2017 Second Al-Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)

Phishing Classification Models

An Empirical Study of Induction Factors For Effective Classification

Hiba Zuhair
Department of Systems Engineering,
College of Information Engineering
Al-Nahrain University
Baghdad, Iraq
hiba.zuhair.pcs2013@gmail.com

Ali Selamat
Center of Information and Communication Technologies,
Software Engineering Department, Faculty of Computing
Universiti Teknologi Malaysia (UTM),
Johor, Malaysia
aselamat@utm.my

***Abstract*—recently, researchers have devoted prominent machine learning-based anti-phishing models to survive a supreme cyber-security versus phishing evolution on the cyberspace. Yet, such models remain incompetent to detect new phish in a real-time application. In this concern, this paper advocates an empirical analysis with the recently published works via a chronological validation. Chronological validation achieved by testing the works on three benchmarking data sets to appraise the causality between their detection outcomes and their limitations. Throughout chronological validation, the tested works have fallen short at detecting new phish web pages with an accessible detection accuracy. High to moderate faults and misclassifications are resulted as implications for their limitations and fixed real-time settings. Accordingly, this paper infers that by elevating the tested models in terms of using new and hybrid features, robust subset of features, and actively learned classifiers; an adaptive anti-phishing model with adjustable settings will be resilient against the up-to-date and scalable web flows. With such inferences, this paper highlights what future trends to develop along with depicting a taxonomy of current status and open problems as a guide to the researchers for their future achievements.**

Keywords*: *phish web page; big data; hybrid machine learning algorithm; active learning; adaptive model.

I. INTRODUCTION

Motivating by the more illegitimate gains, phishers targeting users' credentials and industries reputation on the cyberspace. They deploy social engineering technology to impersonate trustworthy websites with spoofed links for users misleading. Victim users catch the bait, submit their own credentials via spoofed links, and then phishers acquire their credentials for theft and illegal profits [1 and 2]. Day after day, the swift increasing and rapid advancement of phishing activities threaten cyber-security and economy [2]. To mitigate them, many efforts have been made by researchers in academia and industry to achieve effective anti-phishing schemes [1-4]. Among them are machine learning-based phishing classification models that adopt client side filtering with feature vectors and feature-base classifiers (FBCs) to inspect phish websites and warn users online [1, 4]. Even though, they assert accurate phishing classification with least faults among their competitors, they vary in their outcomes and performance against newly emerged phishes (new phishes) [5-8]. This is attributed to their unideal classification of on the training data sets for the testing task [5-8]. Thus, this paper examines the topmost machine learning-based anti-phishing models throughout experiments. Towards obtaining a proficient anti-phishing model, this paper appraises the experimental findings critically and infers what facets need to focus in the future.

To point out the aforesaid, the remaining of this paper is organized as follows: Section II introduces the preliminaries of machine learning algorithms and a background of the related works. Section III presents the experimental setup, evaluation conditions, and the conducted experiments with the results. Section IV discusses the resulted findings, reveals what limitations to boost up. In Section V, conclusions and remarks are provided to contribute the future work.

II. BACKGROUND

A. Machine Learning Classifiers

In the literature, many typical machine learning algorithms were applied in the anti-phishing domain, as they depicted in Table I of Appendix briefly. Furthermore, they were incorporated either in a single feature-base classifier (FBC) or in an ensemble feature-base classifier (EFBC) [6-8]. FBC maps the input feature vector to the output classes by attributing the input feature vector $V = (v_1 \ldots v_n)$ and inducts its relevance to either phish or not phish classes with $Y = f(V, \gamma)$. All input feature vectors that extracted from the m-dimensional training dataset ($V_1, V_2, \ldots, V_m$) are induced in the training phase to classify the incoming instance V_{new} in the testing phase into either phish or legitimate label [9-11]. Whereas, EFBC integrates several FBCs into one assembly such that each constituent FBC has its own features set and induction function that might differ from those of other constituents. Moreover, each constituent FBC fetches its own batch of data from the training data set for learning [12, 13]. EFBC makes its final decision for phishing classification by

averaging the predictions of its ensemble. Therefore, EFBC outperform FBC to classify phishing in practice [12, 13]. Although, the existing anti-phishing models utilize FBCs and EFBCs to tackle phish attacks; they still vary in their performance due to their divergence in induction functions [4, 12, 13].

