International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 4 Apr. 2021, pp: 496-503 www.ijaem.net ISSN: 2395-5252

Detecting Phishing Websites Using Machine Learning

Aniket Garje[1], Namrata Tanwani[2], Sammed Kandale[3], Twinkle Zope[4], Prof. Sandeep Gore[5]

[1, 2, 3, 4, 5] G. H. Raisoni College of Engineering And Management, Pune, India
[5]Professor, G. H. Raisoni College of Engineering And Management, Pune, India

Submitted: 15-04-2021 Revised: 28-04-2021 Accepted: 30-04-2021

ABSTRACT: Phishing is an online fraudulent practice to get hold of sensitive information of users like passwords, credit card numbers, etc. To avoid being a victim of this malpractice, an attempt has been made at researching machine learning techniques to detect Phishing Websites. Therefore, this paper, in its essence, extracts and considers almost all possible ways to find Machine Learning practices and algorithms that will help in detecting phishing websites. The emphasis is on finding concrete solutions as much as possible by going through a bunch of implementations that are mentioned below, that revolve around Machine Learning algorithms such as Random Forest, Support Vector Machine, Neural Networks. The work is conceived in several parts. The introduction describes the focussed zone and the techniques and tools used along. It is followed by the preliminaries section that focuses on the preparation of the information that is required to move further. Later the paper emphasizes the detailed discussion of the sources of information with their advantages and disadvantages.

KEYWORDS: Algorithms, Cybersecurity, Machine Learning, Neural Network, Phishing

I. INTRODUCTION

As mentioned in [1], with high volume and high velocity of data emerging each day, IBM has estimated that each day 2.5 quintillion bytes of data are generated. This brings attention to the challenges and issues related to the private and secure transfer of data. Cybersecurity is one of the most important fields of computer science that concerns the avoidance and protection of data against threats and attacks faced by the user. It makes sure the user is not deceived into being a victim of Cyber-crimes [2].

The term 'crime' means an unlawful act that is offensive and is punishable by the state [3]. It is harmful to a person or society, per se [4]. Cyber is a term that can be used for a thing, person or idea that is a part of computer and information technology [5]. This said, Cyber-crime hardly has anything to do with the law [6]. At its core, it involves a network of computers, mobile phones, laptops that are also referred to as communicating nodes that are involved in the transferring of data from and to the target nodes [7].

The attacks or threats include but are not limited to committing frauds, trafficking child pornography, stealing identities, violating privacy, etc [8]. Amongst them, the most organized crime of the 21st century is Phishing. It has been an effective cybersecurity attack since more than 60% of commercial transactions are done online. Phishing is a cybersecurity attack in which attackers get access to the sensitive information of the use (unknowingly to the user). The sensitive information may contain login credentials of the user, bank-related usernames and passwords, credit card or debit card details, etc. The mails, messages, and clone websites are how the attackers get access to sensitive user information [9].

Numerous methods have been developed to tackle this issue of social engineering called "Phishing" but, a majority of them havefailed to seem promising. One of the most successful methods is the technique of Machine Learning. It has detected patterns and behaviors of the URLs of phishing websites that no other method could [10]. Hence, the aim is to find the solution to detecting the best algorithms under Machine Learning.

II. PRELIMINARIES

A. Methods In Machine Learning

Machine Learning is one of the main topics in research and industries. Machine Learning focuses on the development of computer programs that can access data and use it to learn and make few predictions based on the requirement [11]. Machine Learning algorithms are mainly classifiedinto three methods: Supervised Learning,Unsupervised Learning, and Reinforcement Learning [12].

1) Supervised Learning : In Supervised Learning, there are input variables (x) and output variables (y). Using algorithm a function is mapped:

$$y=f(x) \qquad (1)$$

The goal of a Supervised Learning algorithm is to map function so well that if you introduce any new (x) variable, it can predict (y) perfectly. Thus, from the above example, it is concluded that Supervised Learning is a method of ML in which a set of label data is used to train the model and predict the outcome. Examples of Supervised Learning: Regression, Random Forest, Support Vector Machine [13].

