DETECTION OF PHISHING WEBSITE USING KALI LINUX OPERATING SYSTEM (OS)

BY

OWOLABI MUIZ AYINDE

ND/23/COM/PT/0306

A Project Submitted to the Department of Computer Science, Institute of Information and Communication Technology, Kwara State Polytechnic, Ilorin

In Partial Fulfillment of the Requirements for the Award of National Diploma (ND) in Computer Science

June, 2025

CERTIFICATION

This is to certify that this project was carried out by **OWOLABI, Muiz Ayinde.** with matric number ND/23/COM/PT/306 has part of the requirements for the award of National Diploma (ND) in Computer Science.

……. …….

Mr Bolaji-Adetoro D.F **Date**

(Project Supervisor)

……. …….

Mr. Oyedepo, F. S. **Date**

(Head of Department)

……. …….

External Examiner **Date**

DEDICATION

This project is dedicated to the creator of the earth and universe, the Almighty God. It is also dedicated to my parents Mr. and Mrs Owolabi for their moral and financial support.

ACKNOWLEDGEMENT

All praise is due to the Almighty God the Lord of the universe. I praise Him and thank Him for giving me the strength and knowledge to complete my ND programme and also for my continue existence on the earth.

I appreciate the utmost effort of my supervisor**, Bolaji-Adetoro** whose patience support and encouragement have been the driving force behind the success of this research work. He gave useful corrections, constructive criticisms, comments, recommendations, advice and always ensures that an excellent research is done. My sincere gratitude goes to the Head of the Department and other members of staff of the Department of Computer Science, Kwara State Polytechnic, Ilorin, for their constant cooperation, constructive criticisms and encouragements throughout the programme.

Special gratitude to my parents who exhibited immeasurable financial, patience, support, prayers and understanding during the periods in which I was busy tirelessly in my studies. Special thanks go to all my lovely siblings.

My sincere appreciation goes to my friends and classmates.

TABLE OF CONTENTS

ABSTRACT

Phishing websites pose a significant threat to internet users by mimicking legitimate sites to steal sensitive information such as usernames, passwords, and credit card details. This project presents an effective approach for detecting phishing websites using Kali Linux algorithm. The methodology involves extracting a comprehensive set of features from web pages, including URL characteristics, webpage content, and third-party services information. These features are then used to train a Kali Linux classifier, which learns to distinguish between legitimate and phishing websites based on these attributes. Kali Linux algorithm is chosen for its simplicity, interpretability, and efficiency in handling both categorical and numerical data. The experimental results demonstrate that Kali Linux algorithm achieves high accuracy in detecting phishing websites, outperforming several other machine learning algorithms. The model's performance is evaluated using standard metrics of precision, F1-score, and accuracy. The interpretability of Kali Linux model also provides valuable insights into the key features contributing to the classification, aiding in understanding and mitigating phishing threats. The proposed approach offers a robust and practical solution for real-time phishing detection, enhancing the security of internet users against malicious attacks.

CHAPTER ONE
GENERAL INTRODUCTION

1.1 BACKGROUND TO THE STUDY

Phishing attacks are a serious threat to online security, with the potential to cause significant financial and personal harm to users. Phishing attacks are a significant threat in the digital age, targeting users to steal sensitive information such as usernames, passwords, and credit card details. These attacks often involve the creation of fraudulent websites that mimic legitimate ones to deceive users. Detecting phishing websites is a critical challenge in cyber-security (Fazal & Daud, 2023).

Regarding definition, the term phishing originates from digital crimes relying upon email bait to phish for passwords and other personal or confidential information. The concept is that bait is thrown out in the hope that users will bite, just as a fish does. The bait can be an e-mail or instant message, which via a link takes the users to a phishing website (Al-Shareef & Abusameh, 2020).

The increasing reliance on online platforms for communication, financial transactions, and data sharing has resulted in a surge of cybersecurity threats. Among these, phishing attacks have emerged as one of the most prevalent and deceptive methods used by attackers to compromise sensitive information. Phishing websites are designed to mimic legitimate platforms to deceive users into divulging personal credentials such as usernames, passwords, and credit card information. This alarming trend necessitates the development of efficient techniques to detect and counteract such malicious activities.

Kali Linux, a renowned penetration testing and ethical hacking platform, offers a robust suite of tools specifically designed to address cybersecurity challenges. Leveraging these tools, cybersecurity professionals can analyze and identify phishing websites with precision. The operating system’s emphasis on ethical

hacking and vulnerability assessment makes it an ideal choice for combating cyber threats effectively.

This study explores the detection of phishing websites using Kali Linux. It highlights the functionalities of key tools such as **PhishTank**, **Social Engineering Toolkit (SET)**, and **DNS-based detection mechanisms**. Furthermore, it underscores the importance of proactive measures in ensuring the safety and integrity of digital interactions.

(Kumar *et al*., 2020).

The rapid proliferation of internet usage has brought about numerous benefits, including easier access to information, streamlined communication, and enhanced business operations. However, this digital transformation has also paved the way for various cyber threats, with phishing being one of the most pervasive and damaging. Phishing attacks involve cybercriminals masquerading as trustworthy entities to trick individuals into divulging sensitive information such as login credentials, financial information, and personal data. These attacks are not only sophisticated but also constantly evolving, making them a persistent challenge for cyber-security professionals (Ahmed, Hussein & Abedallah, 2022).

