Machine Learning Based Phishing Attacks Detection Using Multiple Datasets

https://doi.org/10.3991/ijim.v17i05.37575

Ashraf H. Aljammal[1](✉), Salah taamneh [1], Ahmad Qawasmeh[1], Hani Bani Salameh[2]
[1] Department of Computer Science and Applications, The Hashemite University, Zarqa, Jordan
[2] Department of Software Engineering, The Hashemite University, Zarqa, Jordan
ashrafj@hu.edu.jo

Abstract—Nowadays, individuals and organizations are increasingly targeted by phishing attacks, so an accurate phishing detection system is required. Therefore, many phishing detection techniques have been proposed as well as phishing datasets have been collected. In this paper, three datasets have been used to train and test machine learning classifiers. The datasets have been archived by Phish-Tank and UCI Machine Learning Repository. Furthermore, Information Gain algorithm have been used for features reduction and selection purpose. In addition, six machine learning classifiers have been evaluated, namely NaiveBayes, ANN, DecisionStump, KNN, J48 and RandomForest. However, the classifiers have been trained and tested over the three datasets in two stages. The first stage is using all features included in each dataset while the second stage using selected features by IG algorithm. At the first stage RandomForest classifier has shown the best performance over Dataset-1 and Dataset-2, while J48 has shown the best performance over Dataset-3. On the other hand, after features selection, the RandomForest classifier was the superior among the other five classifiers over Dataset-1 and Dataset-2 with accuracy of 98% and 93.66% respectively. While ANN classifier has shown the best performance with accuracy of 88.92% over Dataset-3. Because of the few number of instances as well as features in Dataset-3 comparing to the other two dataset; the performance of the classifiers has been affected.

Keywords—phishing attack, phishing attack detection, cybersecurity, machine learning, web security, network security

1 Introduction

Phishing is considered a cybersecurity threat which is used to commit fraudulent actions; such as stealing users sensitive information which includes user accounts credentials and banking cards information [1-3]. The phishing attack occurs when attackers pretend to be a trusted entity and attracting the victim to open the socially engineered sent message (Email, IM message or text message)[4, 5]. Sometimes, these messages ask the user to input a critical information or even entreating the user for financial gain. Furthermore, the content of the message could be a URL of a rogue version of legitimate webpage created by the attacker and hosted on their own servers

luring user to click the URL. However, it will be hard for the internet user to differentiate between the legitimate and phishing (mimic) webpages. When exploring these webpages it eventually leads the user to reveal her/his own information to the attacker. However, attackers always attract user to explore the phishing website. Therefore, the success of phishing attacks rely on the weaknesses of the user[6]. Individual users as well as organizations are vulnerable to many types of attacks including phishing attacks[7, 8]. According to Anti-Phishing Working Group (APWG)[9] report, the phishing attacks ratio is doubled since early 2020. The report statistics show that the highest number of phishing attacks was in July 2021 with 260,642 attacks. In addition, the most effected sectors in phishing attacks were software-as-a-service and webmail with 29.1% of total attacks. Furthermore, the combined attacks against financial institutions and payment providers with 34.9% of all attacks. Figure 1 shows the most targeted industries in the 3rd quarter of 2021.

