Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020)
IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

Phishing Attacks Detection using Machine Learning Approach

Mohammad Nazmul Alam
Department of Computer Science and Engineering
Royal University of Dhaka
Dhaka, Bangladesh
mna235@yahoo.com

Dhiman Sarma
Department of Computer Science and Engineering
Rangamati Science and Technology University
Rangamati, Bangladesh
dhiman001@yahoo.com

Farzana Firoz Lima
Department of Computer Science and Engineering
East Delta University
Chittagong, Bangladesh
171001212@eastdelta.edu.bd

Ishita Saha
Department of Computer Science and Engineering
University of Science and Technology Chittagong
ishita_saha05@yahoo.com

Rubaiath-E- Ulfath
Department of Computer Science and Engineering
East Delta University
Chittagong , Bangladesh
171000212@eastdelta.edu.bd

Sohrab Hossain
Department of Computer Science and Engineering
East Delta University
Chittagong, Bangladesh
sohrab.h@eastdelta.edu.bd

Abstract- Evolving digital transformation has exacerbated cybersecurity threats globally. Digitization expands the doors wider to cybercriminals. Initially cyber threats approach in the form of phishing to steal the confidential user credentials. Usually, Hackers will influence the users through phishing in order to gain access to the organization's digital assets and networks. With security breaches, cybercriminals execute ransomware attack, get unauthorized access, and shut down systems and even demand a ransom for releasing the access. Anti-phishing software and techniques are circumvented by the phishers for dodging tactics. Though threat intelligence and behavioural analytics systems support organizations to spot the unusual traffic patterns, still the best practice to prevent phishing attacks is defended in depth. In this perspective, the proposed research work has developed a model to detect the phishing attacks using machine learning (ML) algorithms like random forest (RF) and decision tree (DT). A standard legitimate dataset of phishing attacks from Kaggle was aided for ML processing. To analyze the attributes of the dataset, the proposed model has used feature selection algorithms like principal component analysis (PCA). Finally, a maximum accuracy of 97% was achieved through the random forest algorithm.

Keywords—Phishing attack; phishing attack detection; artificial intelligence; machine learning; deep learning; convolutional neural network

I. INTRODUCTION

Phishing attacks have become anxiety for the cyber world. It causes enormous problems for privacy and financial issues of internet users. Scammers, namely fishers, create false websites [1, 2] to feel and look like a genuine to deceive the people. They spoof emails to steal the identity of legitimate users. They gather personal covert information, password, account information, and credit card details for the transaction. Fishers always change their strategy to attack the system. Social engineering [3-6] is one of the essential techniques the fishers use. Using this technique, they gather personal credentials from a trustworthy person. Phishers create false websites and spoof email in such a way that they are very similar and sometimes look like a real company website that comes from a source. Sometimes the attackers act like a real source and force the users to update the system.

Moreover, they threaten the customer to suspend the account and demand ransom. Email spoofing is another technique used for phishing fraud [7]. Customers are usually misled to disclose private information like passwords and credit card number. Thus fishing is mainly used to steal valuable information such as bank account, password, and credit card details [8]. This type of scam is increasing rapidly, and individuals, business-people are losing their trust in online business. Thus, a negative impression of clients on online business was swarmed as they lost faith in online transactions. Even though encryption software is used to protect the information in the computers' storage, they are also vulnerable to attacks [9]. In this paper, the detection of fishing was performed through ML.

II. BACKGROUND AND RELATED WORK

There are various types of phishing attacks used to cheat the users. Besides, various phishing detection techniques and tools are also available to defend phishing attacks. Classification is one of the techniques used to detect website

phishing [10, 11, 29]. Here, common types of a phishing attack and classification techniques are described below:

A. Different types of phishing attacks

Attackers use various methods to explore the vulnerability of internet security. They always try to discover the loopholes in the security system to exploit them. Various phishing assaults that are distinctive from each other are presented below.

