A Review on Machine Learning Based Approaches for Phishing Detection

[1]S Chanti, [2]Chithralekha Balamurugan
[1]Research scholar, [2]Associate Professor
[1]Department of Banking Technology,
[1]Pondicherry University, Puducherry, India

Abstract: Phishing attacks plays are a vital role in Stealing users' Personal information for fraudulent activities in today's digital world. Plenty of anti-phishing solutions are developed and awareness programs have been conducted to prevent the internet users not to fall a victim of Phishing. Targeting a specific set of people or organization through spear phishing will contains a malware that installs automatically and works in the background to provide a remote access to the phisher. Recent survey reports had shown that phishing URL i.e., HTTPS protected to fool the internet users. This paper provides a comprehensive review of machine learning based approaches for phishing detection. The phishing attacks covered by each work as well as some future research direction are discussed.

Index Terms **- Phishing, Machine learning, Malware, Social Engineering, Credential Stealing.**

I. INTRODUCTION

Phishing is a fraudulent activity through which the attacker tries to fool the users, where the final goal of the phisher is to gain access to the personal details of the internet customers. They develop a fake website that looks exactly same to the original one. To do this, they copy the HTML code of the website which is an easy task and then renders small or modifications to the same which still retaining the look and feel of the original site. This makes the user believe that it is the original site. Once the user starts believing the site he/she will provide his personal information like username, password, credit card number, validity, pin code etc., the phisher makes use of these credentials for malicious purpose. This happens mostly with online banking sites which are used to carry out Internet Banking Transactions. Phishing can be performed in many ways, through suspicious emails, by click events, installation of malicious codes in the user system and try to control the system remotely without the user's knowledge. The first phishing scam was done in 1996 to steal the credentials of American Online Users [1]. The phishing attacks are increasing drastically and a large number of Internet users are affected due to phishing. Figure 1 shows the unique phishing attacks and email received from 2013 to 2016 as reported in Anti-Phishing Working Group (APWG) Trend reports.

There are many reasons for performing phishing among them financial gain is the main goal of many phishers. Every day new types of attacks are developing. The researchers are developing many solutions to defeat phishing but the attackers are also finding many loopholes in the existing system. With the help of the new technologies, the attacker is more advanced in stealing the user credentials. User awareness is very important to prevent the phishing

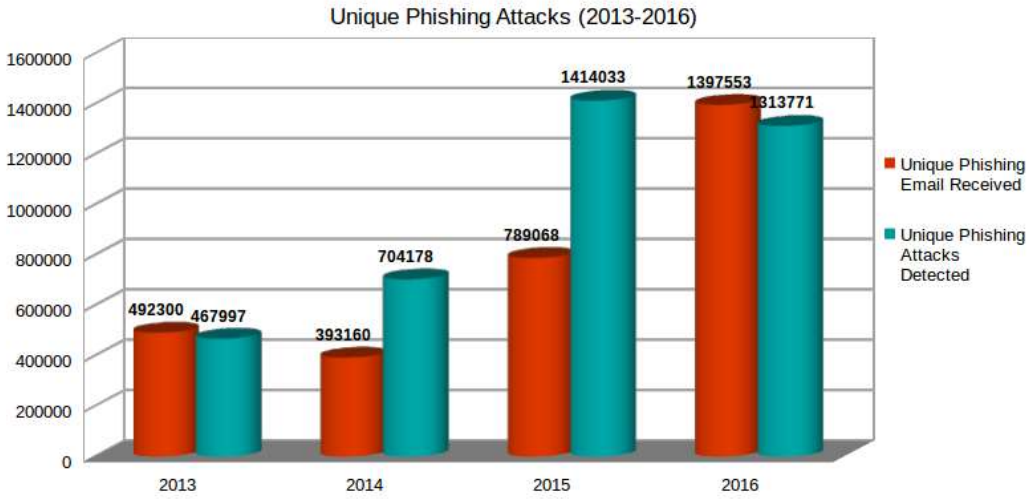

Figure 1: Unique Phishing Attacks Detected from the year 2013 to 2017

The main purpose of phisher is financial gain which can be achieved by stealing user credentials. There are many other factors that motivate the phisher to perform phishing [2]–[4].

