DETECTION OF PHISHING WEBSITE USING MACHINE LEARNING

Vaishnavi Bhoyar[1], Komal Dharak[2],Dipali Gawali[3],Prof.Deepali Patil[4]
Computer Engineering Department[1,2,3,4] ,
Nutan Maharashtra Institute of Engineering And Technology,Pune [1,2,3,4].

Abstract—Phishing, a prevalent cybercrime, involves deceiving individuals into disclosing personal or confidential information under the guise of legitimate websites or emails. Recognizing phishing websites is challenging due to their similarity to authentic ones. Our study focuses on employing machine learning techniques for efficient phishing website detection. We detail the approach encompassing data gathering, preprocessing steps, extracting features, and employing Support Vector Machine (SVM) for classification. SVM stands out due to its capacity to handle high-dimensional data and non-linear relationships, making it robust to overfitting. SVM offers understandable detection capabilities by optimizing the margin between classes to its maximum extent. Our research aligns with broader cybersecurity objectives, aiming to safeguard individuals and organizations against online deception. Through the development of robust detection systems, we contribute to enhancing cybersecurity measures and empowering users with more secure online experiences. This endeavor underscores the importance of proactive measures in combatting evolving cyber threats

Keywords— Phishing, Support Vector Machine, high-dimensional, Feature extraction

I. Introduction

Phishing is an online theft that steals users personal information and credentials, representing a type of fraud where attackers gain full access to private data. Skilled designers can easily create fake websites identical to the originals, making identification challenging and leading people into traps. These fraudulent sites claim reliability, often using HTTPS to convince users of their legitimacy, while instead harvesting their credentials. In today's digital landscape, where most transactions occur online, identifying phishing websites is crucial. According to the Anti-Phishing Working Group, as of September 2018, there were 647,592 reported unique phishing sites.

Phishing is a form of social engineering attack aimed at coercing users into performing actions on behalf of attackers, typically to steal personal or confidential information. Phishers mimic legitimate emails or web pages, often with high visual similarity, accompanied by login forms to gather sensitive data. The content of phishing emails and web pages is crafted using social engineering techniques, persuading victims to follow instructions under false pretenses, such as updating information, finding employment, winning prizes, or receiving discounts.

Phishing poses a significant security concern, using clever tactics to entice individuals into divulging sensitive information through harmful links. Even though they rely on straightforward techniques, these assaults prove remarkably efficient by replicating authentic websites, posing a challenge to their detection. Due to technological advancements, phishing attacks have grown more complex, posing challenges for users in distinguishing between fraudulent and genuine emails or websites. Our Android project, focused on machine learning, aims to address this issue. Users can log in and utilize machine learning algorithms to evaluate URL legitimacy and report phishing attempts. A user's access enables the examination and resolution of grievances, implementing required measures against detected phishing URLs and their origins. By incorporating machine learning technology into our Android app, our goal is to offer a user-friendly solution that bolsters online security, enabling users to recognize and report phishing attempts effectively. This joint endeavor seeks to diminish the occurrence and effectiveness of phishing attempts, providing advantages to both individuals and organizations.

II. Literature Review

J.Shad and S.Sharma, explores a novel approach to phishing detection through a hybrid machine learning system centered on URL analysis. Employing decision tree, random forest, and support vector machine models, the study focuses on feature extraction from URLs. Significantly, the random forest technique is employed due to its resilience and effectiveness in managing extensive data sets. By leveraging multiple models, the system aims to enhance the accuracy and reliability of phishing detection, thus addressing the growing sophistication of cyber threats. [1]

Y.Sonmez, T.Tuncer, H.Gokal and E.Avci presented ensemble architectures which are highlighted for their ability to excel even when faced with a reduction in the number of features within the dataset. Specifically, support vector machine models are incorporated into ensemble frameworks to capitalize on their discriminative capabilities and robustness against complex data patterns. This paper underscores the importance of ensemble techniques in enhancing the efficacy of phishing detection systems, particularly in scenarios where feature availability may be limited. [2]

In the realm of cybersecurity, the detection of malicious URLs poses a significant challenge, prompting researchers to explore various machine learning techniques for effective identification. One such study which is presented by T. Peng, I. Harris and Y. Saw, the application of multiple classification algorithms for this purpose, with a particular focus on the Random Forest Algorithm. Random Forest, a supervised learning algorithm employing the concept of bagging, was utilized in this research to discern malicious URLs. The versatility and robustness of Random Forest in handling classification tasks make it a suitable candidate for detecting potentially harmful URLs. [3]

