www.gphjournal.org
Gph-International Journal Computer Science Engineering

10.5281/zenodo.8059093 Vol. 06 Issue 06 June - 2023 Manuscript ID: #0888

CLASSIFICATION OF PHISHING ATTACKS IN SOCIAL MEDIA USING ASSOCIATIVE RULE MINING AUGMENTED WITH FIREFLY ALGORITHM

HAMMED, Mudasiru & SOYEMI, Jumoke
Computer Science Department, Federal Polytechnic, Ilaro, Ogun State
mudasiru.hammed@federalpolyilaro.edu.ng; jumoke.soyemi@federalpolyilaro.edu.ng

Corresponding: jumoke.soyemi@federalpolyilaro.edu.ng

ABSTRACT

Social media has significantly grown as a preferred medium of communication for individuals and groups. It is also a tool for disseminating information to the public. Social media offers several advantages, most especially contacting millions of people at the same time. Social media attacks such as phishing evolved as a result of messaging and disseminating capabilities of social media network sites. This challenge of continuous attacks has attracted the attention of many researchers to propose different techniques to detect and classify both phishing attacks and legitimate messages. Studies in the literature revealed that some of the models proposed for phishing attacks may not be perfect to stop adversaries and, there are still different phishing attacks that hindered the robust nature of social media. This study proposed associative rule mining augmented with the Firefly algorithm which attained a high degree of accuracy in both phishing attack messages and legitimate messages.

KEYWORDS:

Social media, Phishing attacks, Classification, Associative Rule, Apriori algorithm, Firefly algorithm.

1. Introduction

Social media came with rapid development and adoption as it has been adopted in various ways of modern life and it has a deep influence on interpersonal communication and relationship (Subramanian, 2017). People in their millions use social media network sites to connect, meet, and share information (DiMicco, *et. al.*, 2008). Interaction is required to fulfil the social needs of people and social media provide such a platform to communicate through digital and mobile technologies (Subramanian, 2017). The new technologies, which include smart devices and internet connectivity, have resulted in the increase of internet-based services where some of these services are essential for day-to-day activities (Mohammad, *et. al.*, 2013). Social media has been a tool for an online marketer to catch the attention of consumers' attention amidjumble advertisements (Taylor, 2010). It takes on diverse forms like Internet forums, weblogs, social blogs, microblogging, wikis, podcasts, photographs or pictures, videos, ratings and social bookmarking. The world is being revolutionized by social media making media such as Facebook, Twitter, Orkut, myspace, skype and so on, to be used extensively for communication purposes. The communication forms can be with an individual or a group of individuals (Boyd & Ellison, 2007; Trisha, 2012).Online social networks have also become a primary alternative for communication locally and globally over the internet, e-mail and even mobile phones.

The most popular websites offering social networking currently are MySpace (started in 2003), LinkedIn (started in 2003), Facebook (started in 2004) and Twitter (started in 2006) (Trisha, 2012).The social networks have independently and collectively gathered a huge number of active users, this has made it easy for adversaries to target large numbers of audiences on a single platform. (Boyd & Ellison, 2007). Nonetheless, the unlimited access and usage of social media and other web-based services offer potential for cyber-attacks because of non-uniform cyberspace regulations or processes (Verma, & Das, 2017). The communication processes and social media in terms of information dissemination characteristics have also added to the underlying notion which has attracted the interest of spammers. There are a number of attacksthat are terrorizing social medial which include phishing, social engineering, brand impersonation, site compromise and data theft, spread of malware etc. Adversaries often engage in the use of these accounts during the reconnaissance phase of social engineering or phishing attack. Social media provides a platform for attackers to impersonate trustworthy persons or get information required forfurther attacks which could be social engineering and phishing.