- Financial gain
- Theft of login credentials
- Theft of bank credentials
- Malware Distribution

II. CLASSIFICATION OF PHISHING ATTACKS

Phishing attacks are very dangerous attacks which steals the user personal information. Social Engineering and Malware based phishing are the two main types. Figure 2 explains the complete classification of phishing attacks. The classification is further explained in the upcoming sections.

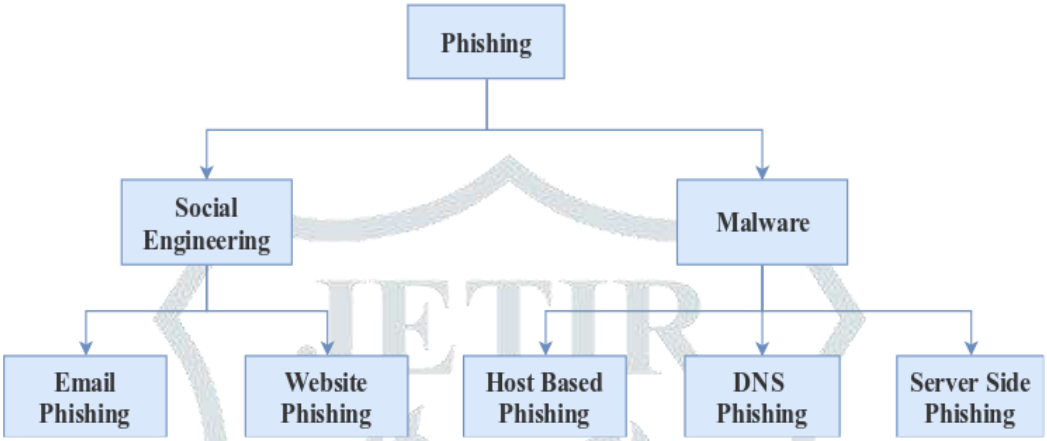

Figure 2: Classification of Phishing Attacks

2.1 Social Engineering based Phishing

Social Engineering are used to fool the internet user by using some tricks to steal their personal information [5]. Social engineering can be done email based and website based. In email-based phishing, the phisher sends a lot of email with malicious content that can steal the user credentials by redirecting to a fraudulent or spoofed site. Website based phishing is through advertisements, pop-ups, a link that redirects to a malicious site etc.

2.1.1 Email Phishing

Email Phishing is a very common way of phishing attack where the attacker send the email to the internet users and ask them to click on the suspicious link provided in the email [6]. Clicking on the link will redirect the victims to a fake or spoofed site to steal their Personal credentials like credit card number, validity, CVV number and so on. Sometimes the suspicious links may download a malware automatically which works in background under the control of the attacker.

2.1.2 Website Phishing

Website based phishing can be performed through advertisements, pop-ups, flash messages. All the business activities are becoming online and most of the companies advertise their products through online advertisements and pop-ups. In this case there is more possibility for the attackers to display the fake advertisements and pop-ups to redirect the internet users to a fake site and steal their credentials for malicious purpose [7], [8].

2.2 Malware based Phishing

Malware Based Phishing is a scam that runs malicious software on user's computers. Malware can be installed in victims system through an email attachment or a downloadable file from the website or by exploiting the vulnerabilities in the system. Most of the Small and Medium Scale Enterprises (SMEs) are affected by malware. The attacker tries to find the loophole in the system (client/server) to install the malware to steal the personal credentials of the internet users [6], [9].

2.2.1 Host based Phishing

In Host based phishing the attacker installs the malware like key loggers, screen loggers, spy ware and observes the user activities remotely. Through keyloggers [10] the attacker can get a log file that contains all the keystrokes of the device which contains the personal information of the user like username, password etc.