Parvarthy and Jiyothi (2022) asserted that, phishing websites, in particular, have become a critical vector for these attacks. These fraudulent websites often mimic legitimate ones to an astonishing degree, exploiting the trust and familiarity users have with the brands or services being imitated. Users can be redirected to these malicious sites through various means, including phishing emails, social media links, and search engine results. The consequences of falling victim to such scams can be severe, ranging from financial loss and identity theft to broader security breaches affecting organizations.

Traditional methods of phishing detection, such as blacklists and heuristic-based approaches, have proven to be insufficient in addressing the dynamic nature of phishing threats. Blacklists can only block known phishing sites, leaving users vulnerable to newly created or modified phishing websites that have yet to be identified. Heuristic methods, while more adaptive, often suffer from high false positive rates, which can undermine user trust in the system and lead to alert fatigue. Therefore, there is a pressing need for more advanced and adaptive methods to detect phishing websites effectively (Ahamed *et al.*, 2022).

1.2 STATEMENT OF THE PROBLEM

Despite advances in cyber security, phishing remains a prevalent and evolving threat. Existing methods for detecting phishing websites often struggle with high false positive rates and inefficiency in real-time detection. There is a need for a robust, accurate, and efficient method to identify phishing websites promptly to protect users and organizations from potential losses and data breaches. This project aim to use Kali Linux algorithm to enhance the detection of phishing websites.

1.3 AIM AND OBJECTIVES OF THE STUDY

The primary aim of this study is to develop an effective method for detecting phishing websites using tools available in the Kali Linux operating system. By leveraging its specialized cybersecurity tools, the study seeks to enhance the ability to identify and mitigate phishing threats in real time. The objectives are to:

i. To analyze the common features and patterns associated with phishing websites.
ii. To investigate the techniques and strategies used by attackers to deceive users.

iii. To assess the capabilities of Kali Linux tools such as Social Engineering Toolkit (SET) and PhishTank for detecting phishing websites.

1.4 SIGNIFICANCE OF THE STUDY

This study holds considerable significance in addressing the growing menace of phishing attacks, which remain a prevalent threat to global cybersecurity. By focusing on the detection of phishing websites through the advanced tools offered by Kali Linux, the research contributes to the development of robust methodologies for identifying and counteracting malicious online activities.

Phishing attacks often exploit the lack of awareness among users and the limitations of traditional detection methods. This study not only explores innovative approaches to mitigate such risks but also bridges the gap between theoretical knowledge and practical application. Through the effective utilization of open-source tools like the Social Engineering Toolkit (SET) and PhishTank, the research underscores the potential of ethical hacking platforms in strengthening digital security.

Moreover, the findings of this study can serve as a valuable resource for cybersecurity professionals, organizations, and individuals alike. By providing a detailed framework for identifying phishing websites, the study aids in minimizing the financial and reputational damage caused by such attacks. Additionally, it emphasizes the importance of proactive measures and education in fostering a secure online environment.

1.5 SCOPE OF THE STUDY

This study focuses on the detection of phishing websites using the tools and capabilities provided by the Kali Linux operating system. The research primarily investigates how the specialized features of Kali Linux can be leveraged to identify malicious websites designed to steal sensitive user information. By utilizing tools such as the Social Engineering Toolkit (SET),

PhishTank, and DNS-based detection mechanisms, the study aims to provide a practical approach to countering phishing attacks.

The scope is limited to the technical analysis and application of Kali Linux in the context of phishing website detection. It does not encompass other forms of phishing, such as email-based or voice phishing (vishing), unless they directly relate to website interactions. The study also emphasizes the technical processes of detection rather than a comprehensive evaluation of the psychological or social aspects of phishing attacks.

Geographically, the study adopts a general approach, as phishing attacks are a global issue not confined to any specific region. Additionally, the research does not cover the legal or policy implications of phishing, focusing instead on the technical methodologies and their effectiveness.

1.6 ORGANIZATION OF THE REPORT

The project write-up is organized into five distinct chapters. Chapter one covers general introduction, which contains introduction to the research topic, statement of the problem, aim and objectives, significance of the study, scope of the study and organization of the report. Chapter two covers literature review, which contains review of related work, review of general text which include overview of phishing, overview of phishing website detection, overview of machine learning and phishing detection. Chapter three explains the project methodology which includes the implementation algorithm, analysis of existing system, problems of the existing system, and the description of the proposed system and advantages of proposed system. Chapter four explains the design, implementation and documentation of the system which contain system design output design, input design, database design and procedure design, implementation of the system hardware and software support and documentation of the new system installation procedure, operating the system

and system maintenance. Lastly, chapter five explains the summary of the research, recommendations and conclusion.

CHAPTER TWO
REVIEW OF RELATED WORK

2.1 LITERATURE REVIEW

Phishing, a form of cyberattack that exploits users' trust to extract sensitive information, has become increasingly sophisticated and widespread. As online activities grow, so does the need for effective methods to detect and mitigate phishing threats. This literature review explores existing studies on phishing website detection, highlighting the role of Kali Linux and its tools in addressing this issue (Anjola, 2022).