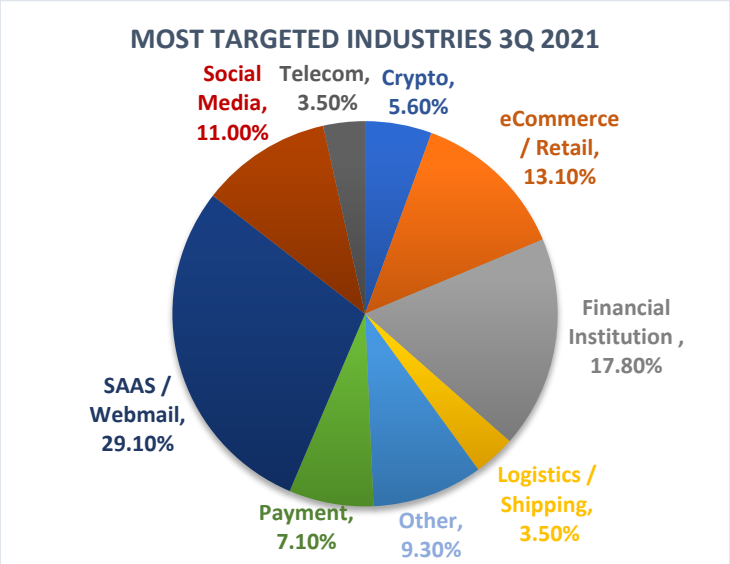

Fig. 1. Most targeted industries in the 3rd quarter-2021

Thirty-seven percent of phishing websites used four generic top-level domains (gTLDs) namely .COM, .ORG, .ASIA and .BIZ. In addition, the .XYZ and .ICU are an examples of new generic top-level domains (nTLDs) representing 9 percent of the total domains. However, s .UK for the United Kingdom and .BR for Brazil country code domains represent 53 percent of the phishing domains in the 3rd quarter of 2021. According to the aforementioned statistics, the need for phishing detection systems is an urgent matter.

The structure of the paper is organized as follows. Section 2 illustrates the literature review. Section 3 presents an overview of the proposed approach. Section 4 discusses the implementation and results analysis. Section 5 concludes our work.

2 Literature review

Recently, many phishing detection techniques have been proposed in the literature. In this section, we will discuss some of the proposed techniques to detect phishing attacks. In [10], the authors proposed an Anti-Phishing-Simulator based on URL Control features which is able to detect phishing and spam emails. The detection process is based on examining the email contents and classifying the spam words and then adding them to a database using Bayesian algorithm. The authors of [11] have proposed a phishing attack system based on natural language processing techniques. The phishing detection process semantically analyzes the text contents of emails to detect some statements indicate whether the email is phishing or legitimate. However, this technique is effective only in detecting a pure text based phishing emails. The SAFE-PC system has been proposed by the authors of [12]. The system is able to detect new forms of phishing attacks. The detection process is based on extracting features from the body and header of the phishing email. The RUSBoost classifier has been used to classify the phishing and legitimate emails. In [13], the authors have proposed a phishing webpages detection framework using deep learning approach. Furthermore, the multilayer perceptron classifier has been used to classify the phishing, suspicious and legitimate webpages. The number of features included in the used dataset is 9 and all of them were used in the classification process. The authors of [14] have proposed a real-time two-level authentication approach to detect phishing attacks based on internet search results and the extracted hyperlinks features. In authentication-level1 uses an independent textual language query to authenticate the target webpage. While in the second level of authentication the hyperlinks are investigated in order to detect phishing and legitimate websites. In [15], the authors have proposed a phishing emails detection system based on recurrent neural network (RNN). The proposed system relies only on the textual contents of the emails to detect the phishing and phishing emails. A phishing attacks detection system has been proposed by the authors of [16] to detect phishing attacks on e-banking and commercial websites. The detection process is based on the hyperlink and visual similarity relations where the keywords, css layout and hyperlinks are considered in the analysis process. Where the css is used to compare the layout similarity among the webpages. Furthermore, the authors used login form and white list filtering to increase the detection accuracy of the proposed system. In [17], authors have proposed the PhishLimiter approach to detect and mitigate phishing attacks. The proposed approach is based on Deep Packet Inspection (DPI) and Software-Defined Networking (SDN) to detect phishing emails and phishing websites communications. In addition, it has a signatures based classification and real-time inspection stages. The ANN has been used to classify the phishing attacks signatures and real-time detection of the phishing activities. The authors of [18] have proposed an approach to detect phishing attacks based on automated white-list. A comparison between the visual link and the actual link is conducted to build the webpages white-list. Eventually, the final detection decision is based on the extracted features from the hyperlink. In [19] the authors have proposed a detection system based on Bayesian classifier to detect to distinguish between the fake emails and real emails. The system

extracts the textual information from the email to be used later in the classification process.