2.2.3 DNS Phishing (Pharming)

DNS based phishing can be done by modifying the host files, browser configuration, rogue DHCP, DNS Hijacking, DNS Spoofing etc. Once the attacker installs the malware successfully in the user system, it provides the access to change the host files and directs the user to a spoofed site controlled by the attacker. In case of DNS servers, the attacker use DNS hijacking or DNS Spoofing technique to replace the IP address of a particular domain. When the users visits that domain will automatically redirects to a spoofed site and even the URL looks same as legitimate site [11]–[13].

2.2.3 Server Side Phishing:

In Server side phishing the attacker gains the unauthorized access to the servers and perform necessary actions to steal the user data. Zero day phishing is one type of server side phishing attack [14], [15].

III. MACHINE LEARNING BASED ANTI-PHISHING SOLUTIONS

In this Session, the existing works on phishing attack detection using machine learning are presented. There are different methods proposed by different authors to detect phishing attacks. In [16], [17] a blacklist are used to store the phishing URL's and alert the users when they visit the site listed in that list. [16], [18] presented a whitelist based approach where the list of trusted sites are stored and used to filter the phishing sites from trusted sites. Visual Similarity-based approach [19], heuristic-based approaches [20], Pattern matching [21], Layout Similarity-based approaches [22], machine learning based approaches and so on. When compared to the other approaches, machine learning based approaches perform well in classifying phishing content from legitimate ones. Machine learning based approaches are mainly two types supervised learning and unsupervised learning. The supervised learning models are trained with a dataset whose output is already known, whereas the unsupervised learning models generate the existing patterns. Figure 3 shows the list of supervised and unsupervised algorithms. Some existing works on machine learning based phishing detection methods are discussed below:

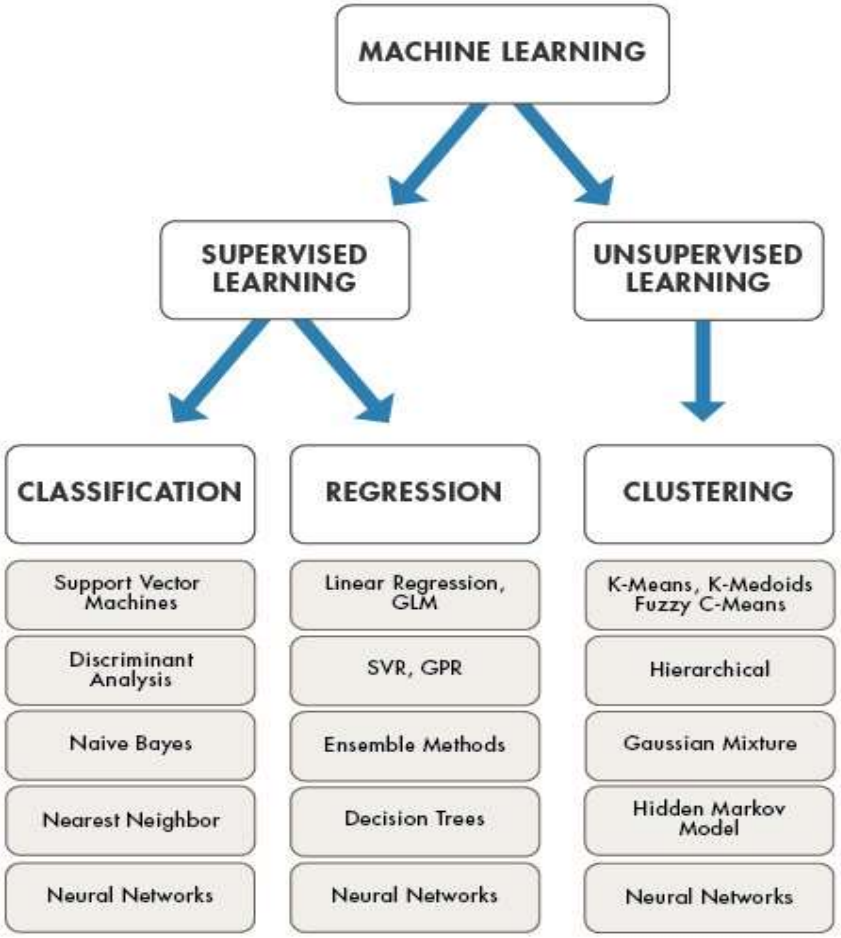

Figure 3: Machine learning algorithms and their classification[23]

