PDMLP: PHISHING DETECTION USING MULTILAYER PERCEPTRON

Saad Al-Ahmadi[1] and Tariq Lasloum[2]

[1]Department of Computer Science, College of Computer and Information Science, King Saud University, Riyadh, Saudi Arabia
[2]Department of Computer Engineering, College of Computer and Information Science, King Saud University, Riyadh, Saudi Arabia

ABSTRACT

A phishing website is a significant problem on the internet. It's one of the Cyber-attack types where attackers try to obtain sensitive information such as username and password or credit card information. The recent growth in deploying a Detection phishing URL system on many websites has resulted in a massive amount of available data to predict phishing websites. In this paper, we purpose a new method to develop a phishing detection system called phishing detection based on a multilayer perceptron (PDMLP), which used on two types of datasets. The performance of these mechanisms evaluated in terms of Accuracy, Precision, Recall, and F-measure. Results showed that PDMLP provides better performance in comparison to KNN, SVM, C4.5 Decision Tree, RF, and RoF to classifiers.

KEYWORDS

MLP, Phishing, machine learning, Features.

1. INTRODUCTION

Phishing is one of Cyber-attack types. This attack takes place by stealing secret information or luring users into giving sensitive information, such as usernames, passwords, and credit card number. Also, phishing attacker mostly lures those victim users to have them enter their privet information into the phishing pages [1]. APWG (Anti-Phishing Working Group) in 2019, a total number of 162,155 was detected as phishing sites in the fourth quarter of 2019, where the third quarter decreased by 266,387, and decreased by 182,465 in the second quarter, and the fourth quarter of 2018 shows an increase of 138,328 [2]. Statistics study from the Kaspersky Lab, in 2019, 19.8% of targeted computer users were attacked by Malware-class. It also showed that web antivirus components identified a total of 273,782,113 URLs as malicious [3].

Phishing URLs are established to provide phishing attacks. Mostly, every legitimate URL has common characterization such as syntax: <protocol>://<hostname><path>. For example:

https://www.linkedin.com/login?fromSignIn=true&trk=guest_homepage-basic_nav-header-signin.

"https://www.linkedin.com" indicates the base URL. To get the requested resource, the first part to be used is the protocol of the URL. For example, HTTP, HTTPS, and FTP are generally the most used protocols. <hostname> represents the webserver identifier on Internet. In the URL shown above, the hostname is www.linkedin.com, while the domain is the LinkedIn name and

the TLD (Top Level Domain) is (.com). <path> in the URL contains different punctuation marks such as /, ., -, ...etc. The path is represented by the content, which happened after the first forward slash and, comes after the <hostname> [4]. Looking to a phishing URL that unoriginal such as Amazon: https://www.co- amazon.us.net.jp.a7w8errq8tcs9bn gt6rxa.net/signin/, where the design is as follows:

- Protocol: https
- Subdomain: co-amazon.us.net
- Domain: a7w8errq8tcs9bngt6rxa.net
- Hostname:amazon.co.jp.a7w8errq8tcs9bngt6rxa.ne
- Free URL: sign in.

The attacker provides a fake website that generally has a login form, therefore when a user opens it and login in with personal information, the attacker will have access to that information [5]. Also, these are steps involved in a phishing attack:

- Attackers establish fake webpage.
- The attacker sends a link to the fake webpage to the victim.
- Users open the fake webpage and submit secret and confidential information.
- Attackers have secret and confidential information about the victim.

Machine learning is the most critical recent phishing URL detection research, which is the phishing URL detection based. The extracted features quality plays an essential role in the machine learning methods' final results. The extraction and selection of the most useful features is a more important thing of recent research before processing them [6].

It's necessary to understand the most fundamentals required for any designed system only so that we could detect phishing URLs. Those fundamentals are features, datasets, and classifiers. Based on studies, each website mostly consists of 31 features. Moreover, the dataset must be obtainable from many websites such as Alexa,Phishtank, and Kaggle website. Dataset classified into two parts: legitimate and phishing. The dataset mostly should be separated into a training dataset and testing data set. However, the classifier used to classify the dataset, for example, Random Forest classifier (RF), Recurrent Neural Networks (RNN), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Multilayer Perceptron (MLP). The results obtained will be based on Accuracy, Precision, Recall, F-measure, and Confusion Matrix.

The proposed work is structured as the following: Section 2 discusses practices and different methods presented in the literature for phishing detection of websites. Section 3 introduces the proposed methodology by merely using a multilayer perceptron. In part 4, we have explained the experiment results, two types of datasets applied to the proposed classifier for the detection of phishing websites—finally, the conclusion is given in Section 5.

