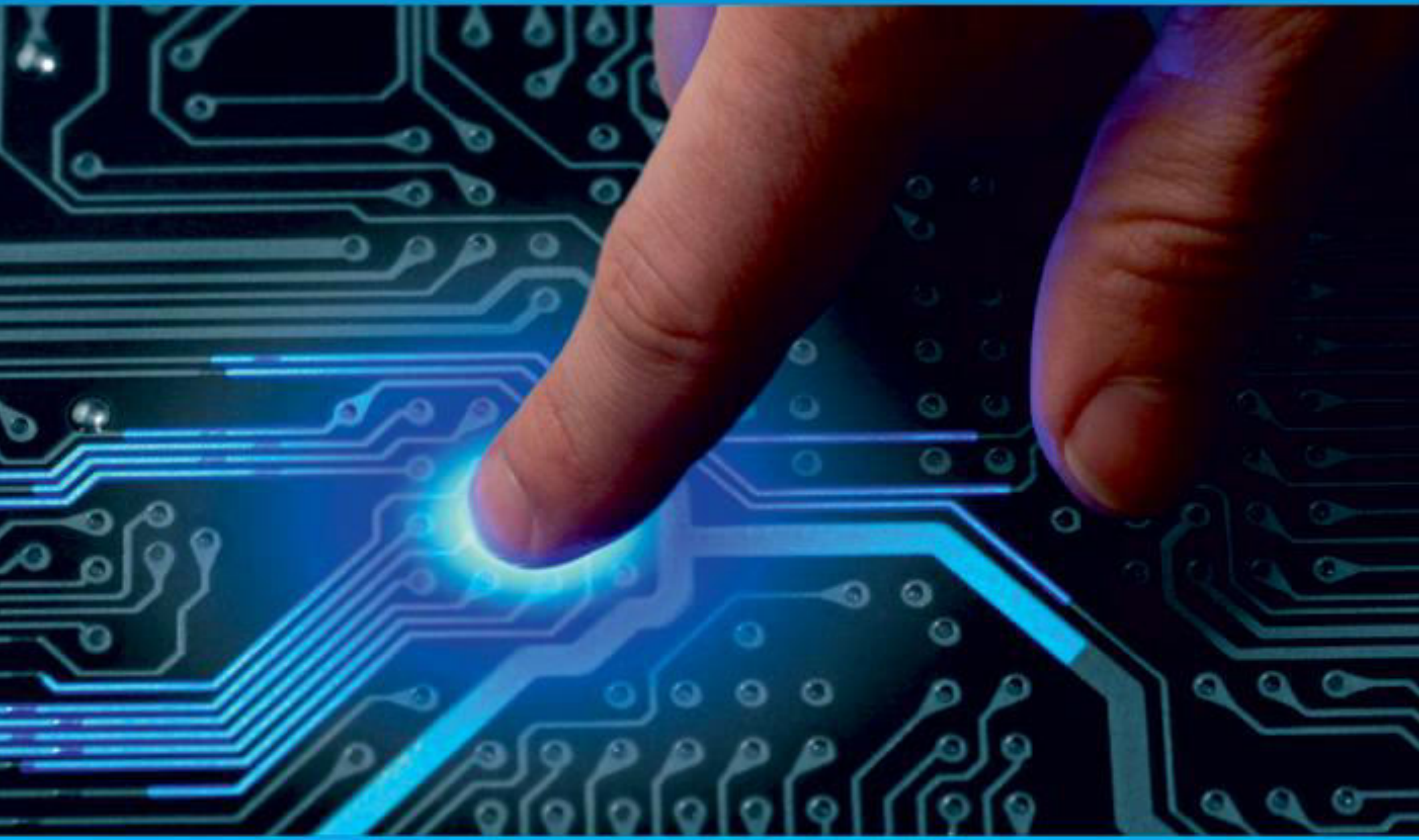
IJIRCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 11, November 2022

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165

Feature Ranking of Phishing Websites based on Machine Learning Techniques

Vidyashree TC, Rekha H

Master of Technology, Department of Computer and Engineering, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India

Assistance Professor, Department of Computer and Engineering, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India

ABSTRACT: In current years, with the growing use of cell devices, there may be a developingfashionto transportnearly all real-international operations to the cyber world. Although this makes smooth our each day lives, it additionally brings many protection breaches because of the namelessshape of the Internet. Used antivirus applications and firewall structures can save youmaximum of the attacks. However, skilled attackers goalat theweak point of the pccustomersthroughseeking to phish them with bogus webpages. These pages imitate a fewfamous banking, social media, e-commerce, etc. web sites to thievea fewtouchystatistics such as, user-ids, passwords, financial institution account, credit score card numbers, etc. Phishing detection is a tough problem, and lots ofuniqueanswers are proposed with inside themarketplace as a blacklist, rule-primarily based totally detection, anomaly-primarily based totally detection, etc. In the literature, it's milesvisible that contemporary works generally tend on usinggadget learning-primarily based totally anomaly detection because of its dynamic shape, specifically for catching the “zero-day” attacks. In this paper, we proposed a gadget learning-primarily based totally phishing detection machinethroughthe use of8unique algorithms to investigate the URLs, and 3unique datasets to examine the outcomes with different works. The experimental outcomes depict that the proposed fashions have an exquisiteoverall performance with achievement.

