International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 411-415 www.ijaem.net ISSN: 2395-5252

Web Vulnerability Scanner Using Machine Learning

Anusha RP*1, Ms. Ashwini C*2

P.G, Student, Department of Master of Computer Applications, University B.D.T Collage of Engineering, Davangere, Karnataka, India

Assistant Professor, Department of Master of Computer Applications, University B.D.T Collage of Engineering, Davangere, Karnataka India

Date of Submission: 10-07-2024 Date of Acceptance: 20-07-2024

ABSTACT: In the modern digital age, maintaining the security of websites is crucial due to the rise in cyberthreats such as malware and phishing. The goal of this project is to create a machine learning-based website scanner that assesses a website's safety and returns a percentage score that represents how likely it is to be safe or unsafe. To thoroughly evaluate the security of websites, the scanner makes use of a number of methods, such as behaviour analysis, SSL/TLS certificate validation, content analysis, URL reputation checks, and threat intelligence feeds. Phishing attacks are a recurring danger to cybersecurity, with the potential to cause substantial financial losses and compromise personal data. In order to improve the precision and dependability of differentiating between secure and malicious websites, this research explores the use of Convolutional Neural Networks (CNNs) for fraud detection. We train the CNN model by extracting relevant features from URLs and webpage content, using a large dataset that includes annotated samples of safe and harmful websites.

Keywords: Domain, Deep learning, Phishing, Authentication.

I. INTRODUCTION

Social engineering attacks, which trick users into thinking something is phony without being detected, are a common security concern that can reveal sensitive and private data. Getting private information, such as usernames, passwords, and account numbers, is the main objective of this assault. According to, phishing attempts can happen through a range of communication methods, such as SMS, messaging, and phony emails. People usually have multiple accounts on various websites, including social media, email, and bank accounts. The most vulnerable targets are innocent web users because most individuals are unaware of their valuable information, which adds to the attack's effectiveness. Social engineering is commonly used in phishing attacks, where a spoof link is sent to the victim that directs them to a fraudulent website. The phony link is sent to the victim by email or prominent websites. The fake website is designed to resemble the original. Consequently, instead of the real web server, the attacker server will get the victim's request.

The firewall, antivirus, and customized software solutions that are currently on the market do not completely stop the web XSS assault. The usage of secure socket layers (SSL) and digital certificates (CA) does not shield the online user against such assaults. An attacker reroutes the request to a fake web server using the web XSS approach. In actuality, a particular type of SSL and CA can be established as long as everything appears real. It is argued that users are not well protected when utilizing a secure connection for surfing, especially from hackers who know how "secure" connections really work. This project's goal is to determine whether a website's URL is good or terrible.

II. LITURATURE SURVEY

1)A Literature Review on web vulnerabilities

The literature on phishing attack detection is surveyed in this article. Phishing attacks aim to exploit weaknesses in systems that result from human interaction. Users are the weakest link in the security chain since many cyberattacks propagate through methods that take advantage of flaws in end users. Since there is no one magic bullet to properly address every vulnerability in the phishing problem, many strategies are frequently used to counteract different types of attacks. This document attempts to survey a large number of phishing mitigation strategies that have been

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 411-415 www.ijaem.net ISSN: 2395-5252

developed recently. A high-level summary of the several types of phishing mitigation strategies is also provided, including offensive defense, detection, rectification, and prevention. It is important, in our opinion, to show how phishing detection approaches fit into the larger picture of mitigation.

2) Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online

Information technology advancements frequently force users to make difficult and important decisions about security and privacy. An expanding corpus of studies has examined people's decisions when faced with trade-offs between privacy and information security, the obstacles to decision-making that stand in the way of such decisions, and strategies to overcome those obstacles. An interdisciplinary evaluation of the literature on privacy and security decision making is given in this article. It focuses on studies that support people's privacy and security decisions by gently guiding users toward better decisions through paternalistic interventions. The article outlines the main ethical, design, and research challenges as well as the possible advantages of those interventions as well as their drawbacks.

3) Priming and warnings are not effective to prevent social engineering attacks.

People have a tendency to trust one another and to divulge personal information with ease. They are therefore susceptible to social engineering scams. The current study examined the efficacy of two interventions priming through cues to increase awareness about the risks of social engineering cyberattacks and cautions against disclosing personal information—in defending users against social engineering attacks. A study was conducted on a sample of people who visited the shopping district of a medium-sized Dutch town. In order to gauge individuals' level of disclosure, questions about their email address, nine digits from their eighteen-digit bank account number, and, for those who had shopped online in the past, what they had bought and from which online retailer were asked. There were several comparatively high disclosure rates: 43.5% of the participants submitted bank account information, and 79.1% of the subjects entered their email address. 91.4% of the subjects who were online shoppers filled in the name of the online store where they made these purchases, and 89.8% of the subjects indicated what kind of product(s) they had bought. Multivariate analysis revealed that the degree of disclosure was unaffected by warnings or priming questions. There were signs of the warning's negative impact. These discoveries' ramifications are examined.

