

**SECURE QR CODE SCANNER TO DETECT MALICIOUS URL USING MACHINE
LEARNING**

Mr. SANDEEP ACHI, Assistant Professor

Y.CHANDANA, T.SRUTHI, E.VAMSHI, D. SRIKANTH

SRI INDU COLLEGE OF ENGINEERING AND TECHNOLOGY

Sheriguda (V), Ibrahimpatnam (M), RangareddyDist – 501 510

ABSTRACT

Q-R codes are utilised for a variety of purposes, including accessing online web-pages and making a settlement. The Internet facilitates a wide range of illegal acts, including unsolicited e-marketing, financial embezzlement, and malicious distribution. Even though all the users identify the presence of Q-R codes visually, the information stored in those codes can only be accessed through an allocated Q-R code decoder. Q-R codes have also been shown to be used as an effective attack vector. For example, techniques include social engineering, phishing, pharming, etc. Harmful codes are distributed under false pretences in congested areas, or malicious Q-R codes are pasted over current ones on billboards. Finally, consumers rely on decoder operating system to determine if a random Q-R code is whether malicious or benign. For the purpose of this report, we consider the identification of malicious Q-R codes as a two-way classification problem in this research, and we test the effectiveness of many well-known ML algorithms, including namely K-Nearest Neighbour, Random Forest, Binary LSTM and Support Vector Machine. This implies that the proposed method might be deemed an optimal and user-friendly QR code security solution. We created a prototype to test our recommendations and found it to be secure and usable in protecting users from harmful QR Codes.

Index : qr code, malicious, knn, random forest, lstm, ml algorithms

1. INTRODUCTION

1.1 INTRODUCTION TO PROJECT:

Quick Response codes have found their way into our daily lives as a result of the Internet's undeniable omnipresence. QR codes are data-encoding two-dimensional matrix barcodes. They were originally designed to track automobile parts, but they have since been adapted for a wider range of applications.

The most typical application is to encode a link or other textual information so that it can be accessed quickly without requiring the user to write the URL manually. Covid-19 has impacted the expansion of internet business like e-commerce and e-banking, primarily because QR-code payment streamlines and secures the payment process's correct execution, saves time, and eliminates data entry problems. Unfortunately, technical breakthroughs are sometimes accompanied by state-of-the-art methods for exploiting users. Phishing websites that hoist any type of personal data, a spammer can use are prominent instances of such attacks. QR Codes have been widely exploited and abused to take advantage of consumers' vulnerabilities. According to data on the rise in the number of dangerous QR Code distributions over time, it indicates a clear necessity to research, implement procedures or ways to prevent and detect them.

In a survey Katharina Krombholz et. al. [1] provides a detailed summary of current studies on QR code usability and security in a survey. They determined the most significant use case as well as the corresponding attack routes. According to the paper, in addition to their many benefits, Social engineers have used QR codes as an attack vector. Malicious links are encoded by attackers in QR code, for example, redirect to run fraudulent code or phishing sites. Social engineering is the most commonly reported assault scenario. Social engineering is a strategy used in information technology (IT) security to manipulate people into revealing secret information to the social engineer.

Katharina Krombholz et al. [2] recognise the importance of QR code security and give a thorough assessment of quick response code reader vulnerability, with an emphasis on practical security techniques. The initial portion of this research concentrated on the decoder software, while another phase is focused upon a user. Their findings reveal and underline the need for security enhancements to ensure a secure user experience when scanning QR codes. Based on an examination of privacy and security problems in mobile software, they suggested a design set of guidelines for an application that takes security, privacy, and usability into account. They demonstrated that when security, privacy, and usability are all taken into account, mobile-phone software can efficiently safeguard consumers from harmful quick response code. Heider and Flaminia [3] have done a study that gives a thorough review of QR code scanning applications in terms of security, usability, and privacy. They identified the shortcomings and found that: Many of these QR code scanners do not meet the consumer privacy and security needs.

Several papers in the literature address this issue from the perspective of Machine Learning. In other words, they construct a list of URLs categorised as harmful or safe along with characterize each uniform resource locator using fixed group of criteria. Classification algorithms are then supposed to learn the border between the decision classes. The extensive literature on classification models provides a variety of alternative solutions and approaches to classification challenges. Among the most common classifiers are K-nearest neighbors, Random Forest, Support Vector Machine, and Bidirectional LSTM, each with its own set of pros and disadvantages. Scammers use a variety of cyber techniques, such as luring consumers to press on phished link, that can cause the system to be compromised. QR codes make it simple to accomplish this. To assist in identifying malicious websites, the online security community has launched blacklisting services. While blacklisting a URL has been shown results effectively in some situations, an scammer can readily trick the system by modifying more or one components of the link. Many harmful websites are never banned, either they were never or incorrectly assessed or they are too new.

2. LITERATURE SURVEY

TITLE: “QR Code Security: A Survey of Attacks and Challenges for Usable Security”,

ABSTRACT: QR (Quick Response) codes are two-dimensional barcodes with the ability to encode different types of information. Because of their high information density and robustness, QR codes have gained popularity in various fields of application. Even though they offer a broad range of advantages, QR codes pose significant security risks. Attackers can encode malicious links that lead e.g. to phishing sites. Such malicious QR codes can be printed on small stickers and replace benign ones on billboard advertisements. Although many real world examples of QR code based attacks have been reported in the media, only little research has been conducted in this field and almost no attention has been paid on the interplay of security and human-computer interaction. In this work, we describe the manifold use cases of QR codes. Furthermore, we analyze the most significant attack scenarios with respect to the specific use cases. Additionally, we systemize the research that has already been conducted and identified usable security and security awareness as the main research challenges. Finally we propose design requirements with respect to the QR code itself, the reader application and usability aspects in order to support further research into making QR code processing both secure and usable.

TITLE: “QR Code Security - How Secure and Usable Apps Can Protect Users Against Malicious QR Code’

ABSTRACT: QR codes have emerged as a popular medium to make content instantly accessible. With their high information density and robust error correction, they have found their way to the mobile ecosystem. However, QR codes have also proven to be an efficient attack vector, e.g. To perform phishing attacks. Attackers distribute malicious codes under false pretenses in busy places or paste malicious QR codes over already existing ones on billboards. Ultimately, people depend on reader software to ascertain if a given QR code is benign or malicious. In this paper, we present a comprehensive analysis of QR code security. We determine why users are still susceptible to QR code based attacks and why currently deployed smartphone apps are unable to mitigate these attacks. Based on our findings, we present a set of design recommendations to build usable and secure mobile applications. To evaluate our guidelines, we implemented a prototype and found that secure and usable apps can effectively protect users from malicious QR codes.