In [24] a email based phishing detection model is proposed by aborting the URLs. Initially, the focus is on the email content and to increase the accuracy they count the number of links present in the email. The email content is analyzed by using some heuristics and header information is not included in that, if the attacker sends the email with an attachment that contains the message then it is into to this method to classify the phishing email.

[25] Proposed a machine learning based approach by using a hybrid classifier i.e., FFA_SVM (Firefly algorithm & Support Vector Machine) for phishing email classification. They used 10 features like presence of IP address, inconsistency between "href" attribute and text within the link, Presence of "Link," "Click," "Here" etc., Number of dots in domain name, Html e-mail, existence of JavaScript, etc. They Collected 3500 ham emails from csmining.org and 500 phishing emails from monkey.org for training and testing the FFA_SVM classifier. To know whether the proposed classifier works efficiently, they performed some experiments using tenfold cross-validation for ten times and found better results compared to the other works. FFA_SVM gave an accuracy of 99.98 % with 0.06 % and 0.04% of false positive (FP) and false negative (FN) rates.

The author [26] explored a method to classify the phishing URLs from legitimate ones. In this works they used two methods i.e., in the first part they extracted 13 lexical and structural features and trained the Random Forest (RF) with this features they built a stronger model using the weaker responses. In the second part, they used long / short term memory (LSTM) model to learn the correlation between more than 10 features for which the existing RNN cannot do. In LSTM model the URL is provided as an input and it validates whether it contains any phishing content or not. By comparing both the methods they found the LSTM model performs better than the first one where RF is used. They used 1 million phishing URLs and 1 million ham URLs for

training and testing purpose. The accuracy and F1-measure of LSTM is 0.98 % and 98.7 %, which is better when compared to RF with 0.93 % F1-measure and 93.5 % of accuracy.

In [27] a robust classifier that detects the phishing emails based on the hybrid features extracted is presented. The Entire classification is divided into four levels. At first feature generator will generate a set of feature vectors from the content of a document and is represented as a Feature matrix. The Feature matrix is an input to learning method, where it uses several machine learning algorithms and finds the best one based on its performance. Induction is used for future predictions. After gaining the information from induction, the feature evaluation is used to select the minimal features to get the optimum results. These two steps are continuous until it finds the best feature vectors. Finally, the refined feature matrix can detect the phishing content with minimal features. To implement this they collected the live emails from Westpac with a total of 659,673 phishing and ham emails. Implemented the above method with different machine learning algorithms and they found that decision trees perform well in classifying the phishing emails.

PILFER [28] a new method for detecting malicious phishing emails. The traditional spam filtering is not capable of detecting the phishing emails. Ten features are extracted by running the scripts using PILFER and these features are used to train the machine learning algorithms. Support vector machine is best suited in PILFER for classifying the malicious emails. They got a better accuracy than the traditional spam filter SpamAssassin. They got 0.12% FP and 7.35% FN and later they combined the PILFER with SpamAssassin and later the FP is 0.30% and FN is 8.40% with an accuracy of 92%. Bu the lifespan of phishing emails are very short and the dataset used for training should not be very old.