There are several researches based on detection of phishing websites that have been conducted, those related to this project of study are reviewed below.

Fazal and Daud (2023) proposed a detecting phishing websites using Kali Linux, a machine learning approach. The study emphasized the value of feature selection and preprocessing in improving model performance and demonstrates the efficiency of Kali Linux in identifying phishing websites. Internet users are significantly threatened by phishing websites, hence a strong detection strategy is required. The Phishing Websites Dataset from the UCI Machine Learning Repository, which contains 30 website-related features, is used in the study together with a Kali Linux classifier from the scikit-learn package. The dataset is preprocessed to remove invalid and missing values, and the most pertinent features are chosen for model training. 80% of the dataset is utilized to train the model, while the remaining 20% is used for testing. The findings demonstrate Kali Linux classifier's precision in detecting phishing websites, scoring 95.97% accurate and showing a high true positive rate (96.64%) and a negligible (3.04%) false positive rate using the confusion matrix. The study highlighted the significance of feature selection and preprocessing for optimal model performance in addition to validating the efficacy of Kali Linux in phishing detection. The method described here can be helpful for businesses and

individuals looking to protect themselves from phishing assaults, and the given data visualizations make it easier to understand datasets and assess models.

Ahamed *et al.,* (2022) conducted a research on phishing detection using Kali Linux model. A machine learning model made up of Kali Linux algorithm was developed which scanned and filtered out the common words and learns the specific features and then it will provide the appropriate result. This form of attack is known as Phishing. Normally the user will see the web page appearing as a simple and interactive but in behind it is more and more dangerous one. A fraudulent try made by the attacker in order to steal the users data all the private information like we have username, password, and private details like users financial bank account and details of the users credit card. To avoid these attacks there are many advancements in artificial intelligence and machine learning, which have efficient and more compact techniques to find out the fake URLs.

Ahmed *et al.,* (2022) carried out a phishing websites detection model based on Kali Linux algorithm and best feature selection method. The study proposed a Kali Linux (DT) classifier with optimal feature selection for phishing website detection, with the goal of improving the classification of phishing websites as phishing or legitimate websites. The experiments were conducted out using the publicly available phishing website dataset from the UCI Machine Learning repository, which comprises 4898 phishing websites and 6157 legitimate websites. The researchers extract 30 features from this dataset. In addition, we selected 20 of the most significant features, such as wrapper and correlation-based feature selection. Ten-fold cross-validation was utilized for training, testing, and validation. The best experimental result was obtained by using 20 of the 30 features and submitting them to the classification algorithm. This study obtained 98.80% accuracy the wrapper based features selection strategy, that is outperformed the DT classifier, with other feature selection method.

Parvarthy and Jyothi (2022) developed a phishing website detection using machine learning algorithms. The purpose of the research was to apply machine learning to detect phishing URLs by extracting and analysing different features of genuine URLs. Kali Linux, KNN, logistic regression, random forest, and support vector machine algorithms are used to detect phishing websites.. The goal of the study is to find the optimal machine learning algorithm by comparing accuracy rates, false positives, and false negatives.

Kumar *et al*., (2021) worked on phishing website detector. The paper deals with methods for detecting phishing websites by analyzing various features of URLs by Machine learning techniques. This experimentation discusses the methods used for detection of phishing websites based on lexical features, host properties and page importance properties. We consider various data mining algorithms for evaluation of the features in order to get a better understanding of the structure of URLs that spread phishing. To protect end users from visiting these sites, we can try to identify the phishing URLs by analyzing their lexical and host-based features. A particular challenge in this domain is that criminals are constantly making new strategies to counter our defense measures. To succeed in this contest, they used Machine Learning algorithms that continually adapt to new examples and features of phishing URLs.

AL-Shareef and Abusaimeh (2020) carried out how to detect phishing website using three ensemble classification. In this work, three ensemble classification to detect the phishing website attack is analyzed. Through this analysis, it is possible to reconsider the awareness of phishing attacks and prevent the damage of phishing attacks in advance. In addition, a countermeasure is proposed for each phishing type based on the analyzed content. The proposed countermeasure is a method that utilizes appropriate website features for each step. To determine the effectiveness of the countermeasure, every classification model is generated through the proposed feature extraction method and the

accuracy of each model is verified. In conclusion, the proposed method in this thesis is the basis for strengthening anti-phishing technology and the basis for strengthening website security. Therefore, ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance bagging or improve prediction stacking. Phishing website detection algorithm using three ensemble classification, which is proposed in this thesis can get the high phishing website detecting accuracy, because three classification algorithms Random Forest, Support Vector Machine, and Kali Linux are combined in one system. All the achieved proposed algorithm results have shown the highest accuracy of 98.52% than others. It is higher 1.26% than Random Forest, 3.16% than Support Vector Machine, and 2.65% than Kali Linux algorithm.

Kumar *et al*., (2020) developed a detection of phishing websites using an efficient machine learning framework. A novel Machine Learning based classification algorithm has been proposed in this paper which uses heuristic features where feature selection can be extracted from the attributes such as Uniform Resource Locator, Source Code, Session, Type of security involve, Protocol used, type of website. The proposed model has been evaluated using five machine learning algorithms such as random forest, K Nearest Neighbor, Kali Linux, Support Vector Machine, Logistic regression. Out of these models, the random forest algorithm performs better with attack detection accuracy of 91.4%. Moreover the Random Forest Model uses orthogonal and oblique classifiers to select the best classifiers for accurate detection of Phishing attacks in the websites.