1) Algorithm-Based Phishing

The first phishing attack was caught by America Online (AOL), which was created using an algorithm. The cheater built an algorithm to match the credit card number of America Online accounts [3].

2) Deceptive Phishing

Deceiver uses sorts of ways to deceit the internet users. Fishers send emails to the users to verify the account. They use links and buttons and request to click them. Behind the links, there has a website where the hackers snatch and store the personal information of the users.

3) URL Phishing

Universal Resource Locator (URL) phishing is another kind of phishing attack done by using a hidden link. The link holds the hackers' website. When the user clicks the link, it is redirected to the hackers' website and stores the user's information [11].

4) Hosts File Poisoning

It is used to poison the host file in the windows operating system platform. When the user locates the desired website, then it rerouted to hackers' site, or sometimes it replies 'The Page Not Found'. If it can redirect to the fake site, then users data are recorded and stolen.

5) Content-Injection Phishing

Hackers target the user and present a fake website as legitimate. The reason is to mislead the user or wrongly present the organization. It is also known as content spoofing. The attackers use this strategy to mislead the user and to collect information on their server.

6) Clone Phishing

Biotechnological term 'clone' means producing individuals the same as the original one. It commonly happens in genetic engineering. Clone phishing is another kind of phishing attack where email is hacked by an opponent party from the sender or receiver email. The malicious attacker creates alike original email and sends it to the first or second party with attachment or link. They request to send an updated version of the original [13].

7) Whaling

This type of phishing is targeted at the higher executives of the organization. The content of the email is about important issues, and it is sent to the executives. The message of the email can be like the customers' complaints.

8) Spear phishing

Spear phishing is one kind of email scam used to target specific persons and organizations. The attacker sends emails to get a reply from the targeted person. The email in such a way that they pretend to know many things about the victims, such as their name, work address, email address, and so on [14, 15].

B. Classification techniques to detect website phishing

There are various techniques to fight against phishing attacks to provide safety measures of internet users [16-18]. Among all types of attacks, email spoofing and URL phishing are hard to identify and to stop because the attacker always changes their strategies. To stop these types of phishing, it was recommended to block malicious emails and bogus URLs [2].

A proactive strategy was taken to identify malicious URLs using lexical analysis [19]. To categories the untruthful URLs, machine learning algorithms were applied to features. The analysis was carried out on these features set to mitigate these URLs. Furthermore, principal component analysis (PCA) technique and random forest algorithm were used for phishing attack detection. PCA was used to identify principal components from the variables. After that, data were analyzed to detect phishing type. Random Forest algorithm was applied to the datasets. It classified the website and detected phishing. This method has high accuracy. It can classify the phishing email about 96 per cent and easy to construct. Moreover, it can handle large volumes of datasets. Labelled datasets were used for classification to detect phishing attacks. Different types of features were used for classification like URL-based features, Email features, and Text-based features. An IP address was used as a URL based feature for classification performed by ML classifies, SVM [20-22], KNN [11], and BNN. Emails were detected using the Classification algorithms. The attacker spoofs the email to collect information. These spoofed emails were detected using a supervised learning algorithm such as Naïve Bayes [23-25]. Using algorithms, emails were classified, and then spam and non-spam emails were separated. There were also hyperlink features used for investigation in ML approaches [26-28].

III. METHODOLOGY

This section explains the methodology to detect Phishing attack using ML, and also explains the proposed framework. The experiment was carried out by using ML approaches. ML approaches can be applied in two ways. The first one is supervised learning, and another is unsupervised learning. Feature selection is crucial for ML algorithms. It reduces the redundancy of data which is irrelevant or unnecessary in the data sets. Another statistical method, principal component analysis (PCA), has been used to identify the components of the variables and classify the datasets. The proposed model was presented in figure 1. To experiment with the phishing website firstly, the dataset was selected. The attributes were then analyzed using a feature selection algorithm. The proposed model has used the REF, Relief-F, IG, and GR algorithm for feature selection. Further, the feature is classified between the weakest and most vigorous. Then PCA was applied for analysis.

