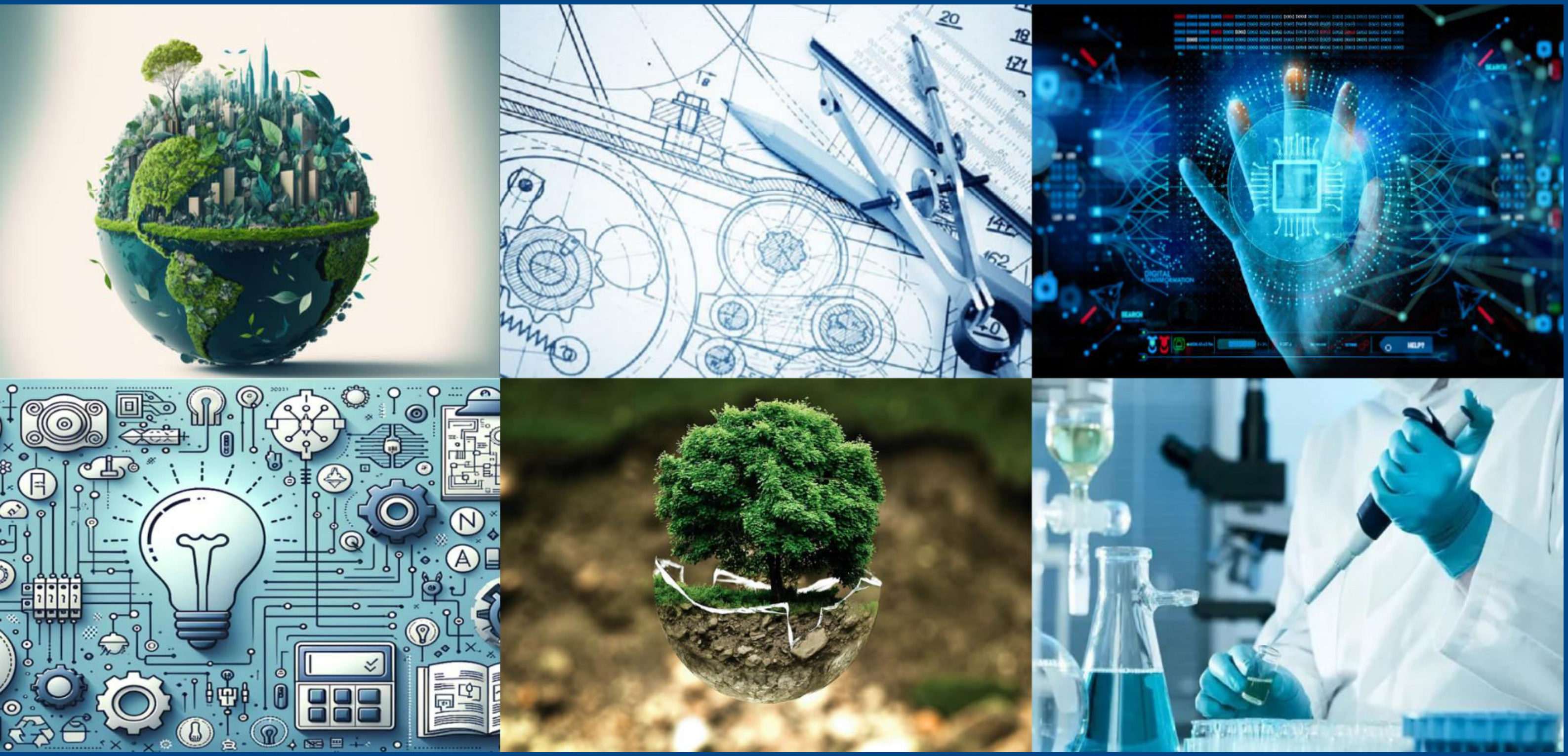
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 3, March 2025

www.ijmrset.com ✉ ijmrset@gmail.com ☎ +91 6381 907 438 ✆ +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Phishing Detection System using Phishtank API

Mrs. Muthuvairaivan Pillai[1], Ravindhar G[2], Sachin D[3], Saravanakumar C[4]

[1]Faculty of Department of Computer Science and Business System, R.M.D Engineering College, Chennai, Tamil Nadu, India

[2-4]Student of Department of Computer Science and Business System, R.M.D Engineering College, Chennai, Tamil Nadu, India