A phishing detection engine using the features extracted from URLs was proposed by M. Karabatak and T. Mustafa. Phishing detection system was developed, again leveraging the Random Forest Algorithm alongside support vector machine (SVM). SVM, renowned for its effectiveness in classification tasks, complemented Random Forest in enhancing the accuracy and reliability of the phishing detection system. Both studies contribute valuable insights into the utilization of machine learning techniques, particularly Random Forest and SVM, in bolstering cybersecurity measures against malicious URLs and phishing threats. [4]

S. Parekh, D. Parikh, S. Kotak and P. S. Sankhe proposed a novel approach for detecting phishing websites through URL detection, focusing on three primary phases: Parsing, Heuristic Classification of data, and Performance Analysis. The Parsing phase presumably involves the extraction and analysis of relevant data from URLs. Following Parsing, the Heuristic Classification phase employs a decision tree algorithm to classify the parsed data, likely based on predefined criteria indicative of phishing behavior. This phase seems crucial in determining the authenticity of websites based on their URLs. Ultimately, the Performance Analysis stage assesses the viability of the suggested approach, likely using measures like accuracy, precision, recall, and F1 score. The decision tree algorithm, being a popular and interpretable classification technique, appears to play a central role in the heuristic classification process. However, the paper lacks detailed descriptions of the specific heuristics utilized or the features considered in classification, which could impact the replicability and generalizability of the proposed method. Additional explanation regarding these elements would improve comprehension and the usefulness of the suggested method. Overall, the paper introduces a promising method for phishing website detection through URL analysis, but it requires more comprehensive elucidation and empirical validation to ascertain its efficacy and robustness in real-world scenarios. [5]

The model proposed by K. Shima et al. reveals a significant emphasis on real-time detection methods to counter the pervasive threat of phishing attacks, with a particular focus on leveraging machine learning techniques. One notable approach highlighted across recent research involves scrutinizing Uniform Resource Locators (URLs) of web pages to differentiate between legitimate and fraudulent websites. By analyzing URL characteristics such as domain structure, length, and presence of suspicious keywords, researchers aim to develop robust detection mechanisms capable of accurately identifying phishing attempts in real- time. The proposed detection technique outlined in the surveyed literature builds upon this foundation, utilizing machine learning algorithms trained on URL features to effectively discern phishing websites from legitimate ones. Evaluation using diverse datasets, including Phistank and Yahoo directory datasets, demonstrates the solution's efficacy in detecting various phishing attacks while maintaining a low false alarm rate, highlighting its practical utility and deployability in real-world scenarios. [6]

In these paper named on feature selection for the prediction of phishing websites by W.Fadheel, M. Abusharkh and I.Abdel-Qader underscores the ongoing challenge of effectively detecting evolving forms of phishing attacks, which often evade traditional detection methods by incorporating subtle structural and semantic alterations in their messages. To address this issue, recent research has introduced innovative approaches, such as the use of machine learning classifiers operating on large datasets of phishing and legitimate emails. One notable contribution is the development of SAFEPC (Semi-Automated Feature generation for Phish Classification),Furthermore, SAFE-PC demonstrates superior performance compared to state-of-the- art email filtering tools like Sophos and SpamAssassin, with its online version showcasing the ability to incrementally improve detection accuracy over time while maintaining a constant retraining time. This research highlights the significance of adaptive and proactive approaches in combating the evolving threat landscape of phishing attacks. [7]

Similarly, X. Zhang, Y. Zeng, X. Jin, Z. Yan and G. Geng, Boosting proposed the Phishing Detection Performance by Semantic Analysis explores novel methodologies for analyzing the dynamics of phishing attacks, aiming to enhance understanding of their occurrence and effectiveness of detection. By investigating changes in the power dynamics of phishing attacks, the study introduces innovative tools rooted in chaos theory and wavelet coherence ideology for analysis. Leveraging real data on phishing attacks, the research expands the scope of analysis by utilizing these advanced tools to probe the intricacies of attack occurrences and effectiveness of detection methods. Through this investigation, the paper contributes to advancing the understanding of phishing attack dynamics and offers insights into improving detection strategies in cybersecurity. [8]