2. LITERATURE REVIEW

Based on cascading style sheet (CSS), Phishing-Alarm is proposed by authors in [1] to detect phishing websites, where this solution based on three phases, feature extraction, compute the similarity, and phishing decision. First, the CSS rule extracted for suspicious page and target page, calculates the similarity between them if the difference between them notified, then the website is phishing. Dataset collected almost 9,307 phishing websites from PhishTank, where testing use approximately 3,115 and training use 6,192. Using Recall, F1 and precision will help determine the performance for this proposed and compared with other approaches, and this is the best when Recall is 97.92%, F1 is 0.99 and, precision is 100%.

Auto-updated white-list of legitimate sites is a system proposed by authors in [5], which is made of multiple modules that prevent phishing attacks. The first one is the DNS and URL matching,which has a white-list that consists of two factors, domain name and its IP address. The second one is phishing identification; it checks if the website is phishing or benign. It's essential to test the hyperlinks features to decide by extracting the hyperlinks from the website and then applying the algorithm that detects phishing. Notice to the user will be shown in the case of website phishing after examination for legitimacy. However, the system will update the white-list by adding the domain if the website is legitimate. Based on three factors of hyperlinks, the algorithm for phishing detection takes the decision. Those factors are a website that does not have hyperlinks, null links in the source code, and foreign links in the source code. As a result, that method is very efficient with protection toward phishing attacks. It has an 86.02 % true positive rate and a false positive rate of 1.48 % and an accuracy of 89.38%.

It is essential to apply deep learning Machine by two types of features: interaction feature and original features, which detect the phishing URL. Deep Belief Networks (DBN) used to discover phishing websites. From ISP (Internet Service Provider), real IPs is used for testing. Fishing features can be extracted from the data set by DBN. Contrastive Divergence (CD) is selected as a training algorithm. However, when (TPR) is being used as an evaluation criterion, the result shows that 90% true positive rate (TPR), while 0.6% is a false positive rate (FPR) [7].

The authors in [8] proposed a new solution to detect phishing attacks called PhishLimiter. This solution used Artificial Neural Network (ANN), which is developed using the PhishLimiter Score (PLS) system to classify phishing signatures. Two-type inspection approaches are used: Store and Forward (SF) as well as Forward and Inspect (FI). To evaluate PhishLimiter, Gruber, Spoon, and Rodney Approaches are used. However, the Gruber approach is the most efficient because of its minimum inspection time needed for every URL. The result shows PhishLimiter accuracy is very high, with an approximate average of 98.39%.

To make the classification, constructing a phishing webpage detection model SSM (SAE-Softmax model) is based on Stacked Auto-encoder (SAE) and uses of the Softmax regression model proposed by authors in [9]. Using the SAE network helps with data reconstruction implementation, while Softmax helps with adjusting the network. A total of 52 extracted features are classified into two categories: URL related features and HTML based features. After many experiments, it's been shown that the best number of hidden layers is 2. Moreover, the width of the first hidden layers is 50, while the second hidden layers are 40. Finally, the result shows a 99,95% accuracy with 0.08 times of computation for one iteration.

The authors in [10] applying machine learning models helps to predict the phishing site by comparing feature-engineering for random forest classifier (RF) and feature-engineering for Long Short Term Memory (LSTM) which is a new method for recurrent neural networks (RNN). First, it extracts a set of 14 features rather than an algorithm to classify and build this model by the random forest (RF) method. In the training process, 2 million URLs were used, 50% were phishing, which was from Phishtank, and 50% of them were legitimate, which were from Crawl. Results showed that the LSTM network is 98.7% average accuracy with a 98% average of F1-Score while in RF, the average accuracy is 93.5%, and the average of F1-Score is 93%.

The authors in [11] proposed a new approach using machine learning detection to classify URLs. This new approach is depending on natural language processing features by using word vector representation and models that called ngram as the most critical features on the blacklist word. Providing classification and criteria with those features extracted from ngram models, word vector representation, and other lexical properties will use the Support Vector Machine (SVM) as a machine learning algorithm. The total number of extracted features from the word2vec model

is 100. However, out of 150,397 URLs, 107,615 were benign, while 42,782 were malicious. The result shows that we can see that SVM achieves the highest level of accuracy rate of 97.1% and 0.95 F1 score while maintaining the classification time of 0.01 second.