TITLE: “Evaluating Security, Privacy and Usability Features of QR Code Readers”,

ABSTRACT: The widespread use of smartphones is boosting the market take-up of dedicated applications and among them, barcode scanning applications. Several barcode scanners are available but show security and privacy weaknesses. In this paper, we provide a comprehensive security and privacy analysis of 100 barcode scanner applications. According to our analysis, there are some apps that provide security services including checking URLs and adopting cryptographic solutions, and other apps that guarantee user privacy by supporting least privilege permission lists. However, there are also apps that deceive the users by providing security and privacy protections that are weaker than what is claimed. We analyzed 100 barcode scanner applications and we categorized them based on the real security features they provide, or on their popularity. From the analysis, we extracted a set of recommendations that developers should follow in order to build usable, secure and privacy-friendly barcode scanning applications. Based on them, we also implemented BarSec Droid, a proof of concept Android application for barcode scanning. We then conducted a user experience test on our app and we compared it with DroidLa, the most popular/secure QR code reader app. The results show that our app has nice features, such as ease of use, provides security trust, is effective and efficient.

TITLE: “Detecting Malicious URLs using Machine Learning Techniques”

ABSTRACT: The World Wide Web supports a wide range of criminal activities such as spam-advertised e-commerce, financial fraud and malware dissemination. Although the precise motivations

behind these schemes may differ, the common denominator lies in the fact that unsuspecting users visit their sites. These blacklists are in turn constructed by an array of techniques including manual reporting, honeypots, and web crawlers combined with site analysis heuristics. Inevitably, many malicious sites are not blacklisted either because they are too recent or were never or incorrectly evaluated. In this paper, we address the detection of malicious URLs as a binary classification problem and study the performance of several well-known classifiers, namely Naïve Bayes, Support Vector Machines, Multi-Layer Perceptron, Decision Trees, Random Forest and k-Nearest Neighbors. Furthermore, we adopted a public dataset comprising 2.4 million URLs (instances) and 3.2 million features.

TITLE: “Malicious URL Detection based on Machine Learning”,

ABSTRACT: Cyber security is a very important requirement for users. With the rise in Internet usage in recent years, cyber security has become a serious concern for computer systems. When a user accesses a malicious Web site, it initiates a malicious behavior that has been pre-programmed. As a result, there are numerous methods for locating potentially hazardous URLs on the Internet. Traditionally, detection was based heavily on the usage of blacklists. Blacklists, on the other hand, are not exhaustive and cannot detect newly created harmful URLs. Recently, machine learning methods have received a lot of importance as a way to improve the majority of malicious URL detectors. The main goal of this research is to compile a list of significant features that can be utilized to detect and classify the majority of malicious URLs. To increase the effectiveness of classifiers for detecting malicious URLs, this study recommends utilizing host-based and lexical aspects of the URLs. Malicious and benign URLs were classified using machine learning classifiers such as AdaBoost and Random Forest algorithms. The experiment shows that Random Forest performs really well when checked using voting classifier on AdaBoost and Random Forest Algorithms. The Random Forest achieves about 99% accuracy.

TITLE: “Malicious URL Detection: A Comparative Study”

ABSTRACT: Malicious uniform resource locator (URL), i.e., Malicious websites are one of the most common cybersecurity threats. They host gratuitous content (spam, malware, inappropriate ads, spoofing, etc.) and tempt unwary users to become victims of scams (financial loss, private information disclosure, malware installation, extortion, fake shopping site, unexpected prize etc.) and cause loss of billions of rupees each year. The visit to these sites can be driven by email, advertisements, web search or links from other websites. In each case, the user must click on the malicious URL. The rising cases of

phishing, spamming and malware has generated an urgent need for a reliable solution which can classify and identify the malicious URLs.

TITLE: “Malicious Qr-Code Threats and Vulnerability of Blockchain”,

ABSTRACT: Today’s rapidly changing world, is observing fast development of QR-code and Blockchain technologies. It is worth noting that these technologies have also received a boost for sharing. The user gets the opportunity to receive / send funds, issue invoices for payment and transfer, for example, Bitcoin using QR-code. The following vulnerabilities were considered: fake QR generators, stickers for cryptomats, phishing using QR-codes, create Malicious QR-Codes for Hack Phones and Other Scanners. The possibility of creating the following malicious QR codes while using the QRGen tool was considered: SQL Injections, XSS (Cross-Site Scripting), Command Injection, Format String, XXE (XML External Entity), String Fuzzing, SSI (Server-Side Includes) Injection, LFI (Local File Inclusion) / Directory Traversal.

TITLE: “Design of an efficient image protection method based on QR code”

ABSTRACT: This paper presents the design and the verification of an efficient image protection method based on the QR code, which is a type of two-dimensional barcode widely used in various fields. For this purpose, we design a new image protection system consisting of a secure image generator and a secure image recognizer. One adds a new pre-processing block to the typical QR code generator and the other combines the existing QR code reader with a new post-processing block. The new architecture provides image de-identification. It is also flexible, allowing the use of text-based compression and encryption

TITLE: “Quick Response Code Validation and Phishing Detection Tool”

ABSTRACT: A Quick Response (QR) Code is a type of barcode that can be read by the digital devices and which stores the information in a square-shaped. The QR Code readers can extract data from the patterns which are presented in the QR Code matrix. A QR Code can be acting as an attack vector that can harm indirectly. In such case a QR Code can carry malicious or phishing URLs and redirect users to a site which is well conceived by the attacker and pretends to be an authorized one. Once the QR Code is decoded the commands are triggered and executed, causing damage to information, operating system and other possible sequence the attacker expects to gain.

TITLE: “A Construction of Fake QR Codes Based on Error-Correcting Codes”

ABSTRACT: QR codes are used for various applications, such as access to web pages and to make a settlement. Although users can visually recognize the existence of the QR codes, they need to use a dedicated QR code decoder for reading the stored information in the codes. A malicious one creates a fake QR code by making bad use of this feature and guides users to a malicious web page with a careless operation by the users.

3. PROBLE STATEMENT

In general, QR codes are divided into sections that are dedicated for specific uses. Some areas of the Quick Response code are working and by error correction it cannot be recovered. Black and white modules are used to encode the data. If you use the QR code as an attack vector, you can change the QR code partially or completely. Partial changes occur when individual modules are being flipped from black to white and conversely. Malicious mods are generally of two different types. Primary method modifies both black white pixels, while the other method only allows white pixel to black pixel changes. Limitations in 5 point.

LIMITATIONS

Dependency on Training Data Quality:

The effectiveness of machine learning algorithms in distinguishing between benign and malicious URLs heavily relies on the quality and representativeness of the training dataset. If the training data is biased or lacks diversity, the system may struggle to accurately identify new and evolving threats.

Dynamic Nature of Threats:

Malicious actors constantly evolve their techniques, making it challenging for any static system to keep up with emerging threats. The system may face limitations in detecting sophisticated and novel attack vectors that were not present in the training data, leading to potential false negatives. **False Positives and**

User Trust:

Striking a balance between sensitivity and specificity is a challenge. A system with high sensitivity might detect more threats but may also produce more false positives, causing inconvenience to users. Balancing this trade-off is crucial to maintaining user trust and preventing unnecessary rejection of legitimate QR codes.

QR Code Obfuscation Techniques:

Malicious actors may employ advanced obfuscation techniques to hide the true intent of a QRcode. This could include encoding URLs in a way that appears benign at first glance but revealsmalicious content after decoding. The system may struggle to detect such obfuscated URLs,posing a challenge to its efficacy.