3 Methodology

Figure 2 illustrates the proposed approach and steps in which each of the datasets (Dataset-1, Dataset-2 and Dataset-3) will go through these steps to test and train the used machine learning classifiers (NaiveBayes, ANN, DecisionStump, KNN, J48 and RandomForest).

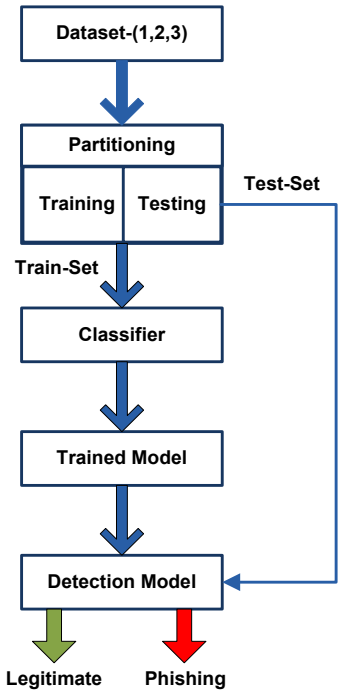

Fig. 2. The proposed approach

3.1 Datasets

Based on the literature, three different phishing attacks datasets have been commonly used to train and test ML algorithms[20].In addition, the three datasets features have been evaluated to select the most significant features. Dataset-1[21] is a dataset of phishing websites was gathered between January and May of 2015 and May and June of 2017. It consists of 10000 instances and divided into two groups. The first group contains 5000 instances reflecting phishing websites while the second group contains 5000 instances reflecting legitimate websites. The dataset set has 48 different

features that could be used for classification purpose. Dataset-1 has a binary labels 0 and 1, where 0 indicates legitimate website and 1 indicates phishing website.

The other two datasets were obtained from University of California, Irvine's Machine Learning Repository. Dataset-2 consists of 11055 different instances where 4898 of them are phishing websites and 6157 legitimate websites[22]. This dataset is labeled as 0, 1 and -1 represent suspicious, legitimate and phishing websites respectively. Moreover, it has 30 different features represented in binary variables. The dataset was collected through Google search engine, MillerSmiles archive, and PhishTank archive. Dataset-3 contains 9 different features and 1353 instances, 702 of them belong to phishing websites, 548 belong to legitimate websites and 103 belong to suspicious websites[23]. Similar to Dataset-2, this dataset includes three labels: 0 for suspicious websites, 1 for legitimate websites and -1 for phishing websites.

The Information Gain (IG) algorithm has been used to evaluate the features importance and reduce the number of features in the three datasets. Formula 1 illustrates the information gain function and Table 1 shows the selected features of the three datasets using IG algorithm. The IG based selected features covering 87% of the total weight of each dataset features.

$$\text{InfoGain}(\text{Class}, \text{Attribute}) = \text{H}(\text{Class}) - \text{H}(\text{Class} \mid \text{Attribute}) \quad (1)$$

Where:
H: represents the Entropy
Class: whether legitimate, suspicious or phishing websites
Attribute: denotes the features

Table 1. Selected Features Using IG algorithm

Feature Rank	Features from Dataset-1	Features from Dataset-2	Features from Dataset-3
1	PctExtHyperlinks	SSLfinal_State	SFH
2	PctExtResourceUrls	URL_of_Anchor	popUpWidnow
3	PctExtNullSelfRedirectHyperlinksRT	Prefix_Suffix	SSLfinal_State
4	PctNullSelfRedirectHyperlinks	web_traffic	Request_URL
5	NumNumericChars	having_Sub_Domain	URL_of_Anchor
6	FrequentDomainNameMismatch	Request_URL	
7	ExtMetaScriptLinkRT	Domain_registeration_length	
8	NumDash		
9	SubmitInfoToEmail		
10	NumDots		
11	InsecureForms		
12	PathLevel		
13	PathLength		
14	NumSensitiveWords		
15	QueryLength		