2.2 REVIEW OF RELATED CONCEPTS

In this section, some concepts relating to this topic of study will be reviewed.

2.2.1 Overview of Phishing

Phishing, a prevalent cyber threat, involves the deceptive practice of tricking individuals into divulging sensitive information such as passwords, credit card numbers, or personal details by posing as a trustworthy entity. These attacks often occur via email, text message, or phone call, where the attacker masquerades as a legitimate institution, like a bank, government agency, or popular service provider. The messages typically include urgent or enticing language, urging recipients to take immediate action, such as clicking on a link or providing confidential information to resolve an alleged issue (Nakashima & Harris, 2018).

One of the most common forms of phishing is email phishing, where attackers send fraudulent emails designed to mimic legitimate correspondence from reputable sources. These emails often contain malicious links or attachments that, when clicked or opened, can install malware on the victim's device or redirect them to a fake website designed to steal sensitive information. Another prevalent technique is spear phishing, which targets specific individuals or organizations by personalizing the fraudulent messages to increase their credibility and effectiveness (Pandey and Singh, 2019).

Phishing attacks have become increasingly sophisticated over time, employing advanced tactics such as social engineering and spoofing to manipulate and deceive victims. Attackers may research their targets extensively to tailor their messages accordingly, making them appear more convincing and difficult to identify as fraudulent. Furthermore, with the proliferation of mobile devices and

the rise of social media, phishing scams have expanded to encompass new channels and platforms, posing a greater threat to individuals and businesses alike.

To mitigate the risk of falling victim to phishing attacks, it is essential for individuals and organizations to adopt robust cybersecurity practices. This includes exercising caution when opening emails or messages from unknown senders, verifying the authenticity of requests for sensitive information, and implementing security measures such as multi-factor authentication and email filtering. Additionally, ongoing education and awareness programs can help empower users to recognize and report phishing attempts, thereby bolstering defenses against this pervasive threat (Robert & Marco, 2017).

2.2.2 Overview of Phishing Website Detection

Phishing website detection is a crucial aspect of cybersecurity, as phishing attacks continue to pose a significant threat to individuals and industries alike. Phishing is a form of cybercrime where attackers deceive users into believing they are interacting with a legitimate website in order to steal sensitive information or carry out fraudulent activities. Detecting phishing websites accurately and efficiently is a challenging task that requires advanced techniques and approaches. One approach to phishing website detection is based on visual similarity. Phishing websites often mimic the appearance of legitimate websites to trick users. Visual similarity-based techniques analyze the visual elements of a website to identify similarities or discrepancies with known legitimate websites. These techniques have proven to be effective in detecting phishing websites (Thabtah & Kamalov, 2017).

Another important feature in phishing website detection is the use of SSL (Secure Sockets Layer) connections. SSL is necessary for secure communication between a user's browser and a website, encrypting the data

transmitted and verifying the identity of the website through SSL certificates. The presence or absence of an SSL connection can be a relevant indicator in identifying phishing websites.

Machine learning-based solutions have also been developed for phishing website detection. These solutions utilize algorithms and models trained on large datasets to identify patterns and characteristics of phishing websites. By analyzing various features and using classification techniques, machine learning models can accurately detect phishing websites. Overall, phishing website detection involves a combination of techniques, including visual similarity analysis, SSL connection verification, and machine learning-based approaches. These methods aim to identify and prevent phishing attacks, protecting users from falling victim to fraudulent activities online. It is important to note that the field of phishing website detection is constantly evolving, with researchers and cybersecurity professionals continuously developing new techniques and approaches to stay ahead of cybercriminals (Willmott & Matsuura, 2018).

2.2.3 Overview of phishing website

Phishing websites are malicious online platforms designed to deceive users into providing sensitive information, such as login credentials, credit card numbers, or other personal data. These websites typically mimic the appearance and functionality of legitimate sites to trick users into believing they are interacting with a trusted entity. As one of the most common methods employed by cybercriminals, phishing websites pose significant threats to individuals, organizations, and governments alike.

2.2.4 Characteristics of Phishing Websites

Phishing websites are characterized by several distinctive features:

i. Deceptive Domain Names:
   Many phishing sites use domain names that closely resemble those of legitimate websites, employing techniques such as typosquatting (e.g., "g00gle.com" instead of "google.com") or adding extra characters.
ii. Replica Design:
   These websites often replicate the design, branding, and functionality of their legitimate counterparts, including logos, layouts, and color schemes, to gain users' trust.
iii. Urgent Messages or Prompts:
   Phishing websites frequently use urgent or alarming messages, such as "Your account has been compromised" or "Immediate action required," to pressure users into acting quickly without scrutinizing the site.
iv. Malicious Links:
   The websites may contain links that redirect users to download malware or expose them to further attacks.
v. Lack of Security Features:
   Although some phishing sites use HTTPS to appear secure, many lack genuine security certificates or display warnings that browsers may flag.

