handler: SFHs that contain a vacant string or ”about:blank” are considered dicey on the grounds that a move ought to be made upon the submitted data. What’s more, assuming the space name in SFHs is unique in relation to the area name of the page, this uncovers that the site page is dubious because the submitted data is seldom taken care of by outside areas.

Website forwarding/iframe redirection: IFrame is a HTML label used to show an extra site page into one that is presently shown. Phishers can utilize the ”iframe” tag and make it imperceptible for example without outline borders. In such manner, phishers utilize the ”frameBorder” property which makes the program render a visual outline. Abnormal URL: This feature can be extracted from WHOIS Database. Most phishing sites live for a brief timeframe. By inspecting our dataset, we observe that the base age of the authentic domain is a half year.

Disabling Right Click: Phishers use JavaScript to handicap the right-click function, so clients can’t view and save the website page source code. This component is dealt with precisely as ”Utilizing onMouseOver to conceal the Link”. The feature will search for event “event.button==2” in the webpage source code and check if the right click is disabled.

Submitting to email: Web form permits a client to submit his own data that is coordinated to a server for handling. A phisher may divert the client’s data to his own email. Keeping that in mind, a server-side content language may be utilized, for example, ”mail()” work in PHP. Another customer side capacity that may be utilized for this intention is the ”mailto:” function.

Age of Domain: This feature can be extracted from WHOIS database (Whois 2005). Most phishing sites live for a brief timeframe. By auditing our dataset, they observed that the base age of the genuine space is a half year.

DNS Record: For phishing sites, either the guaranteed personality isn’t perceived by the WHOIS information base or no records established for the hostname (Pan and Ding 2006). On the off chance that the DNS record is vacant or not observed then the site is delegated ”Phishing”, in any case it is named ”Real”.

On mouseover: Phishers might utilize JavaScript to show a phony URL in the status bar to clients. To extricate this component, we should uncover the site page source code, especially the ”onMouseOver” occasion, and check assuming it rolls out any improvements on the status bar.

Web Traffic: This component estimates the notoriety of the site by deciding the quantity of guests and the quantity of pages they visit. In any case, since phishing sites live for a brief timeframe, they may not be perceived by the Alexa data set (Alexa the Web Information Company., 1996). By inspecting our dataset, we track down that in most exceedingly awful situations legitimate sites positioned among the best 100,000. Besides, assuming that the area has no traffic or isn’t perceived by the Alexa information base, it is delegated ”Phishing”.

Pop Up Window: It is uncommon to observe a genuine site requesting that clients present their own data through a spring up window. Then again, this element has been utilized in some authentic sites and its primary objective is to caution clients about deceitful exercises or broadcast a welcome declaration, however no close to home data was approached to be filled in through these pop-up windows.

Page Rank: PageRank is a worth going from ”0” to ”1”. PageRank means to gauge how significant a website page is on the Internet. The more noteworthy the PageRank esteem the more significant the page. In our datasets, we track down that around 95 percent of phishing website pages have no PageRank. Additionally, we see that the excess 5 percent of

phishing website pages might arrive at a PageRank esteem up to ”0.2”.

Links Pointing to Page: The quantity of connections highlighting the site page shows its authenticity level, regardless of whether a few connections are of a similar area. In our datasets and because of its short life range, we see that 98 percent of phishing dataset things have no connections highlighting them. Then again, genuine sites have somewhere around 2 outer connections highlighting them.

Google Index: This component analyzes whether or not a site is in Google's record. At the point when a website is filed by Google, it is shown on list items (Webmaster assets, 2014). Usually, phishing webpages are accessible for a short period and as a result, many phishing webpages may not be found on the Google index.

Statistical Report: A Several parties, for example, Phish-Tank and StopBadware detail various measurable reports on phishing sites at each given timeframe; some are month to month and others are quarterly.

B. Machine Learning Algorithms

Decision Tree: One of the most generally utilized algorithm in AI advancement. Decision tree algorithm is clear and moreover easy to execute. Decision tree begins its work by picking best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to create tree until it finds the leaf node. Decision tree creates training model which is used to anticipate target value or class in tree representation each internal node of the tree belong to attribute and each leaf node of the tree belongs to class label. In Decision tree algorithm, Gini index and information gain methods are used to compute these nodes [17].

K-Nearest Neighbors: The K-Nearest-Neighbors is a non-parametric grouping calculation, i.e., it doesn't make any assumptions on the rudimentary dataset. It is known for it's straightforwardness and adequacy. It is a regulated learning algorithm. A marked preparing dataset is given where the information focuses are classified into different classes, so the class of the unlabeled information can be anticipated. In Classification, various qualities decide the class to which the unlabeled information has a place. KNN is generally utilized as a classifier. It is utilized to characterize information dependent on nearest or adjoining preparing models in a given district. This technique is utilized for its straightforwardness of execution and low calculation time. For continuous data, it utilizes the Euclidean distance to work out its nearest neighbors.

Random Forest: Random Forest calculation is one of the most remarkable algorithm in AI innovation and it depends on idea of Decision tree algorithm. Random Forest calculation makes the forest with number of Decision trees. High number of trees gives high discovery precision. Creation of trees depend on bootstrap technique. In bootstrap method, highlights and tests of dataset are haphazardly chosen with substitution to build single tree. Random Forest calculation will pick best splitter for the characterization and like Decision Tree Algorithm; Random Forest Algorithm additionally utilizes Gini file and data gain strategies to see as the best splitter This interaction will get proceed until random forest makes n number of trees. Each tree in random forest predicts the objective worth and afterward algorithm will work out the decisions in favor of each anticipated objective. At last Random Forest calculation thinks about considers high voted predicted target as a final prediction [17].