Phishing is a fraudulent method to acquire or retrieve confidential personal data and other vital information by tricking an anonymous person to have believe that the scammer is a legitimate person who can be trusted (Abad, 2005).This method provides hackers with everything needed to access their targets' personal information or social media account (Aman, *et. al.*, 2021). Often phishing attacks may also be categorized as social attacks and people are being targeted by phishing attacks on a daily or regular basis (Faisal Khan & Rana, 2021). Although some social media platform such as Facebook allow users to keep their images and comments private, a phishing attack is still possible because an attacker, often, targeted users' friends or directly send a friend request to a targeted user to access their information.

Spammers are turning to the fastest-growing communication medium to circumvent traditional security infrastructures that were used to detect and prevent attacks. A number of methods have been proposed by researchers such as classification-based, collaborative filtering-based, behavioural analysis-based, and in some cases, friend-graph analysis etc. (Odukoya, 2017). Other various approaches such as Content-Based, Heuristic-Based, and Rule-Based phishing detection were

discussed in (Zuraiq&Alkasassbeh, 2019). However, the behavioural analysis-based which was proposed for spam messages and other malicious content may not be perfect to stop adversaries because some social media users may not often modify their behaviour to evade detection. Even though some of the studies proposed behavioural analysis, the method failed to integrate all behavior characteristics of spammers into their developed framework. This leads to degraded of some classifiers which were proposed in some existing studies. Minor features of a website URL were extracted to determine if the URL is legitimate or illegitimate in the heuristic approach, thus, the heuristic method is capable of detecting new phishing websites (Mohammad, *et. al.,* 2013).

Machine learning base classifiers are trained using both legitimate and illegitimate websites as training datasets which form a base to detect newly developed phishing websites accurately (Ali, 2017). Browsers such as PhishTank, Safe Browsing, and Smart Screenare used by the browser protection mechanisms to provide listing-based (blacklist and whitelist) methods to block and filter any phishing attacks. But, in whitelist databases, some legitimate websites are missing from the databases which might have been listed as victims, and they have been denied access through browsers. However, in the blacklist-based database, phishing URLs are maintained. That is, the blacklist-based method occasionally fails especially when it encounters zero-day phishing websites and updates were not regularly done (Da Silva, *et. al.*, 2020).The literature also revealed that different models that were used for phishing attacks were not perfect. Some of the models that were examined by this study include neural networks, gated recurrent neural networks, convolution neural networks, deep learning, and machine learning and some studies used a heuristics approach.

Neural networks (NN) are very complex structures made up of artificial neurons capable of mimicking biological nerve cells (neurons). All the neurons in neural networks are capable of influencing each other and hence they may be termed as connected (Faisal Khan, and Rana, 2021). But, determining the number of neurons to be used and how neurons can be connected is one of the major challenges when neural networks are to be implemented. Thus, it leads to over fitting where data is less than the number of neurons connected, or data used is greater than the number of neurons. This study also observed that some of the studies that proposed a better technique failed to propose an algorithm to handle misrepresentation of the data that may occur when classifying the datasets. Also, several classifiers which have been used in spam detection failed to choose the right classifier (Odukoya, 2017).

One of the major challenges of machine learning algorithms for phishing detection system is that the detection accuracy depends on the careful extraction of relevant features in the dataset and the extraction process always requires the effort of domain experts (Tao &Chuan, 2020). However, different techniques have been implemented to cater for the robust nature of social media phishing attacks, although there are still different phishing attacks such as spam messages, fake URLs and other tricky messages that hindered the robust nature of social media. In fact, as long as phishing attacks evolve, there is a need for new methods to detect and classify the attack.

This study used an associative rule mining approach (Apriori algorithm) with simplicity and efficiency to reduce phishing attacks. The set of data used to model the system was manually captured from previous users' data stored in the database. The associative rule mining algorithm is augmented with the Firefly algorithm to detect and classify spam messages in online social networks. Thus, the technique enhances security and guarantees users' safety against any form of threats that emanate from phishing attacks.

