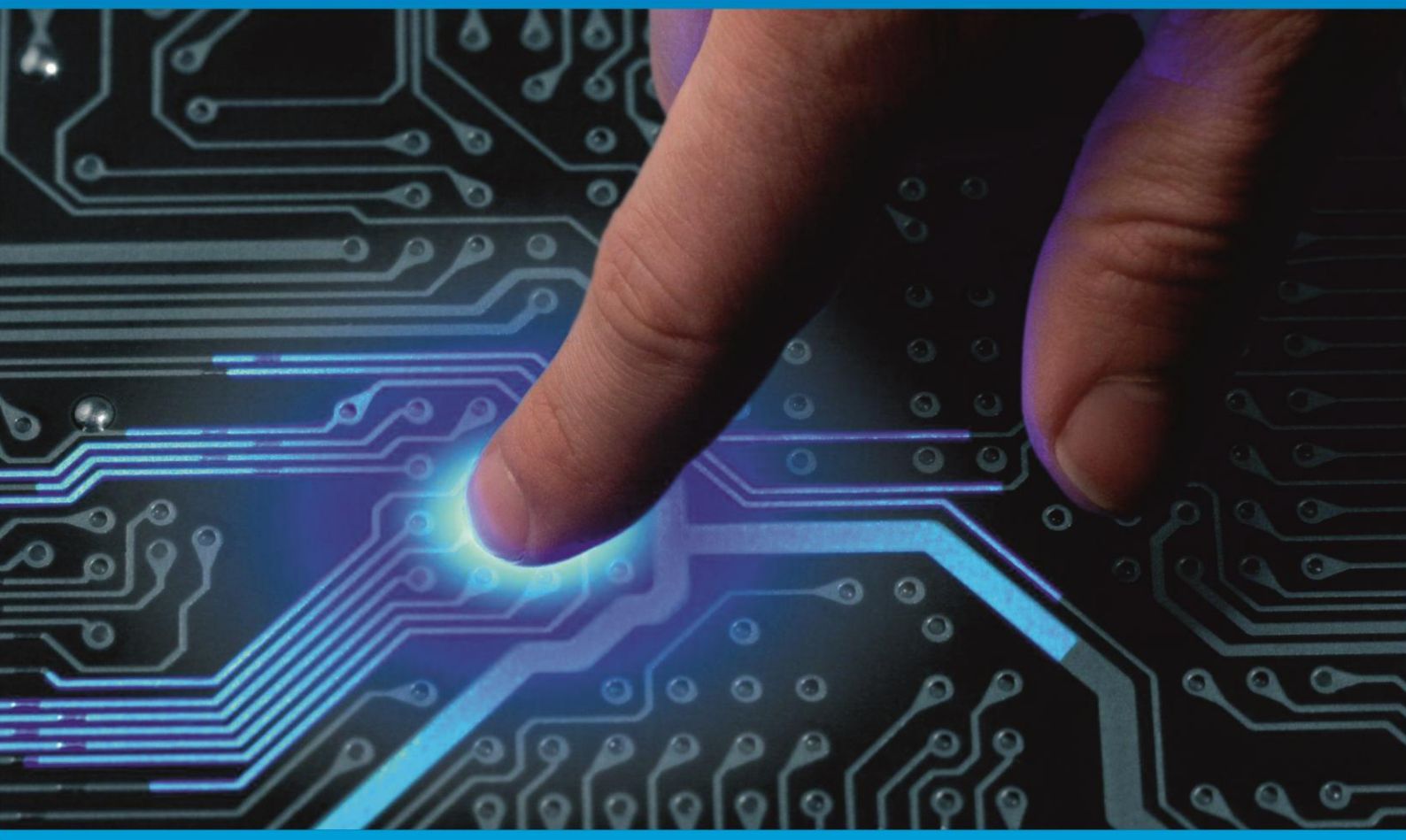
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905093 |

A Review on Detection of Phishing Attacks on Websites Using Deep Learning

Deepali Vaidya, Poonam Dholi

PG Student, Dept. of Computer Engineering, Matoshri College of Engineering and Research Centre, Eklahare, Nashik, India