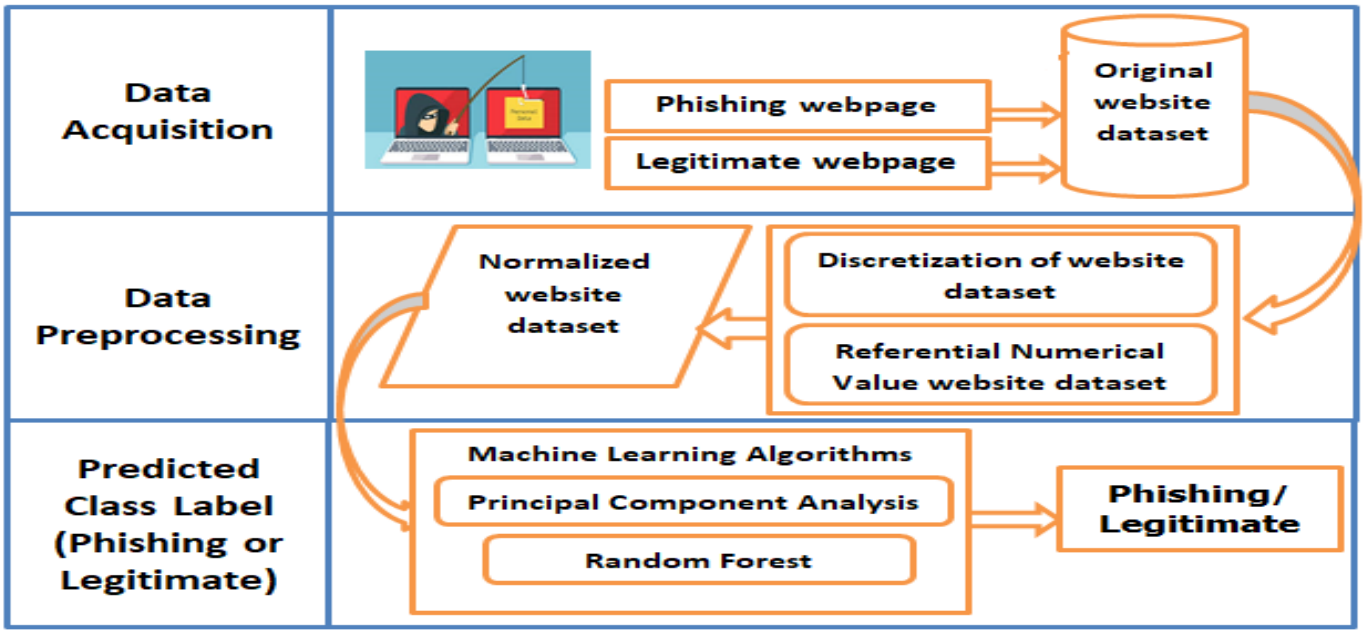

Fig. 1 The proposed framework for phishing detection

A. Data Acquisition

Data acquisition is essential for data analysis—datasets from kaggle.com for our research.