2) Unsupervised Learning : Unsupervised Learning only has (x) in data. There are no output variables present in the dataset. It has to build a structure based on available data to learn more about that data. These are called Unsupervised Learning because unlike Supervised there is no correct answer available, only structures are formed to learn more about data. Examples: Clustering and Association [14].

3) Reinforcement Learning : Reinforcement Learning algorithms allow computers to learn from the experience. The machine trains itself so that when it predicts correctly one reward signals are generated and punishment for the wrong one [15].

4) Natural Language Processing : NLP or Natural Language Processing is a field of Artificial Intelligence that helps the computer to understand words or statements written or spoken in different human languages. Some of its applications are email spam detection, text summarization, information extraction, etc [16].

5) Neural Networks : Deep learning also known as feature mapping, maps the input to an output. This process takes place using multiple connected layers that further contain multiple neurons, forming a network of neurons, called a Neural Network. Each neuron is itself a mathematical unit that aims at learning the relationship between the input features and the output [17].

B. Phishing

Phishing is the most common online security threat all over the world. Phishing involves stealing information, email spoofing, and text messages that instruct users to fill in information that is important to them like passwords, usernames, bank details, etc [18].

1) Phishing Attacks : Following are some types of Phishing attacks[18]:

(1) malware-based phishing: This type of phishing is used to harm the software of the user especially if it is used in a small firm when it is not updated for a long time.

(2) Keyloggers and Screen loggers: This is also a malware attack in which the attacker tracks the input given by the user and sends the required information to the target.

(3) Deceptive phishing: It uses fake social media accounts to lure users and get sensitive information from them.

(4) Data Theft: It is a type of phishing that is practiced mostly with government offices or any large competitive source by stealing their data to jeopardize them by leaking them.

(5) Search Engine Phishing: A type of phishing in which fake or fraudulent websites are created that gives out attractive offers to users to become a victim of fraudulent e-commerce practice.

2) Phishing Websites : Websites are used as a new tool for modern phishing attacks. Thus, phishing websites look the same as the original ones. People with less awareness about these websites are more likely to visit and give their information on those websites. Thus, they look the same as original ones but certain steps can be taken to avoid these attacks effectively using machine learning [19].

III. LITERATURE REVIEW

The dataset used here [20]was self-constructed, where phishing websites are mostly from PhishTank and legitimate websites are from Yandex Search API. It contained a total of 73,575 website data (36,400 legitimate URLs and 37,175 phishing URLs). A URL consists of some meaningful or meaningless words and some special characters, which separates some important components of the address as shown in Fig 1. As for data pre-processing they extracted those words and characters, then added them to the wordlist to be analyzed. The main aim was to detect the word which is similar to brand names, to detect keywords, the words, which are created with random characters. They used modules like “Random word detection module” and “Maliciousness analysis module”, that have helped to detect many possible random words with their length and to detect whether the words in the given/tested URLs are used for fraudulent purposes or not respectively. Different classification algorithms as NaïveBayes, Random Forest, kNN(n=3), Adaboost, K-star, SMO, and Decision Tree with some feature extraction types as NLP-based features, Word Vectors and Hybrid. The Proposed system was able to achieve 98% of accuracy. The use of NLP-based features and word vectors together also increases the performance of the phishing detection system with a rate of 2.24% according to NLP-based features and 13.14% according to word vectors. The Confusion Matrix for all classification Algorithms is tabulated in Table 1.