Assistant Professor, Dept. of Computer Engineering, Matoshri College of Engineering and Research Centre, Eklahare, Nashik, India

ABSTRACT: Phishing sites make to get confidential data of the user that shows the variety of criminal acts through the net and its one among the especially concerns toward numerous areas including e-managing an account and retailing. Phishing site detection is really hit or miss and element issue including numerous components and criteria that don't seem to be stable. This paper proposed an intelligent model for detecting phishing websites pages supported Deep Learning. Forms of web contents are different in terms of their features. Hence, we must use a particular web page features set to forestall phishing attacks. This paper proposed a model supported Deep Learning techniques to detect phishing web pages. This paper suggested some new rules to possess efficient feature classification.

KEYWORDS: Security; Phishing; Deep Learning;

I. INTRODUCTION

Technology is growing rapidly day-by-day and with this rapid growing technology internet has become a necessary part of human's daily activities. Use of internet has grown due to the rapid growth because of ascension technology and intensive use of digital systems and thus data security has gained great importance. The first objective of maintaining security in information technologies is to confirm that necessary precautions are taken against threats and dangers likely to be faced by users during the utilization of those technologies. Phishing is that the fraudulent try to obtain sensitive information like usernames, passwords and MasterCard details by disguising as a trustworthy entity in a transmission. Typically applied by email spoofing or instant messaging, it often directs users to enter personal information at a fake website, the design and feel of which is a twin of the legitimate site. Information security threats are seen and developed through time along development within the internet and information systems. The impact is that the intrusion of knowledge security through the compromise of personal data, and also the victims may lose money or other forms of assets at the tip. Internet users are affected from different types of cyber threats such as like private information loss, fraud, and financial damages. Hence, using of the net may suspect for home and official environments. Identify and defend against privacy leakage efficient analytical tools are required for users to scale back security threats. Effective systems which will improve self-intervention must be formed using artificial intelligence-based information security management system at the time of an attack. Phishing is an Internet-based attack that seduces end users to go to fake websites and provides away personal information.

II. RELATED WORK

With the event of Information and Communication Technology, various varieties of information security threats may be seen. These threats are important within the prevention of damage to person or institution to guard data on computer systems. Studies on various phishing detection methods are seen when the literature is reviewed. In these studies, it is observed that ML is challenging techniques may be used. Santhana Lakshmi and Vijaya they used techniques of Machine learning to verify supervised learning algorithms and modeling the prediction task that Multilayer Perceptron. Decision tree and Naive Bayes classifications were used for observing technique for web phishing detection. It can detect As compared to other learning algorithms the choice tree classifier is more accurate [2]. Zou Futai, Pei Bei and Panli proposed Uses Graph Mt some potential phishing which can't be detected by URL analysis. It uses contact of the user and website. To induce dataset from the real traffic of an oversized ISP. After anonymizing these data, they need cleansing the dataset. Every record that includes eight fields: User node number (AD), Visiting URL (URL), User Agent (UA), User SRC IP (SRC-IP) access time (TS), Reference URL (REF), access server IP (DSTIP), User cookie (cookie)[3]. Kaytan and Hanbay proposed determining phishing websites supported neural network. Around 30 inputs attributes, and output attribute 1 is used for that experiment. The values 1, 0, and -1 were used for input attributes and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905093 |

therefore the values 1 and -1 were used for output attribute. To evaluate the system performance, 5-fold cross-validation method was used. The simplest classification accuracy has been measured as 92.45%. And hence the average accuracy has been measured as 90.61% [1]. Yasin Sonmez, Turker Tuncer perform Extreme Learning Machine (ELM) for 30 feature. That has phishing websites in the database of machine learning repository. They compare ELM with SVM, Naives Bayes. These are other methods of machine learning[6].X. Chen, find the impact of phishing attacks as consider to risk levels and potential market value that downs which is losses experienced by the target companies. It absolutely was analyzed 1030 phishing alerts released on a public database and financial data related to the targeted firms employing a hybrid method. This is the prediction that the attack was survive around 89% accuracy using supervised classification and text phrase extraction It's been identified some important textual and financial variables within the study. Impact the severity of the attacks and potential loss has been investigated [7].