B. Prominent Phishing Classification Models

Among the most salient phishing detection models are those assisted by machine learning classifiers [2-4]. For instance, some researchers at Carnegie Mellon upgraded their former version of an anti-phishing scheme (CANTINA) to a hybrid feature-based scheme CANTINA$^+$. The latter version was developed as an ensemble feature-base classifier including Naïve Bayes (NB), Support Vector Machine (SVM), and Logic Regression (LR) etc. Around 15 textual and structural features were derived from web page URL and web page contents as well as some online features were devoted to accurately classify phish exploits (92% True Positive Rate and 1.4% False Positive Rate) on redirecting web page, login form handler, and web pages hosting in English [14]. However, CANTINA$^+$ encountered a trade-off in leveraging up-to date phish webpages due to the use of limited feature space to English textual features as well as re-learning on defaults settings.

Later, the authors in [15-16] leveraged 17 features to examine login form phish webpages via a developed classification model by using Support Vector Machine (SVM) classifier. Their model achieved a rationale performance with (99.6%) of True Positive Rate and (0.44%) of False Positive Rate. However, it was computationally intensive and time-consuming due to the use of external resources and less adaptive to present training data sets.

On the other hand, the authors in [17] identified phishing on (2,878) Chinese e-business websites via phishing Chinese website detection model. They selected 15 language independent features exclusively to identify Chinese websites. Four machine learning algorithms including Sequential Minimum Optimization (SMO), Logic Regression (LR), Naïve Bayes (NB), and Random Forests (RF) were applied individually in an FBC. Their model performed (95.83%) accuracy rate on Chinese e-business websites solely. Thus, it was not reliable for generic phish websites classification due to its exclusive features and data sets. Oppositely, a phishing classification model was devoted in [18] to catch phishing in e-commerce, login form, and English and French webpages by using 17 ordinary various features and Neural Network (NN) classifier. Even though, achievements yielded up to 94.07% accuracy rates, high misclassification rates were reported. The model scarcely detected novel phish websites due to its inactive learning on imbalanced training data set.

Then, the researchers of [19-20] learnt 212 URL features on an EFBC with multiple machine learning-based classifiers Support Vector Machine (SVM), Random Forest (RF), C4.5, and JRip algorithms. Their EFBC achieved (94.91%) and (1.44%) as detection accuracy and faults resectively. In spite of using big training and testing data sets, the used data set was imbalanced in classes and it included e-Commerce websites exclusively.

Then, this phishing detection model has been examined on present data set that collected during 2015 [21]. As presented in [21] the same model has performed effectively on large and balanced in class distribution data set of (96,018) webpages that aggregated during 2015. However, new experiments have revealed varied outcomes and notable misclassification rates versus new phishes. Furthermore, long execution time and more complex computations have been encountered due to the frequent data query from external resources like GoogleTrends and YahooClues during real-time practice.

Overall, the aforementioned achievements have lacked to attain holistic induction of all phishes (prevalent and new phishes) without deteriorating long-term performed phishing classification model in real-time application. So far, such limitation enabled phishers to intrude existing anti-phishing models with more advanced phish web pages for more damages to both users and enterprises.

III. Empirical Study

In this section, a chronological test of the aforesaid phishing classification models is presented.

A. Experimental Setup

As presented in Table I, three data sets retrieved from three recently published works are used chronologically. As presented in Table I, they vary in terms of size, class distribution, phish samples, legitimate samples, data sources, web page functionalities and hosting languages, aggregation time, and data sources. Such variety attains the objective on which this analysis is conducted.