2. Methods

Apriori is an algorithm used for mining and learning association rules over transactional databases for frequent datasets. It advances to identify the repeated individual items in the database and extend them to larger datasets as long as those data sets appear adequately in the database. The recurrent datasets determined by apriori can be used to establish association rules which emphasize universal trends in the database. This study used an Apriori algorithm to detect and classify any method used by hackers which could provide them with everything needed to access their targets' personal information or social media accounts of individuals.The study assumed that all previous messages that have been received by the user since the creation of social media account with their sources were stored in the social media database. Instances of action taken (that is,read before deletion, deletion without reading, read with responses and messages that have previously caused harm to any of the social media service provider's accounts) at the point of accessing the messages were also stored in the database. The instances of action taken by the user were used to model patterns for each social media account user. Thus, the study, constructed two patterns for each account using either a legitimate pattern or an illegitimate pattern. When frequent pattern mining is applied to the stored data of a particular user, apriori returns a set of instances of action which have previously been taken by the user to computing support and confidence.The apriori first determine support for instance of every message and their source in the database. It will continue to compare the instances of each message and their source with any newly arrived message.

The association rule $X \rightarrow Y$ is interpreted as a set of messages with their sources that satisfies the conditions in X and are also likely to satisfy the conditions in Y. That is, messages with sources that satisfy either read before deletion or read and respond to the messages were classified as legitimate in newly arrived messages. But, messages with their sources that satisfy either deletion without reading and messages that were harmful to the account previously were classified as illegitimate in newly arrived messages. Finding the relationship between previously received messages and newly arrived messages predicted whether the newly arrived messages are legitimate or illegitimate. Both equations 1 and 2 were used to compute the association rule for support and confidence as represented in Figure 1 while system pseudocode depicts system processes.

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \tag{1}$$

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{2}$$

System Pseudocode

*Step 1:*A series of messages and the sources (data) were taken from the user's account.

Step 2: Process the data for finding the frequent item set in social media accounts.

Step 3: The system groups the number of previous messages and their sources into different groups.

Step 4: The system calculates the number of previous messages and their sources that were legitimate and those that were illegitimate.

Step 5: The system classifies the group of messages and their sources into legitimate and illegitimate patterns for each user based on previous messages received.

*Step 6:*Apriori groups the user's data by considering the previous messages' content and their sources.

Step 7: When a new message arrives in the user's inbox, the similarity of both new messages' content and their sources and instances will be compared with the user's previous data stored in the database.

Step 8: The classification, based on whether the content of the messages and their sources is more similar with either legitimate or illegitimate patterns.

