Data Analytics: intelligent anti-phishing techniques based on Machine Learning[1][2]

ABSTRACT

According to the international body Anti-Phishing Work Group (APWG), phishing activities have skyrocketed in the last few years and more online users are becoming susceptible to phishing attacks and scams. While many online users are vulnerable and naive to the phishing attacks, playing catch-up to the phishers' evolving strategies is not an option. Machine Learning techniques play a significant role in developing effective anti-phishing models. This paper looks at phishing as a classification problem and outlines some of the recent intelligent machine learning techniques (associative classifications, dynamic self-structuring neural network, dynamic rule-induction etc.) in the literature that is used as anti-phishing models. The purpose of this review is to serve researchers, organizations' managers, computer security experts, lecturers, and students who are interested in understanding phishing and its corresponding intelligent solutions. This will equip individuals with knowledge and skills that may prevent phishing on a wider context within the community.

Keywords: Classification; Data Mining; Dynamic Self-structuring Neural Network; Intelligent Anti-Phishing; Machine Learning.

[1] Preprint of an article published in [Journal of Information & Knowledge Management, Vol. 18, No. 1 (2019) 1950005 (17 pages)] [Article DOI: 10.1142/S0219649219500059] © [copyright World Scientific Publishing Company] [Journal URL: https://www.worldscientific.com/worldscinet/jikm]

[2] Cite as: Baadel, S., Lu, J. (2019). Data Analytics: intelligent anti-phishing techniques based on Machine Learning. Journal of Information & Knowledge Management, Vol. 18, No. 1 (2019), 1-17.

1.0 Introduction

Phishing is an attempt to gain sensitive personal and financial information such as usernames and passwords, account details and social security numbers with malicious intent via online deception [2][46][3]. Phishing typically employs identity theft social engineering techniques such as creating websites that replicate an existing authentic one and through a seemingly legitimate email sent to users asking them to click on a hyperlink within it that will route them to their fraudulent website where they can persuade unsuspecting users to divulge their private information and credentials [16][11].

Advancements in computer networks and cloud technology in recent years have resulted in an exponential growth of online and mobile commerce where customers perform online purchases and transactions [5]. This online growth has culminated in phishing activities to reach unprecedented levels in recent months. Anti-Phishing Work Group (APWG), an international body that aims to minimise online threats including pharming, spoofing, phishing, malware, etc., published their report [1] in February of 2017 suggesting that there were approximately 1,220,523 phishing attacks in 2016, an increase of more than 65% from the previous year with an average of more than 92,500 phishing attacks per month in the fourth quarter of 2016. As more and more users become prone to information breaches and becoming victims of identity theft, their trust in e-commerce or e-banking websites and platforms will diminish thus resulting in billions in online losses [40][35].

Research by [26][52][53][55] that utilizes the Elaboration Likelihood Model (ELM), a model by Petty and Cacioppo [42] suggests that user's cognitive processing is a key reason why many fall victim to phishing. How a user pays attention to some of the cues in a phishing email (i.e. initial noticing of something fishy in the sender's email or website URL) – what ELM classifies as attention – and consequently digging deeper to search for more cues – what ELM classifies as elaboration process, is a key factor for a user to successfully identify a fraudulent website or falling prey to a phishing scam.

So, why is there an alarming increase in phishing activities and more users becoming susceptible to phishing scams? The answer to this can be summarized as due to the users' naivety and inexperience in interacting with and using online communication channels. According to a phishing survey [58], users don't have security and privacy as their main concern when they are online. Since this is a human problem, software solutions are not able to provide a permanent solution to it. The problem can be minimized by addressing it in two folds; developing more targeted anti-phishing interventions and techniques, and educating the public on how to detect and identify fraudulent phishing websites. As phishing scams and techniques evolve, anti-phishing solutions that adopt Machine Learning (ML) tend to be more practical and effective in combating phishing [36].

In this paper, we investigate the phishing problem and define it in an ML context. Other approaches such as the visual similarity based have been discussed in detail by [57]. Other older reviews such as [58] discuss the countermeasures in a more broad manner. The paper investigates intelligent ML anti-phishing techniques and critically analyses their benefits and disadvantages theoretically. This paper serves researchers, organizations' managers, computer security experts, lecturers, and students who are interested in understanding phishing and its corresponding intelligent solutions.

The rest of the paper is structured as follows: Section 2 presents the phishing attack procedure. Section 3 briefly discusses common anti-phishing techniques and critically analyses the intelligent anti-phishing solutions that employ different strategies in deriving the anti-phishing models. In Section 4 we provide a brief discussion of the solutions discussed in the paper and finally the conclusions in Section 5.

2.0 Phishing Attack Taxonomy and Procedure

Phishing attacks are classified based on the mechanism used by the Phisher to defraud unsuspecting user. The following Figure 1 provides a basic taxonomy.

