



Integrated Intrusion And Phishing Detection Framework Using Advanced Learning Techniques

¹Sowmya Sree.M, ²Sridevi Malipatil,

¹M.Tech Student, ²Associate Prof,

¹Computer Science and Engineering Department,

^{1&2} RYM Engineering College, Ballari, VTU , Belagavi, India

Abstract: The goal of this project “Integrated Intrusion And Phishing Detection Framework Using Advanced Learning Techniques” represents a considerable risk to online security. To address this risk, this project presents Fresh-Phish, an open-source and extensible system aimed at identifying phishing websites by using machine learning. By incorporating 29 unique characteristics, Fresh-Phish strives to accurately discern between phishing and authentic websites, thereby enhancing the protection of internet users against malicious threats. Fresh-Phish's methodology enhances existing techniques by resolving three essential challenges. First, it establishes a systematic framework for feature extraction as well as dataset management, ensuring that the data set is both current and comprehensive. Second, it minimizes dependence on an extensive array of features by focusing on 29 critical characteristics, which are substantiated by evidence of their significance in training the machine learning classifier. Lastly, Fresh-Phish seeks to alleviate dataset bias by ensuring a balanced representation of URL and content-based attributes, thereby improving the classifier's capacity to generalize and effectively identify phishing websites.

Index Terms – Phishing Detection, machine learning, feature extraction, Fresh-Phish.

I. INTRODUCTION

The internet has changed how we live our lives, work environments, and modes of communication, providing unparalleled convenience and connectivity. Attacks by phishers have increased in frequent and sophisticated, exposing internet users to the risks of identity theft, financial loss, and several kinds of cybercrime. Detecting phishing websites presents a significant challenge because to the constantly evolving nature of these attacks. Although numerous countermeasures have been suggested by academic institutions, businesses, and research organizations, phishing attacks remain a considerable threat. Algorithms for machine learning present interesting technique to differentiate between authentic and phishing websites, utilizing patterns in website characteristics to detect malicious intent. The primary limitation is the absence of a comprehensive framework for feature extraction and the maintenance of an up-to-date dataset of both phishing and genuine websites.

This leads to outdated and incomplete datasets, which in turn diminishes the accuracy of phishing detection. Additionally, existing algorithms frequently utilize an excessive number of features, with scant evidence supporting the significance of these features in differentiating between phishing and authentic websites. This not only heightens computational complexity but also complicates the interpretation of results. The field of phishing detection and cybersecurity is essential in the contemporary digital landscape, where the internet is integral to everyday activities. Phishing attacks pose a considerable risk to individuals, businesses, and organizations, as they aim to exploit human weaknesses to gain unauthorized access to confidential information. These attacks are frequently executed through emails, websites, or messages that seem to originate from trustworthy sources, rendering them challenging to recognise and defend against.

Attacks using phishing have progressed over time, becoming increasingly sophisticated and harder to differentiate from genuine communications. Attackers employ various strategies to mislead users, including the creation of counterfeit websites that replicate the look of reputable brands or institutions. These fraudulent websites often encourage users to input sensitive information, which is subsequently captured by the attackers. Consequently, phishing attacks can result in financial losses, identity theft, and other types of cybercrime, underscoring the necessity for effective phishing detection mechanisms.

II. DESIGN

The architecture of the Fresh-Phish system is crafted with a focus on modularity, scalability, and extensibility, facilitating the seamless integration of fresh attributes and algorithms. The architecture consists of several essential components, each fulfilling a distinct role in the phishing detection process:

1. User Interface (UI): The UI offers a web-based platform for user engagement. It enables users to upload websites for analysis, view the results of the analysis, and provide feedback. Developed using Flask, a lightweight Python web framework, the UI incorporates interactive elements to improve user experience and usability.

2. Feature Extraction Module: This module is responsible for obtaining relevant characteristics from websites for the objective of phishing detection. It encompasses components for extracting URL-based features (e.g., domain age, URL length), content-based features (e.g., the presence of specific keywords), and server-based features (e.g., server location, SSL certificate). These features act as input for the machine learning algorithms.

3. Dataset Management Module: This module is tasked with the collection, verification, and maintenance within the dataset comprising phishing and trustworthy websites. It includes systems for adding new data, verifying the genuineness of websites, and regularly updating the dataset to guarantee its accuracy and relevance.

4. Machine Learning Module: This module integrates methods for training with machine learning classifiers that distinguish between phishing and authentic websites. Supporting various algorithms, users can choose the most efficient one for a specific dataset. This module features components for training, testing, and evaluating classifiers, along with systems for algorithm selection and fine-tuning.