Figure 3 illustrates the features ranks in Dataset-1 using IG algorithm. The ranking results show the features with high ranks which will be used in the classification process. In addition, it shows low ranks features which will be eliminated (will not be used in the classification process). For instance **PctExtHyperlinks** and **PctExtNullSelfRedirectHyperlinksRT** have shown the highest ranks indicating the highest importance features among the other features. Whereas, **FakeLinkInStatusBar** and **ImagesOnlyInForm** have shown the lowest ranks indicating the lowest importance features among the other dataset features.

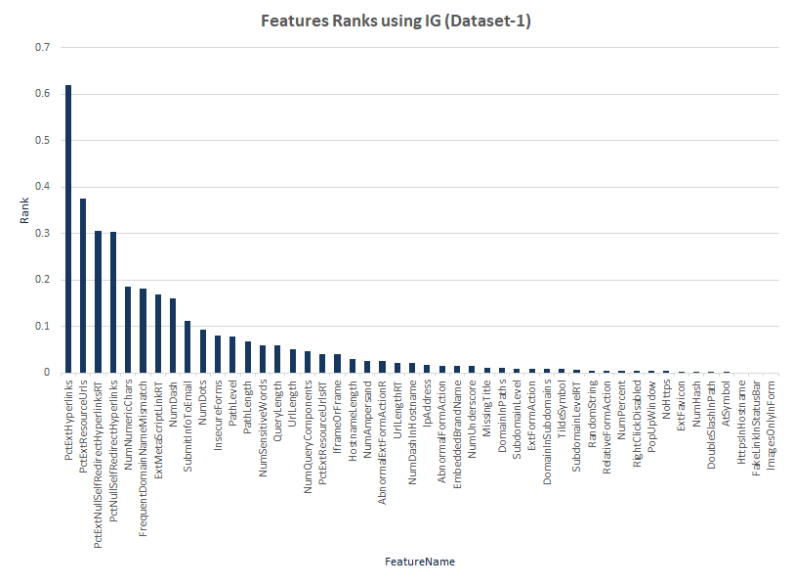

Fig. 3. Features Ranks using IG (Dataset-1)

The features ranks of Dataset-1 using IG algorithm are shown in Figure 4. The features SSLfinal_State and URL_of_Anchor have the highest ranks among the other features in the dataset, therefore, they will be the top two features to be used in the classification process. On the other hand, Favicon and popUpWidnow features have been neglected regarding to their lowest ranks and they will not be used in the classification process.

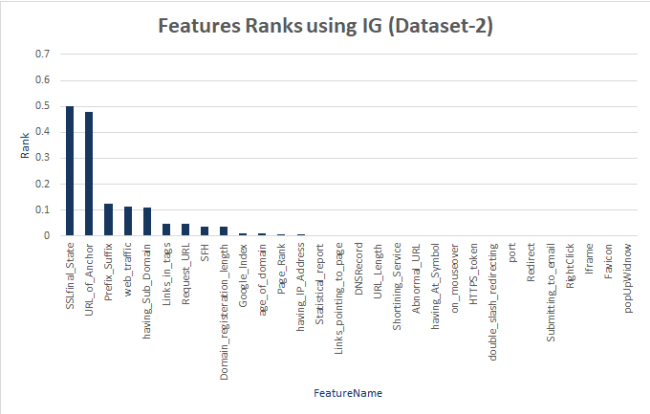

Fig. 4. Features Ranks using IG (Dataset-2)

Since Dataset-3 has only 9 features, all of them have been used in classification process. In addition, omitting any of the features will affect the classification process performance. Figure 5 illustrates the ranks of Dataset-3 features using IG algorithm.