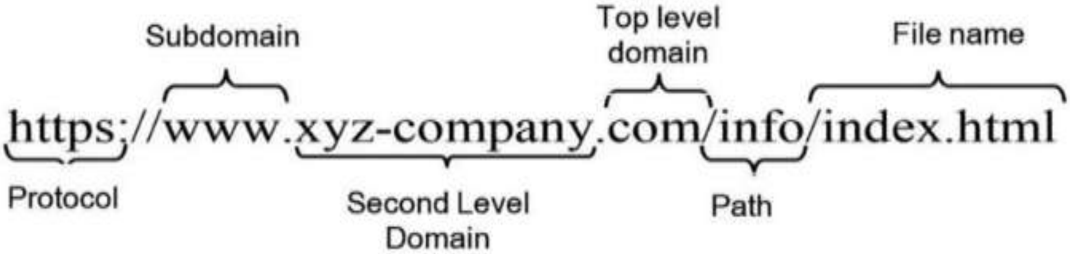

Fig. 1 Components of Address [20]

Algorithm	Confusion Matrix		Predicted	
			P	N
Decision Tree (NLP based)	Actual	P N	36,328 1348	847 35,052
Adaboost (NLP boost)	Actual	P N	35,813 3609	1362 32,791
Kstar[1] (Hybrid)	Actual	P N	3596 227	121 3413
k-NN(n=3) (Hybrid)	Actual	P N	36,214 2082	961 34,318
Random Forest (NLP based)	Actual	P N	36,806 1120	369 35,280
SMO (NLP based)	Actual	P N	36,256 2817	919 33,583
Naïve Bayes (Hybrid)	Actual	P N	27,663 1247	9512 35,153

TABLE 1 CONFUSION MATRIX [20]

The system proposed by [21] used the methods for detection of phishing websites based on lexical features, host properties, and page importance properties are briefly discussed. They considered various data mining algorithms for evaluation of the features to get a better understanding of the structure of URLs that spread phishing. The tuned parameters are used for determining the appropriate machine learning algorithm for separating phishing sites from genuine sites. The techniques of phishing, statistics of phishing attacks are discussed with the evaluation of the various classifying algorithm is done by using the workbench for data mining, Waikato Environment for Knowledge Analysis (WEKA), and using MATLAB. And their performance is tabulated in Table 2 and Table 3 respectively. The classification algorithms they considered were NaïveBayes, J48 Decision Tree, K-NN, and SVM. Decision Tree got a better accuracy percentage than other algorithms, which is 91.08%.

The main function of the system in [22] was to decide the state of a URL if it was legitimate or phishing. The system acts as additional functionality to the browser as an extension. Feature extraction wasbased on URL, Page Content, Page Rank. K combinations of features are made and the combination with higher accuracy and the least number of features was selected. The Random Forest Algorithm was used in the proposed system. Due to its better accuracy of 98.8%. The dataset holds 16,000 URL records. Out of which 12,000 URLs werephishing, were collected from PhishTank.

The remaining 4,000 URLs were legitimate, collected from the daily use of 10 chosen users. However, the final dataset after handling missing data and removing the duplicate was of size

6,116 URLs. The main objective was to achieve higher accuracy with minimum features, they reduced those features to 26 features from 36.

Test Options	Classifier	Confusion Matrix	Success Rate (%)	Error Rate (%)
Percentage Split-60	Naïve Bayes	4438 3578 260 3945	68.60	31.40
	J48	7612 404 428 3777	93.20	6.80
	IBK	7042 974 455 3750	88.30	11.70
	SVM	7511 505 1459 2746	83.93	16.07
Percentage Split-90	Naïve Bayes	1180 792 61 1022	72.08	27.92
	J48	1883 89 101 982	93.78	6.22
	IBK	1756 216 97 986	89.75	10.25
	SVM	1846 126 355 728	84.26	15.74

TABLE 2 CLASSIFIER PERFORMANCE – WEKA [21]

Test Options	Classifier	Confusion Matrix	Success Rate (%)	Error Rate (%)
Percentage Split-60	Naïve Bayes	7281 303 3633 4042	74.20	25.80
	Regression Tree	10856 470 1166 5839	91.08	8.92
	KNN	11299 3025 723 3284	79.55	20.45
	SVM	9871 806 1082 3531	87.65	12.35
Percentage Split-90	Naïve Bayes	13648 1018 2764 5500	83.50	16.50
	Regression Tree	15082 999 2951 8465	85.63	14.37
	KNN	16451 5080 1582 4384	75.77	24.23
	SVM	16416 5848 5 661	74.48	25.52