5. Database: Employs SQLite for effective database management, storing website data, extracted features, and additional metadata such as website URLs and classifications.

6. External Interfaces: This includes interfaces for accessing external resources such as online repositories of phishing websites or supplementary datasets for further analysis.

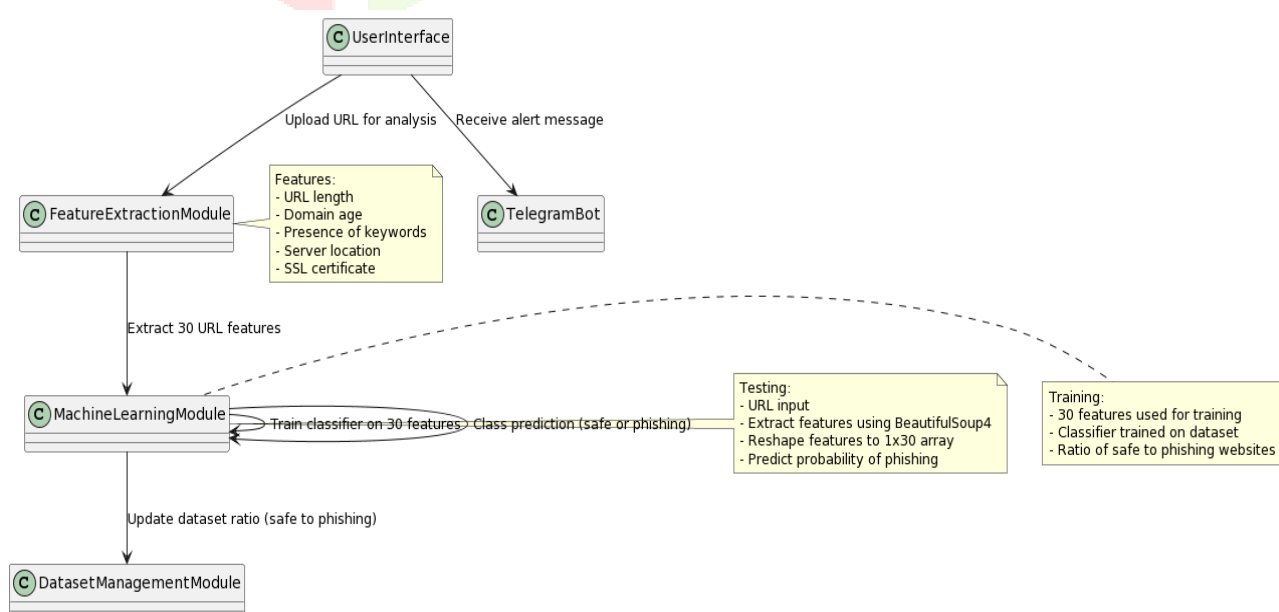


Fig : System architecture

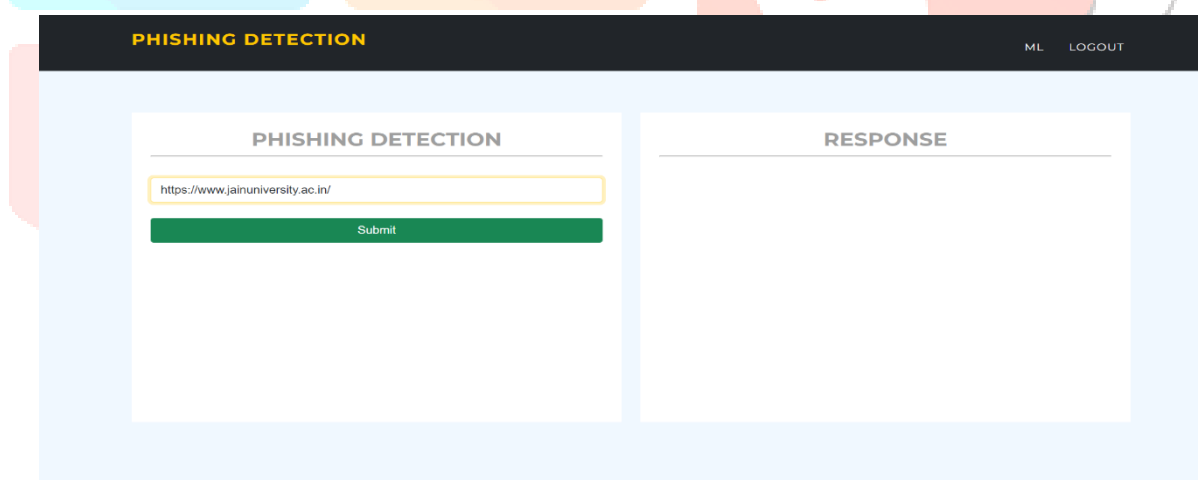
The diagram depicts the fundamental operation of the Fresh-Phish system, emphasizing the training and testing phases, along with the alert mechanism that utilizes a Telegram bot.