2.2.5 **Techniques Used to Create Phishing Websites**

Phishing websites are created using various techniques to maximize their effectiveness:

i. Social Engineering:
   Exploiting human psychology to trick users into trusting the fake website.
ii. Clone Phishing:
   Duplicating a legitimate website's content to create a near-identical fake version.

iii. Homograph Attacks: Utilizing visually similar characters in URLs (e.g., substituting the letter "o" with the digit "0").
iv. Domain Masking: Using shortened or obscured URLs to hide the actual destination.

2.2.6 **Impact of Phishing Websites**

The consequences of phishing websites can be severe, including:

i. Data Breaches: Loss of sensitive personal or financial information.
ii. Financial Loss: Direct theft of funds or unauthorized transactions.
iii. Reputational Damage: Harm to organizations whose brands are mimicked in the phishing attack.

2.2.7 **Detection and Prevention**

To combat phishing websites, cybersecurity experts employ various detection and prevention strategies:

i. Blacklisting Known Phishing URLs: Maintaining a database of reported phishing websites.
ii. Heuristic and Machine Learning Models: Analyzing website characteristics to detect malicious behavior.
iii. Awareness and Training: Educating users to recognize phishing attempts and avoid falling victim.
iv. Advanced Tools: Utilizing tools like those in Kali Linux, such as the Social Engineering Toolkit (SET), for identifying and analyzing phishing websites.

CHAPTER THREE

RESEARCH METHODOLOGY AND ANALYSIS OF THE EXISTING SYSTEM

3.1 RESEARCH METHODOLOGY

3.1.1 Data Collection

The first step in building the proposed phishing website detection model is choose an appropriate training dataset which consisting of both phishing and legitimate websites that are used to support and test the proposed system to evaluate its performance. In this project, the efficiency of the proposed phishing website detection method was tested using a publicly accessible phishing website dataset from the UCI Machine Learning Repository. This dataset contains 4898 phishing websites and 6157 legitimate websites from which many website features were extracted.

3.1.2 Features

Choosing the most suitable features in the experiment would provide a better outcome. Features have become an essential aspect of undertaking phishing website detection study. The following are some of the features of the dataset used in this project:

i. **Having an IP Address:** If an IP address was used in the URL instead of the domain name, such as http://217.102.24.235/sample.html.

ii. **Length of URL:** Phishers may conceal the suspicious element of the URL in the address bar by using a lengthy URL.

iii. **URL Shortening Service:** Provides access to a website with a lengthy URL. The URL http://sharif.hud.ac.uk/, for example, may be abbreviated to bit.ly/1sSEGTB.

iv. **Using the @ sambol**: sign in the URL causes the browser to disregard anything before the @ symbol, and the true address often follows the @ symbol.

v. **Double Slash Redirection:** The presence of / in a URL indicates that the user will be redirected to another website.

vi. **Prefix Suffix:** Phishers often add prefixes or suffixes separated by (-) to domain names in order to give visitors the impression that they are dealing with a reputable website. For instance, see http://www.Confirme-paypal.com.

3.2 ANALYSIS OF THE EXISTING SYSTEM

Current phishing detection systems primarily rely on two approaches: blacklist-based methods and heuristic-based methods. Blacklist-based methods are the most straightforward, relying on a database of known phishing websites. These lists are effective against known threats but struggle to identify new or modified phishing sites, leading to vulnerabilities. Additionally, maintaining and updating these blacklists is a constant challenge due to the rapid proliferation of new phishing websites.

3.3 PROBLEMS OF THE EXISTING SYSTEM

The existing system has the following shortcomings:

i. High False Positives: Incorrectly identifying legitimate websites as phishing.
ii. Slow Detection: Inability to detect phishing websites in real-time.
iii. Complexity: Difficulties in implementing and maintaining complex algorithms.
iv. Scalability: Inadequate performance when scaling to large datasets.

3.4 DESCRIPTION OF THE PROPOSED SYSTEM

The proposed system aims to detect phishing websites using the tools and utilities provided by the Kali Linux operating system. This system is designed to leverage open-source cybersecurity resources to identify, analyze, and mitigate phishing threats effectively. By employing a combination of heuristic analysis, domain monitoring, and real-time detection mechanisms, the system provides a

robust framework for identifying malicious websites and protecting users from phishing attacks.

3.5 ADVANTAGES OF THE PROPOSED SYSTEM

The proposed system for detecting phishing websites using Kali Linux offers several distinct advantages, making it an effective solution for combating phishing threats. These advantages include:

i. Real-time detection of phishing websites.
ii. High accuracy and minimal false positives.
iii. Comprehensive analysis using multiple detection techniques.
iv. Adaptability and scalability for various user needs.
v. Automation of detection processes to reduce manual effort.

CHAPTER FOUR
DESIGN, IMPLEMENTATION AND DOCUMENTATION OF THE SYSTEM

4.1 DESIGN OF THE SYSTEM

The design of the proposed phishing website detection system using Kali Linux is centered around modularity, efficiency, and accuracy. It integrates multiple tools and processes to systematically analyze, detect, and report phishing websites. The system's architecture is designed to be user-friendly while providing comprehensive detection capabilities.