L. MacHado and J. Gadge suggested phishing sites detection based on C4.5 decision tree algorithm which is a novel approach for detecting phishing attacks using web crawling techniques. Phishing remains a significant cybersecurity threat, targeting individuals, organizations, and various online platforms. While hardware-based solutions are prevalent, they are often cost-prohibitive and operationally complex, leading to a preference for software-based approaches. Existing detection methods struggle to address emerging threats such as zero-day phishing attacks effectively. WC-PAD addresses these challenges by employing a three-phase detection process that analyzes web traffic, content, and Uniform Resource Locators (URLs). By leveraging these features, WC-PAD achieves high accuracy in distinguishing between phishing and non-phishing websites, including zero-day attacks, as demonstrated through experimental evaluations using real-world phishing datasets. The suggested method presents a hopeful resolution for effective phishing detection across various digital landscapes. [9]

Sahingoz et al. presented a machine learning-based approach for phishing detection from URLs. Their study focused on extracting relevant features from URLs and employing machine learning algorithms to classify URLs as either legitimate or phishing. The researchers likely utilized techniques such as feature engineering to extract meaningful attributes from URLs, such as domain age, presence of subdomains, and lexical features. They then trained machine learning models, possibly including decision trees, support vector machines, or random forests, on these features to learn patterns indicative of phishing behavior. The research probably assessed the effectiveness of the suggested method by measuring metrics like accuracy, precision, recall, and F1 score to determine its capability in discerning between genuine and phishing URLs. [10]

III. METHODOLOGY

Detecting phishing websites using machine learning typically involves several steps and methodologies.

A. Data Collection:

- Obtain a dataset of URLs labeled as phishing or legitimate. You can acquire such datasets from public repositories, commercial sources, or by scraping relevant websites.
- Ensure the dataset is sufficiently large and diverse to capture the various characteristics of phishing and legitimate URLs.

B. Data Preprocessing:

- Address missing data: Examine the dataset for any gaps and determine the best approach to handle them, such as imputing values or removing incomplete entries. Normalize or scale features: Standardize numerical features to have a mean of 0 and a standard deviation of 1, or scale them to a specific range to prevent features with large values from dominating the model.
- Encode categorical variables: Transform categorical variables into numerical representations through methods such as one-hot encoding or label encoding.

C. Feature Extraction:

- Extract relevant features from the URLs and associated metadata. Some features to consider include:
- URL length: Phishing URLs tend to be longer and more complex.
- Domain-related features: Domain age, presence of IP address in the URL, use of HTTPS, domain reputation, presence of subdomains, etc.
- Content-based features: Analyze the webpage content for suspicious keywords, presence of forms requesting sensitive information, similarity to known phishing templates, etc.
- Use domain-specific knowledge and expertise to identify additional features that may be indicative of phishing behavior.

D. Training:

- Split the preprocessed dataset into training and testing sets (e.g., 70% training, 30% testing).
- Select a machine learning algorithm suitable for binary classification tasks such as logistic regression, random forest, or support vector machines (SVM).
- Train the chosen model on the training data, using the extracted features as input and the phishing label (1 for phishing, 0 for legitimate) as the target variable.

E. Detecting Phishing or Not:

- After training the model, use it to predict the likelihood of a given URL being phishing or legitimate.
- For a new URL, extract the same features used during training and preprocessing.
- Input the extracted features into the trained model in order to generate a prediction score.
- Set a benchmark for the prediction score to differentiate between phishing and legitimate URLs. For example, if the prediction score exceeds a certain threshold (e.g., 0.5), classify the URL as phishing; otherwise, classify it as legitimate.
- Assess the efficiency of the model by analyzing metrics like accuracy, precision, recall, and F1-score on the test dataset to gauge its capability in identifying phishing URLs.
- Phishing content often contains deceptive language, misleading URLs, requests for personal

information, or attempts to impersonate legitimate entities.

- Look for inconsistencies in language, such as grammatical errors or awkward phrasing, which are common in phishing attempts.