To enhance the phishing website prediction, the authors in [12] suggested using the deep neural networks (DNNs) with evolutionary algorithm-based feature selection and weighting procedures during the hybrid intelligent phishing website prediction in order to come up with new suggestions which could be helpful in the reduction of the phishing. First, after that pre-processing phase, collecting phishing and data set of legitimate websites is responsible for extracting features, preparing the training data set, and extracting popular features of the website to be mainly used to construct and train a DNN classifier which is converted into whether numerical or categorical features. However, it's best to use the Genetic algorithm (GA) to get the highest influential features and the best weights after that utilizing the features with DNN to enhance the detection of the phishing webpage. Finally, the DNN training and evaluation phase will take place to finalized,improving the phishing website prediction. By comparing DNNs and other classifiers without and with GA-based feature selecting and weighting the classification accuracies of BPNN, DNN, C4.5, and KNN significantly improved by applying the GA-based feature selection method from 87.14, 88.77, 84.92, and 87.07% to 89.36, 90.39, 85.37, and 87.8%, respectively. The classification accuracies of BPNN, DNN, C4.5, and KNN with GA based feature weighting increased further to 89.28, 91.13, 85.37, and 88.99%, respectively.
To detect the phishing website, the authors in [13] proposed a model called Resource Description Framework (RDF) used as well as ensemble learning algorithms for the classification of a web page. First, construction of the RDF by using 21 features is done. Then, extracting keywords from the website fed into the search engine and obtained the top 10 pages. After that comparing those ten pages with the suspicious website, if they were similar, then the page is legitimate. Otherwise, the page inserts to random forest classifier to decide if this page is phishing or legitimate. To sum up, the 2056 website used with 1256 phishing and 800 legitimate websites. The result shows the accuracy of this method is 98.68%, the true positive rate is 98.8%, and the false-positive rate is 1.5%.

The authors in [14] designed a framework to detect phishing URLs by using fuzzy logic as a classifier. To clarify and check the websites are phishing or legitimate, the extract features will have designed rules, where those rules depend on (If … then).From Phishtank, almost 1000 URLs collected to collect datasets, which the result shows the accuracy of this method is 91.46%. DF.GWO-BPNN module is proposed by authors in [15] to classify the phishing websites. All websites that processed are always caching by using this module. The URL composition extracts and classifies features such as, (Extraction and classification module). In addition to this module, the URL features consist of two criteria: dominant and recessive.URLs feed the DF.GWO-BPNN Classifier for more procession with recessive. Therefore, 3000 phishing sites are tested out of a total of 6000 URL. The result shows the Accuracy of DF.GWO-BPNN is 98.78 %, which the best after comparing it with other famous used classification models such as SVM, PSOBPNN, and the BPNN model.

In [16], the authors used different modules to detect the phishing website and compare them for accuracy. Those modules are the Random Forest Classifier, k-Nearest Neighbour (k-NN), Artificial Neural Networks (ANN), C4.5 Decision Tree, Support Vector Machine (SVM) and Rotation Forest (RoF). They used datasets brought from the available public. The result shows that the Accuracy for Random Forest Classifier is best with 97.36%, where the accuracy for other modules is k-NN 97.18 %, SVM 97.17%, ANN 96.91%, RoF 96.79% and C4.5 95.88%.

The authors in [17] proposed a module to detect phishing websites. A module has been used as the main solution which uses the Random Forest algorithm. However, for accurate results, three

main factors play a role in processing these data, which are Parsing, Heuristic Classification of data, Performance Analysis. Moreover, The RStudio tool and random forest classification algorithm are very beneficial, considering that it specifies how effective it is. Parsing helps analyze the feature set, which represents 8 out of 31. Second, Heuristic Classification of data is done where RStudio was used for the implementation of the Random forest algorithm with 70% training 30% testing. Finally, Performance Analysis, the result showed that the random forest has the best outcome with an accuracy of 95%. Table 1 represents the results of the comparative evaluation.

Table 1. summary of phishing URLs detection approaches

Work	**Algorithm**	**Data Set**				**Performance Acc.,Prec., Rec.**
		Dataset source		**Dataset size**		
		Legitimate	**Phishing**	**Legitimate**	**Phishing**	
[1]	Phishing-Alarm: Based on the cascading style sheet (CSS) to detect the phishing websites.	_	PhishTank	_	9,307	_, 100, 97.92
[5]	white-list: The DNS and URL matching has a white-list that consists of two factors, domain name and its IP address.	Alexa Stuffgate Online payment service provider	PhishTank	200 150 55	1120	89.38, _, _
[7]	CD: Contrastive Divergence (CD) is selected as a training algorithm.	ISP	ISP	_	_	_, _, _
[8]	ANN: Artificial Neural Network. PhishLimiter: a new solution to detect phishing attacks.	_	_	4,150 1,185	11,055 1897 4,559 3,718	98.39, _, _
[9]	SSM: is phishing webpage detection	Alexa	PhishTank	8,848	11,321	99.95, _, _

	model (SAE-Softmax) is based on the Stacked Auto encoder.					
[10]	LSTM: New method for recurrent neural networks (RNN).	Common Crawl	PhishTank	1M	1M	98.76, 98.60, 98.93
[11]	SVM: Support Vector Machine as a machine learning algorithm	DMOZ	MDL+ Mac0de+ CleanMX	107,615	42,782	97.1, 93, 97
[12]	GA: Genetic algorithm to get the highest influential features and the best weights.	UCI	UCI	548	702	91.13, _, _
[13]	RDF: Resource Description Framework is being used as well as ensemble learning algorithms for the classification of webpage	Alexa+ Google's top 1000 most visited Sites	PhishTank + Reasonable-Phishing Web pages List	800	1256	98.68, _, _
[14]	Fuzzy logic: To clarify and check the websites is phishing or legitimate, where those rules depend on (If … then).	_	PhishTank	_	1000	91.46, _, _
[15]	DF.GWO-BPNN: To classify the phishing websites where this	Security Alliance	PhishTank APWG	4300	4500	98.78, _, _

	module have two criteria, dominant feature and the recessive features					
[16]	ANN, KNN, RF and SVM are classifiers used to detect phishing websites.	_	_	_	_	97.36, _, _
[17]	RF: The Random Forest algorithm is a classifier to detect the phishing webpage.	_	PhishTank	_	_	95, _, _