TABLE 3 CLASSIFIER PERFORMANCE – MATLAB [21]

The proposed system in [23] used Machine Learning Classification Algorithms to detect Phishing URLs. The proposed system recognized Phishing URLs, by analyzing the URL structure without visiting the Phishing URL. The data collected was first passed through the training phase where it undergoes feature selection and classification. The dataset contains 4,500 URL records, on which classification was performed. Out of which 2,500 URLs are genuine and the rest 2,000 are the phishing URLs. The 2,500 genuine URLs have been collected from the DMOZ repository. The 2,000 phishing URLs have been picked from PHISHTANK. Data classification after extraction of the relevant features was performed by using the following algorithms namely NaïveBayes, Random Forest, Random Tree, Multi-layer Perceptron, C-RT, J 48 Tree, LMT, C 4.5, ID 3, and K-Nearest Neighbour. After classification, it was observed that the Tree-based classification algorithms had better accuracy, precision, and recall values compared to the other frequently used algorithms. The Random Forest Algorithm had the highest classification accuracy of 99% for different Phishing URL categories. So Tree-based classifiers are best suited for phishing URL classification, in this case**.**

The proposed system in [24] divided the websites into three classes:

(1) Benign: This class contains websites that are safe or legitimate and provide normal services to people.

(2) Spam: These are the websites that practice flooding the user with advertisements, fake surveys, etc.

Malware: These are the websites that are phishing websites, i.e., they look like normal websites but are handled by attackers to misuse sensitive information.Websites were classified based on properties of URL like its length, domain length, host length, number of times dot(.) appears in the given URL, the presence of "//"symbol, ASN number, average token path, and so on. uniform resource locater (URL), length of the URL, the popularity of the site, or the page content itself. The methodologies they've usedwere Random Forest and Support Vector Machine. Random Forest is a supervised learning technique used for classification or regression problems. It contains several decision trees that work parallelly to produce an output class from the given input. In the end, the average of the output of these decision trees was considered. The second method, the Support Vector Machine is used, which can be used to solve regression and classification problems, to work with linear and nonlinear data. In conclusion, SVM gavean accuracy of 91.3% and Random Forest gavean accuracy of 80% on the test data set.

This paper [25]hasanalyzed the various aspects of phishing attacks, including their common defenses, some specific phishing countermeasures at both the user level and the organization level, some recent statistical data on phishing scams, and at the last a multi-layered anti-phishing proposal. The paper helps to get the basics of phishing attacks and how they occur and what are thedefenses and countermeasures to get protected. Briefly presented some commonly seen attacks related to today's business world and showed some corresponding defenses to counter those attacks and some recent statistical results to project the growing problem of phishing spam.

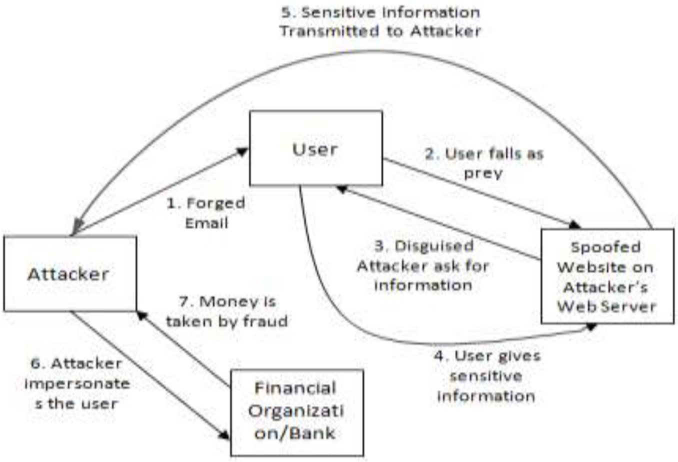
5. Sensitive Information Transmitted to Attacker
User
2. User falls as prey
1. Forged Email
3. Disguised Attacker ask for information
Attacker
Spoofed Website on Attacker's Web Server
7. Money is taken by fraud
4. User gives sensitive information
6. Attacker impersonate s the user
Financial Organizati on/Bank