Table 1. The phishing dataset

Index	UsingIP	LongURL	ShortURL	Symbol@	Redirecting//	PrefixSuffix-	SubDomains	HTTPS	DomainRegLen	Favicon	NonStdPort	HTTPSDomainURL	RequestURL	AnchorURL	LinksInScriptTags	ServerFormHandler	InfoEmail	AbnormalURL	WebsiteForwarding	StatusBarCust	DisableRightClick	UsingPopupWindow	IframeRedirection	AgeofDomain	DNSRecording	WebsiteTraffic	PageRank	GoogleIndex	LinkPointingToPage	StatsReport	class
0	1	1	1	1	1	-1	0	1	-1	1	1	-1	1	0	-1	-1	1	1	0	1	1	1	1	-1	-1	0	-1	1	1	1	-1
1	1	0	1	1	1	-1	-1	-1	-1	1	1	-1	1	0	-1	-1	-1	-1	0	1	1	1	1	-1	1	-1	1	0	-1	-1
2	1	0	1	1	1	-1	-1	-1	1	1	1	-1	-1	0	0	-1	1	1	0	1	1	1	1	-1	1	-1	1	-1	1	-1
3	1	0	-1	1	1	-1	1	1	-1	1	1	1	1	0	0	-1	1	1	0	-1	1	-1	1	-1	-1	0	-1	1	1	1	1
4	-1	0	-1	1	-1	-1	1	1	-1	1	1	-1	1	0	0	-1	-1	-1	0	1	1	1	1	1	1	-1	1	-1	-1	-1	1
5	1	0	-1	1	1	-1	-1	-1	1	1	1	1	-1	-1	0	-1	-1	-1	0	1	1	1	1	1	-1	-1	-1	1	0	-1	-1
6	1	0	1	1	1	-1	-1	-1	1	1	1	-1	-1	0	-1	-1	1	1	0	1	1	1	1	-1	-1	0	-1	1	0	1	-1
7	1	0	-1	1	1	-1	1	1	-1	1	1	-1	1	0	1	-1	1	1	0	1	1	1	1	1	-1	1	1	1	0	1	1
8	1	1	-1	1	1	-1	-1	1	-1	1	1	1	1	0	1	-1	1	1	0	1	1	1	1	1	-1	0	-1	1	0	1	-1
9	1	1	1	1	1	-1	0	1	1	1	1	1	-1	0	0	-1	-1	-1	0	1	1	1	1	-1	1	1	1	-1	-1	1	1
10	1	1	-1	1	1	-1	1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	0	1	1	1	1	-1	-1	-1	-1	1	0	-1	-1
11	-1	1	-1	1	-1	-1	0	0	1	1	1	-1	-1	-1	1	-1	1	1	0	-1	1	-1	1	1	1	-1	-1	-1	1	0	1	-1
12	1	1	-1	1	1	-1	0	-1	1	1	1	1	-1	-1	-1	1	1	1	0	1	1	1	1	-1	-1	0	-1	1	1	1	-1
13	1	1	-1	1	1	1	-1	1	-1	1	1	-1	1	0	1	1	1	1	0	1	1	1	1	1	-1	1	-1	1	-1	1	1
14	1	-1	-1	-1	1	-1	0	0	1	1	1	1	-1	-1	0	-1	1	1	0	1	1	1	1	1	-1	-1	-1	1	0	1	-1
15	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	1	0	-1	-1	-1	-1	0	1	1	1	1	1	-1	0	-1	1	1	-1	-1
16	1	-1	1	1	1	-1	-1	0	1	1	-1	1	1	0	-1	-1	-1	-1	0	1	1	1	1	-1	1	1	-1	1	1	-1	-1

B. Data Preprocessing

Data pre-processing is an essential task for the ML application. It was done from raw data and was formatted using the data mining technique. A clean and noise-free

dataset was needed for analysis of the dataset. Most of the dataset contained incomplete and missing values which are filled and completed for ML processing.

1) Feature Explanation

Feature selection is an important task for the analysis of the dataset. Our dataset contained 32 features. Based on the information on the features assumption can be made whether they were distrustful or malicious. The features are explained in table 2.