TABLE I. DATA SETS WITH THEIR MERITS [1, 17, 19-21]

Merits	Data Set 1	Data Set 2	Data Set 3
Size	52	2878	96,018
Phishes	70%	49%	50%
Legitimates	30%	51%	50%
Data Source	PhishTank /Alexa	Chinese E-Business	PhishTank /DMOZ
Collection Time	25-31/7/2010	2014	2012-2015
Web page Functionality	Login Forms e-Business	e-Business	e-Business/ Homepage Login Forms
Hosting Language	English/ French/ German	Chinese	English/French/ German/Italian/ Spanish etc.

Accordingly, typical evaluation criteria such as True Positive Rate (TPR), False Positive Rate (FPR), and False Negative Rate (FNR); are used for performance evaluation as they depicted in Table II.

TABLE II. EVALUATION CRITERIA [1, 5, 7, 8]

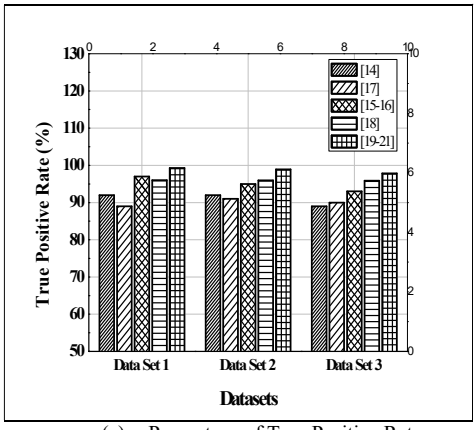
Criterion	Description
TPR	The rate of correctly classified phishing samples: $\frac{N_{P\to P}}{(N_{P\to P}+N_{P\to L})}$ (1)
FPR	The rate of wrongly classified legitimate samples as phishing: $\frac{N_{L\to P}}{(N_{L\to L}+N_{L\to P})}$ (2)
FNR	The rate of wrongly labeled phishing samples as legitimates: $\frac{N_{P\to L}}{(N_{P\to P}+N_{P\to L})}$ (3)
Here $N_{P\to P}$, $N_{L\to P}$, $N_{P\to L}$, $N_{L\to L}$ denote the number of correctly labeled phishing instances, the number of wrongly labeled legitimate instances, the number of phishing instances that are incorrectly recognized as legitimate, and the number of legitimate instances that are identified correctly as legitimate respectively	

B. Results and Discussion

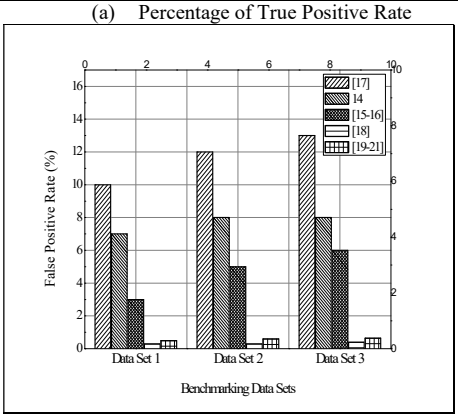
Empirically, the detection performance of the tested models have varied across the three data sets as plotted in Fig 1. That, in turn, demonstrates how the tested models could learn on small and/or large, imbalanced and/or balanced data sets via training and testing tasks. Also, the empirical analysis addresses the issue of filtering typical and new phish web pages in the testing task. More precisely, *Data Set 1* [3] was imbalanced in phish/not phish class distribution despite of its divergence in web page functionalities as presented in Table I. *Data Set 2* was bigger in size than *Data Set 1* and it involved Chinese e-Business websites solely [17]. It was utilized to classify Chinese login forms, redirecting web pages, and e-Business homepages. Whereas, *Data Set 3* was the biggest in size and the most balanced in class distribution among its competitors. In addition, it covered up different web page functionalities and hosting languages.

