https://doi.org/10.46335/IJIES.2023.8.6.10 **e-ISSN: 2456-3463**
Vol. 8, No. 6, 2023, PP. 40-43

International Journal of Innovations in Engineering and Science, www.ijies.net

Detecting Phishing Website Using Machine Learning

Prof. Monika Ingole [1],Achal Meshram[2] , Swati Thengane[3], Asmita Raut[4], Prachi Patil [5]

[1]Assistant Professor, [2,3,4,5] Students,
Department of Information Technology & Engineering
Wainganga College of engineering and management, Nagpur, India, 441108

achalm601@gmail.com

***Received on**: 03 April, 2023* ***Revised on**: 25 April,2023* ***Published on**: 27 April,2023*

Abstract- *Trying to gather personal information through deceptive ways is becoming more common nowadays. In order to assist the user to be aware of the access to such websites, the implemented system notifies the user through email and also pop-up, when trying to access a phishing site. This paper proposes an approach of phishing detection system to detect blacklisted URL also known as phishing websites, so that individual can be alerted while browsing or accessing a particular website. Therefore, it can be utilized for identification and authentication and become a legitimate tool to prevent an individual from getting tricked.*

Keywords— Blacklisted, phishing, Agile Unified Process (AUP), alert, pop-up notification, Email notification, Machine Learning

I -INTRODUCTION

Phishing can be defined as impersonating a valid site to trick users by stealing their personal data comprising usernames, passwords, accounts numbers, national insurance numbers, etc. Phishing frauds might be the most widespread cybercrime used today. There are countless domains where phishing attack can occur like online payment sector, webmail, and financial institution, file hosting or cloud storage and many others. The webmail and online payment sector was embattled by phishing more than in any other industry sector. Phishing can be done through email phishing scams and spear phishing hence user should be aware of the consequences and should not give their 100 percent trust on common security application. Machine Learning is one of the efficient techniques to detect phishing as it removes drawback of existing approach.

The objectives which is the most vital thing in proposed project is to verity the validity of the website by capturing blacklisted URLs. To notify the user on blacklisted website through pop-up while they are trying to access and to notify the user on blacklisted website through email while they are trying to access. This proposed project will allow administrator to add blacklisted URL's in order to alert user during their inquiry.

The two scope of project, which is well known as user scope and system scope. User has some responsibility towards the system. The system includes a few standards and policies that requires to be obliged in order to comply the system. The user can be notified if blacklisted website is being accessed. The admin can capture the blacklisted URL's to alert user.

II-METHEDOLOGY

The proposed algorithm depends on the ML process and automated real-time phishing detection. By using these features phishing URLs are extracted. For a machine learning classification, the extracted features are used to detect phishing websites in real-time. After so much analysis and the survey was done which is due to comparing various classification algorithms[6]. The Waikato Environment for Knowledge Analysis(WEKA) is helping to determine the performance and correctness of every algorithm. To improve efficiency by using ELM as per the classification algorithm and RStudio tool helps us for better analysis[6]. The summary of the proposed method is exposed in

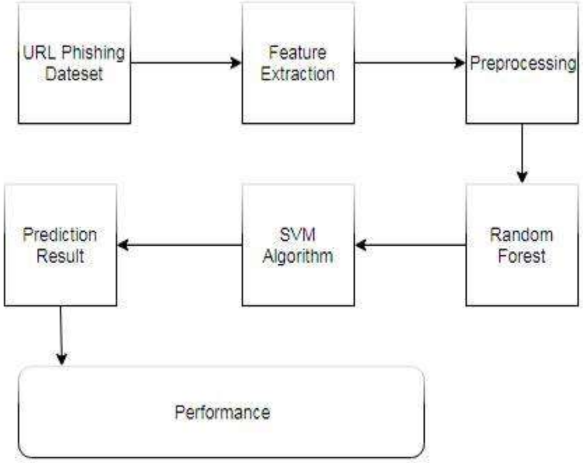

Fig1 . Proposed Architecture

A. Extreme Learning Machine

It is a feed-forward Artificial Neural Network(ANN) and it also has a single hidden layer. ANN is an important tool used in Machine Learning. Neural Network contains input and output layers and it is also hidden layers. Extreme Learning Machine algorithm reduces the time-consuming training speed and over-fitting issues. It depends on its learning process and empirical threat minimization theory. The ELM avoids local minimization and multiple iterations. In the ELM process is different from ANN because it renews its parameters and input weights are accidentally chosen while output weight is calculated analytically. According to generate the cells in the hidden layer of ELM.