ABSTRACT: Phishing attacks have become a significant cybersecurity threat, targeting users by tricking them into revealing sensitive information such as passwords and financial details. Traditional detection methods often fail to identify newly emerging phishing websites, making real-time detection crucial. This project aims to develop an intelligent phishing detection system using machine learning techniques and PhishTank API integration. The system extracts key features from URLs and web content to classify websites as legitimate or phishing. Machine learning models analyze patterns to enhance detection accuracy beyond existing databases. The proposed solution offers a user- friendly interface for real-time phishing analysis and reporting. By leveraging AI-driven techniques, the system minimizes false positives and improves security for individuals and organizations. The implementation ensures high accuracy and adaptability to evolving phishing strategies. The project contributes to enhanced cybersecurity by providing a robust and automated phishing detection framework.

I. INTRODUCTION

Phishing is a deceptive practice where attackers impersonate legitimate entities to steal sensitive information such as passwords, credit card details, and personal data. With the rapid increase in online transactions and digital interactions, phishing attacks have become a major cybersecurity threat. Traditional detection methods, such as manually blacklisting phishing websites, are ineffective against newly emerging threats. This project aims to enhance phishing detection by integrating the PhishTank API with machine learning techniques to identify malicious websites more accurately. By analyzing various URL and webpage features, the system can classify phishing attempts even if they are not yet reported in public databases. The machine learning model continuously improves through data training, ensuring adaptability to evolving phishing strategies. The proposed solution offers real-time analysis, reducing user susceptibility to online fraud. Through this approach, the project strengthens cybersecurity and helps users navigate the internet safely.

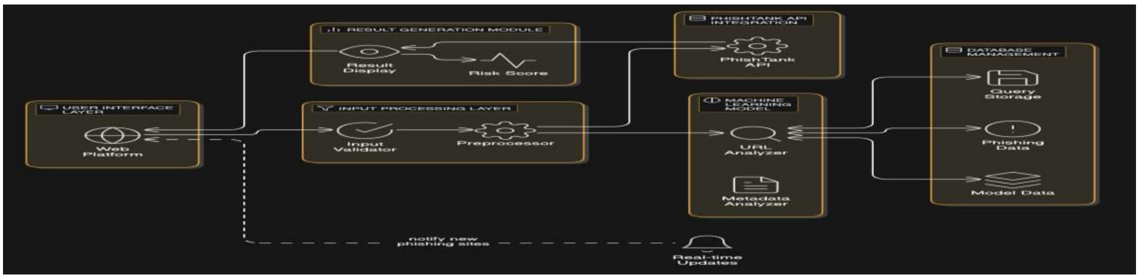


Figure 1 Architecture Diagram

II. LITERATURE REVIEW

The literature review explores existing research on phishing detection methods, highlighting their strengths and limitations. Traditional approaches rely on blacklist-based detection, where databases like PhishTank store known phishing websites. However, these methods struggle with newly emerging phishing sites. Machine learning models have been introduced to analyze URL structures, webpage content, and domain features for improved detection accuracy. Researchers have also explored deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to enhance phishing identification. Comparative studies indicate that hybrid approaches combining blacklist verification with AI-driven analysis provide better results. Despite advancements, challenges remain in reducing false positives and adapting to evolving phishing tactics, necessitating continuous improvements in detection algorithms.

[1] Phishing Detection: Analysis of Visual Similarity Based Approaches
Date of Publication: 10 January 2017 Authors: Ankit Kumar Jain, B. B. Gupta
Phishing is a major cyber threat causing financial losses for both individuals and industries. Detecting phishing attacks with high accuracy remains challenging. Visual similarity-based techniques are effective in detecting phishing websites by comparing their appearance to legitimate websites. These techniques utilize features such as text content, HTML tags, CSS, and images to assess similarity. If a suspicious website's visual features match the legitimate one beyond a threshold, it is flagged as phishing. This paper reviews recent visual similarity-based phishing detection methods and provides a comparative study. It aims to enhance the effectiveness of phishing detection through visual similarity approaches.

[2] Machine learning based phishing detection from URLs
Date of Publication: 12 October 2018
Authors: Ozgur Koray Sahingoz , Ebubekir Buber, Onder Demir, Banu Diri
With the rise of e-commerce, cybercriminals have shifted from traditional crimes to online tactics like phishing, where they use fake websites to steal sensitive information. Detecting phishing websites is challenging due to their deceptive nature and attackers exploiting users' vulnerabilities. This paper proposes a real-time anti-phishing system using seven classification algorithms and natural language processing (NLP) features. The system is language-independent, uses large datasets, and operates in real-time, detecting new phishing sites. A new dataset was created to evaluate the system's performance. Experimental results show that the Random Forest algorithm, using only NLP features, achieves a 97.98% accuracy rate. This approach outperforms other classifiers for detecting phishing URLs.