As presented in Fig. 1(a), learning the past and present data sets by the tested models is still questionable to accurately detect a phish (w) emerged at time ($T+\Omega$) on a data set (W) fetched at time (T). Yet, almost tested classification models need a period of time (Ω) to learn (W) and to build phishing classification settings for the fetched web page on the incoming data set. Escalating accuracies of classification in Figure 1(a) imply that the emerged phish (w) might be short-living and it taken down by its phisher during the period of time (Ω). Furthermore, the emerging time (Ω) could be a long time horizon that misled the detection of the tested models against new phishes. As such, the results plotted in Figure 1(a), point out that (Ω) was a long time spent to re-learn the tested classification models on the incoming data flow. This is due to the divergent aggregating time of all the examined data sets as depicted in Table I. That, in turn, makes the sense to assess the FBCs and EFBCs of the tested models in the term of active learning crosswise scalable web flows.

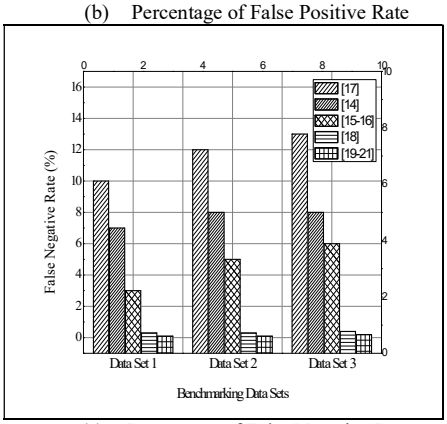
Consequently, Fig. 1(b) shows a variation of FPRs among the tested models from high to low and mild rates across all benchmarking datasets. This is attributed to the Goodness, Stability, Similarity and Phishness Indication Ratio (PIR) of the chosen features [25-28]. Furthermore, almost tested models were devoted without using a robust subset of features that resulted in a features weighting mechanism or feature subset selecting strategy.



(a) Percentage of True Positive Rate



(b) Percentage of False Positive Rate



(c) Percentage of False Negative Rate

Fig. 1. Performance outcomes of chronological validation

Excepting, the tested models that adopted in [17 and 18] applied selection methods but they were limited to leverage maximum relevance of features to phishing class and minimum redundancy of features in the learning data sets as selection criteria [26 and 27]. Indeed, a feature subset of both maximum relevant and minimum redundant features will

afford potential effects that heighten performance and diminish errors despite of what features are included in its compactness [28]. In this concern, the tested model that adopted by [18] achieves high levels of TPR with the lowest level of FPR across all data sets as presented in Fig. 1 (a) and Fig. 1(b). This is because of selecting the most informative features for detection by using three features ranking strategies such as Information Gain (IG), Correlation-Based Feature Selection (CFS) and Chi-Squared (χ^2) which depend on features interdependencies as a selection criterion.

In Fig. 1(c) the tested models perform very high to moderate and then mild FNRs across all data sets. This is attributed to their variations in the amount and the type of features that they used for detection. Almost tested models applied few and conventional features rather than many and new phishing features; i.e. the up-to-minute features that exploited by phishers in their fake web pages. Indeed, fewer features used to train the data sets lead up to fewer phish patterns to be characterized. For example, the tested model developed by [17] detected Chinese e-business websites with the aid of 12 features that explored exclusively for such type of phishes. Thus, it reports the highest FNR across the scaling data sets via the chronological test as plotted in Fig.1(c). That implies it misclassifies the other phish patterns.

On the other hand, the use of conventional features by the tested models led to partial characterization of phish web pages. As can be seen in Fig. 1(c), all tested models achieved misclassification costs against some phishes cross all the testing data sets. This is due to their inabilities to inspect novel phish web pages as the data set growing in the size and advancing in the types of phish web pages. That, in turn, justifies why phishers still mislead phishing induction criteria of the applied classification models, and bypass the existing anti-phishing schemes. Then, phishers cause potential damages to the computer systems while they gain more profits day by day. Overall, as the training data grows in size, evolving in phish/legitimate class distribution, and aging in aggregation; the tested models fall short to classify phishing effectively, see Table II in Appendix.