4) Detection of Phishing Websites Based on Probabilistic Neural Networks and K-Medoids Clustering

Information security researchers are placing an increasing amount of emphasis on studying anti-phishing solutions due to the devastating effects of phishing assaults and their rising frequency. Information leakage, identity theft, money loss, and reputation destruction are a few examples of security hazards. Increasing human knowledge is not a sufficient mitigation strategy; instead, the implementation of complementary technology solutions is essential. While there have been a number of techniques put out in the literature, creating effective phishing detection models remains a difficult challenge for which there is currently no comprehensive solution. In this paper, we introduce a unique probabilistic neural network (PNN) based method for phishing website detection. Additionally, we explore the integration of PNN with K-medoids clustering to achieve substantial reduction in complexity without compromising the accuracy of detection. We performed a thorough analysis to analyze multiple performance metrics using a publicly accessible data set consisting of 11,055 benign and phishing websites in order to determine the viability of the suggested approach. The experimental results demonstrate that higher accuracy models may be constructed and that reduced false errors can be obtained while maintaining >97% accuracy even with a >40% reduction in complexity.

5) Detecting Algorithmically Generated Malicious Domain Name

Lately, DNS-based "domain fluxing" has been utilized by botnets like Conficker, Kraken, and Torpig for command-and-control. Each bot asks whether a set of domain names exist, and the owner just needs to register one of these domain names. Our approach in this work is to identify "domain fluxes" in DNS traffic by searching for patterns in domain names that are generated algorithmically, as opposed to manually. Specifically, we examine the distribution of bigrams and alphanumeric characters across all domains that map to the same set of IP addresses. We showcase and contrast the effectiveness of multiple distance measurements, such as KL-

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 411-415 www.ijaem.net ISSN: 2395-5252

distance, Edit distance, and Jaccard measure. We train on a good data set of domains that we gathered through a crawl of all IPv4 address space mapped domains, and we model bad data sets based on observed and predicted behaviours thus far.

III. METHODOLOGY

In this study, our main goal was to identify phishing (bad URLs) and legal (good URLs) websites. First, we looked into and obtained different public platforms datasets. We created code to extract features from the URLs using the CNN Algorithm. The databases contained a variety of URL kinds, including malware, phishing, benign, and spam URLs. Specifically, we selected 5,000 URLs at random for investigation out of a sample containing 35,000 innocuous URLs. After pre-processing, we combined all of the data into a single frame and split the dataset into training and testing sets. More than 500,000 distinct items made up this data, which was divided into two columns based on how excellent or poor the URLs were. The numerous libraries were used for diverse purposes, such turning URLs into a data frame and showing prevalent words in both good and poor URLs. We then developed the model and divided the data correspondingly. After that, the model was implemented and links were imported for prediction. To identify the top-performing model, we analyze the accuracies of several training datasets using accuracy as our main criterion.