Fig. 2 Stages of Phishing Attacks [25]

Detection of phishing websites was performed in [26] was by using machine learning algorithms like Logistic Regression, Decision tree, Random Forest, Adaboost, Gradient Boosting, Gaussian NB, and Fuzzy pattern tree classifier. In the data collection phase, the data was collected on both phishing and legitimate websites. Then came extracting the useful features of the given dataset. It involved two steps: URL-based features and Domain-based Features. URL-based feature selection involved IP Address, '@' symbol in URL, Dashes in URL, Long URL, presence of unusual number, Dot Count, Sub-domains in URL, etc. Domain-based feature selection includedPage Rank of the Website, Age of the Domain, and Validity of the Website. In Implementation, the first step was to process the data. Dataset Exploration was done along with selecting useful attributes for further process. The dataset was split into training and testing set in the ratio 80:20. The training set with the extracted features was given as input to different machine learning classification algorithms. The accuracy of classification along with precision, recall, and F1 score was determined using different algorithms mentioned above. The results for the different models were then analyzed. The Random Forest algorithm shows 96% of precision and recall along with the highest F1 score of 95%. The Random Forest algorithm classified 94% of the legitimate websites and 96% of phishing websites correctly. Random Forest proved to be the best option to determine phishing websites using machine learning.The accuracy of all Machine Learning Algorithms is tabulated in Table 4.

Algorithm	Independent Accuracy (%)	Accuracy with PCA (%)
Random Forest	95.33	95.82
Decision Tree	94.09	94.26
Gradient Boosting	92.19	92.22
Fuzzy Pattern Tree	91.22	92.23
Adaboost	91.00	90.64
Gaussian NB	83.28	85.17
Logistic Regression	73.78	82.89

TABLE 4 ACCURACY OF ML ALGORITHMS [26]

Moitrayee Chatterjee and Akbar SiamiNamin (2019) [27] chose a lexical signature of a web page consists of several keywords carefully from the webpage and used it to generate robust hyperlinks to find the web page when its URL fails. The proposed system used a Deep Reinforcement Learning based classification model for phishing website analysis. The proposed model took up a dynamic behavior of the phishing websites and found out all the features related to phishing website detection. The dataset was the Ebbu2017 Phishing dataset. The dataset was unavailable, so for detection of phishing websites, the above dataset was created and made public. The dataset that was used for this research contains 73,575 URLs records. Out of which 36,400 URLs were legitimate and the remaining 37,175 were phishing.

A total of 15 research papers have been studied in this [28] research paper. The main components of detecting phishing websites have been explained along with their classification and feature extraction. In this research paper, one method has used five different algorithms that are Decision Tree, Random Forest, Gradient Boosting, Generalized Linear Model, and Generalized Additive Model. From comparing the accuracy of each algorithm Random Forest has given the highest 98.4% accuracy, 98.59% recall, and 97.70% precision. In another method, feature selection algorithms are used to decrease the components and get higher-order execution. Dataset used was taken from the UCI machine learning repository. To reduce the dataset they have used Bayesian Network, Stochastic Gradient Descent, lazy. K.Star, Randomizable Filtered Classifier, Logistic model tree and ID3 (Iterative Dichotomiser). Lazy. K.Star algorithm obtained 97.58% accuracy with 27 reduced features. Another research paper used Machine Learning techniques like logistic regression using bigram, deep learning techniques like convolution neural network and CNN long short-term memory as architecture. The dataset was collected from Phishtank. CNN-LSTM obtained 98% accuracy. Another approach was used starting with a classification method using Logistic Regression and Support Vector Machine. Nineteen best features have been selected from thirty features where SVM shows better results than Logistic Regression. There was one approach where the c4.5 decision tree was used to detect phishing websites. This technique extracted features from the site and calculated heuristic values. This value was used by this algorithm to determine the legitimacy of the site. All this research shows Random Forest Classifier showed the best accuracy and precision over other algorithms.