A new method is proposed for detecting phishing websites using Ball support vector machine (BVM) [29]. To detect the phishing Web sites, the website topology features are analyzed by using DOM Tree. Next, the Topological feature vectors are extracted from the website automatically. These features are provided to BVM to train the classifier. Training the BVM for phishing website detection is done. To perform this, training samples must be obtained through WebCrawler and label the samples as “1” for phishing and “0” for a legitimate site. Collect the Topological features and build the BVM Classifier by solving the minimum enclosing ball of the data set S. Later test the samples using BVM. This gave good results in detecting phishing website using the topological features with F-measure of 0.963%.

[30] Presented novel approach based on semi-supervised learning (Transductive Support Vector Machine). Web images are used to extract the features for phishing detection. The TSVM for phishing Website detection is as follows:

- Web image segmentation
- Extraction of feature vectors from the web image
- Extraction of identity feature vectors from DOM objects
- TSVM based phishing classification

In web images segmentation, the webpage is converted into image to obtain pixel and sub-graph positions. To do this spectral clustering is used to cluster the dataset with any shape and size. The extraction of feature vectors from the web image includes gray scale histogram and spacial relationship between sub-graphs. From the DOM objects, five sensitive identities features are mined from HTML basis of the webpage and it’s URL. Now the TSVM is used for phishing website detection which is a binary classification i.e. Phishing or non-phishing.

Table 1: Summary of Existing Machine Learning Approaches for Phishing Detection

Author and Year	No. of Features	Dataset with size	Classifier used	Result parameters						Phishing Attacks addressed
				FP in %	FN in %	Precision in %	Recall in %	Accuracy in %	F-measure in %	
[25]	13 features	3500 ham emails from csmining.org, 500 phishing emails from monkey.org	FFA_SVM	0.01	0.08	99.94	99.92		99.93	Email Phishing
[24]	4 features that includes first, last and middle name	400 ham emails and 600 phishing emails	Natural Language Processing, WordNet	0.02	0.04	99.6	99.3	99.4		Email Phishing
[26]	13 features	1 million ham URL’s from Common Crawl, 1 million phishing URL’s from PhishTank. Com	Random Forest, Short/long term memory time memory (LSTM)	-	-	0.986	0.989	98.7	0.987	URL Phishing

[31]	7 features	613048 phishing email and 46525 emails from WestPac	Multi-layer perceptron, Decision trees, Support vector machine, Naive Bayes, Random Forest	-	-	-	-	99.8	-	Email Phishing
[32]	10 features with some phishing keywords	279 phishing pages, 100 official pages from world wide web	Support Vector Machine	-	-	-	-	84	-	Website Phishing
[28]	10 features	860 phishing email from monkey.org and 6590 ham emails from spam Assassin	PILFER, SpamAssassin	0.12	7.35	-	-	92	-	Email Phishing
[30]	4 features	Trusted webpages are choosen from web	TSVM	-	-	96.4	90.7	95.5	-	Website Phishing
[29]	11 features	800 phishing and ham websites are collected from Webcrawler	Ball support Vector machine (BVM)	0.037	-	0.996	0.946	-	0.963	Website Phishing
[33]	10 features	4202 ham emails and 4560 phishing emails	Neural Networks	-	-	0.925	0.961	95.5	0.957	Email Phishing
FP= False Positive FN= False Negative										

In [34], a multilayer feed forward neural network for phishing email detection is proposed. They collected 4202 ham emails and 4560 phishing emails and divided them into seven folders, 3 for ham emails and four for phishing emails for training and testing. The features are extracted by using a Perl script and provide the output in MIME format. The Multilayer feed-forward neural network is implemented and compared with other popular machine learning algorithms and the accuracy of 95.5%.

Based on the above works a Table 1 is drawn to show how the different machine learning based approaches are used for phishing detecting and their focus.