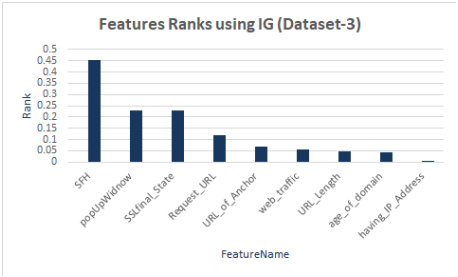

Fig. 5. Features Ranks using IG (Dataset-3)

3.2 Experimental design

WEKA tool[24] has been used to conduct the experiments over the three phishing datasets. Six different ML algorithms: NaiveBayes, ANN, DecisionStump, KNN, J48 and RandomForest have been used along with IG algorithm. The six ML algorithms have been tested over the three datasets where the first run was before applying the IG algorithm, each dataset with its full features. The second run was after applying the IG algorithm on each dataset to evaluate and select the most significant features of these datasets to classify the phishing websites. Each dataset has been split individually into 80% and 20% for training and testing respectively of the used ML algorithms.

4 Results analysis and discussion

Two experiments have been conducted using different six ML algorithms over the three datasets. The first experiment is based on using all features in the datasets, while the second is after features selection and reduction.

4.1 ML algorithms performance using original dataset

The results of the classifiers over dataset-1 are shown in Table 2. RandomForest classifier has shown the best accuracy with 98.45% and J48 in the second place with 97.35% of accuracy. While the DecisionStump classifier has shown the lowest accuracy with 77.75%.

Table 2. Classifiers Performance using all features- Dataset 1

Classifier	Accuracy	TPR	Precision	Recall	F-Measure
NaiveBayes	85.15%	76.4%	92.6%	76.4%	83.7%
ANN	96.59%	98%	96%	98%	97%
DecisionStump	77.75%	57.8%	96%	57.8%	72.2%
KNN	95.85%	96.4%	95.3%	96.4%	95.9%
J48	97.35%	97.5%	97.2%	97.5%	97.3%
RandomForest	98.45%	98.7%	98.2%	98.7%	98.5%

Table 3 illustrates the results of the classifiers over dataset-2. Once again, RandomForest classifier with 97.15% of accuracy has shown the best result in term of accuracy. While ANN and KNN classifiers have shown the same accuracy with 96.69% but with higher precision 97.8% of ANN. On the other hand, DecisionStump classifier has shown the lowest accuracy with 89% among other classifiers.

Table 3. Classifiers Performance using all features- Dataset 2

Classifier	Accuracy	TPR	Precision	Recall	F-Measure
NaiveBayes	92.85%	89.9%	94%	89.9%	91.9%
ANN	96.69%	94.8%	97.8%	94.8%	96.3%
DecisionStump	89%	85.9%	89.4%	85.9%	87.6%
KNN	96.69%	95.2%	97.4%	95.2%	96.3%
J48	96.11%	94.1%	97.2%	94.1%	95.6%
RandomForest	97.15%	95.5%	98.2%	95.5%	96.8%

Table 4 shows the classifiers results over dataset-3. J48 classifier has shown the superior performance in term of accuracy with 89.66%. While ANN and RandomForest classifiers have shown 88.92% of accuracy for both of them. But in term of precision, RandomForest classifier was better than ANN classifier with 91.2%.

Table 4. Classifiers Performance using all features- Dataset 3

Classifier	Accuracy	TPR	Precision	Recall	F-Measure
NaiveBayes	83.39%	87.8%	87.8%	87.8%	87.8%
ANN	88.92%	89.8%	91.7%	89.8%	90.7%
DecisionStump	82.65%	90.5%	82.1%	90.5%	86.1%
KNN	87.08%	86.4%	92.7%	86.4%	89.4%
J48	89.66%	90.5%	92.4%	90.5%	91.4%
RandomForest	88.92%	91.2%	91.2%	91.2%	91.2%

4.2 ML algorithms performance After Features selection and Reduction

The results of the classifiers over dataset-1 are shown in Table 5. It is obvious that RandomForest classifier has the best performance with accuracy of 98% among other classifiers. While DecisionStump classifier has the lowest accuracy with 77.75%. ANN, KNN and J48 classifiers have shown a good performance in term of accuracy with 95.75%, 95.9% and 96.95% respectively.