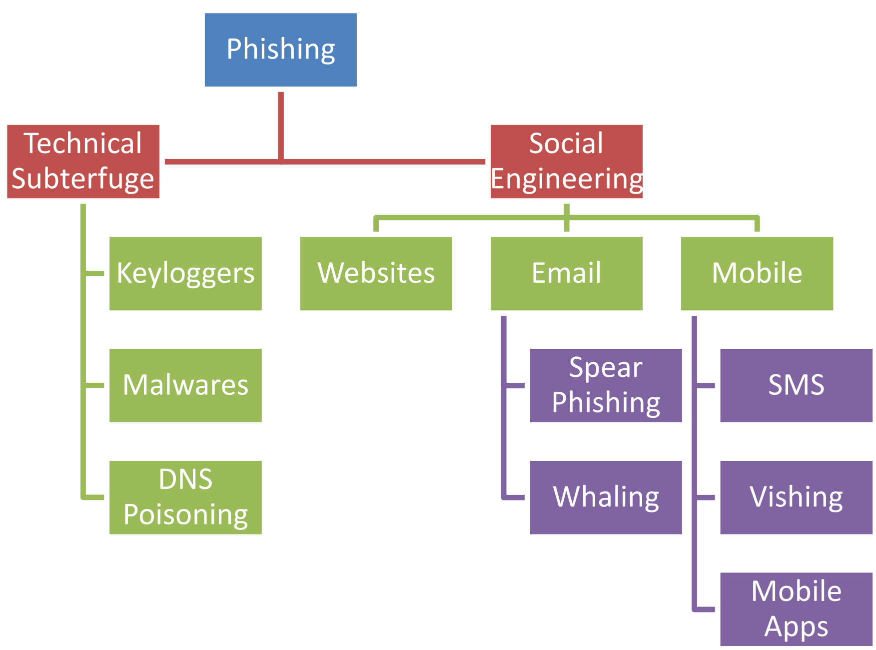

Figure 1. Phishing Attacks Taxonomy

Phishing attacks occur in many forms through malware, keyloggers, DNS poisoning etc. Other social engineering initiation processes include online blogs, short message services (SMS), social media websites using web 2.0 services such as Facebook and Twitter, peer to peer (P2P) file-sharing services, Voice over IP (VoIP) systems where spoofing caller IDs are used by attackers [4] [30]. Each of these phishing methods has a slight variation on how the procedure is done all with the goal of defrauding the unsuspecting user. Email phishing attacks are often initiated when a phisher sends an email to unsuspecting potential victims with a link that can direct them to a phony website that resembles one that is legitimate. To see how the phishers design their scheme, Figure 2 below presents an example of a phishing attack life cycle by email. In this technique, the phisher adds a hyperlink that routes unsuspecting users to a phony website. The process can be summarised as follows:

1) Phishers set up a phony website resembling a legitimate one.
2) A hypertext link is sent via an email asking potential victims requesting them to click it in order to take immediate action such as updating their account information, resetting their password etc. The urgency in such email is a vital element to bait unsuspecting users.
3) Once clicked, the link routes the users to the fraudulent phishing website.
4) The fraudulent website collects vital sensitive information such as username and password, account details, social security numbers etc.
5) Embezzled information can be used for financial gain, identity hiding, or other cybercrimes.

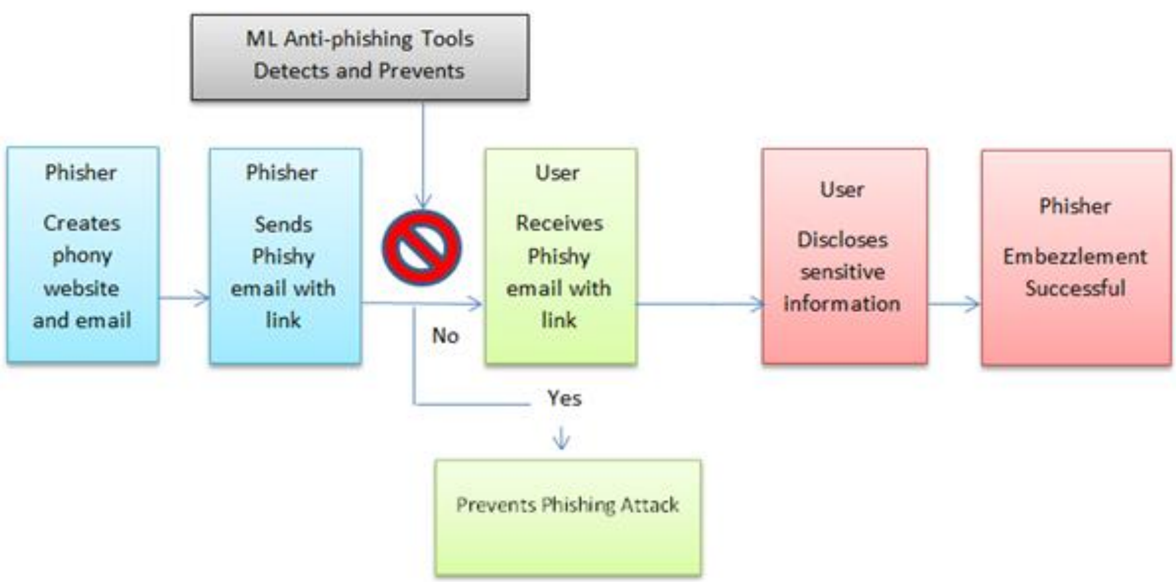

Figure 2. Phishing Attack Life-cycle.