3. THE PROPOSED PDMLP METHOD

The proposed PDMLP is based on the multilayer perceptron (MLP), which is the most important model from the deep neural network. Figure 1 shows the different layers of MLP, which is made of three or more layers. These layers are the input layer, one or more hidden layers, and the output layer. They contain threshold, weight, and transfer function for data to transfer from the anterior to the posterior and then to the output layer. If the error isn't up to the target between the known data and the data of the output layer, the threshold of the layers and weights will be adjusted from the back forward.

By forming a group of linear, the Perceptron is calculating single output from many inputs based on its outputs weights. Producing an output y requires a function called activation function, which multiplies inputs made up of $x_1, x_2, \ldots, x_n$ with corresponding weights made up of $w_1, w_2, \ldots, w_n$. After that, it puts the outputs through a non-linear activation function. Which is written mathematically as followed:

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) = \varphi(w^T x + b) \quad (1)$$

Where w represents the weights vector, x represents the inputs vector, b represents the bias, and φ represents the activation function [18].

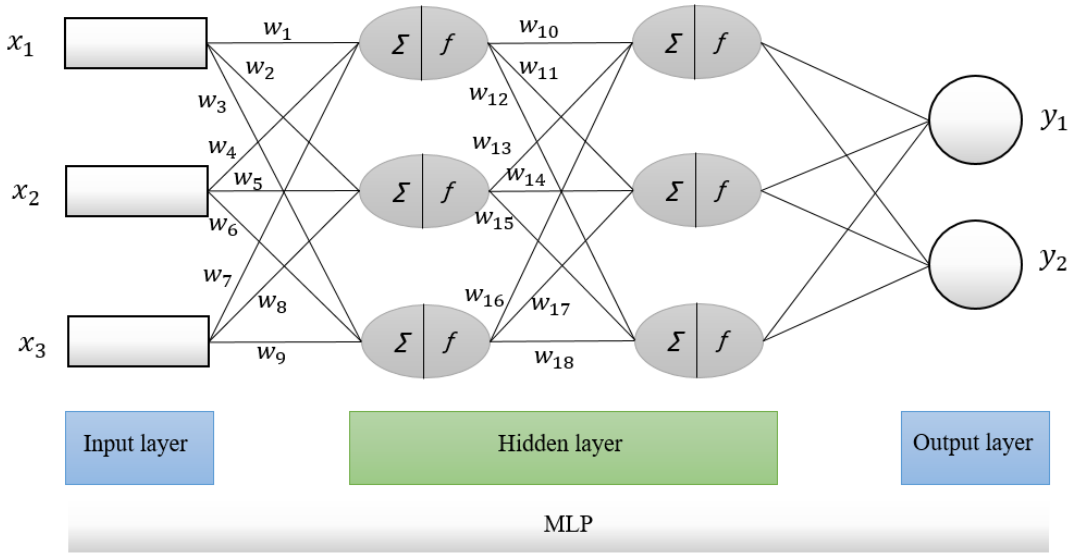

Figure 1. Multilayer Perceptron

We will use the MLP to classify data and detect the phishing website by following the next stages:

3.1 The dataset

Several high-quality datasets are available on many reliable websites. Alexa, Phishtank, UCI, and Kaggle website are well-known sources for a lot of exciting datasets. In this work, the UCI, and Kaggle websites will be used for experimental purposes. Every database consists of up to 31 features, as shown in table 2. The last attribute is " Result " which refers to the presence of a phishing website.

Table 2. Table Type Styles

#	**Attributes**		
1	having_IP_Address	17	Submitting_to_email
2	URL_Length	18	Abnormal_URL
3	Shortining_Service	19	Redirect
4	having_At_Symbol	20	on_mouseover
5	double_slash_redirecting	21	RightClick
6	Prefix_Suffix	22	popUpWidnow
7	having_Sub_Domain	23	Iframe
8	SSLfinal_State	24	age_of_domain
9	Domain_registeration_length	25	DNSRecord
10	Favicon	26	web_traffic
11	port	27	Page_Rank
12	HTTPS_token	28	Google_Index
13	Request_URL	29	Links_pointing_to_page
14	URL_of_Anchor	30	Statistical_report
15	Links_in_tags	31	Result
16	SFH		

We used two types of datasets which are; first, Phishing website 1, which consists of 11,055 rows, and second, Phishing websites 2, which consists of 2456 rows, will be divided into two parts. First part (80%) will be used for training the classifier whereas the second part (20%) will be used for testing purpose. Analysis of the dataset, Figure 2 shows a correlation matrix of the features where is no specific feature has a very high correlation in our target value except for Favicon with popup Window. Moreover, some features have a negative correlation with the target value, and others have positive.