[3] A predictive model for phishing detection
Date of Publication: 12 February 2022
Authors: A.A. Orunsolu, A.S. Sodiya,A.T. Akinwale
Many anti-phishing systems are being developed to detect phishing content, but challenges remain due to zero-day attacks, high computational overhead, and false positives. Despite promising results from machine learning, feature selection and performance limitations hinder effective detection. This work proposes an enhanced machine learning-based predictive model to improve anti-phishing efficiency. The model uses a Feature Selection Module to create an effective feature vector extracted from URLs, webpage properties, and behavior. Support Vector

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine (SVM) and Naïve Bayes (NB) classifiers are trained on a 15-dimensional feature set. Experiments with 2541 phishing and 2500 benign instances show impressive results with 0.04% false positives and 99.96% accuracy using 10-fold cross-validation. Both SVM and NB models achieve these high performance rates.

[4] Analysis of API Driven Application to Detect Smishing Attacks
Date of Publication: 15 Jun 2022
Authors: Phadke, Pranav, Thorpe Christina
The increasing use of mobile smartphones has led to the rise of Smishing attacks, which exploit SMS to deliver malicious URLs, potentially causing harm. While phishing emails are well-known and have detection systems, Smishing is often overlooked due to the small mobile screen size, making URL verification difficult. This research develops an Android app that detects Smishing attacks by integrating existing phishing APIs. The app runs in the background and checks URLs for phishing threats. Five APIs were tested on a dataset of 1500 URLs for accuracy and latency. Results show VirusTotal with 99.27% accuracy but slower response time, while Safe-Browsing offers 87% accuracy with faster response times.

S.NO	TITLE	AUTHOR	METHODOLOGY	PROS	CONS
1.	*Phishing Detection: Analysis of Visual Similarity Based Approaches*	Ankit kumar Jain B. B. Gupta	*Visual similarity-based phishing detection leverages image recognition.*	*It effectively detects phishing sites that closely mimic legitimate websites.*	*It can be computationally expensive and may struggle with phishing sites.*
2.	*Machine learning based phishing detection from URLs*	*Ozgur Koray Sahingoz , Ebubekir Buber, Onder Demir, Banu Diri*	*Machine learning-based phishing detection from URLs involves extracting features such as domain age, URL length.*	*It can process large volumes of data efficiently.*	*It requires a large, labelled dataset for training.*
3.	*A predictive model for phishing detection*	*A.A. Orunsolu, A.S. Sodiya, A.T. Akinwale*	*A predictive model for phishing detection uses machine learning algorithms.*	*It can efficiently handle large datasets.*	*The model's performance heavily relies on the quality and variety of labeled training data.*
4.	***Analysis of API Driven Application to Detect Smishing Attacks***	*Phadke, Pranav, Thorpe Christina*	*In API-driven applications for smishing detection, machine learning models.*	*It allows real-time smishing detection with minimal user.*	*The effectiveness of detection depends on the accuracy of the NLP models.*
5.	*Piracema: a Phishing snapshot database for building dataset features*	*Julio Cesar Gomes de Barros, Carlo Marcelo Revoredo da Silva*	*Piracema collects and categorizes phishing site features such as URL patterns, domain characteristics.*	*It provides a rich, curated dataset for researchers and developers.*	*The quality of the database relies on the effectiveness of the feature.*

III. IMPLEMENTATION METHODOLOGY OF PROPOSED SYSTEM

Setup of Phishing Detection System

The selection of appropriate tools and technologies is crucial for the system's efficiency. The system utilizes the PhishTank API to retrieve updated phishing website data and integrates machine learning techniques to identify phishing attempts not yet reported in public databases. The initial setup involves configuring the development environment, installing necessary dependencies such as Flask for backend development, and setting up the MySQL database for storing collected data. The API integration phase includes obtaining an API key, implementing API requests to fetch phishing data, and normalizing the retrieved information for further analysis. Additionally, a web-based interface is designed using HTML, CSS, and JavaScript to allow users to check website legitimacy by entering URLs.