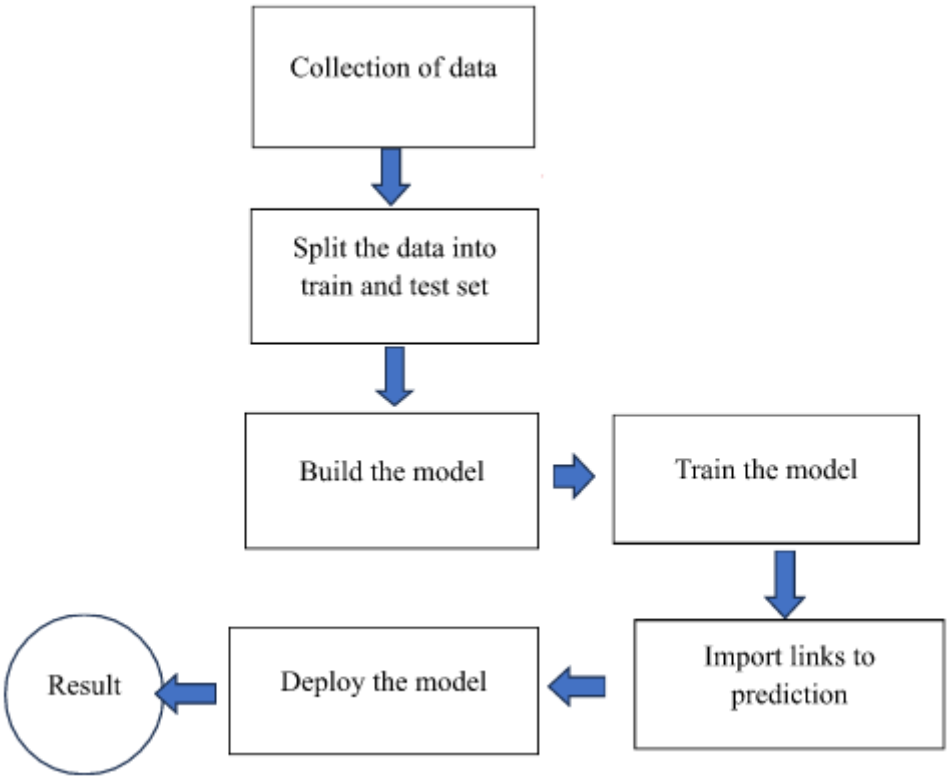
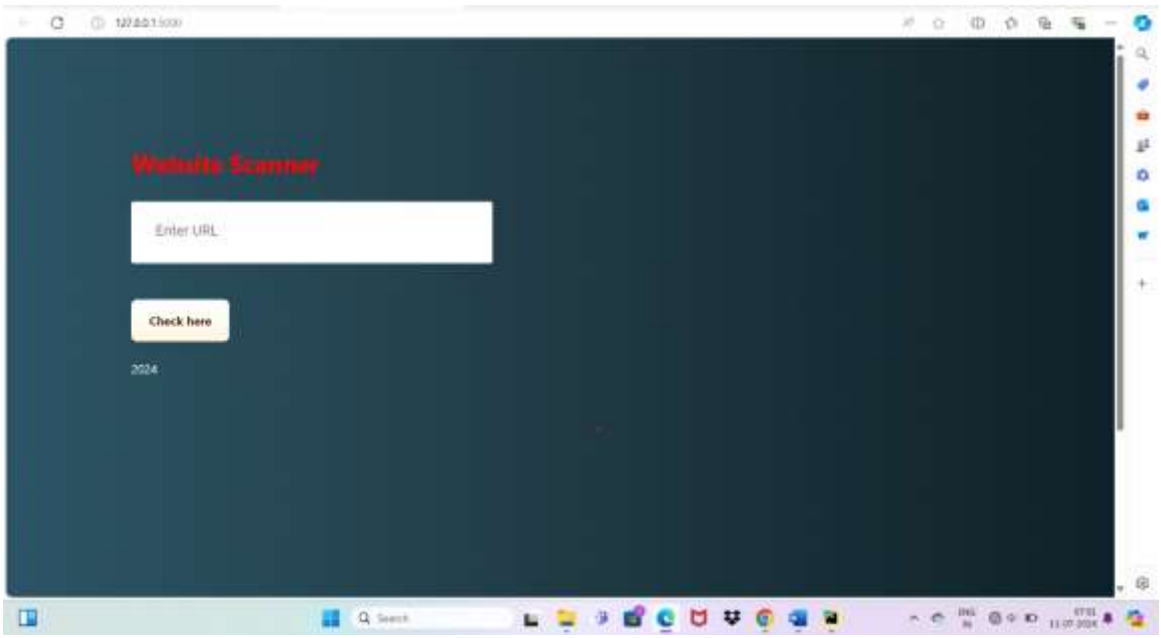