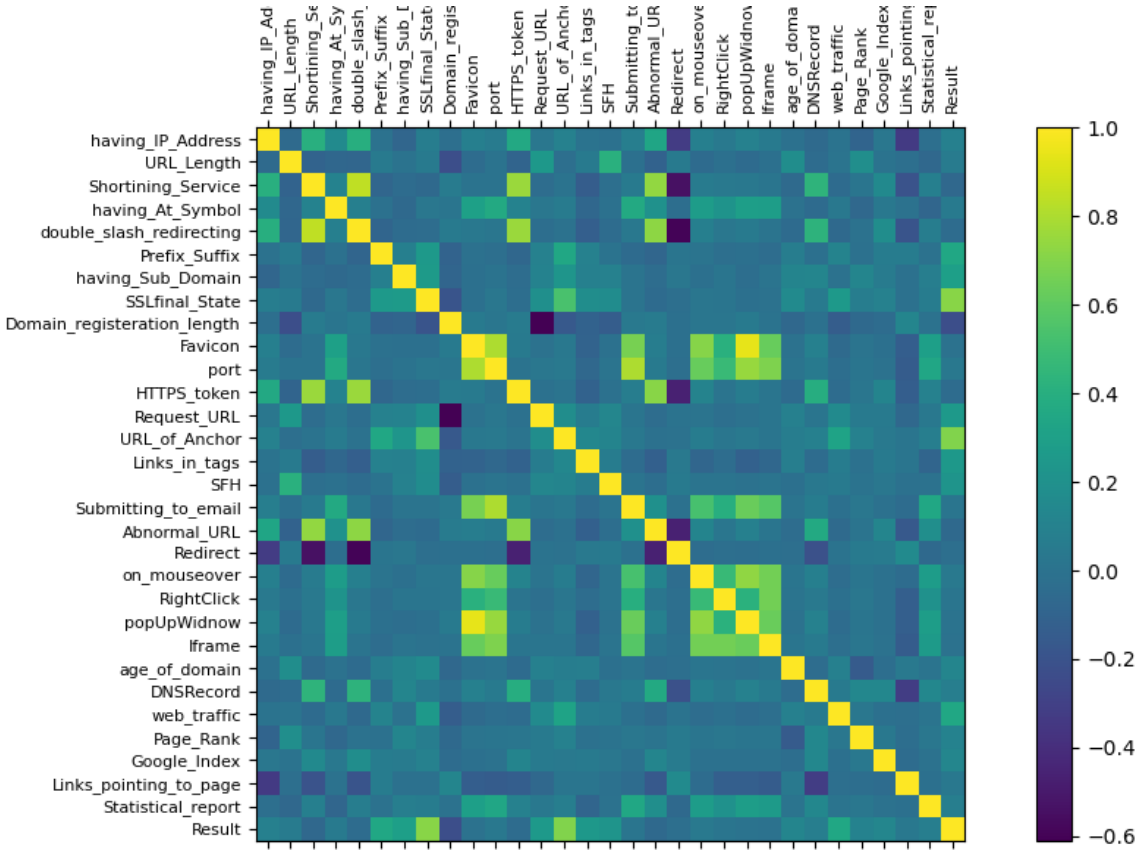

Figure 2. Correlation Matrix of Features

3.2. Result feature

The samples of the dataset which we are working on should be roughly balanced. An unbalanced dataset can obligate the classifier to be biased into one or two classes with many samples, whereas reducing other classes with fewer samples. Our dataset is unbalanced, and it contains more samples for class Phishing, than it does for class Legitimate.

4. EXPERIMENT RESULTS

We implement an MPL classifier with two hidden layers and 100 neurons. We used two types of datasets. The first type contains 11,055 URLs where 6,157 is legitimate, and 4,898 is phishing, the second type is containing 2,456 URLs where 1,094 is legitimate, and 1,362 is phishing. The following parameters are used for reporting results [19]:

4.1. Accuracy

The ratio of instances classified vs. overall number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

4.2. Precision

The ratio of relevant instances properly identified by classifier vs. the overall number of classified relevant instances.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

4.3. Recall

The ratio of relevant instances properly identified by classifier vs. the overall number of relevant instances.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

4.4. F1-score

Also called F-measure or F-score, which means the harmonic mean of Recall and Precision.

$$F1 = \frac{2 * Recall * Precision}{Recall * Precision} \quad (5)$$

One of the challenges that we faced us is to figure out the strength of Accuracy, Precision, Recall, and F1- measure in PDMLP. We added other classifiers such as KNN, SVM C4.5 Decision, RF, and RoF to compare them with PDMLP. Table 3 compares our proposed with several types of machine learning tools, namely KNN, SVM, C4.5 Decision Tree, RF, and Rotation Forest RoF, to classify the phishing website. The performance of these classifiers evaluated in terms of Accuracy, Precision, Recall, F1-measure.