One of the common misconceptions regarding phishing websites is that grammatical errors and typos are typical [45]. While this may be true with novice phishers, it is not necessarily the case with many phishing websites. In a study to understand and evaluate the evolution of techniques used by phishers, Gupta and Kumaraguru [24] concluded that some of the features of phishing emails that compel users to click on phishing links include more legitimate-looking URLs and free subdomains and sending more creative promotional emails to lure users into clicking phishing URLs. As phishers evolve and sharpen their techniques, it is becoming more difficult for novice online users to detect or distinguish phishing websites from legitimate ones.

3.0 Anti-Phishing Methods

Due to the broad nature and severity of phishing scams to individual users, businesses, government entities, and non-profit organizations, there have been different methods proposed in the literature to combat phishing. This paper will focus on some of the common intelligent anti-phishing solutions.

In order to understand the common evaluation metrics where the goal is to detect and identify phishing instances, it important to first mention the four classification possibilities that exist on any given dataset mixture of phishing and legitimate instances. Table 1 below highlights these possibilities.

Table 1. Classification matrix

Instance	Classified as Legitimate	Classified as Phishing
Legitimate	i. correctly identified as legitimate (L_L)	ii. incorrectly identified as phishing (L_P)
Phishing	iii. incorrectly identified as legitimate (P_L)	iv. correctly identified as phishing (P_P)

Computerised anti-phishing techniques use the following common evaluations metrics:

i. True Positive Rate (TPR) – correctly detected phishing instances
ii. False Positive Rate (FPR) – legitimate instances that are incorrectly identified as phishing
iii. False Negative rate (FNR) – phishing instances incorrectly identified as legitimate
iv. True Negative rate (TNR) – correctly detected legitimate instances

Let N_p denote total phishing websites and N_L denote total legitimate websites, then phishing detection performance can be evaluated as follows:

$$\text{TPR} = \frac{P_P}{N_P} \times 100 \quad (1)$$

$$\text{FPR} = \frac{L_P}{N_L} \text{ x } 100 \tag{2}$$

$$\text{FNR} = \frac{P_L}{N_P} \text{ x } 100 \tag{3}$$

$$\text{TNR} = \frac{L_L}{N_L} \text{ x } 100 \tag{4}$$

Precision (*P*) measures the phishing instances identified and detected correctly out of all phishing instances.

$$\text{Precision} = \frac{P_P}{P_{P+} L_P} \text{ x } 100 \tag{5}$$

Accuracy (*A*) measures the rate of phishing and legitimate instances identified correctly out of all instances.

$$\text{Accuracy} = \frac{P_P + L_L}{N_{P+} N_L} \text{ x } 100 \tag{6}$$

Recall (*R*) measures the rate of phishing instances identified correctly out of all correctly identified phishing and legitimate instances.

$$\text{Recall} = \frac{P_P}{P_{P+} L_L} \text{ x } 100 \tag{7}$$

f1 score is the harmonic mean of Precision and Recall

$$\text{f1 score} = \frac{2\ x\ P\ x\ R}{P+R} \tag{8}$$

3.1 Simulated Phishing attacks and Embedded Training

A study by Alsharnouby et al. [12] where improved browser security indicators and visual cues are used to attract attention to users to identify phishing websites found that there was a correlation between users gazing at the visual cues and detecting phishing sites. However, the vast majority of online users are unaware of how phishing attacks start or how visually to recognize and differentiate between a fraudulent phishing site from a legitimate one [36][27].

There are a number of research studies that are done to train users and raise awareness on phishing [14][15][47][28][20][21][33]. These early studies involved either sending unsuspecting participants an email with links and monitoring how they respond to them or making the participants aware that they are participating in a simulated phishing study and are gauged on their abilities to correctly identify phishing emails from legitimate ones. At the end of the training, participants are normally given the training materials and are informed about their vulnerability to phishing. For example, a study by [28] of 921 students from the University of Indiana revealed that students who received an email that was perceived to be from a friend clicked on the link 72% of the time compared to 16% when it was from an unknown address. A similar pilot study was conducted by Arachchilage and Cole [14] using an embedded training methodology to measure phishing awareness at a university. A later study by Arachchilage and Love [13] investigated whether an interactive mobile platform is effective in educating users in contrast to traditional security training. A comparison of user responsiveness to phishing was conducted using a developed mobile game [14], compared to training through a website designed by APWG. Results indicated that users trained through mobile application had a higher success rate of identifying phishing sites compared to their counterparts who only used the APWG website.

3.2 Databases (Blacklist and Whitelist)

A database driven approach to fighting phishing, called blacklist, is a collection of previously identified and detected phishing domain names or URLs developed by several research projects [48]. A blacklisted website significantly loses its user traffic and any potential revenues. Public available blacklists effectiveness depends on;

a) frequency of the database update (fast access time)

b) accurate phishing detection rate i.e. TP

Blacklist solutions tend to have a better frequency of the database update i.e. fast access time but suffer mainly on the detection rate [56]. Google and Microsoft blacklist, commonly used by businesses because of their lower false positive (FP) rates, and due to their database update frequencies have Microsoft's blacklist updated between nine hours to six days whereas Google's blacklist gets updated between twenty hours to twelve days [36]. This is definitely a limitation on the blacklist approach as phishing campaigns take significantly lower times in their attacks before they can be detected and blocked [5][48].