Limited Offline Functionality:

The system's dependency on real-time analysis and connectivity could be a limitation inenvironments with poor or no internet connectivity

4. PROPOSED SYSTEM

The proposed Secure QR Code Scanner system to detect malicious URLs using Machine Learningaims to address the inherent vulnerabilities associated with QR code usage in a dynamic andevolving digital landscape. This system leverages cutting-edge machine learning algorithms,including computer vision and natural language processing, to conduct real-time analysis of QRcodes. The system will be designed to intelligently interpret the content encoded within QR codes,distinguishing between benign and potentially malicious URLs. Anomaly detection mechanismswill be incorporated to identify deviations from expected patterns, enhancing the system's abilityto recognize sophisticated attack vectors. The proposed system also envisions integration withthreat intelligence databases to cross-reference URLs against known malicious domains, ensuringa proactive defense against emerging threats. To mitigate the risk of false positives, the systemwill prioritize a balance between sensitivity and specificity, providing users with clear andimmediate feedback after scanning a QR code. User education features will be integrated toinform individuals about potential risks associated with QR codes, fostering a more informed andsecure user experience. Continuous updates and adaptation to emerging threats will be keycomponents of the proposed system to ensure robust and effective security measures.

ADVANTAGES

Real-time Threat Detection:

The system operates in real-time, enabling immediate analysis of QR codes as they are scanned.This rapid response time enhances the ability to detect and prevent users from interacting withpotentially malicious content, reducing the window of exposure to cyber threats.Adaptive **Machine Learning Models:**

Leveraging machine learning algorithms, the system continuously learns and adapts to evolving threat landscapes. This adaptability allows the system to stay ahead of emerging attack vectors and improve its accuracy over time, providing robust protection against new and sophisticated threats.

Comprehensive Analysis with Computer Vision:

Integration of computer vision techniques allows the system to extract and comprehensively analyze visual and textual information within QR codes. This multi-faceted approach enhances the system's capability to identify malicious content, even when hidden through various obfuscation techniques.

Anomaly Detection for Advanced Threats:

The incorporation of anomaly detection mechanisms enables the system to identify deviations from expected behavior, allowing it to detect and mitigate advanced threats that may not be apparent through traditional signature-based methods. This proactive approach enhances the overall security posture.

User-Friendly Feedback and Education:

The system provides clear and immediate feedback to users after scanning a QR code, indicating whether the content is considered safe or potentially malicious. This user-friendly approach empowers individuals to make informed decisions and fosters a safer digital environment. Additionally, user education features help raise awareness about QR code security, promoting responsible and secure usage.

5. METHODOLOGY

QR Code Parsing Module:

This module is responsible for extracting data from scanned QR codes. It involves the implementation of computer vision techniques to interpret visual information and retrieve the encoded content. The parsed data is then passed on to subsequent modules for analysis.

Machine Learning Analysis Module:

The heart of the system, this module employs machine learning algorithms to analyze the parsed QR code data. It categorizes URLs as either benign or potentially malicious based on the learned patterns from a diverse dataset. The module should be adaptive, continuously updating its models to address evolving cybersecurity threats.

Anomaly Detection Module:

Focused on identifying irregularities or unexpected patterns in the QR code content, this module enhances the system's capability to detect sophisticated attack vectors. By flagging anomalies, the system can raise alerts and take preventive measures against potential threats that may not conform to known malicious patterns.

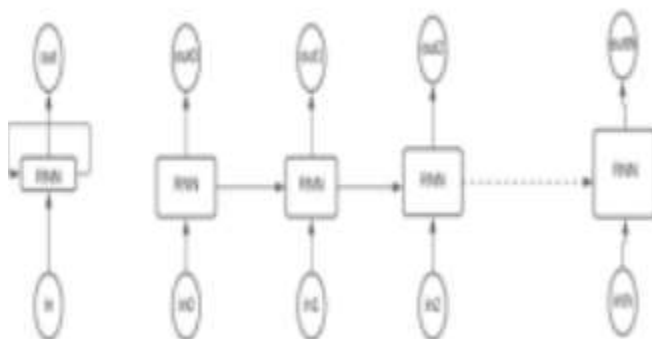
Threat Intelligence Integration Module:

This module integrates with external threat intelligence databases to cross-reference URLs against known malicious domains. By leveraging up-to-date threat information, the system can enhance its accuracy in identifying URLs associated with active cyber threats, providing an additional layer of defense.

User Interface and Feedback Module:

The user interface module is responsible for presenting a user-friendly experience. It provides clear and immediate feedback to users after scanning a QR code, indicating the security status of the content. This module also includes features for user education, offering information about potential risks associated with QR codes and promoting responsible usage.

6. SYSTEM ARCHITECTURE



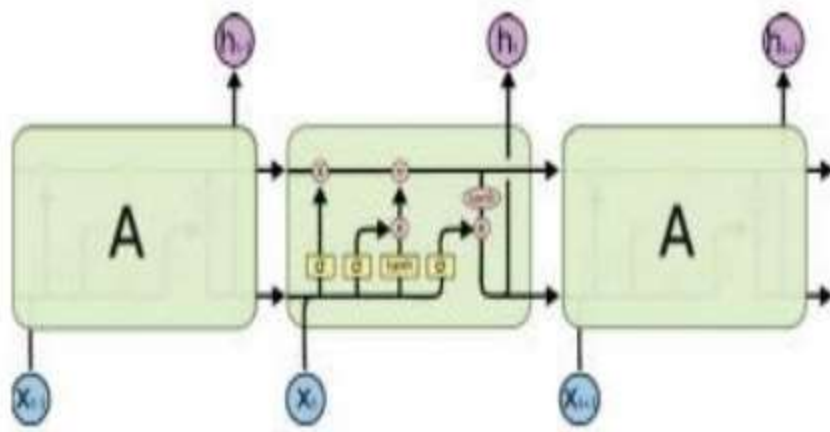


Fig 1

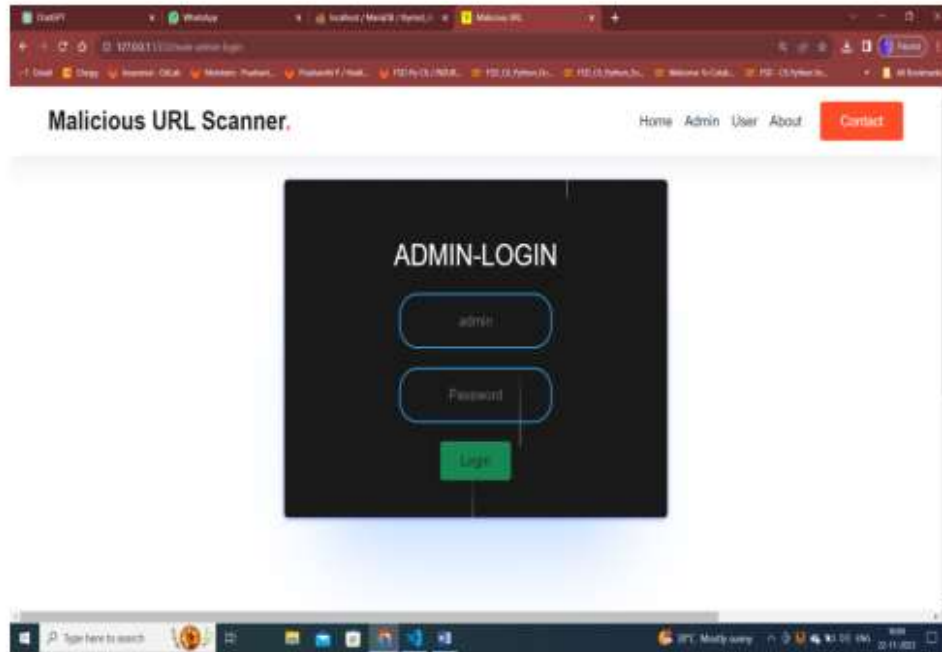
7. OUT COME RESULTS

URL-PROJECT

Home Page:

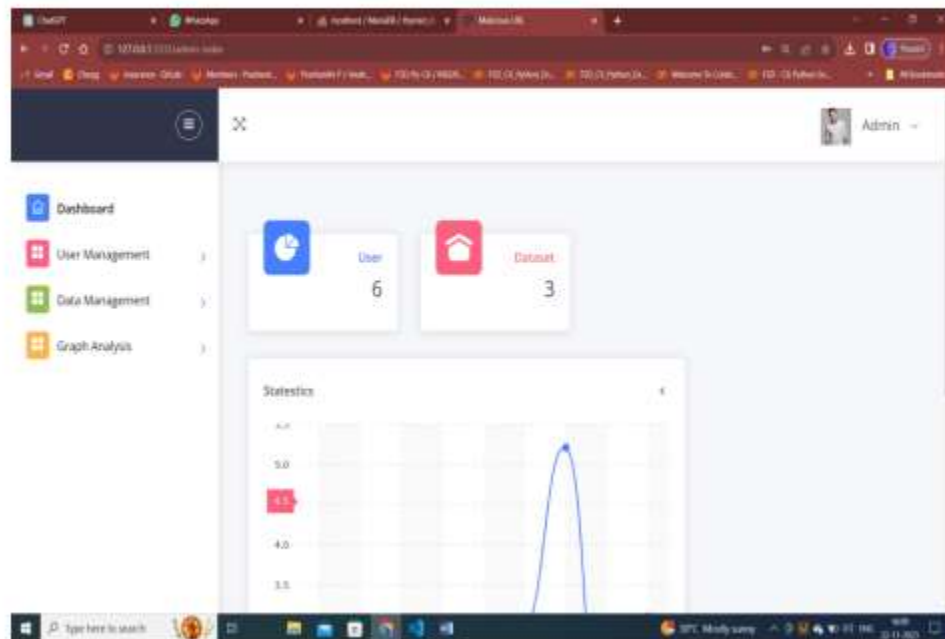


Admin Page:



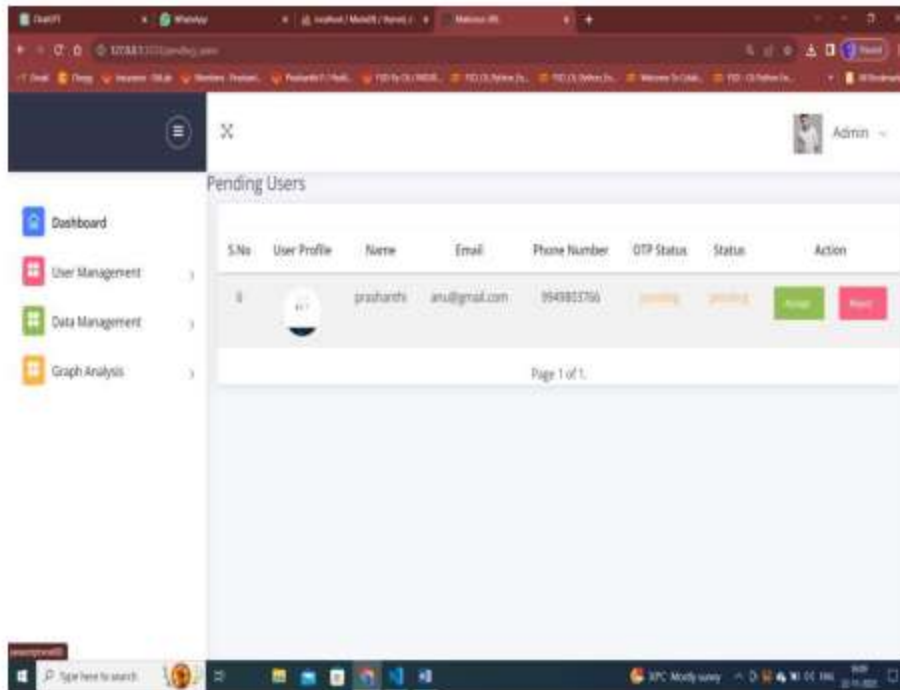
Screen 2

Admin Dashboard:



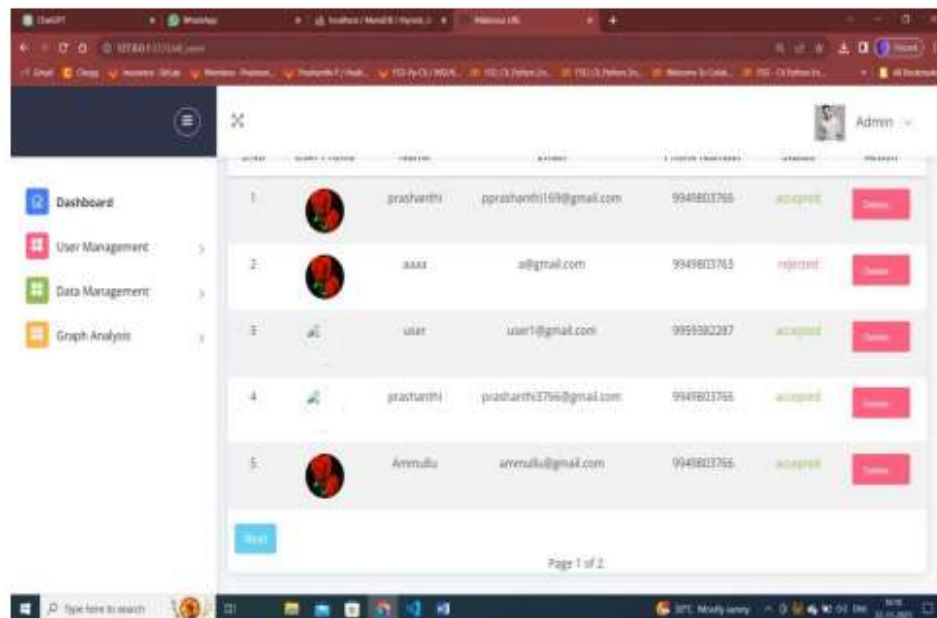
Screen 3

Pending Users:



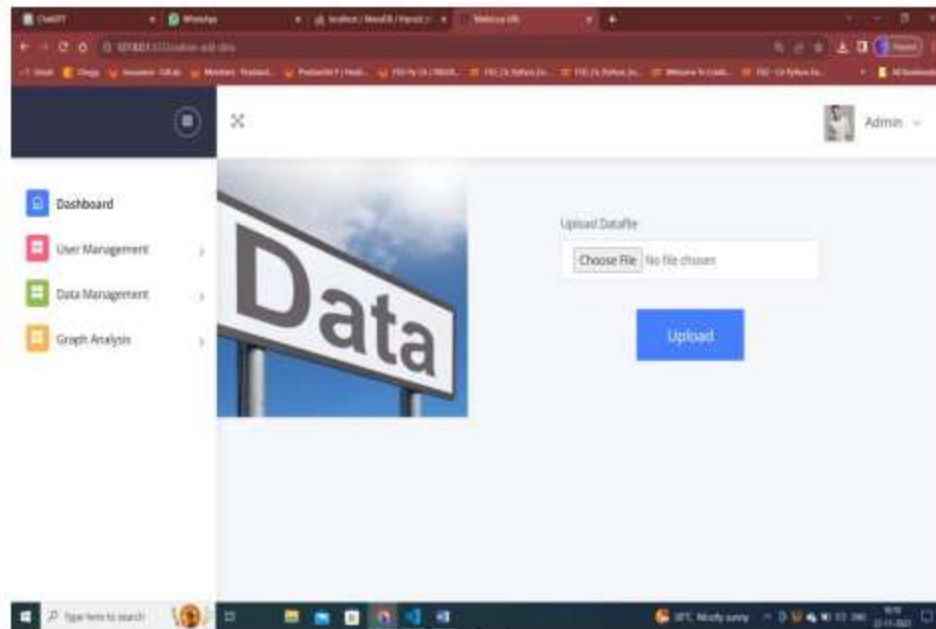
Screen 4

All Users:



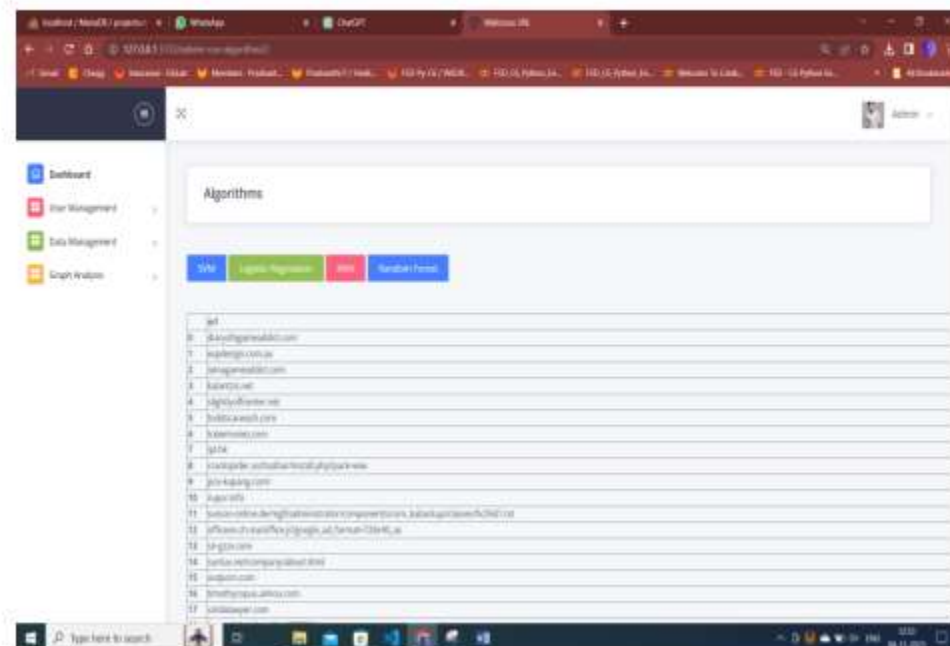
Screen 6

Upload dataset:



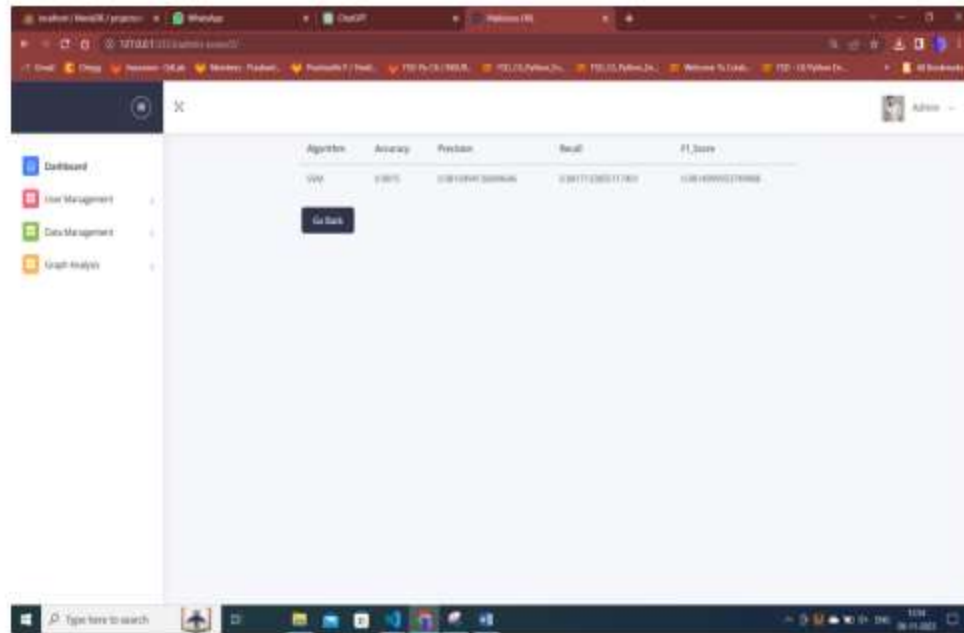
Screen 7

Algorithms Page:-



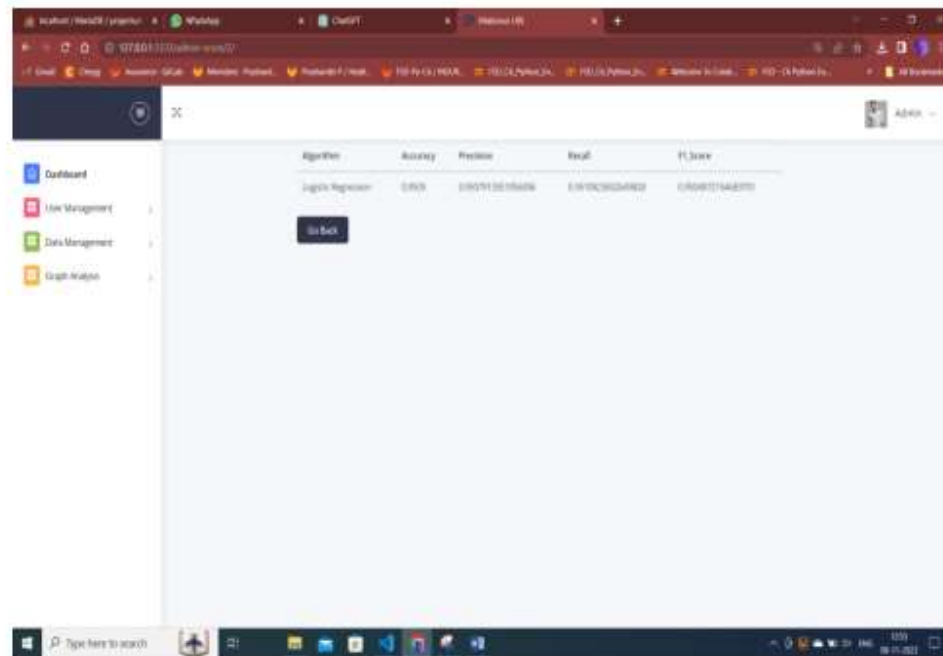
Screen 8

SVM:-



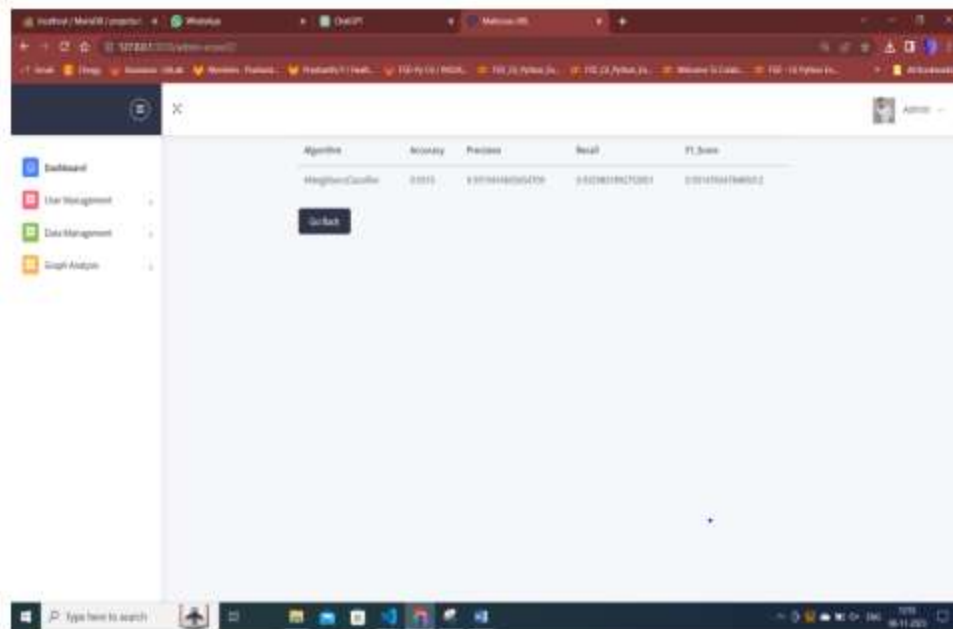
Screen 9

Logistic Page:-



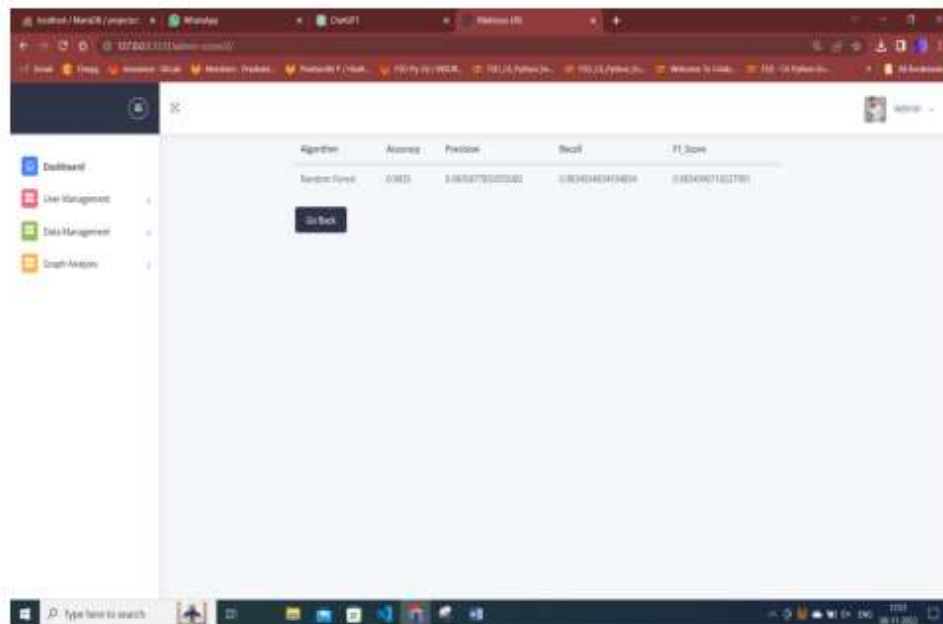
Screen 10

KNN :-



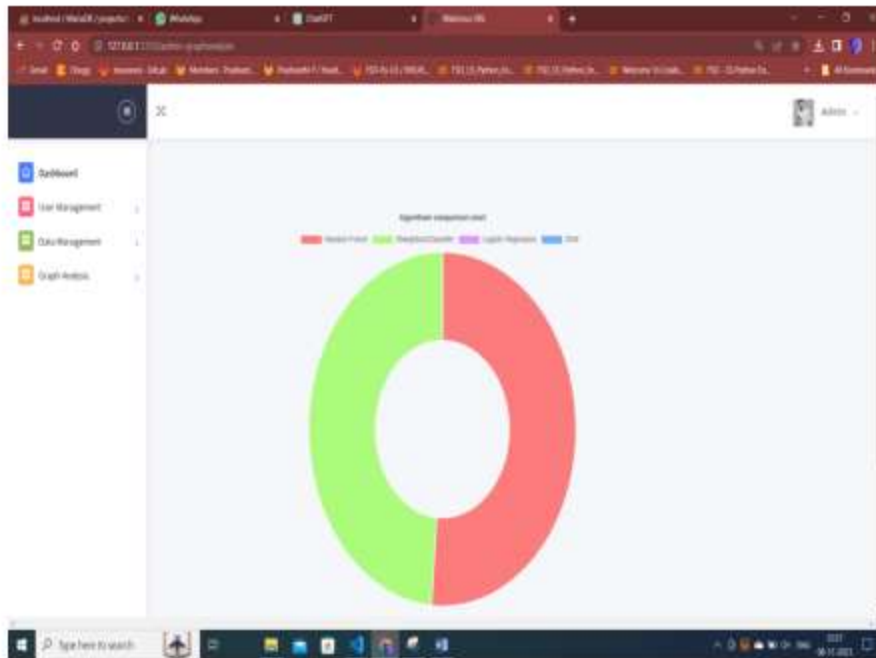