Table 3. Results

Type	Dataset name	Accuracy	Precision	Recall	F1-measure
PDMLP	Phishing website1	**0.9665**	**0.9665**	**0.9665**	**0.9665**
	Phishing Websites2	0.9573	0.9578	0.9573	0.9572
KNN	Phishing website1	0.9647	0.9647	0.9647	0.9647
	Phishing Websites2	0.9532	0.9533	0.9532	0.9532
SVM	Phishing website1	0.9181	0.9184	0.9181	0.9179
	Phishing Websites2	0.9390	0.9390	0.9390	0.9389
C4.5 Decision Tree	Phishing website1	0.9544	0.9548	0.9544	0.9540
	Phishing Websites2	0.9489	0.9491	0.9489	0.9470
RF	Phishing website1	0.9560	0.9563	0.9560	0.9552
	Phishing Websites2	0.9450	0.9452	0.9450	0.9435
RoF	Phishing website1	0.9476	0.9480	0.9476	0.9440
	Phishing Websites2	0.9380	0.9395	0.938	0.9325

Figure 3 shows a comparison between the accuracy of PDMLP and the accuracy of other classifiers such as KNN, SVM, C4.5 Decision Tree, RF, and RoF in two types of phishing websites. In Phishing website 1, The accuracy of PDMLP is better than the accuracy of the other classifiers; also, the accuracy of PDMLP is better than the accuracy of the different classifiers in Phishing websites 2. Regarding the increase of the dataset, an improvement in the accuracy will be seen.

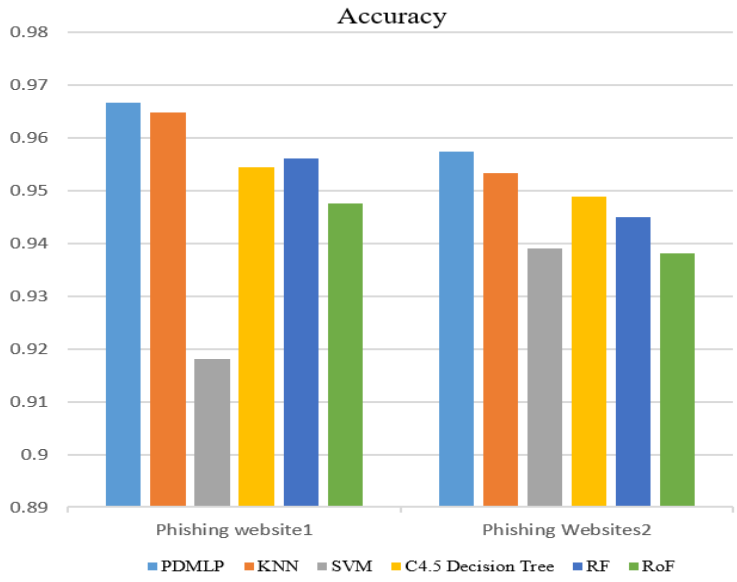

Figure 3. The accuracy for PDMLP and other classifiers

In the PDMLP precision compressed to the other classifiers, we can say PDMLP precision is better than the precision of the KNN, SVM, C4.5 Decision Tree, RF, and RoF, as shown in Figure 4.

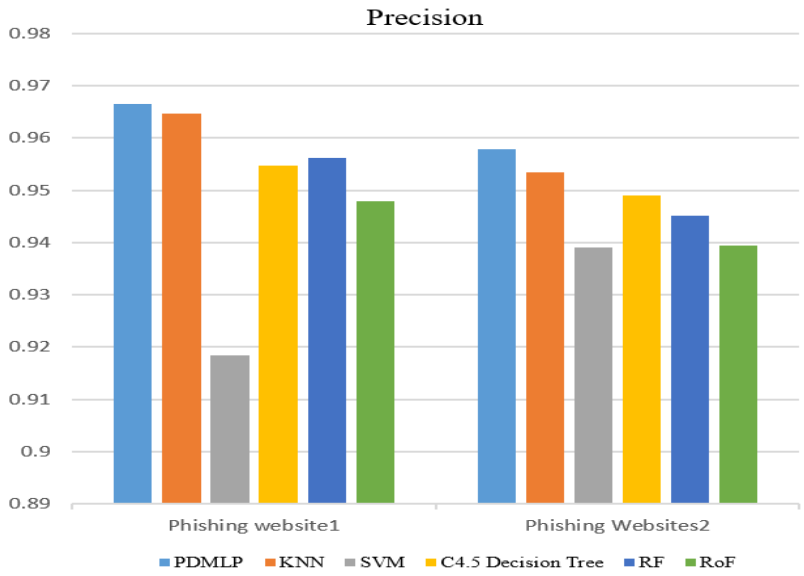

Figure 4. The precision for PDMLP and other classifiers

Figure 5 shows the difference between the Recall of the PDMLP and other classifiers. So, we notify the Recall of PDMLP is better than Recall of the different classifiers.