IV. DISCUSSION

In This work, we discussed different machine learning based phishing detection solutions and most of these works focused on a specific type of phishing attacks i.e. email-based phishing detection, website phishing detection, malicious URLs or attachment detection etc. Machine Learning techniques for another type of phishing attacks are required. There is no any solution that can cover all phishing attacks. Solutions for malware based phishing detections should be focused. The accuracy of the machine learning algorithms depends on the feature vectors. The machine learning based anti-phishing solutions can't detect the attacks in which the text is replaced with images. From the existing works, Support vector machine performs well. As [3] said phishing site that uses embedded objects is still an open challenge as if the solutions available are not up to the mark. The challenges that required further research is listed as follows:

- There is no solution that covers all phishing attacks.
- The existing solutions focus on specific type of attacks only (limited works that cover several phishing attacks).
- There are several factors that can affect the performance of a learning model
    - Selection of feature set.
    - Learning algorithm used.
    - Type of training provided.
- Only limited works focus on detecting the presence of JavaScript.
- Phishing websites with embedded objects are still a challenge.

V. CONCLUSION

Phishing is the most perilous threat in the current internet world. In this paper, we discussed phishing, types of phishing and various machine learning approaches for detecting phishing attacks. Most of the machine learning based solutions focus only on some particular type of phishing attacks (for example, Email phishing filters, malicious URL detection, phishing website detection etc.). Few works focused on email-based phishing attacks, few works on website based phishing detection, Analyzing malicious URLs and malicious attachment using machine learning. There is less number of work has been done in detecting the presence of JavaScript. More research is required to address the DNS phishing attacks, where the user had nothing to do with it.

REFERENCES

[1] Wikipedia, "Phishing," 2018. [Online]. Available: https://en.wikipedia.org/wiki/Phishing.

[2] A. Jain, "Implementing a Web Browser with Phishing Detection Techniques," vol. 1, no. 7, pp. 289–291, 2011.

[3] A. Tewari, A. K. Jain, and B. B. Gupta, "Recent survey of various defense mechanisms against phishing attacks," vol. 6548, no. May, 2016.

[4] C. Wilson and D. Argles, "The fight against phishing: Technology, the end user and legislation," *Inf. Soc. (i-Society), 2011 …*, pp. 501–504, 2011.

[5] Wikipedia, "Social engineering attack." [Online]. Available: https://en.wikipedia.org/wiki/Social_engineering_(security). [Accessed: 01-Dec-2016].

[6] Innovateus, "What are the Different Types of Phishing Attacks?" [Online]. Available: http://www.innovateus.net/science/what-are-different-types-phishing-attacks. [Accessed: 01-Oct-2016].

[7] A. Singh and S. Tripathy, "TabSol: An efficient framework to defend tabnabbing," *Proc. - 2014 13th Int. Conf. Inf. Technol. ICIT 2014*, pp. 173–178, 2014.

[8] Security.stackexchange, "Tabnabbing." [Online]. Available: http://security.stackexchange.com/questions/136227/what-is-tabnabbing. [Accessed: 01-Dec-2016].

[9] P. M. Numerous *et al.*, "Types of Phishing Attacks," pp. 1–2, 2016.

[10] Wikipedia, "Hardware keyloggers." [Online]. Available: https://en.wikipedia.org/wiki/Hardware_keylogger. [Accessed: 01-Dec-2016].

[11] G. Ollmann, "The Pharming Guide," *Security*, pp. 1–37, 2005.

[12] "Types of DNS attacks reveal DNS defense tactics." [Online]. Available: http://searchsecurity.techtarget.com/tip/Types-of-DNS-attacks-reveal-DNS-defense-tactics.

[13] R. W. Steve Jaworski, "Using Splunk to Detect DNS Tunneling," *SANS Inst. InfoSec Read. Room*, 2016.

[14] Pctools, "Zero-Day-Vulnerability." [Online]. Available: http://www.pctools.com/security-news/zero-day-vulnerability/. [Accessed: 01-Oct-2016].

[15] S. Gastellier-prevost, G. G. Granadillo, and M. Laurent, "A dual approach to detect pharming attacks at the client-side," 2011.

[16] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A Survey of Phishing Email Filtering Techniques," *IEEE Commun. Surv. TUTORIALS*, vol. 15, no. 4, 2013.