Fig (1) System Architecture

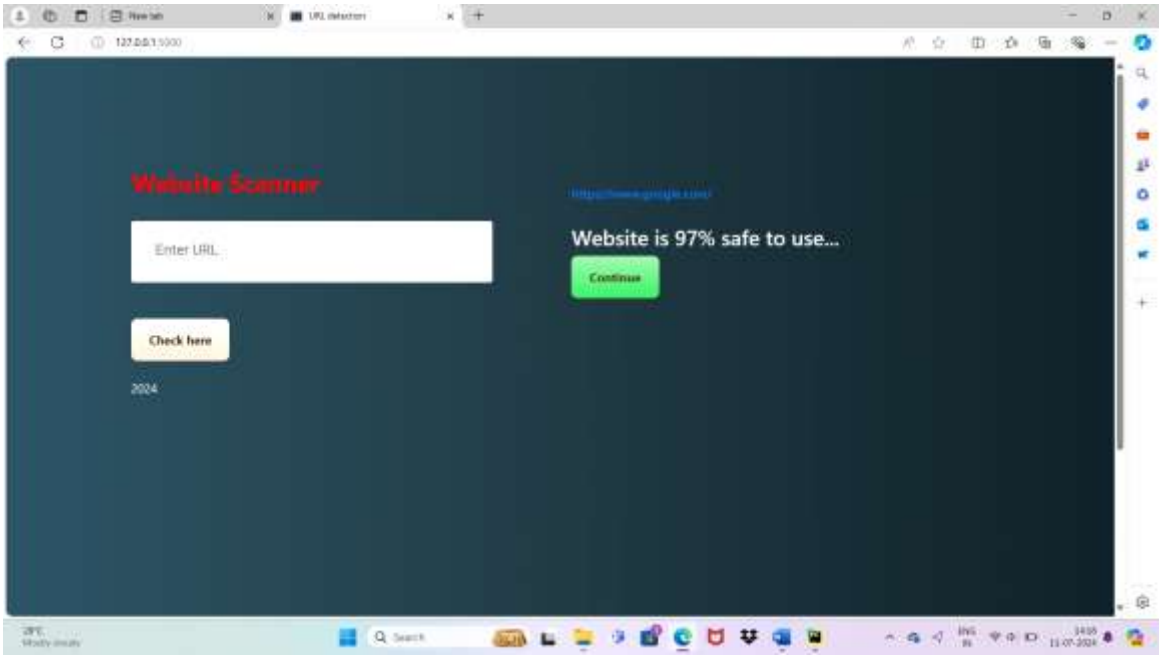
International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 411-415 www.ijaem.net ISSN: 2395-5252

IV. RESULT AND DISCUSSION



Fig(2)

In above image we are showing the dashboard of our system which will be use by user to enter the URL.



Fig(3)

In above image we are detecting the weather the URL is safe or not.

V. CONCLUSION

The most crucial defense against phishing attacks for users is education and awareness. Users of the internet should be aware of all security advice provided by professionals. Users should also be taught not to click on links indiscriminately that take them to websites that need them to provide sensitive personal data. Verifying the URL is

crucial before visiting the website. The technology may eventually be upgraded to automatically identify websites and determine whether an application is compatible with a given web browser. More effort is put on the web browser. In order to differentiate the phony websites from the authentic ones, more work can be done by adding additional features. The CNN algorithm is employed by the system to attain a high level of accuracy and resilience in detecting phishing attempts.

REFERENCE

[1]. Website Detection Using URL-Assisted Brand Name Weighting System," 2014 IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), December 1-4, 2014. Choon Lin Tan, Kang LengChiew, San Nah Sze.

[2]. "Mining the web to detect unsecured urls," Proceedings of the International Conference on Machine Learning and Applications, vol. 1, pp. 568-573, December 2012. R. B. Basnet and A. H. Sung.

[3]. Mohiuddin Ahmed, AbdunNaser Mahmood, and Jiankun Hu, "A study of network anomaly detection strategies," J. Netw. Comput. Appl., vol.60, no. C, pp. 19-31, 2016.

[4]. Predicting unsafe websites based on self-structuring neural network," Neural Computing and Applications, vol. 25, no. 2, pp. 443-458, 2013-B. R. M. Mohammad, F. Thabtah, and L. McCluskey.

[5]. Wenyin Liu et al., "Discovering scam target via semantic link networks," Future Generation Computer Systems, vol. 26.3, no. 3, pp. 381- 388, 2010.

[6]. "An Online Back propagation Algorithm with Validation Error-Based Adaptive Learning Rate," in Artificial Neural Networks – ICANN 2007, Porto, Portugal, 2007. S. Duffner and C. Garcia.

[7]. "Detection and analysis of drive-by-download assaults and malicious javascriptcode," Proceedings of the 19th International Conference on World Wide Web, pp. 281-290, 2010. Marco Cova, Christopher Kruegel, Giovanni Vigna.

[8]. "Improved Phishing Detection Using Model-Based Features," by A. Bergholz, J. H. Chang, G. Paass, F. Reichartz, and S. Strobel. 2008, CEAS.

[9]. Helena Matute, Mara M. Moreno-Fernández, Fernando Blanco, Pablo Garaizar I'm looking for phishers. To combat electronic fraud, Internet users' sensitivity to visual deception indicators should be improved. pp.421-436 in Computers in Human Behavior, Vol.69, 2017.

[10]. F.J. Overink, M. Junger, L. Montoya. Preventing social engineering assaults with priming and warnings does not work. pp.75-87 in Computers in Human Behavior, Vol.66, 2017. 2017.