Screen 11

Random Forest:-



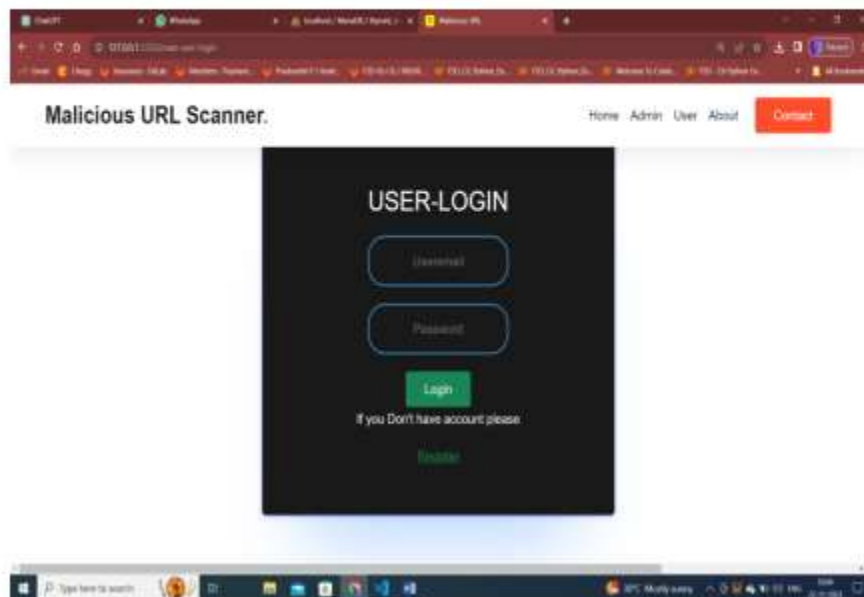
Screen 12

Graph:-



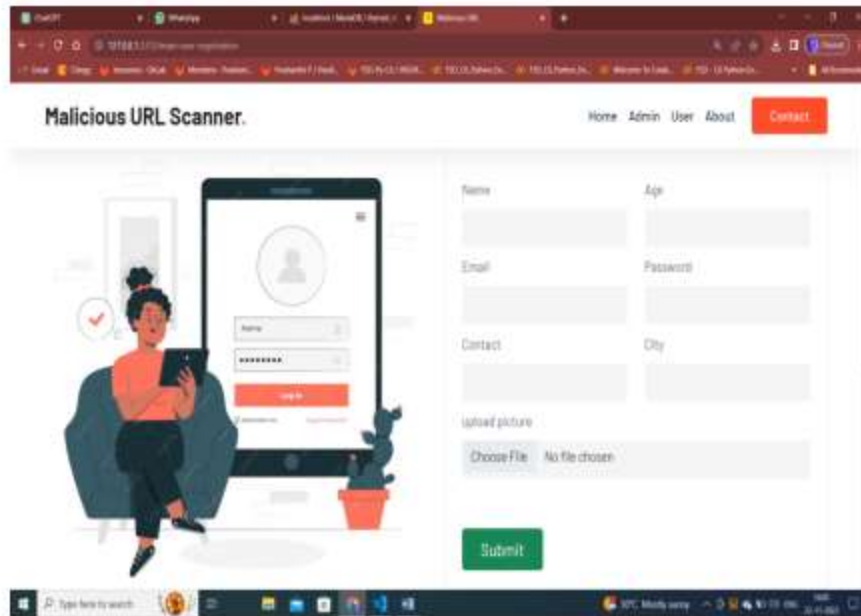
Screen 13

User Page:



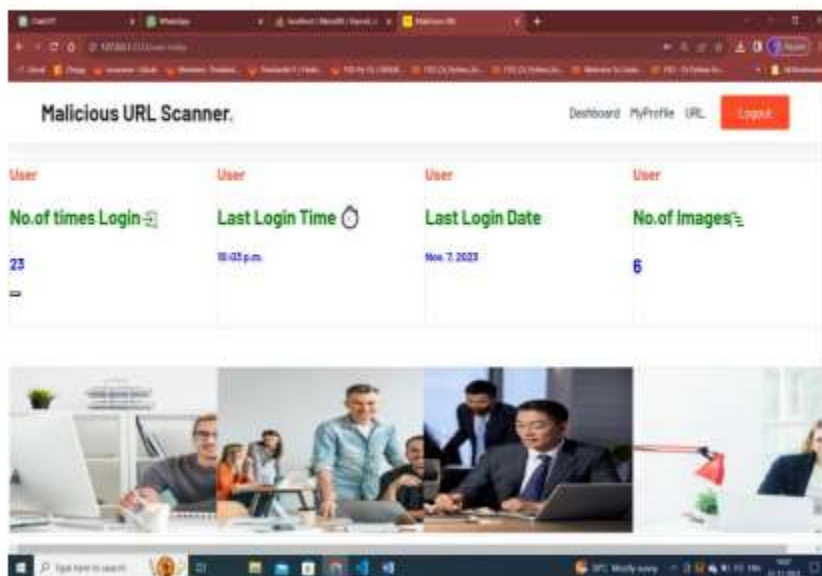
Screen 14

Register Page:



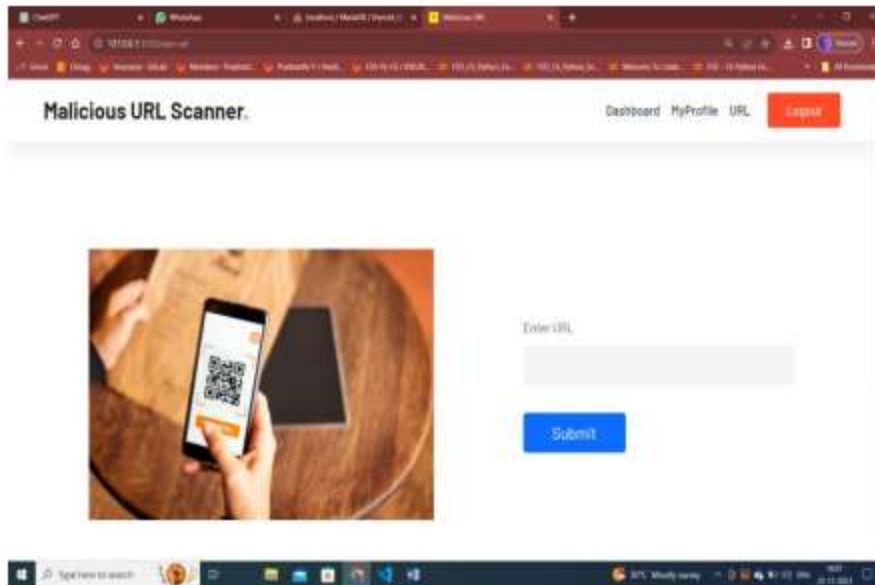
Screen 15

User dashboard:



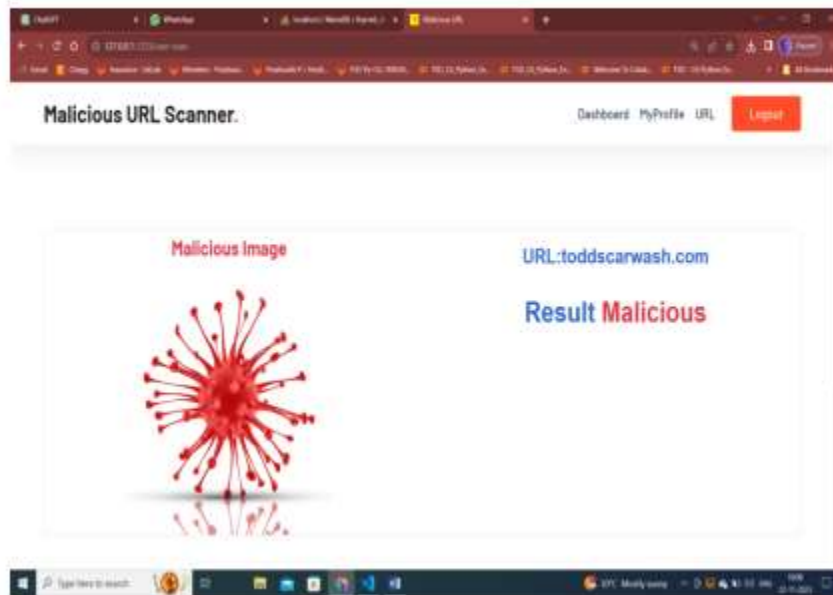
Screen 17

Url Page:



Screen 19

Result Page when bad:



Screen 20

Result page when good:

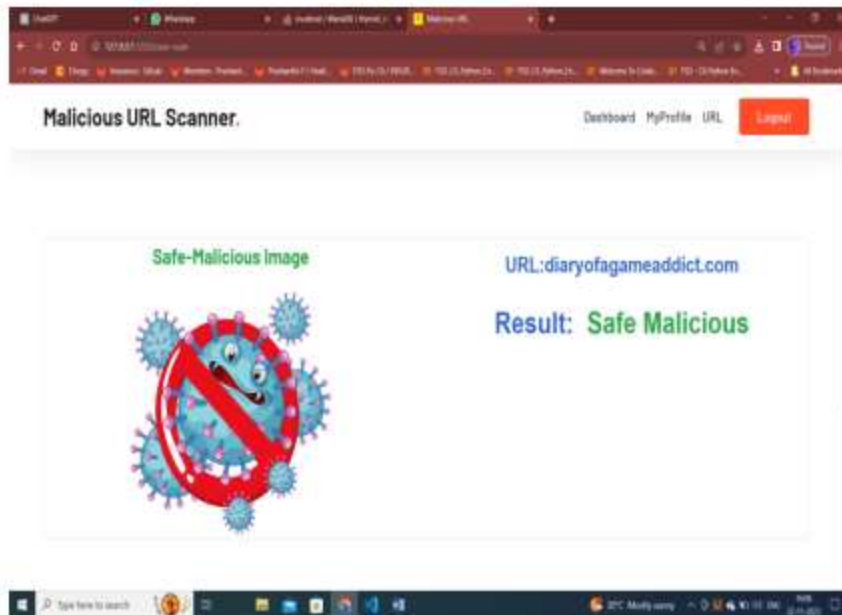


Fig 21

8.CONCLUSION

To summarise, phishing is an online platform used to steal personal details from internet users, including personal information as well as details related to online transactions, such as credit card information and net banking passwords. QR Codes are being used to carry out phishing attacks. It's because QR codes are more adaptable, convenient, and visually appealing. However, a lack of focus on QR Code security leads to issues like confidential data loss and inadequate prevention against malicious assaults. The development of QR code phishing detection tools raises the security of quick response codes to a higher level and prevent attackers from abusing QR Codes. This research is essentially for users who scan QR codes into their day to day activity, such as employees or students, and this tool will ensure that their personal details is not compromised or is not leaked while scanning a quick response code. The present detection methodology has the advantage of detecting aberrant and harmful links embedded in QR codes that can be utilised as a vector attack. Before being presented, this phishing detection tool for QR code was built and tested. The protocols that have been utilised to ensure that the scanning is done precisely and that the phishing and malicious page is detected, that are Telnet, NNTP, HTTPS, HTTP, RTSP, SFTP, FTP

and Gopher. Support Vector Machines, Random Forest, Binary LSTM, and k_Nearest Neighbor algorithms were used. The project was also built to check domain, subdomain, IP addresses, prefix and if any symbol is present is also scanned before producing results. This solution has aided in the detection and security of the application, allowing for safer scanning. This programme will allow users to read QR codes more quickly and safeguard them from fraudulent URLs. QR Codes can be used by an attacker to commit fraud. Whenever a user scans the quick response code, they may unintentionally be directed towards a bogus website. With the development of this phishing detection tool, users will be able to perform secure scans and improve the security of their QR codes.

9. REFERENCES

- [1] Katharina Krombholz, Peter Frühwirth, Peter Kieseberg et. al. “QR Code Security: A Survey of Attacks and Challenges for Usable Security”, In International Conference on Human Aspects of Information Security, Privacy, and Trust.
- [2] Bhowmick, A., & Prasad, C. G. V. N. (2017). Time and cost optimization by grid computing over existing traditional IT systems in business environment. *Int J*, 5, 93-98.
- [3] Heider A. M. Wahsheh and Flaminia L. Luccio, “Evaluating Security, Privacy and Usability Features of QR Code Readers”, In ICISSP 2019 - 5th International Conference on Information Systems Security and Privacy.
- [4] Frank Vanhoenshoven Nagamani, K., Prasad, C. G., & Chatrapati, K. S. (2016, February). A novel framework for optimal component based data center architecture. In *2016 International Conference on Information Communication and Embedded Systems (ICICES)* (pp. 1-10). IEEE.
- [5] Cho Do Xuan¹, Hoa Dinh Nguyen et. al. “Malicious URL Detection based on Machine Learning”, In (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 1, 2020.

[6] CHOLLETI, N., & HIRWARKAR, T. (2020).BIOMEDICAL DATA ANALYSIS IN PREDICTING AND IDENTIFICATION CANCER DISEASE USING DUO-MINING. *Advances in Mathematics: Scientific Journal*, 9, 3487-3495

[7] Andrey Averin, Natalya Zyulyarkina, “Malicious Qr-Code Threats and Vulnerability of Blockchain”, In 978-1-7281-8075-5/20/\$31.00 ©2020 IEEE.

[8] R. Shobarani, R. Sharmila, M. N. Kathiravan, A. A. Pandian, C. Narasimha Chary and K. Vigneshwaran, "Melanoma Malignancy Prognosis Using Deep Transfer Learning," *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*, Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICAIA57370.2023.10169528

[9] SafwatiIsmail,AlvinEbenazer Kumar et. al. “Quick Response Code Validation and Phishing Detection Tool”, In 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE) | 978-1-6654-0338-2/21/©2021 IEEE.

[10] Makoto Takita, Hiroya Okuma, et. al. “A Construction of Fake QR Codes Based on Error-Correcting Codes”, In 2018 Sixth International Symposium on Computing and Networking

(CANDAR) IEEE.