[17] M. Mishra and A. Jain, "Anti-Phishing Techniques : A Review," vol. 2, no. 2, pp. 350–355, 2012.

[18] N. Vaishnaw and S. R. Tandan, "A Bird's Eye View of Anti-Phishing Techniques for Classification of Phishing E-Mails," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 3, no. 6, pp. 263–275, 2015.

[19] M. Dunlop, S. Groat, and D. Shelly, "GoldPhish: Using images for content-based phishing analysis," *5th Int. Conf. Internet Monit. Prot. ICIMP 2010*, pp. 123–128, 2010.

[20] Y. Zhang, J. Hong, and L. Cranor, "Cantina: a content-based approach to detecting phishing web sites," *... Conf. World Wide Web*, pp. 639–648, 2007.

[21] J. Hajgude and L. Ragha, "Phish mail guard: Phishing mail detection technique by using textual and URL analysis," *Proc. 2012 World Congr. Inf. Commun. Technol. WICT 2012*, pp. 297–302, 2012.

[22] A. P. E. Rosiello, E. Kirda, C. Kruegel, and F. Ferrandi, "A layout-similarity-based approach for detecting phishing pages," *Proc. 3rd Int. Conf. Secur. Priv. Commun. Networks, Secur.*, pp. 454–463, 2007.

[23] Kiran, "Machine Learning Classifiaction," *CodeBytez*, 2018. [Online]. Available: http://www.codebytez.com/machine-learning-basics/. [Accessed: 10-Aug-2018].

[24] S. Aggarwal, V. Kumar, and S. Sudarasan, "Identification and Detection of Phishing Emails Using Natural Language

Processing Techniques," *SIN '14 Proc. 7th Int. Conf. Secur. Inf.*, 2014.

[25] O. A. Adewumi and A. A. Akinyelu, "A hybrid firefly and support vector machine classifier for phishing email detection," *Kybernetes*, vol. 45, no. 6, pp. 977–994, 2016.

[26] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and F. A. Gonz, "Classifying Phishing URLs Using Recurrent Neural Networks," 2017.

[27] L. Ma, B. Ofoghi, P. Watters, and S. Brown, "Detecting phishing emails using hybrid features," *UIC-ATC 2009 - Symp. Work. Ubiquitous, Auton. Trust. Comput. Conjunction with UIC'09 ATC'09 Conf.*, pp. 493–497, 2009.

[28] I. Fette *et al.*, "Learning to Detect Phishing Emails," vol. 0389, no. June, 2006.

[29] Y. Li, L. Yang, and J. Ding, "A minimum enclosing ball-based support vector machine approach for detection of phishing websites," *Opt. - Int. J. Light Electron Opt.*, vol. 127, no. 1, pp. 345–351, 2016.

[30] Y. Li, R. Xiao, J. Feng, and L. Zhao, "A semi-supervised learning approach for detection of phishing webpages," *Opt. - Int. J. Light Electron Opt.*, vol. 124, no. 23, pp. 6027–6033, 2013.

[31] S. B. Liping Ma, Bahadorrezda Ofoghi, Paul Watters, "Detecting phishing emails using hybrid features," *UIC-ATC 2009 - Symp. Work. Ubiquitous, Auton. Trust. Comput. Conjunction with UIC'09 ATC'09 Conf.*, pp. 493–497, 2009.

[32] Y. Pan and X. Ding, "Anomaly Based Web Phishing Page Detection," *Comput. Secur. Appl. Conf.*, pp. 381–392, 2006.

[33] H. Zhang, G. Liu, T. W. S. Chow, and W. Liu, "Textual and visual content-based anti-phishing: A Bayesian approach," *IEEE Trans. Neural Networks*, vol. 22, no. 10, pp. 1532–1546, 2011.

[34] and Y. Y. Zhang, Ningxia, "Phishing Detection Using Neural Network- CS229 lecture notes," in *CS229 lecture notes*, 2012.