Machine Learning-Based Phishing Detection

Once the model is trained, it is deployed for real-time phishing detection. When a user submits a URL, the system first checks the PhishTank database to determine whether the site is already flagged as phishing. If the URL is not found in the database, the machine learning model evaluates the website based on extracted features. Custom classification rules are applied to detect suspicious patterns, such as long URLs with excessive special characters or domain mismatches. The system continuously updates its model by incorporating new phishing data, improving detection accuracy over time.

Real-Time Threat Analysis

To enhance security, the system provides real-time analysis of submitted URLs. If a website is detected as phishing, an immediate warning is displayed to the user. Additionally, the system logs details of detected phishing attempts for further review. Advanced techniques, such as deep learning-based anomaly detection, can be integrated to identify evolving phishing tactics that bypass traditional detection methods. The use of neural networks enables pattern recognition in complex datasets, helping the system detect zero-day phishing threats.

Reporting and Alerting

When a phishing website is detected, the system generates a detailed report containing information about the identified threats, including URL analysis, risk score, and potential attack vectors. Email alerts can be configured to notify security teams or users about suspicious websites in real time. The system also provides downloadable reports for cybersecurity professionals, allowing them to take preventive measures. To further improve security, the system can integrate with browser extensions that automatically warn users when they visit malicious sites.

Evaluation and Continuous Improvement

The effectiveness of the phishing detection system is evaluated based on key performance metrics such as accuracy, precision, recall, and F1-score. False-positive and false-negative rates are analyzed to refine classification rules and improve detection accuracy. Security assessments, including penetration testing, are conducted to identify potential vulnerabilities and enhance system security. Feedback mechanisms are integrated, allowing users to report undetected phishing sites, which helps in continuously improving the system. The project's future scope includes expanding the system's capabilities by incorporating AI-driven threat intelligence and blockchain-based URL verification for enhanced phishing detection.

IV.CONCLUSION AND FUTURE WORK

With the integration of machine learning techniques and the PhishTank API, a centralized phishing detection platform has been developed to enhance cybersecurity by identifying malicious websites in real time. This system enables users to verify URLs efficiently, ensuring protection against phishing attacks. The backend, built using Flask and Python, processes website URLs, extracts key features, and cross-verifies them with known phishing databases. The integration of the PhishTank API strengthens detection capabilities, providing real-time verification and automated updates on emerging phishing threats. Additionally, machine learning models are employed to analyze website structures, domain age, and other indicators to detect phishing attempts beyond existing blacklists.

To ensure comprehensive threat analysis, various detection mechanisms have been implemented, including URL feature extraction, domain reputation checks, and content analysis. The system automatically categorizes URLs based on their risk levels, allowing for prompt action against suspicious links. Moreover, user feedback is incorporated to refine detection models continuously. A reporting module is also integrated, enabling users to generate detailed security reports on analyzed websites, which aids in cybersecurity awareness and threat mitigation. By combining heuristic analysis with data-driven machine learning techniques, the system improves accuracy in identifying evolving phishing tactics.

In summary, the phishing detection system provides a proactive defense against cyber threats by leveraging API-based validation and AI-powered phishing identification. By integrating threat intelligence sources, real-time alerting mechanisms, and continuous model training, the system enhances online security. Future enhancements may include integrating browser extensions for real-time URL analysis, deep learning models for enhanced classification, and blockchain-based validation to strengthen authenticity checks, ensuring robust protection against phishing attacks.

REFERENCES

[1] Sharma, R., & Gupta, P. "Real Estate Price Prediction using Machine Learning Techniques," in Proceedings of the International Conference on Data Science and Applications, 2021.

[2] Thomas, L., & Wang, H. "Integrating AI for Accurate Property Valuation in Dynamic Markets," Journal of Real Estate and Urban Economics, 2020.

[3] Kumar, S., & Patel, A. "AI-Driven Market Valuation for Real Estate," IEEE Transactions on Artificial Intelligence, 2022.

[4] Li, H., & Zhang, Y. "Deep Learning for Real Estate Price Estimation," International Journal of Data Science and Analytics, 2021.

[5] Williams, J. "A Comparative Study on Machine Learning Algorithms for Property Price Prediction," ACM Transactions on Intelligent Systems, 2020.

[6] Banerjee, P., & Choudhary, R. "Big Data in Real Estate Market Analysis," Springer Journal of Computational Intelligence, 2021.

[7] Singh, M. "Neural Networks for Real Estate Pricing," International Conference on AI and Smart Cities, 2022.

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com