IV. APPLICATIONS

These days users face security issues due to phishing attacks whenever they connect to the internet through browsing websites, opening mails, or using other web applications. As prevention, an attempt is made to find the solution to the problem of phishing using machine learning. By upgrading this approach into an application it can be used by any user[29]. It can work as an alarm while unknowingly copying phishing websites or receiving spear-phishing emails or while basic form-filling applications. This will prevent users from accessing malicious links[30]. Transactions became insecure for E-commerce users due to phishing attacks as phishing is done by mass-mailing malicious mails, making it easy to disclose users' data. So, in this field also this model is useful[31].

V. CONCLUSION

Phishing has been a major problem for us ever since it was acknowledged. Phishing attacks became the easy way for phishers to get access to user's legal data illegally and manipulate it successfully. Phishers set their fake website which looks exactly like the original website, creating web servers and the same web pages. Although laws have been enacted, education is the best defense against phishing. The defense techniques like data mining and heuristics, blacklisting, machine learning, and soft computing algorithms are doing wonders by preventing users from phishing attacks. The defense mechanism which can detect phishing websites or URLs with low false-positive does have the good capability. Like the different machine learning algorithms used for the detection, the model was able to achieve 91% to 98% accuracy approximately. The main focus here was to study phishing and machine learning practices and algorithms against phishing attacks.

REFERENCES

[1] R. Devakunchari, "Analysis on big data over the years," International Journal of Scientific and Research Publications (IJSRP), vol. 04, no. 01, January 2014.

[2] G. Nikhita Reddy, G.J. Ugander Reddy, "A Study Of Cyber Security Challenges And Its Emerging Trends On Latest Technologies," International Journal of Engineering and Technology, vol. 4, no.1, January 2014.

[3] Larry Sanger,"Crime",en.wikipedia.org/wiki/Crime , September 20, 2001.

[4] Esther Ramdinmawii, Seema Ghisingh, Usha Mary Sharma, "A Study on the Cyber-Crime and Cyber Criminals: A Global Problem ," International Journal of Web Technology, vol 04, pp. 53-57, June 2015.

[5] "Cyber",merriam-webster.com/dictionary/cyber, January 2021.

[6] TechTarget Contributor,"Cyber", searchsoa.techtarget.com/definition/cyber , April 05, 2005.

[7] Sharma, Ushamary and Ghisingh, Seema and Ramdinmawii, Esther, "A Study on the Cyber - Crime and Cyber Criminals: A Global Problem," International Journal of Web Technology, vol 03, pp. 172-179 , June 2014.

[8] Andrewa, "Cybercrime", http://en.wikipedia.org/wiki/Computer_crime , October 15, 2003.

[9] Vayansky, I. and Kumar, S. , "Phishing – challenges and solutions.", Computer Fraud & Security, vol 2018, no. 1, pp. 15-20, January 2018.

[10] Vahid Shahrivari, Mohammad Mahdi Darabi, Mohammad Izadi , "Phishing Detection Using Machine Learning Techniques,"unpublished.

[11] Taiwo Oladipupo Ayodele, Introduction to Machine Learning, InTech February 2010.

[12] Essinger, Steve and Rosen, Gail. "An introduction to machine learning for students in secondary education," Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE, February 2011 , pp. 243 - 248.

[13] Iqbal Muhammad and Zhu Yan, "Supervised Machine Learning Approaches: A Survey," ICTACT , vol 05, pp. 946-952, April 2015.

[14] Osvaldo Simeone, Fellow, "A Very Brief Introduction to Machine Learning With Applications to Communication Systems," IEEE, vol. 04, no. 04, pp. 648 - 664, November 2018.