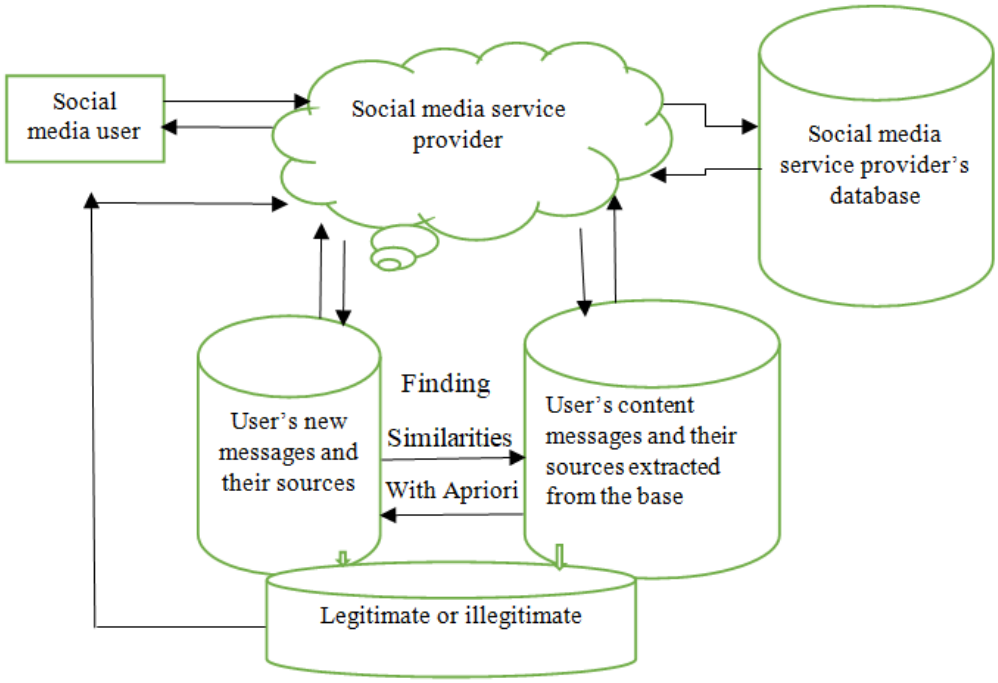

Figure 1: Proposed System Architecture

Time Complexity for computing support

There must be considerable time taken for an Apriori algorithm to compute support for a given system to avoid unnecessary processing which delays the computational time of the algorithm. This study used the firefly algorithm to optimize the computational time of the Apriori algorithm. The techniques enhanced the processing time of the algorithm when computing support for finding the relationship between the previous messages stored in the social media database and the newly received messages. The stored messages (data) in the database are divided into several groups. In each group, there is a result of the computation. The fitness value of the messages (data) outlines the hierarchy, the message (data) with the best fitness will be the result of support computation in each group. The results of groups are appropriately combined to make decisions. This process continues to update whenever new messages arrive. The firefly strategies are depicted in algorithm 1.

Algorithm 2

Step1: Generate a random solution set, $\{x_1, x_2,...,x_n\}$.

Step2: Compute intensity for each solution member, $\{r_1, r_2,...,r_n\}$.

Step3: Move each firefly i towards other brighter fireflies, and if there is no other brighter firefly, move it randomly.

Step4: Update the solution set.

Step5: Terminate if a termination criterion is fulfilled otherwise go back to step 2.

Mathematically, each firefly has a location $X = (x_1, x_2, ...x_n)$ in a n-dimensional space and a light intensity $I(x)$ or attractiveness $\beta(x)$ which are proportional to the objective function $f(x)$.

Therefore, the attractiveness of a firefly is modelled with Equation 3

$$\beta = \beta_0 e^{-kr2} \tag{3}$$

The r isdefined as the distance between any two fireflies i and j at x_i and x_j respectively which is modelled in Cartesian distance in equation 4

$$r_{ij} = \left\| x_i - x_j \right\| \tag{4}$$

For any given two fireflies x_i and x_j, the movement of the firefly i is attracted to another more attractive (brighter) firefly j is modelled in equation 5

$$x_i^{t+1} = x_i^t + \beta_0 e_{ij}^{-kr2}(x_j^t - x_i^t) + \alpha(Rand - \frac{1}{2}) \tag{5}$$

The study hoped that the approach used gained a better performance and optimal solutions to phishing attacks with the minimum effort and time.

Similarity Measurement

The study used the Jaccard similarity coefficient for measuring the similarity of user's messages. Only the instances of shared ratings between the user's new messages and the previous messages were taken into account when calculating the similarity. The similarity technique compares both new messages' content and their sources and instances will be compared with the user's previous data stored in the database. The similarity technique enhances the system's performance for detecting any changes in the user's new messages and the previous messages in the social media database. The technique is mathematically modelled in equation 6.

$$JCS(x, y) = \frac{|I_x| \cap |I_y|}{|I_x| \cup |I_y|} \tag{6}$$

4. Results and Discussion

Every user has an account that was created in the social network provider's platform and all user data such as names, user ID, Passwords, phone numbers, and photographs are stored in the social network service provider's database. All messages and sources that enter the user's account including the instances of action taken by the user when the messages were received are also stored in the database. Figure 2 depicts the graphical User Interface (GUI) of social media service providers where every user who wishes to have an account will create an account.