A different approach was to create a whitelist database of legitimate URLs as opposed to blacklists. A proposal by Chen and Guo [18] was that if a user's login attempt to a certain website was successful then the site is assumed to be legitimate and the URL can be added to the whitelist database. Phishzoo [11] is a technique that constructs a website profile using a fuzzy hashing approach. The website profile is contrasted with existing profiles in their whitelist and if an identical match is found or the security certificate matches then the website is added to the list, otherwise flagged as suspicious.

3.3 Intelligent Anti-Phishing Techniques based on ML

According to Witten and Frank [54], one of the main tasks of ML is the prediction of a target variable within datasets based on other variables. This prediction occurs in an automated manner using a classification model referred to the classifier. Given a test data, the classifier tries to predict a target variable as accurately as possible. In supervised learning, this is classification. Abdelhamid and Thabtah [5] defined classification as the ability to "accurately" predict class attributes for a test instance using a predictive model derived from a training dataset.

In classification context, website phishing can be viewed as involving automatic categorization of websites into a predefined set of class values based on a number of available features (variables) and the class variable. ML anti-phishing techniques rely on website features to derive knowledge that can assist in identifying phishing websites and minimizing the problem. Due to the numerous amounts of features linked with a website, it becomes necessary to pre-process the feature set in order to pick the most effective features in order to enhance the predictive process. The effectiveness of these features can be measured using computational intelligence methods such as correlation analysis, information gain, and Chi-Square etc. [34].

Many ML and data mining (DM) algorithms for classification that have been developed using one of the following major classification approaches in deriving their predictive systems:

1) Decision trees (C4.5, and their successors) [44].
2) Rule-based classification such as associative classification (AC) [51].
3) Neural Networks (NN) methods and their successors [23].
4) Support Vector Machine (SVM) [31]
5) Fuzzy Logic (FL)

In the following subsections, we critically analyze intelligent anti-phishing attempts based on ML for the five outlined approaches. We show how these approaches derive a classification anti-phishing system along with some of their benefits and weaknesses.

3.3.1 Decision Trees and Rule Induction

A Random Forest method called Phishing Identification by Learning on Features of Email Received (PILFER) that utilized the C4.5 decision tree classifier including the Random Forest, SVM, and Naïve Bayes was developed [22]. The authors conducted an experiment on a set of 860 phishy and 695 ham emails where various features for distinguishing phishing emails were identified such as IP URLs, time of space, HTML messages, number of connections inside the email, and JavaScript among others. The authors suggested that PILFER can be improved towards grouping messages by joining all ten features discovered in the classifier aside from "Spam filter output".

Mohammed et al. [38] developed a special handcrafted rule to collect data based on statistical analysis of a security dataset that contains 2500 instances and 16 features. This was used to investigate a number of rule induction algorithms on the problem of website phishing classification which was then compared to RIPPER [19], C4.5 (Rules) [44], CBA [34], and PRISM [17]. The investigation of the four rule-based classification methods suggested that there were eight effective features that can be employed by the classification algorithm in combating phishing: SSL and HTTPS, Domain-age, Site-traffic, Long-URL, Request-URL Sub-domain, Multi—sub-domain, Suffix-prefix, and IP-address.

After studying the problem of email-based phishing a proposal of combining a RIPPER classifier with fuzzy logic was suggested [32]. The authors envisioned the role of fuzzy logic to pick the main features of the email and rank them based on a probability score while RIPPER was to automatically use those features to classify the type of emails as ham or phishy. They utilized two components of the email; content data and metadata. The content data or the email message to look for spelling errors, embedded links etc., and the metadata or URL to investigate the IP address, length, long URL, Suffix Prefix, Crawler URL, Nonmatching URL etc. The experiment had very limited data consisting of only 100 instances from phishtank using the WEKA software tool. Results showed that there were twelve rules generated by RIPPER from the dataset with 85.4% prediction rate. However, no comparison with other fuzzy logic or rule-based classifications was conducted by the authors in their experiment.

Aburrous et al. [9] classified web features into six criteria. Using WEKA, [7] investigated rule induction methods to seek their applicability for categorizing websites based on phishing features in their earlier study by Aburrous et al. in [9]. Many experiments with four classification algorithms (RIPPER, PART, PRISM, C4.5) were conducted. The focus of the experiments was the classification accuracy of the classifiers produced. They concluded that rule induction was able to detect 83% of phishing websites. The authors suggested that the results could be further enhanced when careful feature selection is employed.

3.3.2 Associative Classification (AC)

A number of research have been conducted to evaluate the applicability of two AC methods named CBA and multi-class classification based on association rule (MCAR) in phishing using phishtank dataset [34][51][8][43]. For example, Aburrous et al. [8] used a dataset with 27 different features and applied CBA, MCAR, and four other rule-based classifiers using the WEKA tool. Their aim was to assist security managers within organizations by building an intelligent anti-phishing tool within browsers that can detect phishing as accurately as possible. Experimental results revealed that the AC methods despite having higher predictive classifiers generated more rules than the rest of the algorithms. The AC systems showed higher correlations among features linked with three major criteria: URL, Domain Identity, and Encryption. However, the massive number of rules derived by MCAR and CBA may overwhelm end-users that may not be able to control the anti-phishing system. The authors did not implement the AC rules within a browser to evaluate its real performance making it difficult to measure the success or failure of their classification systems.