KEYWORDS:. Machine learning Algorithms

I. INTRODUCTION

In our day by day life, we perform maximum of our paintings on virtual platforms. Using a laptop and the net in lots of regions helps our commercial enterprise and personal life. It lets in us to finish our transaction and operations fast in regions consisting of trade, health, education, communication, banking, aviation, research, engineering, entertainment, and public services. The customers who want to get admission to a neighborhood community were capable of without difficulty connect with the Internet everywhere and each time with the improvement of cellular and Wi-Fi technologies. Although this example gives brilliant convenience, it has discovered extreme deficits in phrases of facts security. Thus, the want for customers in our on-line world to take measures in opposition to viable cyber-assaults has emerged. The approach of achieving goal customers in phishing assaults has constantly expanded because the remaining decade. This approach has been accomplished with inside the Nineties as an algorithm-primarily based totally; with inside the early 2000s primarily based totally on e-mail, then as Domain Spoofing and in current years thru HTTPs. Due to the dimensions of the mass attacked in current years, the price and impact of the assaults at the customers were high. The common monetary price of the information breach as a part of the phishing assaults in 2019 is $ 3.86 million, and the approximate price of the BEC (Business Email Compromise) terms is expected to be around $ 12 billion. Also, it's far recognised that approximately 15% of individuals who are attacked are at the least one extra goal [5]. With this result, it could be stated that phishing assaults will maintain to being accomplished with inside the ongoing years. Figures 1 additionally helps this concept and displays the wide variety of phishing web sites in 2019, and as may be visible from it, there's a growing fashion on this kind of attack.

II. THERESEARCHMETHOD

In this section, it becamementioneda number of the strategies which primarily based totally on listing, rule, visible similarity, and gadget learning. A. List Based Phishing Detection Systems These structures use lists to categories phishing and non-phishing web sites. These are known as whitelist and blacklist. The whitelist carriessecure and validweb sites, even as the blacklist consists ofweb sitescategorized as phishing. In [7], researchers used the whitelist to discover phishing sites. In the study, get admission toweb sites takes areahandiestat thecircumstance that the URL is with inside the whitelist. Another approach is the blacklist approach. In the literature, aside fromprogramsinclusive of Google Safe Browsing API, Phish Net, there also area fewresearchesthe use of blacklists like [8]. In blacklist-primarily based totallystructures, the URL is checked from the listing and get admission tothe URL if it isn'tprotectedwith inside thelisting. The largestdownsideof thosestructures is that the small alternatewith inside the URL prevents matching with inside thelisting. Additionally, the most recent attacks, which might be named zero-day attacks, can't be catches with thosekindsafetystructures. B. Rule-Based Phishing Detection Systems In thosestructures, functions are receivedprimarily based totally on relational rule mining. The policies are envisionedto emphasizefunctionswhich might begreaternot unusual place in phishing URLs [9]. In researchthe use of this sort ofgadget, it's miles aimed to applypowerfulfunctionsgreater actively with inside the classification. In thosestructures, a hard and fast of policies are determined. Thus, the gadgetoffers a better.

III. THE REFLECTIVE PROCESS

- The factors leading to phishing website change over period since they are reliant on upon multiple party-political and social reasons.
- Hence the classify of the phishing website is necessary for saving the people of the country to upload the data in website.

RESOURCES NEED FOR THE PROJECT

3.1 H/W System Configuration:

Processor	Dual Core.
Speed	1.1 G Hz.
RAM	8 GB (min).
Hard Disk	20 GB.

3.2 S/W System Configuration:

Operating System	Windows 10.
Technology	ML.
Front End	GUI-tkinter
IDLE	Python 3.9