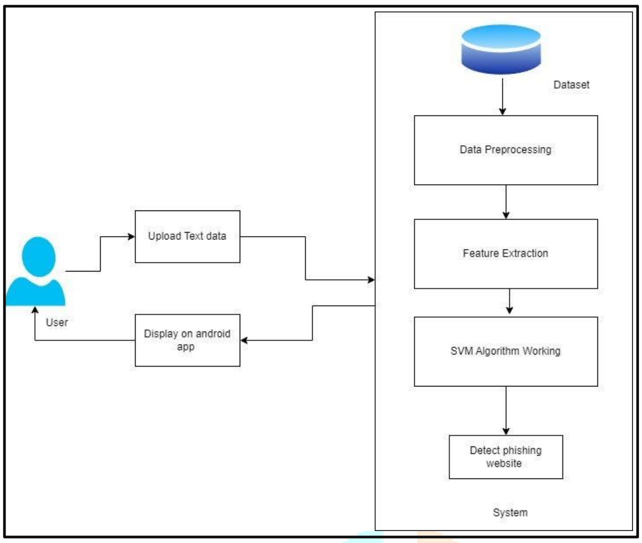


Fig. 1. System Architecture of Phishing Detection

IV. Modeling and Analysis

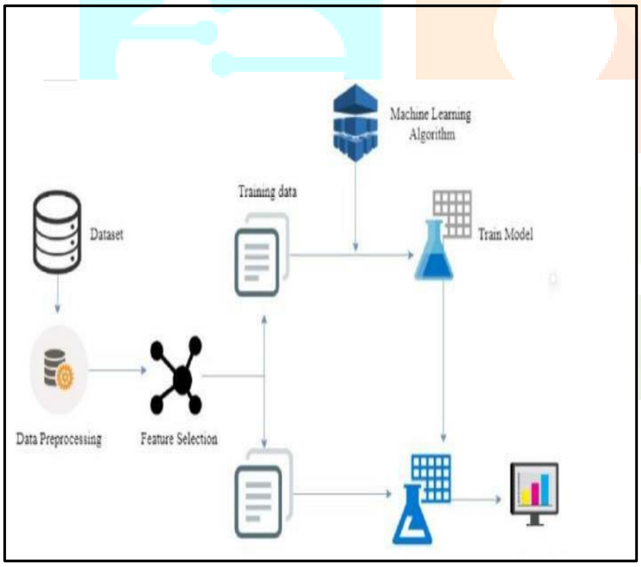


Fig. 2. Modeling of Machine Learning

A. *Algorithms:*

- The effectiveness of Support Vector Machine (SVM) in the suggested system for detecting phishing websites stems from its adeptness at managing complex data, capturing non-linear patterns, and addressing class imbalances. SVM aims to find a clear decision boundary while maximizing the margin between classes, making it robust to overfitting and providing interpretable results. With efficient training and classification, SVM can effectively detect phishing websites in real-time, making it a suitable choice for the system.
- Feature Extraction: Relevant features such as URL length, presence of HTTPS, domain age, and other characteristics are extracted from the URLs. These characteristics act as inputs to the SVM algorithm.
- Training Phase: During training, the SVM algorithm learns to construct a hyperplane that best separates the phishing URLs from legitimate ones in the feature space. Its objective is to optimize the margin, specifically the space between the hyperplane and the closest data points (support vectors) of each class.
- Classification: When a new URL is submitted for phishing detection, its features are extracted, and the SVM model predicts whether it belongs to the phishing or legitimate class based on its position relative to the learned hyperplane. If the web address is positioned on one side of the hyperplane, it's labeled as phishing; if it's on the opposite side, it's categorized as legitimate.
- Boundary Decision: The boundary decision produced by SVM delineates the optimal division between phishing and legitimate URLs within the feature space. This division is important to note that SVMs can manage non-linear decision boundaries through the utilization of kernel functions, enabling them to grasp intricate relationships among characteristics.
- Assessing the performance : SVM model in identifying phishing involves analyzing various metrics including accuracy, precision, recall, and F1-score using either a validation or test dataset. The model's performance can be further refined through hyperparameter tuning and cross-validation techniques.

V. Proposed System

Proposed System for Phishing Website Detection Using Machine Learning:

A. Registration/Login:

- Users can register as new users with basic details like email and password.
- Existing users can log in using their credentials.
- Two user roles: "User" and "Officer".

B. User Interface:

- Upon login, users are presented with an interface where they can enter a URL to check for phishing.
- Users have the option to submit a report if the URL is identified as phishing.

- Officers have access to a separate interface for monitoring reports and taking action.

C. Phishing Detection:

- When a user enters a URL and clicks the "Check" button, the system performs phishing detection using machine learning.
- Extract relevant features from the URL (e.g., URL length, domain age, presence of HTTPS, etc.) for input into the trained machine learning model.
- The model predicts whether the URL is attempting phishing or not.
- Display the result to the user indicating Determine if the URL is attempting phishing or not.

D. Reporting Phishing:

- If the URL is identified as phishing, the user can choose to submit a report.
- The report includes basic details (e.g., user's email) and the URL of the phishing website.
- Submitted reports are forwarded to the officer for further action.