Table 5. Classifiers Performance using IG based selected features- Dataset 1

Classifier	Accuracy	TPR	Precision	Recall	F-Measure
NaiveBayes	83.65%	72.9%	92.7%	72.9%	81.7%
ANN	95.75%	93.9%	97.5%	93.9%	95.7%
DecisionStump	77.75%	57.8%	96%	57.8%	72.2%
KNN	95.9%	97.1%	94.8%	97.1%	95.9%
J48	96.95%	97.4%	96.5%	97.4%	97%
RandomForest	98%	98.1%	97.9%	98.1%	98%

Table 6 shows the results of the classifiers over dataset-2. Almost all classifiers except DecisionStump classifier have shown a good performance with close results. Whilst, the supremacy was for RandomForest with 93.66% in term of accuracy. On the other hand, DecisionStump classifier has shown the lowest performance among the other classifiers with an accuracy of 98%.

Table 6. Classifiers Performance using IG based selected features- Dataset 2

Classifier	Accuracy	TPR	Precision	Recall	F-Measure
NaiveBayes	91.81%	88.5%	93.1%	88.5%	90.7%
ANN	93.08%	91.1%	93.4%	91.1%	92.3%
DecisionStump	89%	85.9%	89.4%	85.9%	87.6%
KNN	93.35%	91.3%	93.8%	91.3%	92.6%
J48	93.35%	91.2%	93.9%	91.2%	92.5%
RandomForest	93.66%	91.5%	94.3%	91.5%	92.9%

The results of the classifiers over dataset-3 are illustrated in Table 7. ANN and RandomForest classifiers have shown the best performance among the other classifiers with accuracy of 88.92% and 88.56% respectively, however, ANN was the superior. The DecisionStump classifier has shown the lowest performance among the other classifiers with 82.65% of accuracy.

Table 7. Classifiers Performance using IG based selected features- Dataset 3

Classifier	Accuracy	TPR	Precision	Recall	F-Measure
NaiveBayes	85.23%	91.2%	85.9%	91.2%	88.4%
ANN	88.92%	93.2%	87.8%	93.2%	90.4%
DecisionStump	82.65%	90.5%	82.1%	90.5%	86.1%
KNN	87.82%	91.2%	88.2%	91.2%	89.6%
J48	87.08%	91.8%	86.5%	91.8%	89.1%
RandomForest	88.56%	92.5%	87.7%	92.5%	90.1%

5 Conclusion

In this paper two experiments have been conducted using six machine learning classifiers using three phishing datasets. The first experiment has conducted using all datasets features and the second has conducted after reducing the number of features. The RandomForest classifier has shown the best accuracy among the other classifiers over Dataset-1 and Dataset-2 with accuracy of 98.45%, 97.15% respectively while J48 classifier was the best with 89.66% of accuracy over Dataset-3. On the other hand, the results of the second experiment showed that RandomForest classifier has the best performance over Dataset-1 and Dataset-2 with 98% and 93.66% of accuracy respectively. While using Dataset-3 the ANN classifier has shown an accuracy of 88.92%. It is obvious that reducing the number of used features has affected the performance of the classifiers. However, there has been a slight drop in accuracy using RandomForest classifier with 0.45% using Dataset-1. Whereas, using Dataset-2 it has a noticeable dropping in the accuracy with 3.49 %. In addition, J48 classifier has shown a manifest drop in the accuracy with 2.58% using Dataset-3. Oppositely, ANN classifier has shown the same performance with the same detection accuracy.