| DOI: 10.15680/IJIRCCE.2022.1011020 |

IV. PROPOSED FAULT INJECTION METHOD

SYSTEM DESIGN

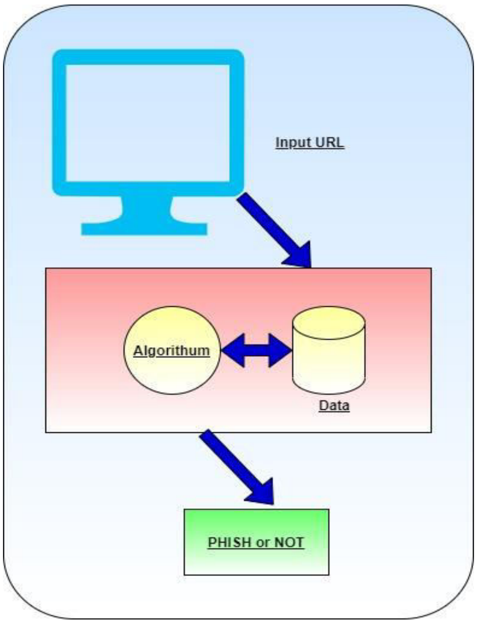


Fig. 1 Architecture

Steps1: We Taken dataset included 30 parameter with legitimate, suspicious and phishing.
Step 2: Then we send this data to the multiple algorithms like naive bayes ,random forest and Decision Tree.
Step3: We analyses those data regarding accuracy
For URL Detection:
Step1: Input Valid URL
Step2: Conect with algorithm with dataset
Step3: Finding the that url is safe or not

V. CONCLUSIONS

In latest years, because of the evolving technology on networking now no longer best for conventional internet programs however additionally for cellular and social networking tools, phishing assaults have come to be one of the crucial threats in cyberspace. Although maximum of protection assaults goal on gadget vulnerabilities, phishing exploits the vulnerabilities of the human end-users. Therefore, the primary protection shape for the organizations is informing the personnel approximately this kind of attack. However, protection managers can get a few extra safety mechanism which may be done both as a selection guide gadget for the consumer or as a prevention mechanism at the servers.

In this paper, we aimed to put in force a phishing detection gadget with the aid of using the usage of a few device getting to know algorithms. The proposed structures are examined with a few latest datasets with inside the literature and reached outcomes are as compared with the most modern works with inside the literature. The contrast outcomes display that the proposed structures decorate the performance of phishing detection and attain excellent accuracy rates. As destiny works, firstly, it's far aimed to create a brand new and large dataset for URL primarily based totally Phishing Detection Systems.

REFERENCES

[1] Samuel Marchal, Jérôme François, Radu State, and Thomas Engel, “PhishStorm: Detecting Phishing With Streaming Analytics,” IEEE Transactions on Network and Service Management, vol. 11 , issue: 4 , pp. 458-471, December 2014

[2] Mohammed NazimFeroz,SusanMengel, “Phishing URL Detection Using URL Ranking,” IEEE International Congress on Big Data, July 2015

[3] MahdiehZabihimayvan, Derek Doran, “Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection,” International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, June 2019

[4] MoitrayeeChatterjee,Akbar-SiamiNamin, “Detecting Phishing Websites through Deep Reinforcement Learning,” IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), July 2019

[5] Chun-Ying Huang,Shang-Pin Ma,Wei-Lin Yeh,Chia-Yi Lin,ChienTsung Liu, “Mitigate web phishing using site signatures,” TENCON 2010-2010 IEEE Region 10 Conference, January 2011

[6] Aaron Blum,BradWardman,ThamarSolorio,Gary Warner, “Lexical feature based phishing URL detection using online learning,” 3rd ACM workshop on Artificial intelligence and security, Chicago, Illinois, USA, pp. 54-60, August 2010

[7] Mohammed Al-Janabi,Ed de Quincey,PeterAndras, “Using supervised machine learning algorithms to detect suspicious URLs in online social networks,” IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, Sydney, Australia, pp. 1104-1111, July 2010

[8] ErzhouZhu,YuyangChen,ChengchengYe,XuejunLi,Feng Liu, “OFSNN:An Effective Phishing Websites Detection Model Based on Optimal Feature Selection and Neural Network,” IEEE Access(Volume:7), pp. 73271-73284, June 2019

[9] AnkeshAnand,KshitijGorde,Joel Ruben Antony Moniz,NoseongPark,TanmoyChakraborty,Bei-Tseng Chu, “Phishing URL Detection with Oversampling based on Text Generative Adversarial Networks,” IEEE International Conference on Big Data (Big Data), December 2018

[10] Justin Ma,Lawrence K. Saul,StefanSavage,Geoffrey M. Voelker, “Learning to detect malicious URLs,” ACM Transactions on Intelligent Systems and Technology (TIST) archive Volume 2 Issue 3, April 2011

[11] Youness Mourtaji,Mohammed Bouhorma,Alghazzawi, “Perception of a new framework for detecting phishing web pages,” Mediterranean Symposium on Smart City Application Article No. 11, Tangier, Morocco, October 2017

[12] Akihito Nakamura,FumaDobashi, “Proactive Phishing Sites Detection,” WI '19 IEEE/WIC/ACM International Conference on Web Intelligence), pp. 443-448, October 2019

INTERNATIONAL JOURNAL
OF INNOVATIVE RESEARCH
IN COMPUTER & COMMUNICATION ENGINEERING