E. Officer Interface:

- Officers can log in using their credentials to access the officer interface.
- The officer interface displays a list of submitted reports, showing details such as the user who submitted the report and the URL of the phishing website.
- Officers can take appropriate action on reported phishing URLs, such as blocking the website or sending warnings to users.

F. Model Training and Updating:

- Periodically retrain the machine learning model using newly collected data to improve accuracy and adapt to evolving phishing techniques.
- Implement mechanisms to update the model seamlessly without disrupting the user experience.

G. Security Measures:

- The system will implement strong encryption and secure protocols for user authentication and data transmission.
- User data will be handled with care, adhering to privacy regulations to protect user privacy and confidentiality.
- Access controls will be in place to prevent unauthorized access to sensitive user data and administrative functionalities.

By following this proposed system, users can easily detect phishing websites using machine learning while ensuring their data security and privacy. Additionally, officers can efficiently manage reported phishing incidents and take timely actions to protect users from online threats.

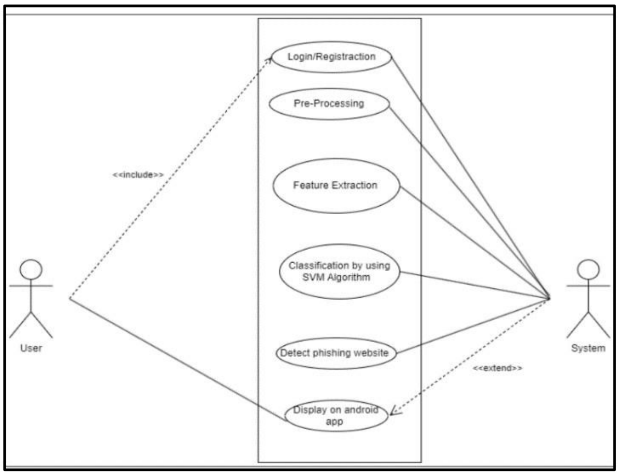


Fig. 3. Use Case Diagram of Phishing Detection App

VI. FUTURE SCOPE

Real-time Detection: Advancements in hardware capabilities and optimization techniques could facilitate the deployment of real-time phishing detection systems on resource-constrained devices, such as smartphones or IoT devices. This would allow for proactive safeguarding against phishing dangers on various platforms and devices.

User Feedback Integration: Integrating user feedback mechanisms into the phishing detection system could enable continuous learning and adaptation to emerging phishing tactics. By incorporating feedback from users who encounter suspicious URLs, the system can dynamically update its detection capabilities and improve its effectiveness over time.

Ensemble Methods: Combining multiple machine learning models, including SVM, into ensemble methods such as random forests or gradient boosting could improve overall detection accuracy. Ensemble techniques harness the capabilities of individual models while offsetting their limitations, resulting in enhanced and dependable phishing detection systems.

Deep Learning Architectures: Exploring various deep learning structures, like convolutional neural networks (CNNs) or recurrent neural networks (RNNs), for the detection of phishing could provide advantages in identifying complex patterns and connections within URL attributes and webpage content. These architectures have demonstrated significant success in various domains and may yield further improvements in phishing detection accuracy. Deep learning architecture refers to the design and structure of neural networks used in deep learning models. Deep learning is a subset of machine learning that utilizes artificial neural networks with multiple layers to extract features from data and make predictions.

VII. Advantages

1) Real-time Analysis: The app conducts real-time analysis of website URLs, swiftly detecting suspicious characteristics indicative of phishing attempts.
2) Dynamic Database Updates: Continual updating of its database with the latest phishing patterns ensures proactive protection against emerging threats.
3) Enhanced User Security: By promptly identifying phishing attempts, the app reduces the risk of users falling victim to fraudulent schemes, thus bolstering overall online security.
4) Proactive Defense: The app's proactive approach to detecting and blocking phishing websites helps prevent potential financial loss, identity theft, and other cyber threats before they occur.
5) User-Friendly Interface: Many phishing detection applications provide interfaces that are easy to use, allowing individuals, even those with minimal technical expertise, to navigate and employ the application efficiently.
6) Multi-Platform Compatibility: Numerous phishing detection applications work seamlessly across a range of devices and operating systems, ensuring broad protection across diverse platforms.
7) Educative Features: Some apps provide educational resources and tips on how to recognize and avoid phishing attempts, empowering users to develop a heightened awareness of online security practices.
8) Customizable Settings: Users often have the option to customize settings based on their preferences, allowing for tailored protection suited to their specific needs and risk tolerance.
9) Privacy Protection: Phishing detection apps prioritize user privacy by implementing robust data encryption measures and transparent data handling practices, ensuring sensitive information remains secure
10) Regular Audits and Compliance: Reputable phishing detection apps undergo regular audits and adhere to industry standards and regulations, providing users with confidence in the app's reliability and adherence to best practices.