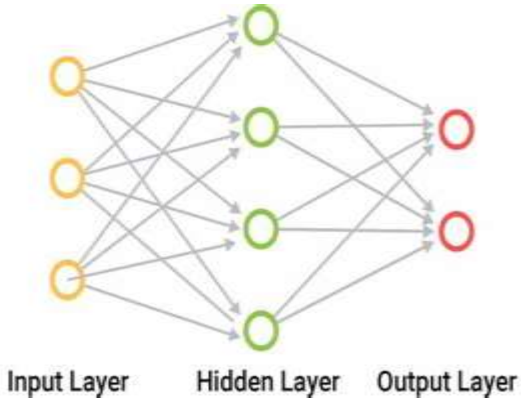

Fig2. An ANN model with a single hidden layer[25]

The proposed algorithm depends on the ML process and automated real-time phishing detection. By using these features phishing URLs are extracted. For a machine learning classification, the extracted features are used to detect phishing websites in real-time. After so much analysis and the survey was done which is due to comparing various classification algorithms[6]. The Waikato Environment for Knowledge Analysis(WEKA) is helping to determine the performance and correctness of every algorithm. To improve efficiency by using ELM as per the classification algorithm and RStudio tool helps us for better analysis[6]. The summary of the proposed method is exposed in Figure 4. Fig 3. Structure of the proposed work A. Extreme Learning Machine It is a feed-forward Artificial Neural Network(ANN) and it also has a single hidden layer. ANN is an important tool used in Machine Learning. Neural Network contains input and output layers and it is also hidden layers. Extreme Learning Machine algorithm reduces the time-consuming training speed and over-fitting issues. It depends on its learning process and empirical threat minimization theory. The ELM avoids local minimization and multiple iterations. In the ELM process is different from ANN because it renews its parameters and input weights are accidentally chosen while output weight is calculated analytically. According to generate the cells in the hidden layer of ELM.

A. Support Vector Machine

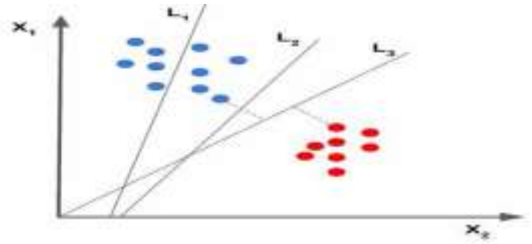

Fig3-Comparative Study Between SVM and Supervised Learning

Support Vector Machine follows supervised learning. SVM is helped to avoid the use of an Internet from a victim of phisher do not loss personal and financial information. Identify the right

B. Anti-Phishing Approach

This approach is a knowledge base service that helps to prevent illegitimate access to secure and sensitive information. Anti-phishing services protect a different type of data in other ways beyond the variety of stages. Ant phishing software comprises computer programs that try to determine phishing content.

III –CONCLUSION

After reviewing and researching for appropriate monitoring tools, proposed system has been identified and chosen to address the complexity of monitoring requirement for current situation. This software is designed to show awareness of the extensive level of its functionality, features that can be displayed in the monitoring era. The system fosters many features in comparison of other software. Its unique features such as capturing blacklisted URL's from the browser directly to verify the validity of the website, notifying user on blacklisted websites while they are trying to access through pop-up, and also notifying through email. This system will assist user to be alert when they are trying to access a blacklisted website. In conclusion, this system is designed for resources are used as intended, prevents from valuable information from leaks out, produce better control mechanism and alerts the user to keep their private information safe. Like any other programs, there are improvements which could be made into this system. Based on the capabilities which the current system processes, text message integration would a great recommendation that could be made to improve the program in the future. The future version of the application could also implement an option to directly notify the blacklisted website with a text message. The program could be made to access the list as an attachment. This text message integration function would further the usability of the user.