XGBoost: XGBoost is a refined and tweaked form of a Gradient Boosting to give better execution and speed. The most significant component behind the achievement of XG-Boost is its adaptability in all situations. The XGBoost runs in excess of multiple times quicker than popular solutions on a solitary machine and scales to billions of models in distributed or memory-limited settings. The adaptability of XGBoost is because of a few significant algorithmic enhancements. These advancements incorporate an novel tree learning algorithm for taking care of scanty information; a theoretically justified weighted quantile sketch methodology empowers taking care of instance weights in approximate tree learning. Parallel and distributed computing make learning quicker which empowers faster model exploration [14].

Logistic Regression: Logistic Regression is an arrangement calculation used to allocate perceptions to a discrete arrangement of classes. Unlike linear regression which outputs continuous number values, Logistic Regression changes its result utilizing the calculated sigmoid capacity to return a likelihood esteem which would then be able to be planned to at least two discrete classes. Logistic regression functions admirably when the relationship in the information is practically direct not withstanding on the off chance that there are perplexing non-straight connections between factors, it has poor performance. Besides, it requires more measurable presumptions prior to utilizing different procedures [14].

V. Implementation

Scikit-learn tool was used to import and implement Machine Learning Algorithms. The dataset was divided into the training and testing sets of data in 70:30 ratio. Each of the machine learning algorithm is used in evaluate the performance accuracy of the algorithm.

TABLE I
Accuracy Performance of ML algorithms

Sr. no.	*Algorithms*	*Train Accuracy*	*Test Accuracy*
1	XGBoost	0.986	0.970
2	K Nearest Neighbors	0.987	0.957
3	Random Forest	0.932	0.929
4	Logistic Regression	0.928	0.928
5	Decision Tree	0.922	0.922

VI. Conclusion

The proposed system will help users to defend their private credentials from leaking and falling into the wrong hands.

However, a challenge still exists in this domain is that the hackers or cyber criminals are constantly evolving their strategies to overcome the defence mechanisms of phishing detection. This results in increased chance of getting suspicious website being left unrecognized. In order to succeed in this context, there need algorithms that will keep on adapting with the new features and examples of phishing websites. Using different approaches altogether, might help in strengthen the accuracy of detection and provide an efficient defensive system.

ACKNOWLEDGMENT

I would like to take this opportunity to record the deepest sense of appreciation to every one of the individuals who helped me in accomplishing the objective. Most importantly, I might offer my thanks towards Dr. Priya Swaminarayan for her support, inspiration and significant ideas. Then I would thank Prof. Dharmendrasinh Rathod for guiding me as best as possible as research guide.

I'm additionally thankful to our Head of Department, Prof. Vivek Dave, for extending all the facilities needed to carrying out this project and reviewing the entire part of it with a great attention.

I extend my thanks to all other staff members of the institute.

REFERENCES

[1] https://www.cisco.com/c/en_in/products/security/email-security/what-is-phishing.html.

[2] S. Baadel and J. Lu, "Data Analytics: Intelligent Anti-Phishing Techniques Based on Machine Learning," *J. Inf. Knowl. Manag.*, vol.18, no. 1, 2019, doi: 10.1142/S0219649219500059.

[3] https://info.phishlabs.com/blog/the-definition-of-phishing%0A.

[4] https://www.mcafee.com/blogs/consumer/family-safety/what-is-a-keylogger.

[5] https://www.csoonline.com/article/3295877/what-is-malware-viruses-worms-trojans-and-beyond.html.

[6] https://www.kaspersky.co.in/resource-center/definitions/dns.

[7] https://www.imperva.com/learn/application-security/social-engineering-attack.

[8] https://www.csoonline.com/article/3334617/what-is-spear-phishing-why-targeted-email-attacks-are-so-difficult-to-stop.html.

[9] https://searchsecurity.techtarget.com/definition/whaling.

[10] https://www.csoonline.com/article/3538831/what-is-smishing-how-phishing-via-text-message-works.html.

[11] https://www.threatmark.com/on-vishing-attacks-and-how-to-prevent-them.

[12] I. Vayansky and S. Kumar, "Phishing – challenges and solutions," *Comput. Fraud Secur.*, vol. 2018, no. 1, pp. 15–20, 2018, doi: 10.1016/S1361-3723(18)30007-1.

[13] V. Patil, P. Thakkar, C. Shah, T. Bhat, and S. P. Godse, "Detection and Prevention of Phishing Websites Using Machine Learning Approach," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–5, 2018, doi: 10.1109/ICCUBEA.2018.8697412.

[14] J. Rashid, T. Mahmood, M. W. Nisar, and T. Nazir, "Phishing Detection Using Machine Learning Technique," *Proc. - 2020 1st Int.Conf. Smart Syst. Emerg. Technol. SMART-TECH 2020*, pp. 43–46, 2020, doi: 10.1109/SMART-TECH49988.2020.00026.

[15] R. Kiruthiga and D. Akila, "Phishing websites detection using machine learning," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 11, pp. 111–114, 2019, doi: 10.35940/ijrte.B1018.0982S1119.

[16] L. Mohammad, R.M., Thabtah, F. and McCluskey, "Phishing websites features," *Sch. Comput. Eng. Univ. Huddersfield.*, 2015.

[17] R. Mahajan and I. Siddavatam, "Phishing Website Detection using Machine Learning Algorithms," *Int. J. Comput. Appl.*, vol. 181, no.23, pp. 45–47, 2018, doi: 10.5120/ijca2018918026.