Giovanni Armano and Samuel Marchal [4] proposed a system which is based upon minimum enclosing ball support vector machine (BVM) to find out phishing website. It has been aimed toward achieving high speed and high accuracy to detect the phishing website. Studies were exhausted order to reinforce the integrity of the feature vectors. Firstly, an analysis of the topology structure of the website was performed consistently with Document Object Model (DOM) tree. Then, the net crawler was accustomed extract 12 topological features of the web site. Finally, the BVM classifier detects the feature vectors. When the proposed method is getting compared with DVM then observed that the proposed method has relatively high precision of detecting. Additionally it absolutely was observed that the proposed method complements the disadvantage of slow speed of convergence on large-scale data. It is been shown that the proposed method has better performance than SVM within the experimental results. Finally, the proposed systems accuracy and validity has been evaluated.

Gowtham and Krishnamurthi[5] studied the characteristics of legitimate and phishing web content thorough. Heuristics were proposed to extract 15 features from similar kinds of web pages supported the analysis. The heuristic results which were proposed are fed as an input to a trained machine learning algorithm to find out phishing websites. Before the applying the heuristics to the net pages, two preliminary screening modules were employed in the system. By the preapproved site identifier that is the primary module, sites were checked against a non-public white-list maintained by the user. By the login form finder that's the second module, web pages were classified as legitimate when no login forms present. Unnecessary computation within the system was reduced by helping the used modules. Additionally, the speed of false positives without compromising on the false negatives was reduced by helping the used modules. To detect new output algorithms uses historical data as a input. The extreme module websites having 0.4% false positive rate and 99.8% precision. It's been shown that the proposed method is efficient for safeguarding users from online identity attacks. The primary topic is about the computation of required thresholds to describe the three email groups. And also the second topic is that the interpretations of the cost-sensitive characteristics of spam filtering. They calculate the decision-theoretic rough set model continue which are based on thresholds.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905093 |