A more domain-specific AC anti-phishing systems that took modified the phishing problem into three class values of legitimate, phishy, and a much harder to detect case of “suspicious” label was developed by Abdelhamid et al. [3][4]. Instances that cannot be fully phishy nor legitimate are very hard to detect by typical ML algorithms, thus increasing their false positive (FP) rates. The authors have expanded the current intelligent classification systems by including two distinct advantages:

1) Extending the phishing problem to include suspicious cases, making it more realistic.
2) Proposing a new multi-label learning phase that can discover disjunctive in addition to conjunctive rules. These additional disjunctive rules are tossed out by existing AC methods. This new multi-label phase enhances predictive power and provides more useful knowledge to the end-user.

The experimental results on the data indicated a higher performance of the new multi-label associative classifiers compared with CBA, MCAR, rule induction, and decision trees.

An AC mining classifier called Fast Associative Classification Algorithm (FACA) that employs a vertical mining approach called Diffset to discover frequent itemsets and uses the All Exact Match prediction method to classify unseen instances was proposed by [25]. Diffset keeps track of only the transactions IDs in which a rule item does not occur. The authors use the algorithm to investigate a dataset of phishing websites using the 10-fold cross-validation testing method. Using a min support and min confidence threshold of 2% and 50% in WEKA, the authors used chi-square feature selection filter method on the phishing websites dataset and compared their algorithm with CBA, MCAR among others and concluded that FACA outperformed all mentioned AC Algorithms.

3.3.3 Neural Network (NN)

An experimental study contrasting five ML algorithms namely Classification and Regression Trees (CART), NN, Random Forests (RF), Bayesian Additive Regression Trees (BART), and Logistic Regression (LR) on the problem of classifying emails to ham or suspicious in order to measure the most successful approaches in email phishing detection was conducted by Abu-Nimeh et al. [6]. A training dataset consisting of 2889 emails and 43 email's features was used. The authors employed a ten-fold cross-validation to test their experiment using the evaluation measures of precision, recall, and harmonic mean. Results revealed that RF achieved a lower error rate while NN generated the highest error rate among classifiers. However, despite RF generating the highest predictive classifiers, it also derived the least false positive (FP) rate among all contrasted algorithms. It was concluded that a more carefully chosen features set may improve the performance of the anti-phishing email tool.

Mohammad et al. [39] tested the ANN Back Propagation algorithm to measure the correlation between the features and target attributes using simple univariate statistical analysis (frequency of features values and the target attribute values) to derive anti-phishing models. The authors used a dataset with over 2000 instances from different legitimate and phishing sources. The experiment showed an increased accuracy of the models generated from the Back Propagation algorithm when compared with other classification algorithms.

An implementation of a multilayer Feed Forward NN (FWNN) based on Back-Propagation was applied on an email phishing classification problem to differentiate suspicious from legitimate emails was proposed [29]. The authors used eighteen binary features (0, 1) extracted from the email metadata and content data (header and HTML body) as the training dataset attributes. These features were given values based on human rules developed by security domain experts. To derive the NN models, 6000 ham and suspicious emails were used. The results obtained showed that FFNN is able to categorize emails with less than a 2% error rate and with high speed. However, the authors did not embed their FFNN into browsers for live testing.

Mohammad et al. [37] developed a self-structuring NN classification algorithm that dealt with the vitality of phishing features that improved the learning phase based on previous training experience. The algorithm employed validation data to track the performance of the constructed network model and made the appropriate decision based on results obtained against the validation dataset. During the training process, when the achieved error against the network is smaller than the minimum achieved error, the algorithm saved the network's weights and did not save the weight if the error was larger. Experimental results obtained against a phishing dataset of thirty features and over 10000 instances showed that the self-structuring NN model was able to generate anti-phishing models more accurately than traditional classification approaches such as C4.5.

One of the common ways to train an NN is through trial and error. This methodology suffers a major drawback due to the fact that a lot of time is required to tune the parameters and also a domain expert may be needed to decipher the dataset. Instead of the trial and error, an improved self-structuring NN anti-phishing model was proposed by Thabtah et al. [50]. Their algorithm would update several parameters such as the learning rate in a more dynamic way prior to adding a new neuron to the hidden layer. These NN features are updated during the building of the classification model and are based primarily on the computed error rate, desired error rate, and the network environment. A large dataset from UCI with over 11000 websites was utilized. The experimental results showed the dynamic NN anti-phishing model had a better predictive accuracy compared to Bayesian network and decision trees.

3.3.4 Support Vector Machine (SVM)

The SVM classification method proposed by Joachim [31] evaluated the discrepancy between a website's identity, its HTTP transactions, and structural features. Once a new website identity and its structural features were captured (Abnormal URL, Abnormal anchors, Server Form Handler, Abnormal certificate in SSL, Abnormal DNS, Abnormal cookies), an SVM algorithm is trained on a historical dataset consisting of the features in order to derive the new website type. Experimental results on six features using the proposed SVM indicated that the first layer that involves the website's identity extraction helps toward increasing the detection rate since malicious websites are not correlated. The SVM model achieved a little over 83% prediction rate indicating that the feature selection phase needs to include other features that may improve the performance of the classifier.