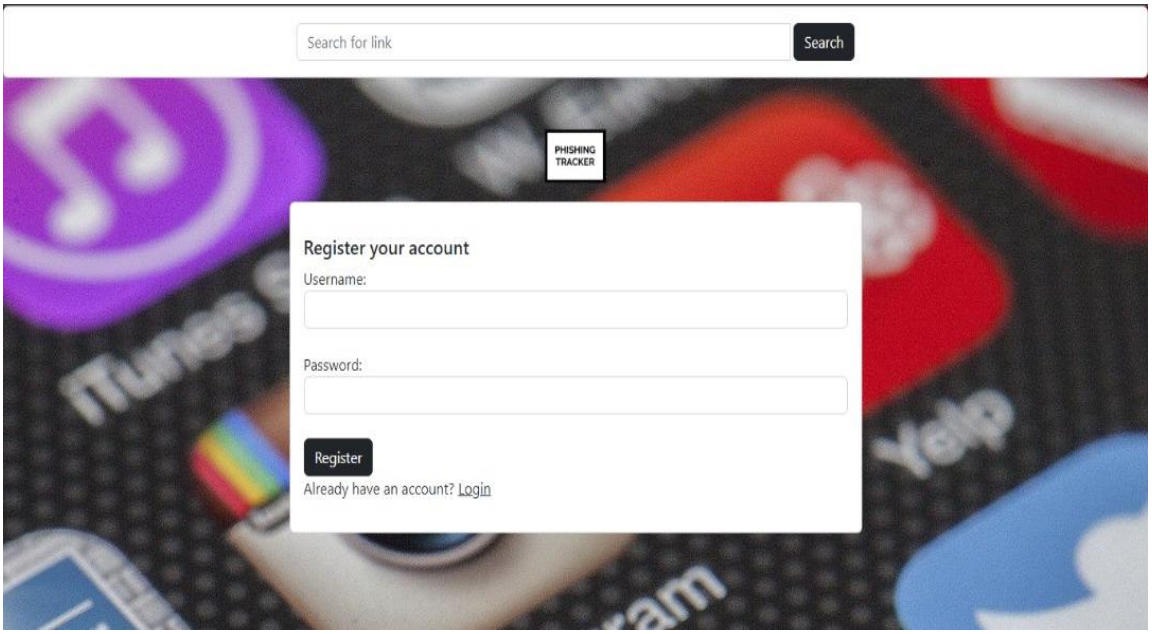

Figure 2: Registration Page

Each time a particular user wants to login into his/her social media account, the user's account details with the content of all previously received messages that were stored will be quickly extracted from the social media service provider's database to compare them with the newly received messages and their sources. The instances of previous actions taken by the user which were stored in the database will also be extracted to validate the new message. If there are similarities, the message will be classified as legitimate, and the user will successfully receive the message. But the user will not receive the message if there are no similarities, and the message has been classified as illegitimate message. Again, any message that was harmful to any of the accounts in the database will also be classified as illegitimate because it has been listed among the malicious messages. A message that the user did not read before it was deleted because the content looks suspicious, is also listed among the malicious messages. Each time these types of messages enter the user's account they will be classified as illegitimate messages. Only messages that the user received and responded to the sender will be classified as legitimate messages. All data used in the study were users' data captured from a database of social media service providers. There was no misrepresentation of datasets which may lead to misclassification. The firefly optimized the performance of the Apriori algorithm to quickly extract users' data and find the relationship between newly received messages and all messages stored in the database. This study achieved a 99.3% detection accuracy because of its classification precision, the same as the study done by Tao &Chuan, (2020). However, this research recorded a better performance in terms of quick response time in the extraction of user data compared to other machine learning algorithms reviewed.

The study extracted 80,000 datasets from a Google phishing tank to test the accuracy of the Apriori algorithm augmented with fireflies. The set of data in Figure 2 are set of attributes used in the study to fill in missing values after the data cleaning process, it was also used to classify the data into legitimate and phishing.