IV. INFERENCES

Based on empirical study, this paper infers that all the tested models are still insufficient to induce an effective phishing classification in a chronological test. Thus, this section restates what factors are needed to promote an effective classification as taxonomy in Fig. 2 highlights. Factors could be as follows:

- New features along with the conventional ones should be utilized in the tested anti-phishing models. Byproduct, deploying the new features that crafted by phishers in phishing induction will reveal a misclassification cost-effective anti-phishing model. In addition, the variety and big amount of features will promote a holistic characterization of phish exploitations, i.e. phishing induction on all kinds of phish web pages [22-24].
- Selecting the robust subset of informative features according to their interdependencies will highly constrain the power to feature values' heterogeneity, features' relevance and redundancy to the trained data sets as well as inducing phishness on the testing web page flows [25-28].
- Training Big Data. Learning the anti-phishing models on big data sets will mitigate the problems of class imbalance that may cause misleading induction on the incoming web flows in the real-time practice [29]. If the classification model applies fixed induction settings that might classify phishes mistakenly as legitimates. Further, it uses classic features that are in common between phishes and other attacks like ham, and spam [26, 28]. Then, a sub-optimal real-time phishing detection will be achieved. Thus, fixed and inadequate induction settings will lead to inaccurate classification of phish web pages among other types of attacks and then non-zero learning faults will be the byproduct.

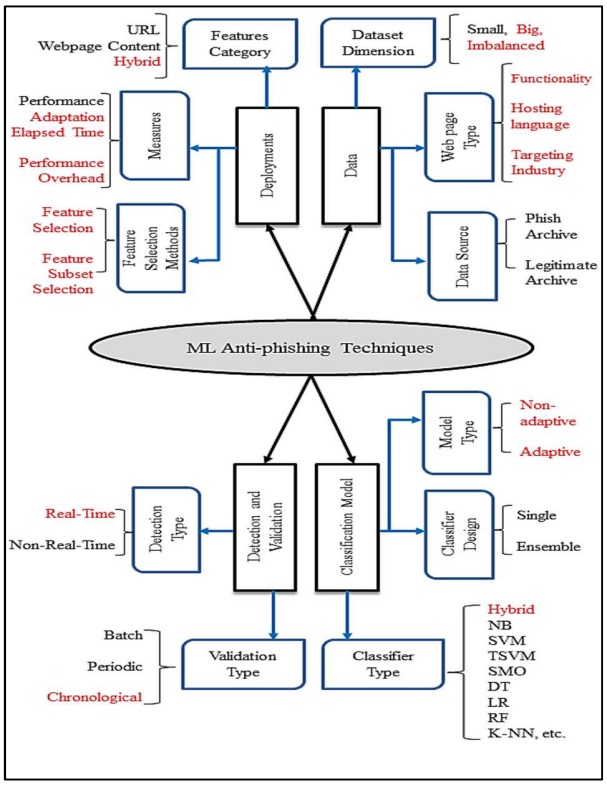


Fig. 2. Taxonomy of phishing classification models

- Hybrid machine learning algorithm. It refers to the integration of different decision making criteria. As such each criterion exploits a different set of induction parameters or settings that work in a different manner. Hybrid algorithm is significant to better classify phishes on time with zero-false classification rates [30]. This is attributed to the hybridity of machine learning algorithms

delegating their merits without compromising their demerits that might incur impressions to the induction of phishing.

- Actively learned classifiers can instate the expected future errors by selecting the batch of instances that expected to decrease the future error [30]. As such, minimal detection faults and misclassifications will be obtained due to the frequent alteration to the induction settings. However, almost tested phishing classification models applied inactive classifiers that train batches of data chronologically. Therefore, active learned FBCs or EFBCs are required to re-train data sets that reveal the best and up-to-date induction settings.
- Adaptive model is that reconfigures its induction settings dynamically [11, 30]. Adaptable induction confirmed by inspecting new and unknown patterns of phishing during the real-time experience [30]. Throughout the empirical study, the tested models achieved mild to moderate misclassifications because the induction biases were still default and unalterable during the testing phase. They leverage the same features and the same functions of their classifiers frequently across the three data sets. Byproduct, they are not adaptive to new phishes.