6 Acknowledgment

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

7 References

[1] Q. Cui, G.-V. Jourdan, G. V. Bochmann, R. Couturier, and I.-V. Onut, "Tracking phishing attacks over time," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 667-676. https://doi.org/10.1145/3038912.3052654

[2] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommunication Systems,* vol. 76, pp. 139-154, 2021. https://doi.org/10.1007/s11235-020-00733-2

[3] P. Ghann, E. D. Tetteh, K. Asare Obeng, and M. Elias, "Preserving the Privacy of Sensitive Data Using Bit-Coded-Sensitive Algorithm (BCSA)," *International Journal of Recent Contributions from Engineering, Science & IT (iJES),* vol. 10, pp. pp. 4-16, 12/07 2022. https://doi.org/10.3991/ijes.v10i04.35023

[4] A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list,"*EURASIP Journal on Information Security,* vol. 2016, pp. 1-11, 2016. https://doi.org/10.1186/s13635-016-0034-3

[5] H. Y. Kadhim, K. H. Al-saedi, and M. D. Al-Hassani, "Mobile Phishing Websites Detection and Prevention Using Data Mining Techniques," *International Journal of Interactive Mobile Technologies (iJIM),* vol. 13, pp. pp. 205-213, 09/25 2019. https://doi.org/10.3991/ijim.v13i10.10797

[6] S. Gupta, A. Singhal, and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," in *2016 international conference on computing, communication and automation (ICCCA)*, 2016, pp. 537-540. https://doi.org/10.1109/CCAA.2016.7813778

[7] A. H. Aljammal, H. Bani-Salameh, A. Alsarhan, M. K. Kharabsheh, and M. Obiedat, "Node Verification to Join the Cloud Environment Using Third Party Verification Server," *Int. J. Interact. Mob. Technol.,* vol. 11, pp. 55-65, 2017. https://doi.org/10.3991/ijim.v11i4.6501

[8] A. H. Aljammal, H. Bani-Salameh, A. Qawasmeh, A. Alsarhan, and A. F. Otoom, "A new technique for data encryption based on third party encryption server to maintain the privacy preserving in the cloud environment," *International Journal of Business Information Systems,* vol. 28, pp. 393-403, 2018. https://doi.org/10.1111/tme.12467

[9] A.-P. W. Group. (2021, 1-December). *Phishing Activity Trends Report*. Available: https://docs.apwg.org//reports/apwg_trends_report_q3_2021.pdf

[10] M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," in *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, 2018, pp. 1-5. https://doi.org/10.1109/ISDFS.2018.8355389

[11] T. Peng, I. Harris, and Y. Sawa, "Detecting phishing attacks using natural language processing and machine learning," in *2018 ieee 12th international conference on semantic computing (icsc)*, 2018, pp. 300-301. https://doi.org/10.1109/ICSC.2018.00056

[12] C. N. Gutierrez, T. Kim, R. Della Corte, J. Avery, D. Goldwasser, M. Cinque*, et al.*, "Learning from the ones that got away: Detecting new forms of phishing attacks," *IEEE Transactions on Dependable and Secure Computing,* vol. 15, pp. 988-1001, 2018. https://doi.org/10.1109/TDSC.2018.2864993

[13] I. Saha, D. Sarma, R. J. Chakma, M. N. Alam, A. Sultana, and S. Hossain, "Phishing Attacks Detection using Deep Learning Approach," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 1180-1185. https://doi.org/10.1109/ICSSIT48917.2020.9214132

[14] A. K. Jain and B. B. Gupta, "Two-level authentication approach to protect from phishing attacks in real time," *Journal of Ambient Intelligence and Humanized Computing,* vol. 9, pp. 1783-1796, 2018. https://doi.org/10.1007/s12652-017-0616-z

[15] L. Halgaš, I. Agrafiotis, and J. R. Nurse, "Catching the Phish: Detecting phishing attacks using recurrent neural networks (RNNs)," in *International Workshop on Information Security Applications*, 2019, pp. 219-233. https://doi.org/10.1007/978-3-030-39303-8_17

[16] A. K. Jain and B. B. Gupta, "Detection of phishing attacks in financial and e-banking websites using link and visual similarity relation," *International Journal of Information and Computer Security,* vol. 10, pp. 398-417, 2018. https://doi.org/10.1504/IJICS.2018.100–16392