VIII. Applications

A phishing website detection app that utilizes machine learning (ML) can offer several valuable applications in enhancing cybersecurity. Here are some ways such an app could be beneficial:

1. Real-time Detection: ML algorithms can be trained to analyze website features and user behavior to identify potential phishing websites in real-time. This helps users avoid falling victim to phishing attacks by warning them before they enter sensitive information.

2. Email Security: Many phishing attacks originate from fraudulent emails. ML algorithms can be employed to scan emails for suspicious links and attachments, thereby enhancing email security and preventing users from clicking on harmful links.

3. Browser Extensions: Integrate the phishing detection capability into browser extensions, enabling users to receive alerts when visiting potentially malicious websites. These extensions can leverage ML models to continuously learn and adapt to new phishing techniques and variations.

4. Mobile Security: Develop mobile apps that utilize ML for phishing detection, providing users with on-the-go protection against phishing attempts on their smartphones and tablets.

5. Fraud Prevention: Beyond just detecting phishing websites, ML algorithms can be trained to recognize patterns of fraudulent behavior, helping financial institutions and e- commerce platforms prevent fraudulent transactions

6. Training Data Enhancement: Phishing detection models require large amounts of labeled data for training. ML techniques can help in automatically labeling phishing websites, thereby reducing the manual effort required for dataset creation.

7. User Education: The app can also be used to educate users about common phishing tactics and how to recognize them. By providing real-world examples and explanations, users can become more vigilant and less likely to fall for phishing attempts.

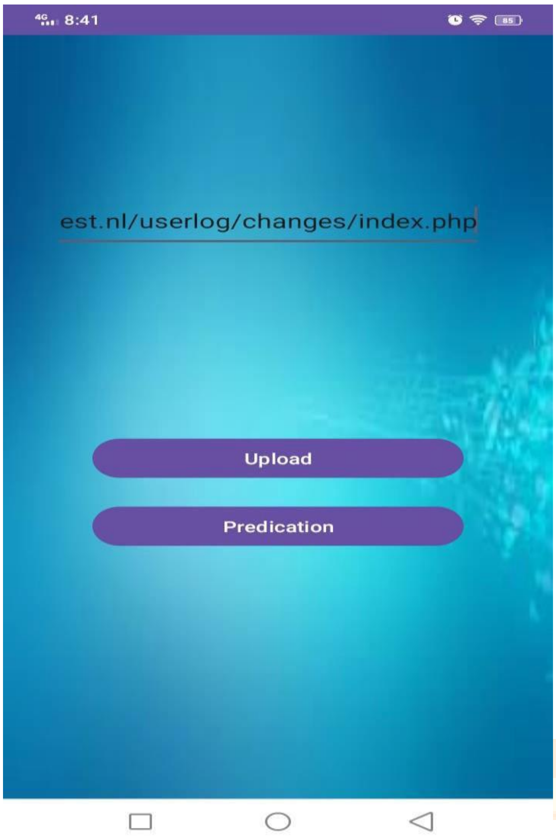
8. Adaptive Learning: ML algorithms can adapt and evolve over time based on new data and emerging phishing techniques. This ensures that the detection capabilities of the app remain effective even as attackers develop more sophisticated methods.

9. API Integration: Integrate the phishing detection capabilities into other security products and services through APIs. This allows organizations to leverage the app's detection capabilities within their existing security infrastructure.

10. Reporting and Feedback : Allow users to report suspected phishing websites and provide feedback on the app's detection accuracy. This data can be used to further improve the app's performance and stay ahead of evolving phishing threats.