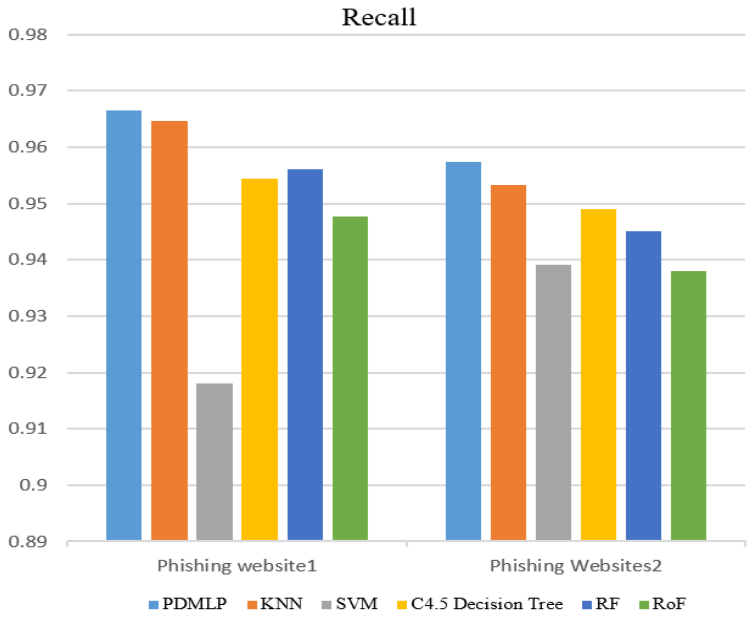

Figure 5. The difference of the recall for PDMLP and other classifiers

Figure 6 provides a comparison of the F1-measure of the two different datasets for the PDMLP and other classifiers where the F1-measure with PDMLP has better results than with KNN, SVM, C4.5 Decision Tree, RF, and RoF also the F1-measure of the PDMLP is better if datasets are large.

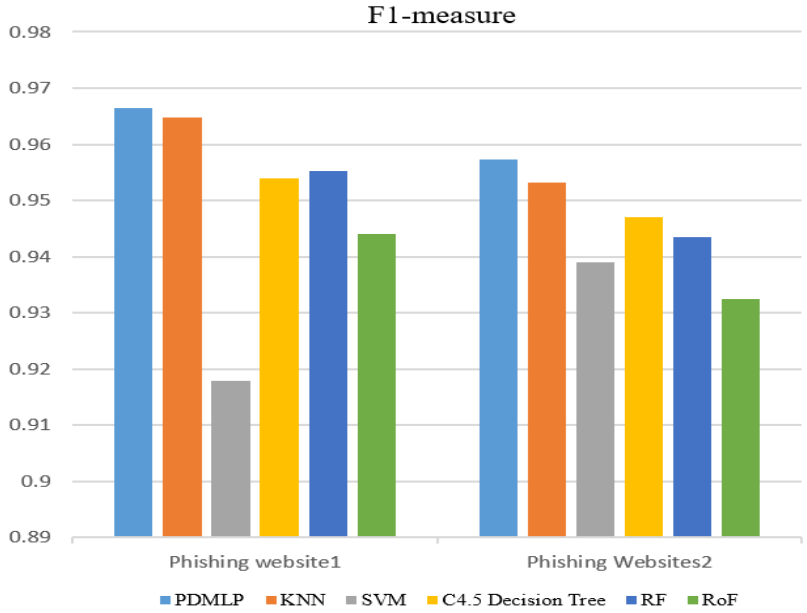

Figure 6. The comparison of F1-measure for PDMLP and other classifiers

5. CONCLUSIONS

We have presented a new method for phishing website detection called PDMLP, which is based on MLP classifier. We used two types of datasets. The first one is 11,055, and the second one is 2456, where 31 features are used. The performance of PDMLP has been evaluated in terms of Accuracy, Precision, Recall, and F-measure. From experiment results, the performance of our proposal is better than that of KNN, SVM, C4.5 Decision Tree, RF, and RoF.

REFERENCES

[1] J. Mao, W. Tian, P. Li, T. Wei, and Z. Liang, "Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity," IEEE Access, vol. 5, pp. 17020–17030, 2017, doi: 10.1109/ACCESS.2017.2743528.