IV-ACKNOLEDGEMENT

Authors are grateful to the School of Graduate Studies and the Faculty of Information Sciences and Engineering, Management and Science University, Malaysia for their support.

REFFERENCE

[1] *Srushti Patil, and Sudhir Dhage, "A Methodical Overview On Phishing Detection Along With An Organized Way To Construct an AntiPhishing Framework", 2019 5th International Conference On Advanced Computing & Communication System(ICACCS), pp. 1-6.*

[2] *Huaping Yuan, Xu Chen, Yukun Li, Zhenguo Yang and Wenyin Liu, "Detecting Phishing Websites and Targets Based On URLs and Webpage Links", 2018 24th International Conference on Pattern Recognition(ICPR) Beijing, China, August 20-24, 2018.*

[3] *Vaibhav Patil, Pritesh Thakkar, Cjirag Shah, Tushar Bhat, Prof. S. P.Godse, "Detection and Prevention of Phishing Websites using Machine Learning Approach", 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.*

[4] *Mustafa AYDIN and Nazifa BAYKAL, "Feature Extraction and Classification Phishing Websites Based on URL", 2015.*

[5] *C. Emilin Shyni, Anesh D Sundar and G. S. Edwin Ebby. "Phishing Detection In Websites Using Parse Tree Validation", 2018 Recent Advances On Engineering , Technology and Computational Sciences(RAETCS).*

[6] *Shraddha Parekh, Dhwanil Parikh, Srushti Kotak and Prof. Smita Sankhi, "A New Method For Detection of Phishing Websites: URL Detection", Proceedings of the 2nd International Conference on Inventice Communicastion and Computational Technologies(ICICCT 2018) IEEE Explorer Complaint-Part Number: CFP18BAC-ART:ISBN: 978-1- 5386-1974-2*

[7] *Anu Vazhayil, Vinaya Kumar R and Soman KP, "Comparative Srudy Of The Detcetion Of Malicious URLs Using Shallow and Deep Netoworks ", 9th ICCCNT2018 July 10-12,2018,IISC,Bangluru,India.*

[8] *Martyn Weedon, Dimitris Tsaptsinos and James*

Denholm-Price, "Random Forest Explorations for URL Classification", 2017

[9] *Mehek Thakar, Mihir Parikh and Preetika Shetty. "Detecting Phishing Websites Using Data Mininh", Proceeding of the Second International Conference On Electronics, Communication and AEROSPACE Technology(ICECA 2018).*

[10] *Chuan Pham, Luong A.T. Nguyen, Nguyenh. Tran, Eui-nam Huh and Choong Seon Hong, "Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks", DOI 10.1109/TNSM. 2018. 2831197, IEEE Transactions on Netowork and Service Management.*

[11] *Mohhamed Alqahtani." Phishing Websites Classification Using Association Classification(ATWCAC)", 2019 International Conference On Computer and Information Sciences(ICCIS).*

[12] *Varsharani Ramdas Hawanna, V. Y. Kulakarni and R.A. Rane, "A Novel Algorithm to Detect Phishing URL's", 978-1-5090-2080- 5/16/2016 IEEE.*

[13] *Xueni Li, Guanggang Geng, Zhiwei Yan, Yong Chen and Xiaodong Leee, "Phishing Detection Based on Newly Registered Domains", 2016 IEEE International Conference On Big Data(Big Data).*

[14] *Ebubekir Buber, Onder Demir and Ozgur Koray Sahingoz, "Feature Selections For The Machine Learning Based Detection of Phishing Websites", 978-1-5386-1880-6/17/2017 IEEE.*

[15] *Amani Alswailem, Bashayr Alabdullah and Norah Alrumayh , "Detecting Phishing Websites Using Machine Learning", 978-1- 7281.*