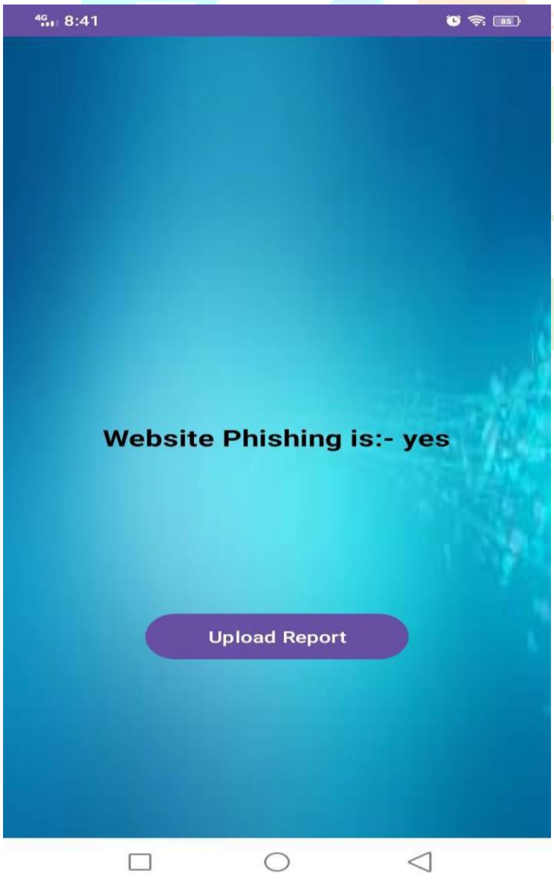
IX. Result

Detecting phishing websites is essential for protecting users from online threats by recognizing and preventing access to fraudulent sites aimed at stealing personal data. Upon detecting a phishing website, the detection system typically alerts the user, either through browser warnings, email notifications, or within the app interface itself. This quick response assists individuals in preventing accidental disclosure of personal or financial details to malicious parties. Moreover, by reporting and blacklisting phishing URLs, detection systems contribute to the broader cybersecurity ecosystem, preventing other users from falling victim to the same fraudulent schemes. Overall, the result of phishing website detection is a fortified defense against online scams, bolstering user confidence in navigating the digital landscape securely.