Table 2. Feature explanation

Feature Name	Feature explanation
Index	Indexing could be used for the webpage to display in the search engine
Using IP	Phishing website used IP instead of DNS
LongURL	LongURL considered as more than a hundred characters
ShortURL	ShortURL is shorter form of URL. bit.ly is an example of ShortURL
Symbol@	Special character used in the phishing URL
Redirecting//	Redirecting was used to get intended destination from the existing link
PrefixSuffix	Prefix was used the beginning of the original word. For example happy become unhappy. On the other suffix was used the end of the word. For example, run become running
SubDomains	Subdomains were used before the main domain to separate the website into a section, and it was used in many ways and purposes. For example, blog.mna235.com, the blog was a subdomain
HTTPS	Most of the modern browsers use https for secure website handling.
Domain Reg Len	It was the number of the year that have registered the domain for a website. It was paid in advance for registration to start the website.
Favicon	Favicon was an icon used on the website to save space.
NonStdPort	
HTTPSDomainURL	It was secure HTTP which was used widely with TLS/SSL Protocol.
RequestURL	It was used to request the resource from the server from the client-side.
AnchorURL	It was a clickable text form used to hyperlink.
LinksInScriptTags	It was used to link at script tag to manipulate the image, form validation and dynamic content
ServerFormHandler	It was used to process the item in the server from the client.
InfoEmail	Info email could be used with its domain or business website.
AbnormalURL	Opposite of normalURL unlikely to occur
WebsiteForwarding	It means multiple sources redirect to a single web address.
StatusBarCust	It showed information about the system at the bottom of the screen.
DisableRightClick	It prevents saving the content from the website. It can be an image or other content.
UsingPopupWindow	It appears with a menu on the screen and disappears after a click
IframeRedirection	Used to survey your website and then redirection
AgeofDomain	Duration of the domain that has existed
DNSRecording	Provides important information especially IP address.
WebsiteTraffic	Define web users who visited the website.
PageRank	Used to rank the webpage used by google search engine
GoogleIndex	Add webpages in google search
LinksPointingToPage	Used to rank the webpages
StatsReport	Provides information about all the downloaded files.
class	Contains attributes and behave

C. Machine Learning Classification Based Modeling

Supervised learning techniques identify the classification and used to detect cyber-attacks. In this method, data was trained previously to predict the new data. In this research, two popular machine learning algorithms are used to detect phishing attacks.

1) Principal Component Analysis.

Principal Component Analysis (PCA) was first invented by Karl Pearson in 1901 in the application of mechanics and then independently developed by Harold Hotelling in 1930. It transforms the large datasets into smaller datasets by reducing the dimensionality of the dataset. Increasing data demands a reduction of dimensionality of variables that are not important to use for prediction. This paper has used PCA to reduce the variance by preserving the information the dataset intact. PCA is much easier to explore the solution and faster to analyze in the machine learning algorithm. For that, it has performed to standardize the variable first so that they could turn into the same scale of measurement. Secondly, it has performed a covariance matrix computation to verify the dependency or correlation of the variables. This feature of space computation was essential to remove the correlated variable, which contained additional information. In the third step, the eigenvectors and eigenvalues of the covariance matrix are calculated to find out principal components. Figure 2 showed the two principal components about the website, and figure 3 showed three dimensions of the correlated data from the dataset.