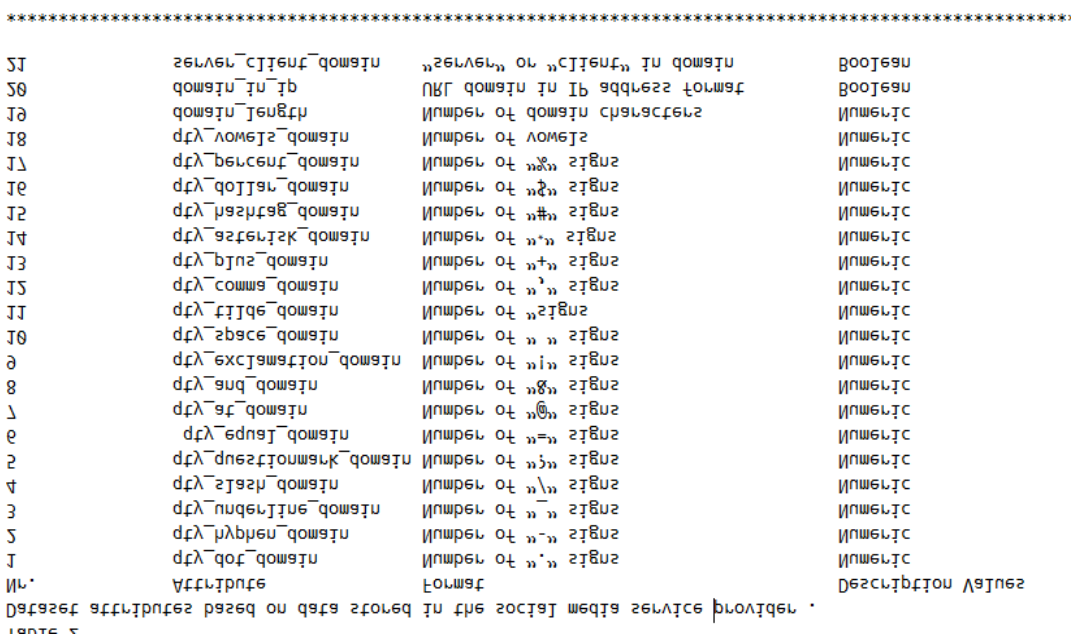

Figure 2: Dataset Attributes

The proposed system achieved a better performance in terms ofquick extraction of user's data from the social media service provider database and finding the relationship between newly received messages and user's data previously stored in the database. The data in Table 2 were captured from the system implementation which shows the system response time to find the relationship between newly arrived data and data stored in the database to classify whether legitimate or illegitimate. The study found the standard deviation of the system response time using QM for Windows as it is shown in Figure 3.

Table 2: System response time for computing classification

phis_messg_in_database	time_response	domain_spf
0	0.207316	0
0	0.207316	-1
0	0.207316	0
0	0.207316	0
0	0.207316	0
0	0.207316	-1
0	0.207316	0
0	0.207316	0
0	0.207316	0
0	0.207316	1
0	1.799702	0
0	0.381044	1
0	1.024603	-1
0	0.127553	-1
0	0.310562	0
0	0.372825	1
0	1.188645	0
0	0.535979	0
1	0.855594	-1
0	0.327323	0
0	0.804266	0
0	0.400331	0