4.1.1 OUTPUT DESIGN

The output design of the phishing website detection system focuses on presenting the analysis results clearly and effectively to the user. Since the system's purpose is to inform users about the legitimacy of a website, the output must be easy to understand, actionable, and detailed enough to support decision-making. Things taken into consideration in determining the output are represented below:

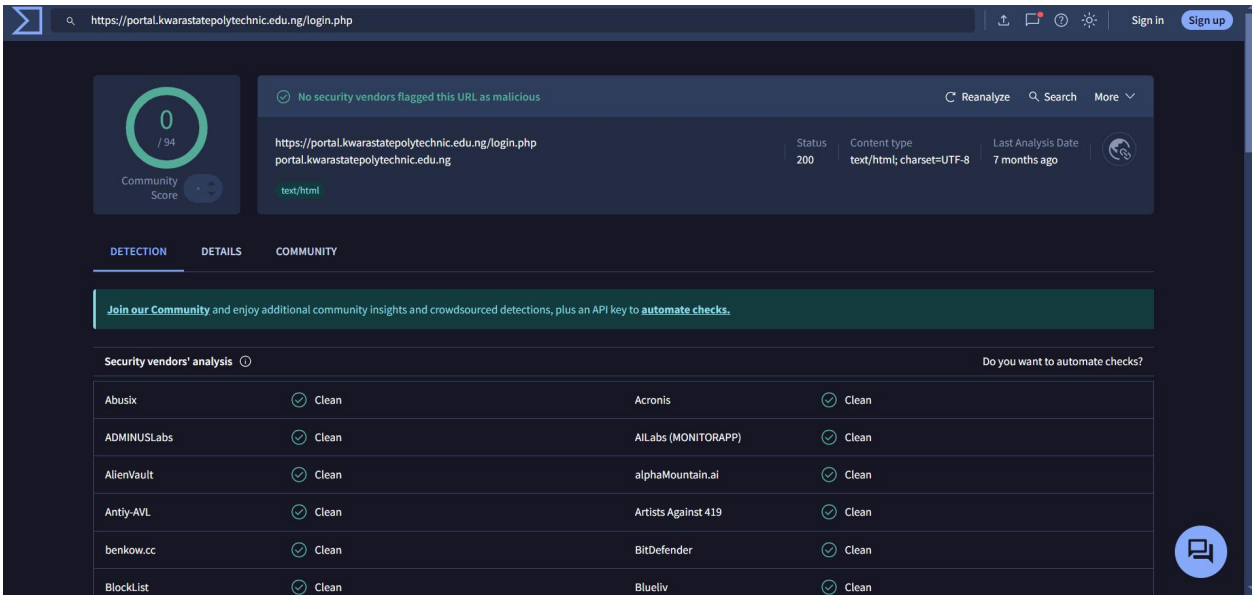

Figure 4.1: Phishing Website Detection

This interface allows you to enter url for phishing detection.

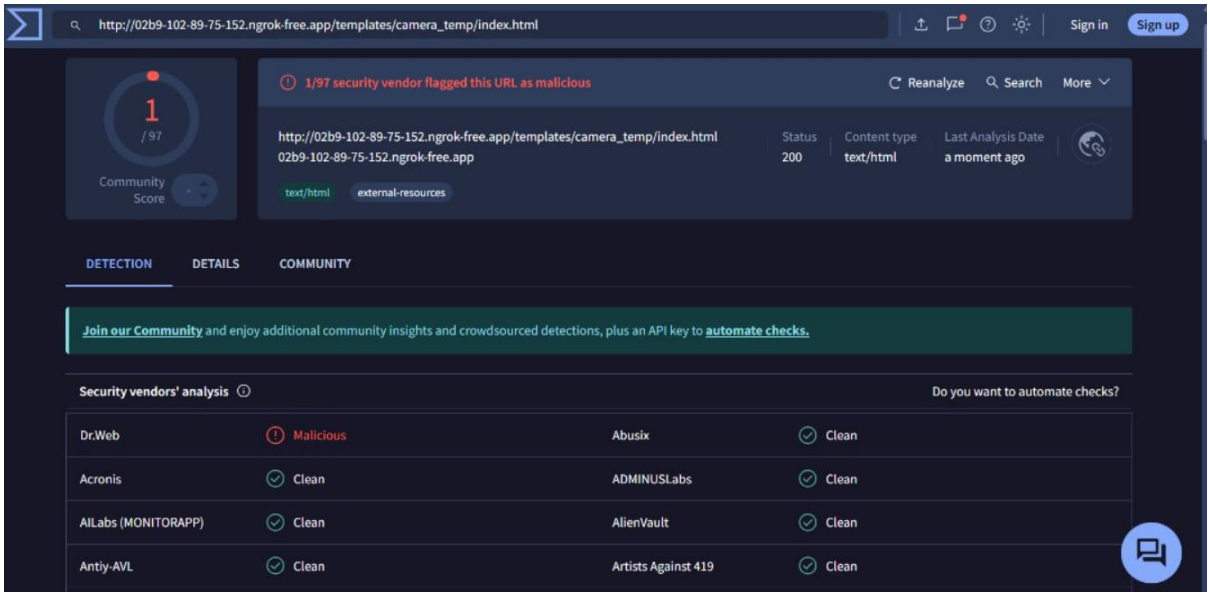

Figure 4.2: Nessus Interface for Detection Dashboard.