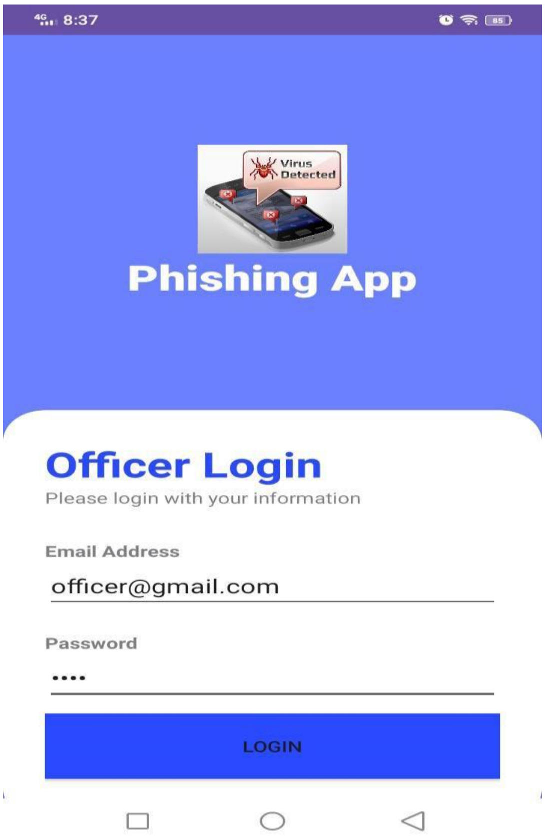
est.nl/userlog/changes/index.php

Fig. 4.1 Input URL



Website Phishing is:- yes

Fig. 4.2. Phishing Detection



Phishing App

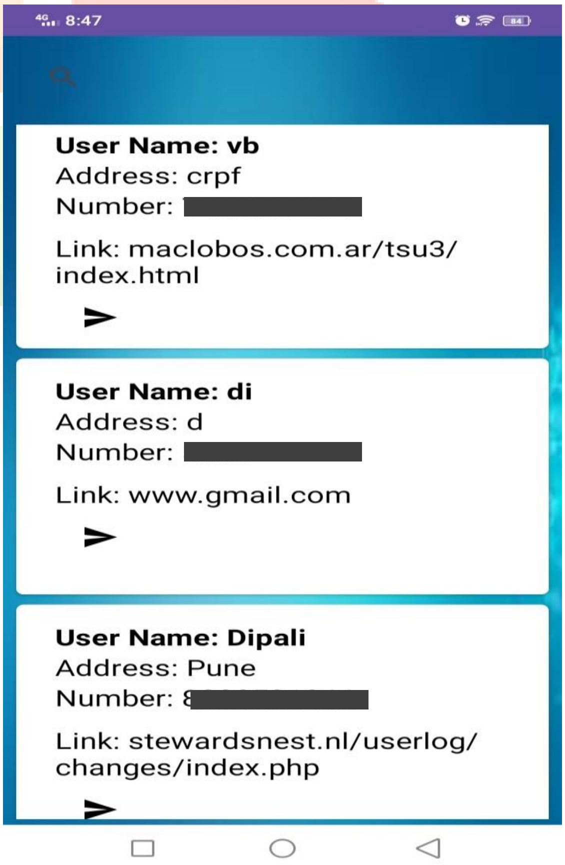
Officer Login

Please login with your information

Email Address: officer@gmail.com

Password: ••••

Fig. 4.3. Officer Login



User Name: vb
Address: crpf
Number: [redacted]
Link: maclobos.com.ar/tsu3/index.html

User Name: di
Address: d
Number: [redacted]
Link: www.gmail.com

User Name: Dipali
Address: Pune
Number: [redacted]
Link: stewardsnest.nl/userlog/changes/index.php

Fig. 4.4. Reported Data

ACKNOWLEDGMENTS

Acknowledgment of phishing website detection signifies a critical appreciation of the multifaceted efforts involved in fortifying cybersecurity measures against evolving online threats. It encompasses recognition of the sophisticated algorithms and machine learning models employed to swiftly identify and categorize potentially harmful websites. Moreover, acknowledgment extends to the tireless endeavors of cybersecurity experts and software developers who continually refine detection mechanisms through rigorous testing, analysis of phishing trends, and proactive updates to databases. This collective effort underscores a commitment to staying ahead of cybercriminals by leveraging innovative technologies and fostering a culture of collaboration within the cybersecurity community. Additionally, acknowledgment implies an understanding of the pivotal role played by end-users in remaining vigilant and informed about phishing tactics, thereby empowering them to recognize and report suspicious activity effectively. By acknowledging the significance of phishing website detection, stakeholders affirm their dedication to mitigating risks, safeguarding digital assets, and preserving trust in online interactions.

REFERENCES

[1] J. Shad and S. Sharma, A Novel Machine Learning Approach to Detect Phishing Websites Jaypee Institute of Information Technology, pp. 425-430, 2018.

[2] Y. Sonmez, T. Tuncer, H. Gokal and E. Avci, "Phishing web sites features classification based on extreme learning machine", *6th Int. Symp. Digit. Forensic Secur. ISDFS 2018 - Proceeding*, vol. 2018- Janua, pp. 1-5, 2018.

[3] T. Peng, I. Harris and Y. Sawa, "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning", *Proc. - 12th IEEE Int. Conf. Semant. Comput. ICSC 2018*, vol. 2018-Janua, pp. 300-301, 2018.

[4] M. Karabatak and T. Mustafa, "Performance comparison of classifiers on reduced phishing website dataset", *6th Int. Symp. Digit. Forensic Secur. ISDFS 2018 - Proceeding*, vol. 2018-Janua, pp. 1-5, 2018.

[5] S. Parekh, D. Parikh, S. Kotak and P. S. Sankhe, "A New Method for Detection of Phishing Websites: URL Detection", *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, vol. 0, pp. 949-952, 2018.

[6] K. Shima et al., "Classification of URL bitstreams using bag of bytes", *2018 21st Conference on Innovation in Clouds Internet and Networks and Workshops (ICIN)*, vol. 91, pp. 1-5, 2018.

[7] W. Fadheel, M. Abusharkh and I. Abdel-Qader, "On Feature Selection for the Prediction of Phishing Websites", *2017 IEEE 15th Intl Conf Dependable Auton. Secur. Comput. 15th Intl Conf Pervasive Intell. Comput. 3rd Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr.*, pp. 871-876, 2017.

[8] X. Zhang, Y. Zeng, X. Jin, Z. Yan and G. Geng, Boosting the Phishing Detection Performance by Semantic Analysis, 2017.

[9] L. MacHado and J. Gadge, "Phishing Sites Detection Based on C4.5 Decision Tree Algorithm", *2017 International Conference on Computing Communication Control and Automation ICCUBEA 2017*, pp. 1-5, 2018.

[10] Sahingoz, O. K., Buber, E., Demir, O. & Diri, B. Machine learning based phishing detection from URLs. *Expert Syst. Appl.* 2019(117), 345–357 (2019).