V. CONCLUSIONS

This paper studies the prominent machine learning-based phishing classification models via a chronological test across three different data sets. Findings attributed the variations in performance of the tested models to their overlooking of some induction factors such as novelty and amount of features, robustness of the selective features subset, big data leveraging, active learning of classifiers, and hybridity of applied machine learning algorithm as well as adaptive modelling. Altogether deficiency could assert misclassification-costly and ineffective phishing classification models in real-time application. From the insights of the empirical stud, this paper restates how such factors could be boosted up for future developments. Then, substantial outcomes of phish web page detection will be attained versus the vast data of the Web. It is hoped that the recommended directions of research will serve as a navigating taxonomy to the reseachers in the future.

REFERENCES

[1] M. Khonji, Y. Iraqi & A.Jones, "Phishing detection: a literature survey," Comm. Surveys & Tutorials, Vol. 15, No. 4, pp. 2091–2121, 2013.

[2] H. Z., Zeydan, A. Selamat, M. Salleh, "Survey of anti-phishing tools with detection capabilities," *In the proceedings of 14 Int. Symposium on Biometrics and Security Technologies (ISBAST'2014)*, Kuala Lumpur, Malaysia, 2014.

[3] H. Shahriar, "Trustworthiness testing of phishing websites: a behavior model-based approach," Future Generation Comput. Syst., Vol. 8, No. 28, pp. 1258–1271, 2012.

[4] H. Z., Zeydan, A. Selamat, M. Salleh, "Current state of anti-phishing approaches and revealing competencies," *Journal of Theoretical and Applied Information Technology*, Vol. 70, No. 3, pp. 507-515, 2014.

[5] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," *Proceedings of 17th Annual Internet Society on Networks and Distributed System Security Symposium* (NDSS2010). San Diego, California, USA, 2010.

[6] B. Wardman, J. Britt, and G. Warner, "New tackle to catch a phisher," *International Journal of Electronic Security and Digital Forensics*. Vol. 6, No. 1, pp. 62-80, 2014.

[7] A. Abbasi, and H. Chen, "A comparison of fraud cues and classification methods for fake escrow website detection," *Information Technology and Management*, Vol. 10, No. 2-3, pp. 83-101, 2009.

[8] R. Islam, and J. Abawajy, "A multi-tier phishing detection and filtering approach," *Journal of Network and Computer Applications*, Vol. 36, No. 1, pp. 324-335, 2013.

[9] G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence based techniques for intrusion detection: a review," *Artificial Intelligence Review*, Vol. 34, No. 4, pp. 369-387, 2010.

[10] C. M. Bishop, "Pattern recognition and machine learning," Vol. 4, Springer Verlag, New York.

[11] T. T. Nguyen, and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *Communications Surveys & Tutorials*, *IEEE,* Vol. 10, No. 4, pp. 56-76, 2008.

[12] "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions On Systems, Man And Cybernetics*, *Part C, Applications and reviews,* Vol. 42, No. 4, pp. 463-484, 2012.

[13] A. Shabtai, R. Moskovitch, Y., Elovici, and C. Glezer, "Detection of malicious code by applying machine learning classifiers on static features: a state-of-the-art survey." *Information Security Technical Report*. Vol. 14, No. 1, pp.16-29, 2009.

[14] G. Xiang, "Towards a phish free world: a cascaded learning framework for phishing detection," Doctoral Dissertation, *Carnegie Mellon University*, Pittsburgh, PA 15213, 2013.

[15] R. Gowtham, and I. Krishnamurthi, "A comprehensive and efficacious architecture for detecting phishing webpages," *Computers & Security*, Vol. 40, pp. 23-37, 2014.