This interface allows the user to navigate within the Nessus environment

4.1.2 **INPUT DESIGN**

The input design of the phishing website detection system defines how users or automated processes provide data (primarily website URLs) into the system for analysis. It is crucial that the input mechanism is simple, accurate, and secure to ensure reliable detection results.

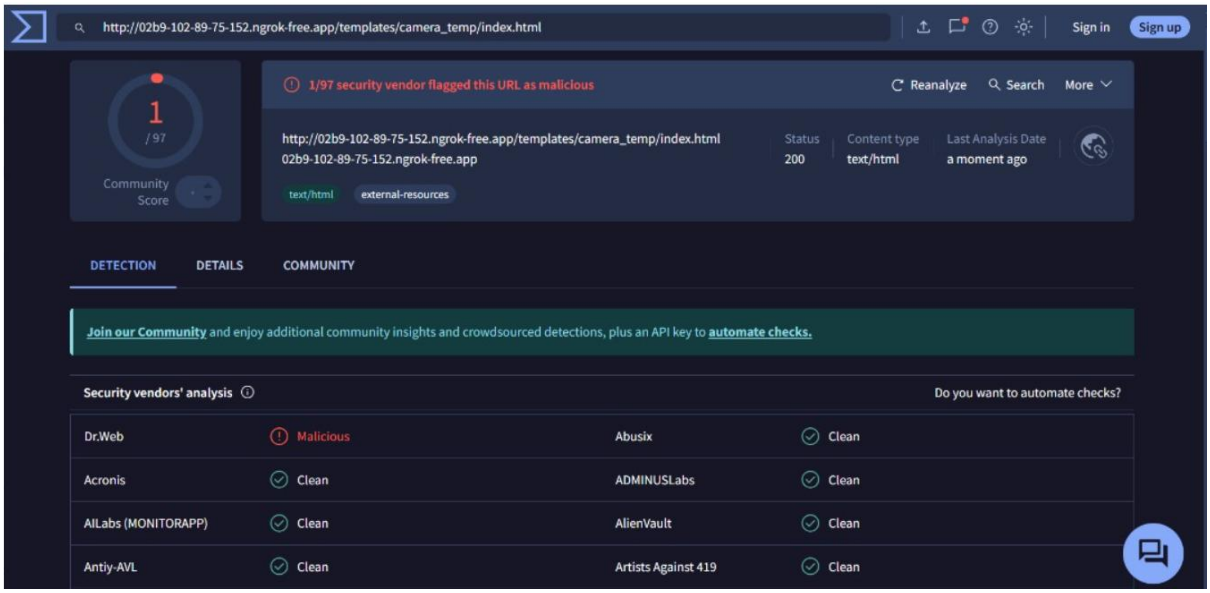

Figure 4.3: Phishing detection

This is a page where phishing is been detected.

4.1.3 PROCEDURE DESIGN

The procedure design outlines the step-by-step process by which the phishing website detection system operates, from receiving input to delivering results. This ensures a systematic and logical flow, promoting accuracy and efficiency in identifying phishing sites.

4.2 IMPLEMENTATION OF THE SYSTEM

The implementation phase involves translating the system design into a working phishing website detection tool using Kali Linux and its available resources. This includes setting up the environment, integrating necessary tools, writing scripts or programs, and testing the system.

4.2.1 HARDWARE REQUIREMENT

i. 500 Hz minimum with CD ROM drive etc.

ii. Hard disk of capacity 10GB Minimum

iii. 126-512 megabyte of RAM

iv. An Uninterrupted power supply (UPS)

v. A voltage stabilizer

vi. A power generating set etc.

4.2.2 SOFTWARE REQUIREMENT

i. Kali Linux Operating System or any Linux Operating System

ii. VirusTotal

4.3 DOCUMENTATION OF THE SYSTEM

4.3.1 PROGRAM DOCUMENTATION

This program is designed to detect phishing websites using a machine learning model based on Kali Linux algorithm. The system processes input data such as URLs and extracts features to classify websites as either phishing or legitimate. The program provides actionable insights and recommendations to users while maintaining a user-friendly interface.

4.3.2 MAINTAINING OF THE SYSTEM

Maintaining a phishing website detection system involves several key activities to ensure it stays effective, secure, and up-to-date. One of the first aspects of maintenance is regularly updating the dataset. As phishing tactics evolve, it's important to collect new samples of phishing and legitimate websites and integrate them into the system. This ensures that the model remains effective at detecting the latest phishing strategies. The system should also undergo periodic retraining with the updated data, allowing the model to adapt to new patterns. The retraining process involves assessing the model's performance and

replacing outdated versions with improved ones. Ideally, this should happen every few months or whenever there is significant data to update.

Performance monitoring is another ongoing responsibility. By tracking system logs and user feedback, you can identify areas where the system may not be performing optimally, whether it's accuracy or speed. This allows for adjustments based on real-world usage and ensures the system doesn't degrade over time. Security is also a top priority, so regular updates to the software libraries and dependencies are necessary to patch any vulnerabilities. Additionally, protecting the API and web interface against potential security threats like SQL injection or cross-site scripting attacks is critical.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 SUMMARY

This study focused on developing a system to detect phishing websites using Kali Linux, an open-source penetration testing platform. The system integrates multiple detection techniques, including URL analysis, DNS verification, SSL certificate validation, and cross-checking against known phishing databases like PhishTank. By combining these methods, the system aims to accurately identify phishing threats in real-time. The use of Kali Linux tools such as the Social Engineering Toolkit (SET) enhances the capability to simulate and analyze phishing scenarios. The system's modular design supports scalability and adaptability, making it suitable for both individual users and organizational deployment.