[17] T. Chin, K. Xiong, and C. Hu, "Phishlimiter: A phishing detection and mitigation approach using software-defined networking," *IEEE Access,* vol. 6, pp. 42516-42531, 2018. https://doi.org/10.1109/ACCESS.2018.2837889

[18] N. Azeez, S. Misra, I. A. Margaret, and L. Fernandez-Sanz, "Adopting Automated Whitelist Approach for Detecting Phishing Attacks," *Computers & Security,* p. 102328, 2021. https://doi.org/10.1016/j.cose.2021.102328

[19] P. K. Sahoo, "Data mining a way to solve Phishing Attacks," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 2018, pp. 1-5. https://doi.org/10.1109/ICCTCT.2018.8550910

[20] S. A. Khan, W. Khan, and A. Hussain, "Phishing attacks and websites classification using machine learning and multiple datasets (A comparative analysis)," in *International Conference on Intelligent Computing*, 2020, pp. 301-313. https://doi.org/10.1007/978-3-030-60796-8_26

[21] C. L. Tan, "Phishing dataset for machine learning: Feature evaluation," *Mendeley Data,* vol. 1, p. 2018, 2018.

[22] L. M. Rami Mustafa A Mohammad, Fadi Thabtah. (2012, 12- September). *UCI Machine Learning Repository*. Available: http://archive.ics.uci.edu/ml/datasets/phishing+websites

[23] N. Abdelhamid, "Website Phishing Data Set," ed: Retrieved from UCI Machine Learning Repository: https://archive. ics. uci …, 2016.

[24] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: An update. ACM SIGKDD Explorations. 2009; 11 (1): 10–18," ed. https://doi.org/10.1145/1656274.1656278

8 Authors

Ashraf H. Aljammal is currently an Associate Professor at the Department of Computer Science and Applications, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan. Dr.Aljammal received the B.S. degree in computer science from Albalqa' Applied University, Al-Salt, Jordan, in 2006, the master's degree from Universiti Sains Malaysia, USM, Malaysia, in 2007, and the PhD degree from Universiti Sains Malaysia, USM, Malaysia, in 2011. His research interests include but not limited to network security, cyber security, IoT security, network monitoring, cloud computing, Machine learning and Data mining, ashrafj@hu.edu.jo.

Salah Taamneh is currently an Associate Professor at the Department of Computer Science and its Applications, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan. He received the B.S. degree in computer science from Jordan University of Science and Technology, Irbid, Jordan, in 2005, the M.S. degree in computer science from Prairie View A&M University, Prairie View, Texas, in 2011 and the Ph.D. degree in computer

science from University of Houston, Houston, Texas, USA, in 2016. He. His current research interests include parallel and distributed computing, machine learning and human- computer interaction, taamneh@hu.edu.jo.

Ahmad Qawasmeh is a native of Jordan where he studied Computer Engineering. He obtained his M.S. degree in Computer Science in 2010 and completed his Ph.D. on performance analysis support for HPC applications in Computer Science from the University of Houston in 2015. His research interests include parallel programming languages, performance analysis, and machine learning. He joined The Hashemite University, Zarqa, Jordan in 2016 as an assistant professor in the Dept. of Computer Science, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, ahmadr@hu.edu.jo.

Hani Bani-Salameh is a Full Professor in the Software Engineering Department at Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan. He holds a BSc in Computer Science, MSc in Computer Science from the New Mexico State University (NMSU), and PhD in Computer Science from the University of Idaho (UI). His research interests include software engineering, computer supported cooperative work (CSCW), software development environments, collaborative software development in virtual environments, and social networking and social media. He studies social interactions in social networks and online environments, hani@hu.edu.jo.

Article submitted 2022-12-21. Resubmitted 2023-01-16. Final acceptance 2023-01-17. Final version published as submitted by the authors.