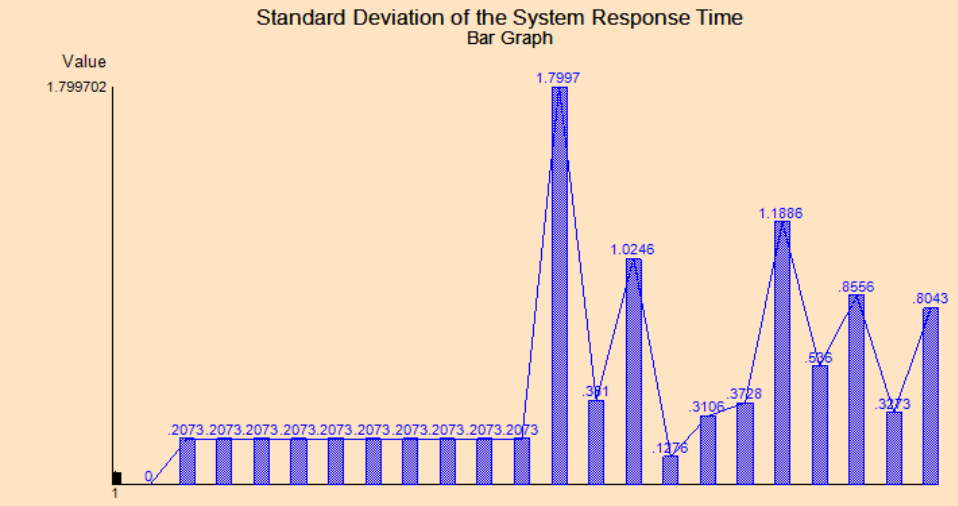

Figure 3: **Standard Deviation of System Response Time**

Conclusion

The literature revealed that one of the major challenges of machine learning algorithms for the implementation of phishing detection systems is that detection accuracy depends on the careful extraction of relevant features in the dataset. In this study, all data used in the study were users' data captured from the database of social media service providers. The study used the firefly algorithm to enhance the performance of rule mining (Apriori). Thus, the optimization technique helped anApriori to quickly extract data from the social media service provider's database to find the relationship between stored data and newly arrived data. The system achieved a high degree of accuracy in the classification of both phishing attacks and legitimate messages.

References

1. Subramanian, K.R., (2017). Influence of social media in Interpersonal Communication. *International Journal of Scientific Progress and Research (IJSPR). 38(109), 70-75*

2. DiMicco, J., Millen, D. R., Geyer, W., Dugan, C., Brownholtz, B., & Muller, M. (2008, November). Motivations for social networking at work. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 711-720).

3. Mohammad, R., Thabtah, F. &Mccluskey, T. (2013). Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications*, 25, 443–458.

4. Taylor, C. (2010). Integrated marketing communications in 2010 and beyond. *International Journal of Advertising*, 29 (2), 161-164.

5. Boyd, D.M. & Ellison, N.B. (2007). Social network sites: Definition, history, and scholarship. *Journal of CMC*, 13(1), 210-230.

6. Trisha, D. B., (2012). Effectiveness of social media as a tool of communication and its potential for technology enabled connections: A micro-level study. *International Journal of Scientific and Research Publications, 2(5),1-10.*

7. Verma, R. & Das, A. (2017). What's in a URL: Fast feature extraction and malicious URL detection? *In Proceedings of the 3rd Association for Computing Machinery (ACM) on International Workshop on Security And Privacy Analytics, IWSPA '17, (New York, NY, USA), pp 55–63.*

8. Abad, C., (2005). The economy of phishing: A survey of the operations of the phishing market. *Retrieved May, 29 2023 from: https://firstmonday.org/ojs/index.php/fm/article/view/1272/1192*

9. Aman, R. Tarun, K. & Dhan (2019). Phish-Defense: Phishing Detection Using Deep Recurrent Neural Networks. ACM Cryptography and security, Cornel University, Retrieved May 9, 2023 from: https://doi.org/10.48550/arXrv.2110.13424

10. Faisal-Khana M. D. &Ranab, B. L. (2021). Detection of Phishing Websites Using Deep Learning Techniques. *Turkish Journal of Computer and Mathematics Education.* 12(10), 3880-3892

11.Odukoya, O.H. (2017). A social network spam detection model. *International Journal of Scientific Engineering and Research*, *8*, 11

12. Zuraiq, A. A., &Alkasassbeh, M. (2019, October). Phishing detection approaches. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)* (pp. 1-6). IEEE.

13. Ali, W. (2017). Phishing website detection based on supervised machine learning with wrapper features selection. *International Journal of Advanced Computer Science and Applications*, *8(9).*

14. Da-Silva, C. M. R., Feitosa, E. L. & Garcia, V. C. (2020). Heuristic-based strategy for Phishing prediction: A survey of URL-based approach. *Computational Security*. 88(pp. 101613).

15. Tao, F., &Chuan, Y., (2020). Visualizing and Interpreting RNN Models in URL-based Phishing Detection. Session 1: Assessment and Detection of Security Threats, *SACMAT '20, June 10–12, 2020, Barcelona, Spain*