III. SYSTEM ARCHITECTURE

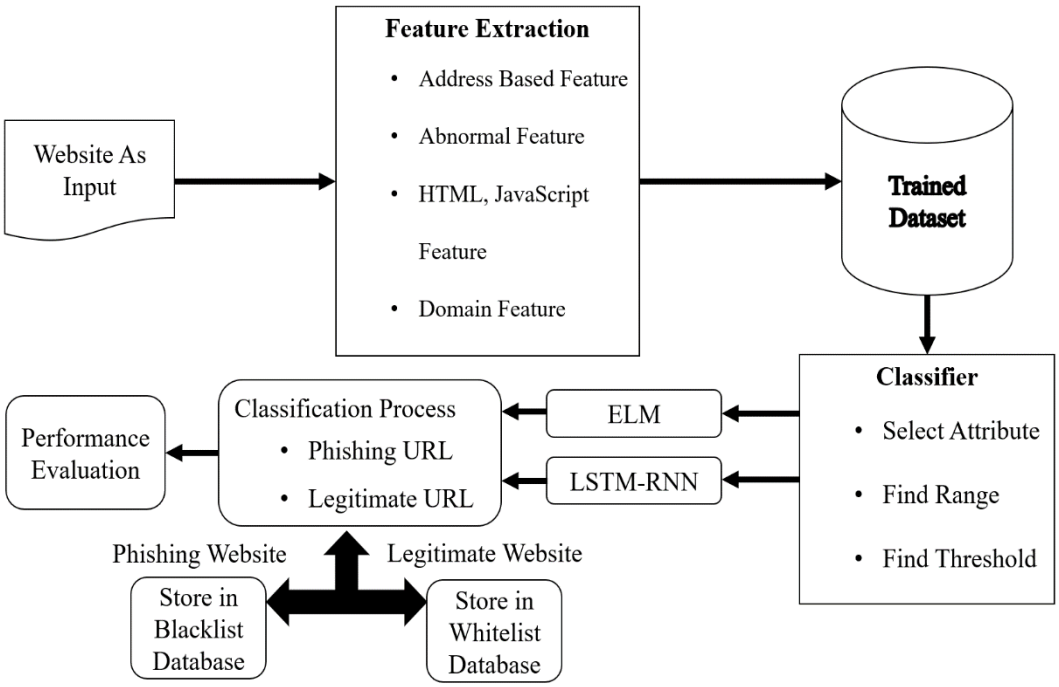


Fig 1: System Architecture

IV. PROPOSED ALGORITHM

A. *Extreme Deep Learning:*

Deep Learning is such permits software applications to become more a lot of correct at predicting outcomes while not being expressly programmed to try and do therefore. Deep Learning Machine area unit feed forwards networks for regression, classification, clustering, compression, sparse approximation and have feature learning with a one layer or multiple layers of hidden nodes, where the parameters of hidden nodes (not simply the weights connecting inputs to hidden nodes) needn't be tuned. These hidden nodes will be arbitrarily assigned and never updated (i.e. they are random projection however with nonlinear transforms), or will be genetic from their ancestors while not being modified. Guang-Bin Huang gives name "extreme learning machine" (ELM) to such models. Many cases contain the output weights of hidden nodes are learned in a single step, that is consider the amounts to learning a linear model.

B. *LSTM-RNN:*

Long Short Term Memory (LSTM) is a model of the recurrent neural network. Output from the model is again fed as input this is major difference in RNN. Problem of RNN of not predicting on word store for long time was solved with this change. LSTM was proposed by Hochreiter & Schmid Huber. Due to this change RNN became more accurate and was able retain the information for long time. Thus, it can be used for processing, predicting and classifying on the basis of time series data. Here LSTM-RNN is used for detecting phishing websites.

V. CONCLUSION AND FUTURE WORK

Systems are different from data and its processing applications which are made through websites. The entered information is going to process and getting as output. Normally web sites are used in many fields such as education, technical, scientific, business, economy and many more. It also can be used as a tool for attacker to prevent malicious attack. One of the malicious purposes emerges as a phishing attack. A website or a web page can be imitated by phishing attacks and using various methods. Some information such as user's credit card information, identity

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905093 |

information can be obtained with these fake websites or the web pages. The motive to make this paper is that determines the types of attack as cyber treats which is called as phishing.

REFERENCES

[1] Mustafa KAYTAN and Davut HANBAY, "Effective Classification of Phishing Web Pages Based On New Rules By Using Extreme Learning Machines" , Anatolian Journal of Computer Sciences, Vol:2 No: 1, pp: 15-36, 2017

[2] V. Santhana Lakshmi and M. Vijaya, "Efficient prediction of phishing websites using supervised learning algorithms", Procedia Engineering, 30, pp.798-805, 2012.

[3] Zou Futai, Gang Yuxiang, Pei Bei , Pan Li, Li Linsen "Web Phishing Detection Based on Graph Mining " 2nd IEEE International Conference on Computer and Communications 978-1-4673-9026-2116 ©20 16 IEEE

[4] Giovanni Armano, Samuel Marchal, N. Asokan "Real-Time Client-Side Phishing Prevention Add-on" IEEE 36th International Conference on Distributed Computing Systems 1063-6927/16 © 2016 IEEE

[5] Ramesh Gowtham, Ilango Krishnamurthi "A comprehensive and efficacious architecture for detecting phishing webpages" researchgate.net/publication/259118063

[6] Yasin Sonmez, Turker Tuncer, based Huseyin Gokal, Engin Avci, "Phishing Web Sites Features Classification Based On Extreme Learning Machine " IEEE 2018 6th International Symposium on Digital Forensic and Security (ISDFS), DOI: 10.1109/ISDFS.2018.8355342

[7] X. Chen, I. Bose, A. C. M. Leung and C. Guo, "Assessing the severity of phishing attacks: A hybrid data mining approach", Decision Support Systems, 50(4), pp.662-672, 2011

[8] Yi, P., Guan, Y., Zou, F., Yao, Y., Wang, W., & Zhu, T. (2018). Web phishing detection using a deep learning framework. Wireless Communications and Mobile Computing, 2018.

[9] Singh, C. (2020, March). Phishing Website Detection Based on Machine Learning: A Survey. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 398-404). IEEE.

[10] Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2020). Intelligent phishing detection scheme using deep learning algorithms. Journal of Enterprise Information Management.

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details