[16] R. Gowtham, and I. Krishnamurthi, "PhishTackle-a web services architecture for anti-phishing," *Cluster Computing*, Vol. 17, No. 3, pp. 1051-1068, 2014.

[17] D. Zhang, Z. Yan, H. Jiang, H., and T. Kim, "A domain-feature enhanced classification model for the detection of Chinese phishing e-Business websites," *Information & Management*, Vol. 51, No. 7, pp. 845-853, 2014.

[18] R. M. Mohammad, F. Thabtah, F., and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," *Neural Computing and Applications*, Vol. 25, No. 2, pp. 443-458, 2014.

[19] S. Marchal, J. François, R. State, and T. Engel, "PhishScore: hacking phishers' minds," *Proceedings of 10th International Conference on Network and Service Management* (CNSM2014). Rio de Janeiro: IEEE, pp. 46-54, 2014.

[20] S. Marchal, S., J. François, R. State, and T. Engel, "PhishStorm: detecting phishing with streaming analytics," *IEEE Transactions on Network and Service Management*, Vol. 11, No. 4, pp. 458-471, 2014.

[21] S. Marchal, "DNS and semantic analysis for phishing detection," Doctoral Dissertation. *University of Luxembourg*, 22 June 2015.

[22] E. Uzun, H. V. Agun, and T. Yerlikaya, "A hybrid approach for extracting informative content from web pages," *Int. Journal of Information Processing and Management: an International Journal.* Vol. 49, No. 4, pp. 928-944, 2013.

[23] H., Zuhair, M. Salleh, and A. Selamat, "New Hybrid Features for Phish Website Prediction," *International Journal of Advances in Soft Computing & Its Applications*, Vol. 8, No. 1, 2016.

[24] H., Zuhair, M. Salleh, and A. Selamat,"Hybrid features-based prediction for novel phish website," Jurnal Teknologi, Vol. 78, No. 12-3, 2016.

[25] H. Zuhair, A., Selamat, M. Salleh, "Feature Selection for phishing detection: a review of research," *Int. Journal of Intelligent Systems Technologies and Applications.* Vol. 15, No. 2, pp.147-162, 2016.

[26] F. Toolan, and J. Carthy, "Feature selection for Spam and Phishing detection," *Proceedings of eCrime Researchers Summit* (eCrime), 2010.

[27] H. Zuhair, A. Selamat, M. Salleh, "The effect of feature selection on phish website detection: an empirical study on robust feature subset

selection for effective classification," *Int. Journal of Advanced Computer Science and Applications*. Vol. 6, No. 10, pp.221-232, 2016.

[28] H. Zuhair, A. Selamat, M. Salleh, "Selection of robust feature subsets for phish webpage prediction using maximum relevance and minimum redundancy criterion," *Journal of Theoretical and Applied Information Technology*, Vol. 81, No. 2, pp.188-205, 2015.

[29] O. Kwon, and J. M. Sim, "Effects of data set features on the performances of classification algorithms," *Expert Systems with Applications*, Vol. 40, No. 5, pp.1847-1857, 2013.

[30] C. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: a review," *Expert Systems with Applications*, Vol. 36, No. 10, pp. 11994-12000, 2012.

Appendix

TABLE I. EXAMPLES OF MACHINE LEARNING ALGORITHMS USED IN ANTI-PHISHING DOMAIN[5, 7-9, 14-18, 27]