1. Training Phase: At the outset, the system permits users to submit URLs for examination through the user interface. These URLs are analysed by the Feature Extraction Module, which identifies 30 distinct features from each URL. These features encompass URL length, domain age, keyword presence, server location, and SSL certificate details. This training phase guarantees that the classifier can proficiently distinguish between safe and phishing URLs based on their characteristics.

2. Testing Phase: During the testing phase, when a user provides a URL for analysis, the system utilizes BeautifulSoup4 to take out the identical 30 characteristics of the URL. These features are then reformatted into a 1x30 array and input into the trained classifier. The classifier estimates the possibility of the URL being phishing according to its features. If the likelihood surpasses a specified threshold, the URL is categorized as phishing, triggering an alert.

3. Alert Mechanism: Should a URL be identified as phishing, the system dispatches a cautionary note for the user through a Telegram bot. The alert message notifies the user of the potential phishing attempt and recommends that they don't participate with the URL. This alert mechanism ensures that users are swiftly informed of potential threats, enabling them to take necessary measures to safeguard themselves against phishing attacks.

III. .RESULT



The screenshot displays the web application interface for 'PHISHING DETECTION'. The header is dark blue with the title 'PHISHING DETECTION' in yellow and 'ML' and 'LOGOUT' in white. The main content area is light blue and divided into two panels. The left panel, titled 'PHISHING DETECTION', contains a text input field with the URL 'https://www.jainuniversity.ac.in/' and a green 'Submit' button. The right panel, titled 'RESPONSE', is currently empty.