The integration of the Firefly Algorithm (FFA) with SVM, FFA-SVM, to construct a robust hybrid classifier for parameter optimization was proposed by Adewumi and Akinyelu [10]. The authors compared the results with random forest, Clustering, and SVM and concluded the FA-SVM produced 99.98% accuracy with a False Negative and False Positive rate of 0.08 and 0.01 respectively.

3.3.5 Fuzzy Logic

Aburrous et al. [9] utilized Fuzzy Logic to investigate phishing in electronic banking (Ebanking) applications. After obtaining the necessary authorizations, the authors sent a simulated phishing email with the help of the security manager of a bank to measure security indicators of phishing among a sample of 120 employees. The email urged the selected participants to log in and reactivate their accounts since previous server maintenance necessitated their reactivations. Their experimental study yielded interesting results where about 37% of the targeted employees readily submitted their credentials without any hesitation, of which 7% was Information Technology employees. The simulated email was used to determine features that users may look for when they suspect phishing. The authors used FL as an anti-phishing model to help classify websites into legitimate or phishy. Their proposed FL classification model was built manually to categorize websites using the six criteria listed in Table 2. Each of those criteria contains a number of phishing indicators as. Each feature in the dataset was assigned three possible values by the authors: Phishy, Genuine, and Doubtful. They concluded that Domain Identity and URL were the two effective indicators to distinguish phishiness in websites.

A fuzzy-based phishing technique that combines fuzzy and NN (neuro-fuzzy without rule set) is proposed by Nguyen et al. [40] to classify websites based on a smaller set of phishing features related to the website's URL and rank. The technique is categorized into 4 layers. The first layer uses 6 heuristics (primary domain, subdomain, path domain, pagerank, alexarank, alexareputation) which are considered the input layer. The values of these heuristics are calculated and given fuzzy values to determine if they are legitimate or phishing. The third layer calculates the weighted sum of nodes from the second layer, and returning the mean legitimate (ML) and mean phishing (MP). The last layer is the output node with the value of layer 3 (between 0 and 1). This layer classifies the output into 2 classes, phishing if the value passed is less than 0.5, and legitimate if the value is greater than 0.5. Results were compared to that of [7] on fuzzy techniques and found their technique was able to slightly enhance the phishing detection rate.

4.0 Summary: Anti-phishing Solutions

As phishers' techniques evolve and their phishing attacks become more sophisticated, their systematic attack strategies make it harder for even security experts to keep up. This makes ordinary users vulnerable. 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. It is therefore imperative that advance intelligent ML approaches are used and are a necessity to combat the phishing menace. Table 2 below summarizes some of the common anti-phishing methods based on ML that are discussed in this paper.

Table 2. Summary of anti-phishing methods based on ML

Method name	ML technique	Approach Limitation	Reference
PILFER	Decision tree	Rule pruning scheme to reduce the number of rules and increase generalization of the classifier Limitation: High time and storage complexity	[22]
Enhanced Dynamic rule induction	Rule induction and covering approaches		[49][50]
RIPPER with Fuzzy	Rule induction with Fuzzy		[32]
Classification based association	AC		[7][8]
Multi-label Classifier based Associative Classification	AC		[3][4]
Fast Associative Classification Algorithm	AC		[25]
Self-structuring neural network	NN	Limitations: a Large amount of misclassifications. Need to improve the performance of the classifier	[37][49]
Neural Network trained with Back-Propagation	NN		[39]
Feed Forward Neural Network	NN		[29]
Fuzzy DM	Fuzzy logic		[9]
Neuro-Fuzzy	Fuzzy with NN		[40]
Page classifier	SVM		[31][41]
Hybrid Firefly Algorithm	SVM		[10]

5.0 Conclusions

In this paper, phishing has been described in the classification context where website phishing is viewed as involving automatic categorization of websites into a predefined set of class values based on a number of available features (variables) and the class variable. ML anti-phishing techniques rely on website features to derive knowledge that can assist in identifying phishing websites. The phishing problem cannot be eradicated completely but rather be minimized by addressing it in two folds;

developing more targeted anti-phishing interventions and techniques, and educating the public on how to detect and identify fraudulent phishing websites. ML anti-phishing techniques are needed to combat the ever-evolving and sophistication of phishing attacks and strategies. Thus, the focus of the paper was on predictive models produced by ML anti-phishing techniques; Rule induction, decision trees, associative classification, SVM, NN, and computational intelligence. The paper critically analyzed the ways these anti-phishing methods work and showed their positive and negative aspects of the user and performance perspective.

In future work, it is planned to present an anti-phishing framework that integrates automated knowledge produced by computational intelligence in visual cues besides using human expert knowledge as a base.

References

[1] Aaron, G., Manning, R. *APWG Phishing Reports*. 2017. http://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf [Accessed August 20th, 2017].

[2] Aaron G, Rasmussen R. *Global phishing survey: trends and domain name used in 2H 2009*. Lexington, MA: Anti-Phishing Working Group (APWG). 2010.

[3] Abdelhamid N. *Multi-label rules for phishing classification*. Applied Computing and Informatics 11 (1), 29-46. 2015.

[4] Abdelhamid N., Thabtah F., Ayesh A. *Phishing detection based associative classification data mining.* Expert systems with Applications Journal. 41 (2014) 5948–5959. 2014.