Algorithm	**Description**
Decision Tree (DT)	In a rooted tree, instances of unknown class are ordered according to their feature values by labeling the nodes with features and the edges with feature values. An instance is classified by starting up at the root node, approaching to the next nodes, and ending up at a leaf that is labeled with the final decision. Examples: C4.5 and JRip.
Naïve Bayes (NB)	A probabilistic judgment done conditionally with independent attributes of all instances belonging to a given class: $P(C\|X) = P(C\|x_1, \dots, x_n) = \frac{P(C)P(x_1,\dots,x_n\|C)}{P(x_1,\dots,x_n)}$ (1) Where X is an instance with a vector of *n* features$(x_1, \dots, x_n)$, C is the class label that the classifier seeks for.
Support Vector Machine (SVM)	A separating hyper-plane maximizes the margins between closest points of two classes to estimate the induction function: $min\frac{1}{2}w^T w + C\sum_i \xi_i$ (2) That subjects to: $y_i\big((w^T \cdot x_i) + b\big) \geq 1 - \xi_i, \xi \geq 0, i = 1, 2, \dots, m$ (3) $max\sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j K\big(x_i, x_j\big)$ (4) Which is Subject to :$0 \leq \alpha_i \leq C, i = 1, 2, \dots, m \; and \; \sum_{i=1}\alpha_i y_i = 0$ (5) Where: x_i is m-dimensional data vector $x_i \in R^m$ with samples belong to either one of two classes labeled as $y \in \{-1, +1\}$that it is separated by a hyper-plane of $(w \cdot x) + b = 0$, $\alpha_i$denotes the lagrange multipliers for each vector in the training dataset.
Logistic Regression (LR)	Use probabilistic induction that evaluates relationship between a categorical dependent variable and a continuous independent variable (s): $\pi(x) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta_1 x)}+1} = \frac{1}{e^{-(\beta_0+\beta_1 x)}+1}$, (8) $g(x) = ln\frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x,$ (9) $\frac{\pi(x)}{1-\pi(x)} = e^{(\beta_0+\beta_1 x)}$ (10) Where: $g(x)$ is the logistic function of a given predictor *X*, ln and, $\pi(x)$ denote natural logarithm and case probability, $\beta_0$and β_1 denote criterion of *X* , and $\beta_1 x$ is the regression coefficient.
Random Forests (RF)	Forest constructed for randomly selected set of instances on training data set. Given *n, p* and k where *n* is the number of training observations, *p* is the number of features in the training data set and *k* is the number of selected features such that $k \ll p$. A boot strap sample is selected from n and used to estimate the error of the tree in the testing task. At a certain node in the tree, *k* of features are selected randomly and used as decision to calculate the best split in the training data set.
Sequential Minimal Optimization (SMO)	It solves the optimization problem caused during classification iteratively and analytically:- $\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j K\big(x_i, x_j\big)\alpha_i \alpha_j$ (11) Where$0 \leq \alpha_i \leq C$, for $i = 1, 2, \dots, n$ and$\sum_{i=1}^{n} y_i \alpha_i = 0$. C is the classifier's hyper-sphere, $K\big(x_i, x_j\big)$ refers to the kernel function provided by user, and $\alpha_i$is the Lagrange multiplier.
Neural Network (NN)	$f(x) = g\big[\sum_i v_i g\left(\sum_j w_{ij} x_j + b_i + b_0\right)\big]$ (12) Where x, v_i, g, w_{ij} and $b_{i,o}$ are the input vector, the weight of output neuron, the activation function, the weight of hidden neuron and the bias respectively.

TABLE II. INDUCTION ISSUES OF NOTABLE MACHINE LEARNING-BASED PHISHING CLASSIFCATION MODELS

Related Work / **Issues**	**[14]**	**[15-16]**	**[17]**	**[18]**	**[19-21]**
Machine Learning Algorithm	SVM, LR, DT	SVM	SMO, LR, RF, NB	NN	SVM, RF, C4.5, JRip,
Amount of Features	15	17	15	17	**212**
New Features	**3**	**7**	Not	Not	**12**
Features Selection Mechanism	Not	Not	**χ^2**	**CFS, IG, χ^2**	Not
Train Big Data	Not	**Yes**	Not	Not	**Yes**
Actively Learned Classifier(s)	Not	Not	Not	Not	**Active**
Adaptive Modelling	Not	Not	Not	Not	Not
Hybrid Machine Learning- Based Algorithm	Not	Not	Not	Not	Not