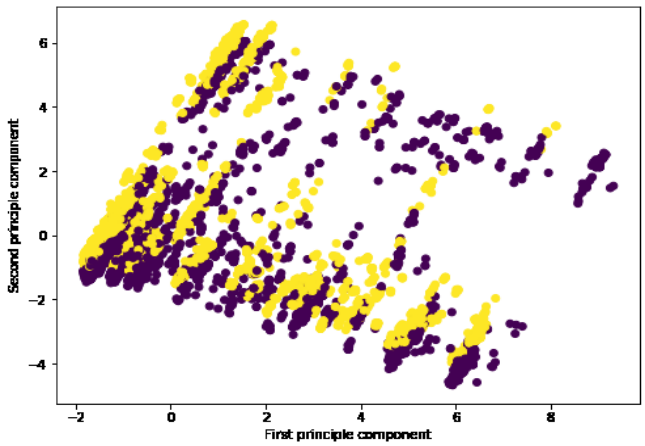

Fig. 2 Two Principle components

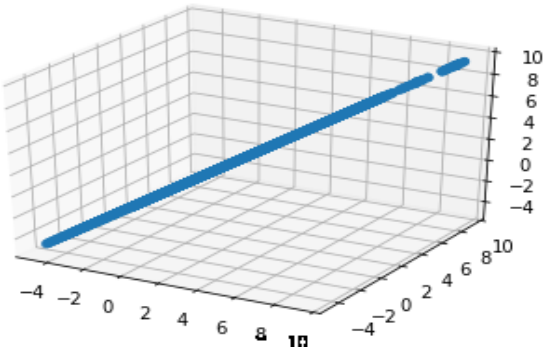

Fig. 3. 3D scatter by PCA

2) Decision Tree

Decision tree (DT) is one of the most popular algorithms in machine learning for binary classification. It results from the decision very fast by creating a small tree and can predict upon training dataset. As its name implies tree, it holds nodes and attribute denotes a test. The branch is the consequence of the test, and each terminal or end node, which is called leaves are the labels of the classification. Determining the best attribute is the most important in this algorithm. Ross Quinlan developed the decision tree ID3 algorithm. It was primarily used in data mining and information theory. Now it is used in machine learning and natural language processing. The proposed model has used the ID3 algorithm in this paper to classify the website, whether it was an official or phishing website. The following steps are followed to get the outcome of the classification of this algorithm:

1. Start with the training data set. Give it the name 'S' and it should have attributes and classification

2. Determine the best attribute of the data sets

3. Divide the 'S' which each have a value of the best attributes

4. Build a decision tree node which holds the best attribute.

5. Use iteration from step 3 and construct a new decision tree unless you cannot classify any more. Represent the leaf node as the outcome of the classification.

Information gain can be obtained using entropy. Split information and gain ratio were used to select the alternative attributes which contained numerical values.

3) Random Forest

Random Forest (RF) is one of the robust algorithms in machine learning. It is a supervised learning algorithm used for classification and regression. It uses the bagging method to combine the learning model and average the overall result for better prediction. RF is used to classify the website between legitimate and phishing. As the random forest is a combination of many single trees, it can produce high accuracy of the result. Firstly it selected samples randomly. Then a decision tree on each sample of the dataset was built where the results are obtained from each decision tree. Then the method was applied to predict the result and to select the highest voted result for final prediction. RF produced high accuracy over a single decision tree, even the data was missing. It could overcome the over-fitting problem.

ML algorithm	Accuracy	Recall	Precision	F1 Score
DT	0.9194	0.9384	0.8804	0.9084
RF	0.9696	0.4216	0.9689	0.5874

IV. RESULTS AND DISCUSSIONS

Phishing attack detection based on feature analysis, data analysis on the selected dataset was carried in this paper.
The confusion matrix shows the performance table on accuracy when compared with the actual classifications in the dataset. Accuracy, precision, recall, and F1 score were used for performance evaluation which was calculated based on the confusion matrix. Confusion matrix used specific table layout for the projection of the performance, as shown in Table 3 and computed according to the following equations:

$$\text{Precision} = TP/(TP + FP)$$

$$\text{Recall} = TP/(TP + FN)$$

$$\text{Accuracy} = (TP + TN)/(TP + FP + TN + FN)$$

$$\text{F1 Score} = 2*((\text{precision}*\text{recall})/(\text{precision}+\text{recall}))$$

True Positive (TP): Correctly predicted phishing URLs were detected with the actual phishing URLs. False Negative (FN): The actual phishing URLs were false classified and detected as legitimate URLs. False Positive (FP): The actual legitimate URLs were classified as false values and detected as phishing URLs. True Negative (TN): The actual class and the predicted class was the same as it showed here the actual

legitimate URLs were correctly predicted as legitimate URLs.

Table 3 presented a confusion matrix on the DT algorithm for actual class on phishing and legitimate website with the predicted class, which was classified with phishing and legitimate class. In table 4, a confusion matrix on the RF algorithm was shown, which was classified in the same way as table 3. In table 5, showed the evaluation matrix on DT and RF for performance analysis.