[5] Abdelhamid N., Thabtah F. *Associative Classification Approaches: Review and Comparison.* Journal of Information and Knowledge Management (JIKM). Vol. 13, No. 3 (2014) 1450027. 2014.

[6] Abu-Nimeh, S., Nappa, D., Wang, X. and Nair. *A Comparison of Machine Learning Techniques for Phishing Detection*. The 2nd annual Anti-Phishing Working Groups eCrime researchers, eCrime '07. New York, NY, USA. ACM. 2007.

[7] Aburrous M., Hossain M., Dahal K.P. and Thabtah F. E*xperimental Case Studies for Investigating E-Banking Phishing Techniques and Attack Strategies.* Journal of Cognitive Computation, Springer Verlag, 2 (3): 242-253. 2010.a

[8] Aburrous M., Hossain M., Dahal K.P., and Thabtah F. *Associative Classification techniques for predicting e-banking phishing websites.* Proceedings of the 2010 International Conference on Information Technology, Las Vegas, Nevada, USA, 2010, pp. 176-181. 2010.b

[9] Aburrous M., Hossain A., Dahal K., Thabtah F. *Intelligent Quality Performance Assessment for E-Banking Security using Fuzzy Logic.* Proceedings of the 7th IEEE International Conference on Information Technology (ITNG 2008). Las Vegas, USA. 2008.

[10]Adewumi, O., Akinyelu, A. *A hybrid firefly and support vector machine classifier for phishing email detection.* Kybernetes, Vol. 45 Issue 6 pp. 977 – 994. 2016.

[11]Afroz, & Greenstadt, R. *PhishZoo: Detecting Phishing Websites by Looking at them.* Proceedings of the Fifth International Conference on Semantic Computing. Palo Alto, California, USA. IEEE. 2011.

[12]Alsharnouby M., Alaca F., Chiasson S. *Why phishing still works: User strategies for combating phishing attacks.* International Journal of Human-Computer Studies. Vol. 82, 69-82. 2015.

[13]Arachchilage N., Love S. *A game design framework for avoiding phishing attacks.* Computers in Human Behavior 29 (3), 706-714. 2013.

[14]Arachchilage N., Cole, M. *Design a mobile game for home computer users to prevent from "phishing*

attacks". International Conference on Information Society (i-Society) pp. 485-489. 2011.

[15] Arachchilage N., Rhee Y., Sheng S., Hasan SH., Acquisti A., Cranor L. F., Hong J. *Getting users to pay attention to anti-phishing education: evaluation of retention and transfer.* Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit. Pittsburgh, PA, USA. ACM. 2007.

[16] Atkins, B., Huang, W. *A study of social engineering in online frauds*. Open J Soc Sci, 1 (03), pp. 23-32. 2013.

[17] Cendrowska, J. P*RISM: An algorithm for inducing modular rules.* International Journal of Man-Machine Studies, Vol.27, No.4, 349-370. 1987.

[18] Chen, J., Guo, C. *Online Detection and Prevention of Phishing Attacks (Invited Paper).* In First International Conference on Communications and Networking in China. ChinaCom '06. Beijing. IEEE. 2006.

[19] Cohen, W. *Fast Effective Rule Induction.* In In Proceedings of the Twelfth International Conference on Machine Learning. Tahoe City, California, Morgan Kaufmann. 1995.

[20] Downs, J., Holbrook, M., Cranor, L. *Decision strategies and susceptibility to phishing.* Symposium on Usable Privacy and Security, Pittsburgh, PA. 2006.

[21] Downs, J., Holbrook, M., Cranor, L. *Behavioral response to phishing risk.* 2nd annual eCrime researcher's summit. Pittsburgh, PA. The USA. 2007.

[22] Fette I., Sadeh N., Tomasic A. *Learning to detect phishing emails.* Proceedings of the 16th international conference on World Wide Web. 649-656. 2007.

[23] Grossberg. *Nonlinear neural networks: Principles, mechanisms, and architectures.* Neural Networks, 1(1), p.17–61. 1988.

[24] Gupta, S., Kumaraguru, P. *Emerging Phishing Trends and Effectiveness of the Anti-Phishing Landing Page.* eCrime Researchers Summit, eCrime. 2014.

[25] Hadi, W., Aburub, F., Alhawari, F. *A new fast associative classification algorithm for detecting phishing websites.* Applied Soft Computing. Vol. 48 Pg. 729-734. 2016.

[26] Harrison, B., Vishwanath, A., Yu, J. Ng and Rao, R. *Examining the impact of presence on individual phishing victimization.* 48th Hawaii International Conference on System Sciences (HICSS), pp. 3483-3489. 2015.

[27] Huang H., Tan J., Liu L. *Countermeasure techniques for deceptive phishing attack.* International Conference on New Trends in Information and Service Sciences. Pg 636-641. 2009.

[28] Jagatic, T., Johnson, N., Jakobsson, M., Menczer, F. *Social phishing.* Communications of the ACM, 50 (10), pp. 94-100. 2007.

[29] Jameel N. Gh., George L. *Detection of Phishing Emails using Feed Forward Neural Network.* Journal of Computer Applications 77(7):10-15. 2013.

[30] James, L. *Phishing Exposed*. Syngress Publishing. 2005.

[31] Joachim H. *Large-scale support vector machine learning practical, Advances in kernel methods: support vector learning*, MIT Press, Cambridge, MA. 1999.