[15] Leslie Pack Kaelbling, Michael L. Littman, Andrew W. Moore, "Reinforcement Learning: A Survey," Journal of Artificial Intelligence Research, vol. 04, pp. 237-285, May 1996.

[16] Diksha Khurana1, Aditya Koli1, Kiran Khatter, and Sukhdev Singh, "Natural Language Processing: State of The Art, Current Trends and Challenges," unpublished.

[17] Adrian Iustin Georgevici1 and Marius Terblanche, "Neural networks and deep learning: a brief introduction," Intensive Care Medicine, vol. 45, no. 5 ,pp. 712–714, February 2019.

[18] PhirashishaSyiemlieh, Golden Mary Khongsit, Usha Mary Sharma, Bobby Sharma, "Phishing-An Analysis on the Types, Causes, Preventive Measures and Case Studies in the Current Situation," National Conference on Advances in Engineering, Technology & Management (AETM'15), vol. IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, January 2015, pp. 01-08.

[19] Dr.RadhaDamodaram, "Study on Phishing Attacks and Antiphishing Tool," International Research Journal of Engineering and Technology (IRJET), vol. 03, no. 01, January 2016.

[20] Sahingoz, O. K., Buber, E., Demir, O., &Diri, B. "Machine Learning-Based Phishing Detection from URLs," Expert Systems with

Applications, vol. 117, pp. 345-357, January 2019.

[21] J. James, Sandhya L. and C. Thomas, "Detection of phishing URLs using machine learning techniques," International Conference on Control Communication and Computing (ICCC), December 2013.

[22] Alswailem, B. Alabdullah, N. Alrumayh, and A. Alsedrani. "Detecting phishing websites using machine learning," 2nd International Conference on Computer Applications Information Security (ICCAIS), pages 1–6, 2019.

[23] Pradeepthi, K. V., & Kannan, A. "Performance study of classification techniques for phishing URL detection," Sixth International Conference on Advanced Computing (IcoAC), December 2014.

[24] K.Venkateswara Rao, Jagan Mohan Reddy D, G L Vara Prasad, "An Approach For Detecting Phishing Attacks Using Machine Learning Techniques,"JCR, vol. 07, no. 18, pp. 321-324, June 2020.

[25] Biju Issac, Raymond Chiong and Seibu Mary Jacob, "Analysis of Phishing Attacks and Countermeasures," 6th IBIMA International Conference on Managing Information in Digital Economy (IBIMA 2006) At: Bonn, Germany, January 2006, pp.339-346.

[26] Dipayan Sinha, Dr. MinalMoharir, Prof. Anitha Sandeep, "Phishing Website URL Detection using Machine Learning," International Journal of Advanced Science and Technology, vol. 29, no. 3, pp. 2495-2504, 2020.

[27] Moitrayee Chatterjee, Akbar SiamiNamin, "Detecting Phishing Websites through Deep Reinforcement Learning," 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), July 2019, pp. 227-232.

[28] R. Kiruthiga, D. Akila, "Phishing Websites Detection Using Machine Learning," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 2S11, pp. 11-114, September 2019.

[29] P.A. Barraclough, M.A. Hossain, M.A. Tahir, G. Sexton, N. Aslam, "Intelligent phishing detection and protection scheme for online transactions," Expert Systems with Applications, vol. 40, no. 11, pp. 4697-4706, September 2013.

[30] Higashino, Masayuki; Kawato, Toshiya; Ohmori, Motoyuki; Kawamura, "An Anti-phishing Training System for Security Awareness and Education Considering Prevention of Information Leakage," 2019 5th International Conference on Information Management (ICIM), March 2019, pp. 82-86.

[31] Ge Wang, He Liu, Sebastian Becerra, Kai Wang, Serge Belongie, HovavShacham, and Stefan Savage, "Verilogo: Proactive Phishing Detection via Logo Recognition," unpublished.