Table 3: Confusion Matrix DT Algorithm

	Classified Phishing	**Classified legitimate**
Actual Phishing	884	58
Actual legitimate	120	1149

Table 4: Confusion Matrix RF Algorithm

	Classified Phishing	**Classified legitimate**
Actual Phishing	904	38
Actual legitimate	29	1240

Table 5: Metrics for Evaluating DT and RF Performance

Accuracy is the ratio of absolute predicted class value to the all measurement value.

Recall is the ratio of correctly predicted positive measurement to the total measurement of actual yes class

Precision is the ratio of the predicted positive measurement to the total predicted positive measurement.

F1 score means the weighted average of Precision and Recall. Hence, this set takes both False Positives and False Negatives into the record to deliver a proportion betwixt precision and recall.

V. CONCLUSION

A machine learning (ML) based phishing attack detection was proposed in this paper. The investigation utilizes many strategies to identify phishing detection. Standard datasets of phishing attacks from kaggle.com were used as input for the ML algorithms. Two popular machine learning algorithms, namely decision tree and random forest, are implemented to analyze and select datasets for classification and detection. Principal component analysis (PCA) was applied to identify and classify the components of the datasets. DT was used to classify the website and RF was used for classification. Finally, the confusion matrix was drawn to evaluate the performance of these two algorithms. RF had less variance, and it could handle the over-fitting problem. The random forest tree achieved an accuracy of 97%. In our future work, fishing attacks will be predicted from the logged dataset of attacks by using a convolution neural network (CNN). It will be added as a tool for intrusion detection system (IDS).

REFERENCES

[1] S. Mahdavifar and A. A. Ghorbani, "DeNNeS: deep embedded neural network expert system for detecting cyber attacks," (in English), *Neural Computing & Applications,* Article; Early Access p. 28.

[2] S. Mishra and D. Soni, "Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis," (in English), *Future Generation Computer Systems-the International Journal of Escience,* Article vol. 108, pp. 803-815, Jul 2020.

[3] K. S. Adewole, T. Hang, W. Q. Wu, H. B. Songs, and A. K. Sangaiah, "Twitter spam account detection based on clustering and classification methods," (in English), *Journal of Supercomputing,* Article vol. 76, no. 7, pp. 4802-4837, Jul 2020.

[4] B. A. Eduardo, F. D. Walter, and S. G. Sandra, "An Experiment to Create Awareness in People concerning Social Engineering Attacks," (in Spanish), *Ciencia Unemi,* Article vol. 13, no. 32, pp. 27-40, Jan-Apr 2020.

[5] S. Farrugia, J. Ellul, and G. Azzopardi, "Detection of illicit accounts over the Ethereum blockchain," (in English), Expert *Systems with Applications,* Article vol. 150, p. 11, Jul 2020, Art. no. 113318.

[6] C. D. Xuan, H. D. Nguyen, and T. V. Nikolaevich, "Malicious URL Detection based on Machine Learning," (in English), *International Journal of Advanced Computer Science and Applications,* Article vol. 11, no. 1, pp. 148-153, Jan 2020.

[7] J. C. Gonzalez, V. Garcia-Diaz, E. R. Nunez-Valdez, A. G. Gomez, and R. G. Crespo, "Replacing email protocols with blockchain-based smart contracts," (in English), *Cluster Computing-the Journal of Networks Software Tools and Applications,* Article; Early Access p. 7.

[8] S. M. Mironova and S. S. Simonova, "Protection of the Rights and Freedoms of Minors In The Digital Space," (in Russian), *Russian Journal of Criminology,* Article vol. 14, no. 2, pp. 234-241, 2020.

[9] D. Sarma, "Security of Hard Disk Encryption," ed, 2012

[10] S. Shabudin, N. S. Sani, K. A. Z. Ariffin, and M. Aliff, "Feature Selection for Phishing Website Classification," (in English), *International Journal of* Advanced *Computer Science and Applications,* Article vol. 11, no. 4, pp. 587-595, Apr 2020.

[11] A. Zamir *et al.*, "Phishing web site detection using diverse machine learning algorithms," (in English), *Electronic Library,* Article vol. 38, no. 1, pp. 65-80, Jan 2020.