[32] Khadi A., Shinde S. *Detection of phishing websites using data mining techniques.* International Journal of Engineering Research and Technology, Volume 2(12). 2014.

[33] Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L., et al. *Getting users to pay attention to anti-phishing education: evaluation of retention and transfer.* 2nd annual eCrime researchers summit, Pittsburgh, PA. The USA. 2007.

[34] Liu, H., Setiono, R. *Chi2: Feature Selection and Discretization of Numeric Attribute.* Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence, November 5-8, 1995, pp. 388. 1995.

[35] McCall, Gartner, Inc. http://www.gartner.com/newsroom/id/565125 [Accessed August 20th, 2017]. 2011.

[36] Mohammad R., Thabtah F., McCluskey L. *Tutorial and critical analysis of phishing websites methods.* Computer Science Review Journal. Volume 17, August 2015, Pages 1–24 Elsevier. 2015.

[37] Mohammad R., Thabtah F., McCluskey L., *Predicting Phishing Websites based on Self-Structuring Neural Network.* Journal of Neural Computing and Applications, 25 (2). pp. 443-458. ISSN 0941-0643. Springer. 2014.a

[38] Mohammad R., Thabtah F., McCluskey L., *Intelligent Rule based Phishing Websites Classification.* Journal of Information Security (2), 1-17. ISSN 17518709. IET. 2014.b

[39] Mohammad, R. M., Thabtah, F. & McCluskey, L. *Predicting Phishing Websites using Neural Network trained with Back-Propagation.* Las Vegas, World Congress in Computer Science, Computer Engineering, and Applied Computing, pp. 682-686. 2013.

[40] Nguyen L. A. T., To B. L., and Nguyen H. K. *An Efficient Approach for Phishing Detection Using Neuro-Fuzzy Model.* Journal of Automation and Control Engineering Vol. 3, No. 6. 2015.

[41] Pan Y., and Ding X. *Anomaly Based Web Phishing Page Detection.* The 22nd Annual Computer Security Applications Conference (ACSAC). Miami Beach, Florida, USA, 2006. IEEE. 2006.

[42] Petty, R.E. and Cacioppo, J.T. T*he elaboration likelihood model of persuasion.* L. (Ed.), Advances in Experimental Social Psychology, Vol 19, Academic Press, New York, NY, pp. 123-205. 1986.

[43] Qabajeh I., Thabtah F., Chiclana F. D*ynamic Classification Rules Data Mining Method.* Journal of Management Analytics. Vol 2, Issue 3, pp. pages 233-253. Wiley. 2015.

[44] Quinlan, J. *C4.5: Programs for machine learning.* San Mateo, CA: Morgan Kaufmann. 1993.

[45] Rader M., Rahman S. *Exploring historical and emerging phishing techniques and mitigating the associated security risks.* International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.4. 2015.

[46] Ramanathan V, Wechsler H. *Phishing detection and impersonated entity discovery using conditional random field and latent Dirichlet allocation.* Computers & Security, 34 pp. 123-139. 2013.

[47] Ronald, D.J., Curtis, C., Aaron, F.J., *Phishing for user security awareness.* Computers & Security, 26(1), pp. 73-80. 2007.

[48]Sheng S., Holbrook M., Arachchilage NAG., Cranor L. Downs J. *Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions.* Proceedings of the 28th international conference on Human factors in computing systems. New York, NY, USA, ACM. 2010.

[49] Thabtah F., Mohammad R., McCluskey L. *A Dynamic Self-Structuring Neural Network Model to Combat Phishing.* Proceedings of the 2016 IEEE World Congress on Computational Intelligence. Vancouver, Canada. 2016.

[50] Thabtah F., Qabajeh I., Chiclana F. *Constrained dynamic rule induction learning.* Expert Systems with Applications, 63, 74-85. 2016.

[51] Thabtah, F., Cowling, P., and Peng, Y. *MCAR: Multi-class classification based on association rule approach.* Proceedings of the 3rd IEEE International Conference on Computer Systems and Applications, 1-7. 2005.

[52] Vishwanath, A., Harrison, B. and Ng, Y.J. S*uspicion, cognition, automaticity model (SCAM) of phishing susceptibility.* Proceedings of the Annual Meeting of 65th International Communication Association Conference, San Juan. 2015.

[53] Vishwanath, A., Herath, T., Chen, R., Wang, J., Rao, H.R. *Why do people get phished? Testing individual differences in phishing vulnerability within an integrated information processing model.* Decision Support Systems, Vol. 51 No. 3, pp. 576-586. 2011.

[54] Witten I. H. and Frank E. *Data Mining: Practical Machine Learning Tools and Techniques.* 2005.

[55] Workman, M. *A test of intervention for security threats from social engineering.* Information Management & Computer Security, Vol. 16 No. 5, pp. 463-483. 2008.

[56] Jain, K, Gupta, B. A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURAPIS Journal on Information Security. 2016.1, pp 1-11. 2016.

[57] Jain, K, Gupta, B. A novel approach to protect against phishing attacks at client side using auto-updated white-list. Security and Communications Networks. pp 1-20. 2017.

[58] Purkait, S. Phishing countermeasures and their effectiveness – literature review. Information Management & Computer Security. Vol. 20 No. 5. Pp 382 – 422.