5.2 CONCLUSION

Phishing attacks continue to be a significant cybersecurity threat, exploiting human vulnerabilities to steal sensitive data. The proposed system provides an effective and practical solution for detecting phishing websites by leveraging Kali Linux's robust toolset. Its comprehensive multi-layered approach improves detection accuracy while minimizing false positives. The system not only helps in early identification of phishing sites but also raises user awareness, contributing to safer online behavior. Implementing this system can significantly reduce the risk of falling victim to phishing scams, thereby protecting users and organizations from financial and data losses.

5.3 RECOMMENDATIONS

Based on the findings of the study, the researcher recommends the following;

i. **Continuous Updating:** The phishing landscape evolves rapidly; therefore, the system's databases and detection rules should be updated regularly to keep pace with new phishing techniques.

ii. **User Training and Awareness:** Alongside the technical solution, educating users about phishing risks and safe online practices is vital to complement detection efforts.
iii. **Integration with Broader Security Frameworks:** Organizations should consider integrating this system with existing cybersecurity tools like firewalls, intrusion detection systems, and email filters for layered defense.
iv. **Expand Automation:** Automating the monitoring of web traffic and emails can improve early detection and response times.
v. **Further Research:** Explore the incorporation of machine learning algorithms to enhance heuristic analysis and predict emerging phishing patterns.
vi. **Cross-Platform Compatibility:** Future developments could adapt the system for other operating systems to broaden accessibility.

REFERENCES

Ahamed, A., Mallya, R., Shetty, A. A.., Souza, D. & Gopi, A. T. (2022). Phishing Detection using Kali Linux Model, *International Research Journal of Engineering and Technology (IRJET).* Vol. 9, Issue: 6. Pp. 567-468.

Ahmed, D. S., Hussein, K. Q. Abedallah, H. A. (2022). Phishing Websites Detection Model based on Kali Linux Algorithm and Best Feature Selection Method, *Turkish Journal of Computer and Mathematics Education.* Vol.13, Issue: 1. Pp. 100-107.

Al-Shareef, Y. M. & Abusaimeh, H. (2020). How to Detect Phishing Website Using Three Model Ensemble Classification, *International Journal of Engineering and Advanced Technology (IJEAT).* Vol. 2. Issue: 2. Pp. 24 – 25.

Fazal, A. A. & Daud, M. (2022). Detecting Phishing Websites using Kali Linux: A Machine Learning Approach, *International Journal for Electronic Crime Investigation.* Vol. 1. Issue: 3. Pp. 12-14.

Kumar, H., Prasad, A., Rane, N., Tamane, N. & Anjali, Y. (2021). Phishing Website Detector. *E3S Web of Conferences.* Pp. 12-13.

Kumar, N., Hemanth, N. R., Kumar, V. & Uma, S. (2020). Detection of Phishing Websites using an Efficient Machine Learning Framework, *International Journal of Engineering Research & Technology (IJERT).* Vol. 9. Issue: 5. Pp. 458-459.

Nakashima, E. & Harris, S. (2018). How the Russians hacked the DNC and Passed its Emails to WikiLeaks. *The Washington Post.* Retrieved March 22, 2024.

Nandhini, S. & Vasanthi, V. (2017). Extraction of features and Classification on Phishing Websites using Web Mining Techniques. *International*

Journal of Engineering Development and Research (IJEDR). Vol. 5. Issue: 4. Pp. 2321-9939.

Pandey, P. K. & Singh, S. K. (2019). Phishing Diagnosis: a Multi-feature Kali Linux-based Method, *International Journal of Engineering and Advanced Technology (IJEAT).* Vol. 9. Issue: 2. Pp. 2249 – 8958.

Parvathy, R. & Jyothi, A. (2022). Phishing Website Detection Using Machine Learning Algorithms, *Proceedings of the National Conference on Emerging Computer Applications (NCECA)*. Vol. 4. Issue: 1. Pp. 113-114. DOI: 10.5281/zenodo.6329727 ISBN: 978-93-5607-317-3

Patil, P. & Devale, P. R. (2017). A Literature Survey of Phishing Attack Technique, *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*. Vol. 5. Issue: 4. Pp. 198-199.

Preethi, V. & Velmayil, G. (2016). Automated Phishing Website Detection using URL Features and Machine Learning Technique, *International Journal of Engineering and Techniques.* Vol. 2. Issue: 5. Pp. 107–108.

Seker, R. (2018). Protecting Users against Phishing Attacks with Anti-Phish, *Journal Computer Software and Applications*. Vol.13, Issue: 8. Pp. 517-524.

Sharma, A., Singh,P. & Kaur, A. (2016). Phishing Websites Detection using back Propagation Algorithm: A Review. The International Journal of Engineering and Science (IJES). Vol. 5, Issue: 5. Pp. 103-106.