Figure:Providing Input

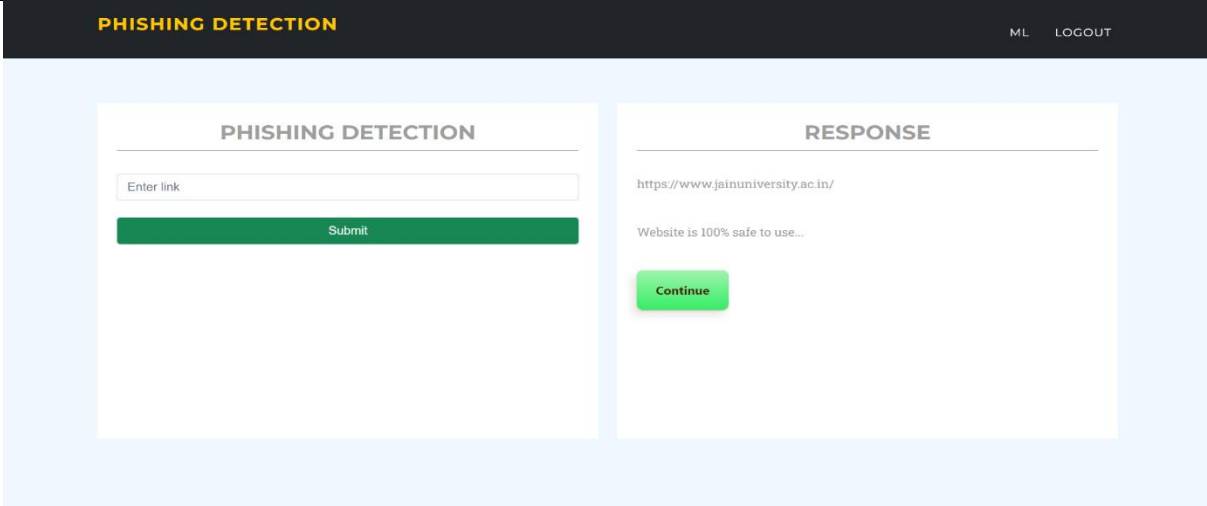


Figure : Output Result Safe

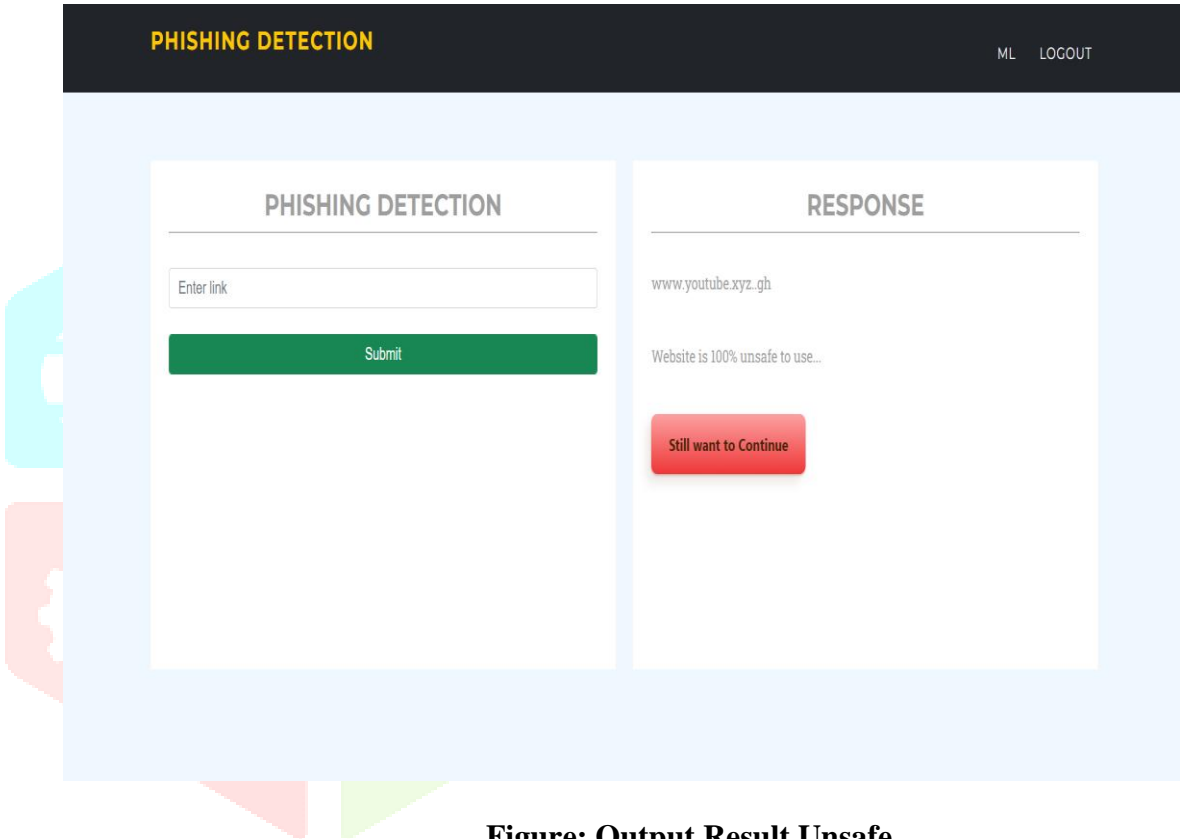


Figure: Output Result Unsafe

PHISHING DETECTION

Enter link

Submit

RESPONSE

<https://www.ijcrt.net/archives/V7/i5/IJRET-V7I5345.pdf>

Website is 93% safe to use...

Continue

Figure : Moderate Results

IV. REFERENCES

- [1] U. S. Musa, M. Chhabra, A. Ali, and M. Kaur, "Intrusion Detection System using Machine Learning Techniques: A Review," 2020 International Conference on Computational Performance Evaluation (ComPE), pp. 1170–1175, 2020.
- [2] A. A. Halimaa and K. Sundarakantham, "Machine Learning Based Intrusion Detection System," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 2S8, pp. 652–655, 2019.
- [3] C.-H. Lee, Y.-Y. Su, Y.-C. Lin, and S.-J. Lee, "Machine Learning Based Network Intrusion Detection," IEEE International Conference on Big Data (Big Data), pp. 3613–3621, 2018.
- [4] Ali, A., Shaukat, S., Tayyab, M., Khan, M. A., Khan, J. S., Arshad, K., and Ahmad, J., "Network Intrusion Detection Leveraging Machine Learning and Feature Selection," IEEE Access, vol. 8, pp. 176826–176838, 2020.
- [5] Taher, K. A., Jisan, B. M. Y., and Rahman, M. M., "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection," 2020 IEEE International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), pp. 638–642, 2020.
- [6] Jain, A., Kumar, V., & Sharma, S.(2020).Detection of Phishing Websites Utilizing Machine Learning Techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(2),1-5.