[2] Phishing Attack Trends Report-4Q 2019.Online. Available: https://apwg.org/trendsreports.

[3] Kaspersky Security Bulletin:Overallstatisticals for 2017. [Online]. Available: https://securelist.com/kaspersky-security-bulletin-2019-statistics/95475/, accessed December 12, 2019.

[4] S. G. Selvaganapathy, M. Nivaashini, and H. P. Natarajan, "Deep belief network based detection and categorization of malicious URLs," Inf. Secur. J., vol. 27, no. 3, pp. 145–161, 2018, doi: 10.1080/19393555.2018.1456577.

[5] A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list," Eurasip J. Inf. Secur., vol. 2016, no. 1, 2016, doi: 10.1186/s13635-016-0034-3.

[6] P. Yang, G. Zhao, and P. Zeng, "Phishing website detection based on multidimensional features driven by deep learning," IEEE Access, vol. 7, no. January, pp. 15196–15209, 2019, doi: 10.1109/ACCESS.2019.2892066.

[7] P. Yi, Y. Guan, F. Zou, Y. Yao, W. Wang, and T. Zhu, "Web phishing detection using a deep learning framework," Wirel. Commun. Mob. Comput., vol. 2018, 2018, doi: 10.1155/2018/4678746.

[8] T. Chin, K. Xiong, and C. Hu, "Phishlimiter: A Phishing Detection and Mitigation Approach Using Software-Defined Networking," IEEE Access, vol. 6, pp. 42513–42531, 2018, doi: 10.1109/ACCESS.2018.2837889.

[9] J. Feng, L. Zou, and T. Nan, "A phishing webpage detection method based on stacked autoencoder and correlation coefficients," J. Comput. Inf. Technol., vol. 27, no. 2, pp. 41–54, 2019, doi: 10.20532/cit.2019.1004702.

[10] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and F. A. Gonzalez, "Classifying phishing URLs using recurrent neural networks," eCrime Res. Summit, eCrime, pp. 1–8, 2017, doi: 10.1109/ECRIME.2017.7945048.

[11] Q. T. Hai and S. O. Hwang, "Detection of malicious URLs based on word vector representation and ngram," J. Intell. Fuzzy Syst., vol. 35, no. 6, pp. 5889–5900, 2018, doi: 10.3233/JIFS-169831.

[12] W. Ali and A. A. Ahmed, "Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting," IET Inf. Secur., vol. 13, no. 6, pp. 659–669, 2019, doi: 10.1049/iet-ifs.2019.0006.

[13] V. Muppavarapu, A. Rajendran, and S. K. Vasudevan, "Phishing detection using RDF and random forests," Int. Arab J. Inf. Technol., vol. 15, no. 5, pp. 817–824, 2018.

[14] H. Chapla, R. Kotak, and M. Joiser, "A Machine Learning Approach for URL Based Web Phishing Using Fuzzy Logic as Classifier," Proc. 4th Int. Conf. Commun. Electron. Syst. ICCES 2019, no. Icces, pp. 383–388, 2019, doi: 10.1109/ICCES45898.2019.9002145.

[15] E. Zhu, D. Liu, C. Ye, F. Liu, X. Li, and H. Sun, "Effective phishing website detection based on improved BP neural network and dual feature evaluation," Proc. - 16th IEEE Int. Symp. Parallel Distrib. Process. with Appl. 17th IEEE Int. Conf. Ubiquitous Comput. Commun. 8th IEEE Int. Conf. Big Data Cloud Comput. 11th IEEE Int. Conf. Soc. Comput. Netw. 8th IEEE Int. Conf. Sustain. Comput. Commun. ISPA/IUCC/BDCloud/SocialCom/SustainCom 2018, pp. 759–765, 2019, doi: 10.1109/BDCloud.2018.00114.

[16] A. Subasi, E. Molah, F. Almkallawi, and T. J. Chaudhery, "Intelligent phishing website detection using random forest classifier," 2017 Int. Conf. Electr. Comput. Technol. Appl. ICECTA 2017, vol. 2018-January, pp. 1–5, 2017, doi: 10.1109/ICECTA.2017.8252051.

[17] S. Parekh, D. Parikh, S. Kotak, and S. Sankhe, "A New Method for Detection of Phishing Websites: URL Detection," Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018, no. Icicct, pp. 949–952, 2018, doi: 10.1109/ICICCT.2018.8473085.

[18] I. Tereikovskyi, I. Subach, O. Tereikovskyi, L. Tereikovska, S. Toliupa, and V. Nakonechnyi, "Parameter Definition for Multilayer Perceptron Intended for Speaker Identification," no. 1, pp. 227–231, 2020, doi: 10.1109/atit49449.2019.9030504.

[19] A. El Aassal, S. Baki, A. Das, and R. M. Verma, "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs," IEEE Access, vol. 8, pp. 22170–22192, 2020, doi: 10.1109/ACCESS.2020.2969780.