[12] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, "Intelligent phishing detection scheme using deep learning algorithms," (in English), *Journal of Enterprise Information Management,* Article; Early Access p. 20.

[13] G. Sonowal and K. S. Kuppusamy, "PhiDMA - A phishing detection model with multi-filter approach," (in English), *Journal of King Saud University-Computer and Information Sciences,* Article vol. 32, no. 1, pp. 99-112, Jan 2020.

[14] N. A. Azeez, B. B. Salaudeen, S. Misra, R. Damasevicius, and R. Maskeliunas, "Identifying phishing attacks in communication networks using URL consistency features," (in English), *International Journal of Electronic Security and Digital Forensics,* Article vol. 12, no. 2, pp. 200-213, 2020.

[15] E. D. Frauenstein and S. Flowerday, "Susceptibility to phishing on social network sites: A personality information processing model," (in English), *Computers & Security,* Article vol. 94, p. 18, Jul 2020, Art. no. Unsp 101862.

[16] M. Binjubeir, A. A. Ahmed, M. A. Bin Ismail, A. S. Sadiq, and M. K. Khan, "Comprehensive Survey on Big Data Privacy Protection," (in English), *Ieee Access,* Article vol. 8, pp. 20067-20079, 2020.

[17] G. D. Parra, P. Rad, K. K. R. Choo, and N. Beebe, "Detecting Internet of Things attacks using distributed deep learning," (in English), *Journal of Network and Computer Applications,* Article vol. 163, p. 13, Aug 2020, Art. no. 102662.

[18] R. T. Pashiri, Y. Rostami, and M. Mahrami, "Spam detection through feature selection using artificial neural network and sine-cosine algorithm," (in English), *Mathematical Sciences,* Article; Early Access p. 7.

[19] S. Anwar *et al.*, "Countering Malicious URLs in Internet of Things Using a Knowledge-Based Approach and a Simulated Expert," (in English), *Ieee Internet of Things Journal,* Article vol. 7, no. 5, pp. 4497-4504, May 2020.

[20] A. S. Bozkir and M. Aydos, "LogoSENSE: A companion HOG based logo detection scheme for phishing web page and Email brand recognition," (in English), *Computers & Security,* Article vol. 95, p. 18, Aug 2020, Art. no. Unsp 101855.

[21] S. E. Raja and R. Ravi, "A performance analysis of Software Defined Network based prevention on phishing attack in cyberspace using a deep machine learning with CANTINA approach (DMLCA)," (in English), *Computer Communications,* Article vol. 153, pp. 375-381, Mar 2020.

[22] T. A. Tuan, H. V. Long, L. Son, R. Kumar, I. Priyadarshini, and N. T. K. Son, "Performance evaluation of Botnet DDoS attack detection using machine learning," (in English), *Evolutionary Intelligence,* Article vol. 13, no. 2, pp. 283-294, Jun 2020.

[23] F. Ahmed, et al., "A Combined Belief Rule based Expert System to Predict Coronary Artery Disease," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 252-257.

[24] S. Hossain, et al., "A Belief Rule Based Expert System to Predict Student Performance under Uncertainty," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1-6.

[25] S. Hossain et al. "A Critical Comparison between Distributed Database Approach and Data Warehousing Approach." International Journal of Scientific & Engineering Research, Article 5.1 (2014): 196-201.

[26] K. Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258-263.

[27] S. Hossain, D. Sarma, R. J. Chakma, W. Alam, M. M. Hoque, and I. H. Sarker, "A Rule-Based Expert System to Assess Coronary Artery Disease Under Uncertainty," in Computing Science, Communication and Security, Singapore, 2020, pp. 143-159: Springer Singapore.

[28] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.

[29] H. Alqahtani et al., "Cyber Intrusion Detection Using Machine Learning Classification Techniques," in Computing Science, Communication and Security, Singapore, 2020, pp. 121-131: Springer Singapore.