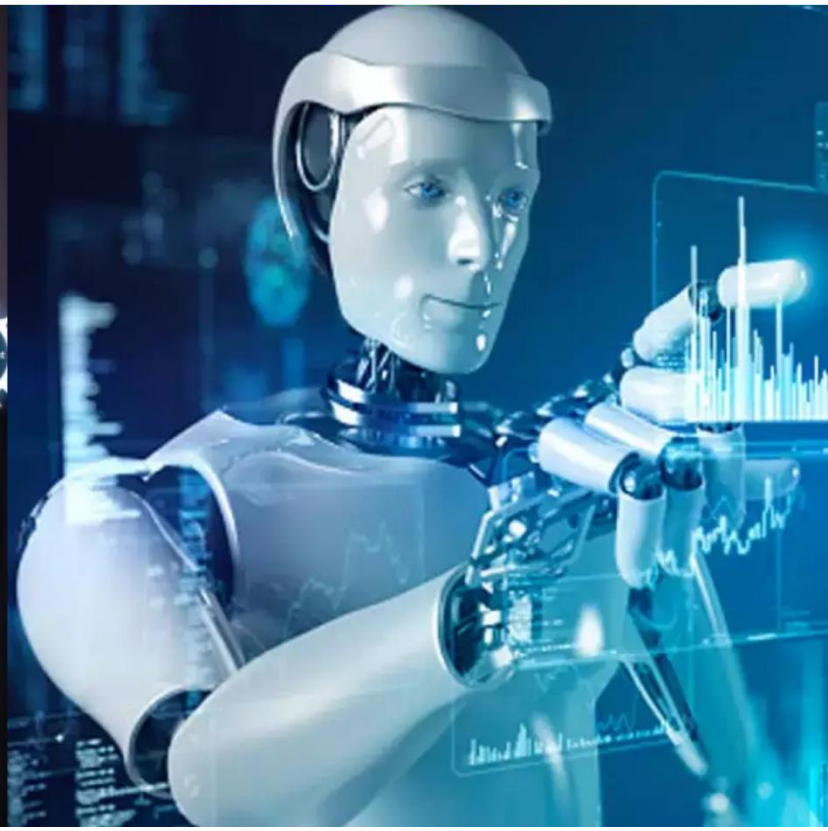
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504208|

Machine Learning Based Phishing URL Detection System

Yuvarani C [1], Vaitheeswari M [2], Nandhini.M [3], Karthik R [4]

UG Student, Department of AI&DS, Jayalakshmi Institute of Technology, Thoppur, Tamil Nadu, India[1,2,3]

Assistant Professor, Department of AI&DS, Jayalakshmi Institute of Technology, Thoppur, Tamil Nadu, India[4]

ABSTRACT: With the rapid growth of online services and digital transactions, phishing attacks have emerged as a serious cybersecurity threat, targeting users by imitating legitimate websites to steal sensitive information. This project presents a machine learning–based phishing URL detection system that aims to identify and prevent access to fraudulent websites. The proposed approach utilizes a Multilayer Perceptron (MLP) classifier to distinguish between phishing and legitimate URLs based on a set of extracted website and URL features. A labelled dataset containing both genuine and malicious URLs is used to train the model, enabling it to learn patterns commonly associated with phishing behaviour. When a user submits a URL, the system analyses its features and classifies it as either safe or phishing in real time. The experimental results demonstrate that the proposed system effectively detects phishing URLs with high accuracy, thereby enhancing user security and reducing the risk of online fraud. This approach provides a reliable and automated solution for improving cybersecurity in web-based environments.

I. INTRODUCTION

The rapid expansion of the internet and digital technologies has transformed the way individuals interact, communicate, and conduct financial transactions. Online platforms such as banking services, e-commerce websites, and social media applications have become integral parts of daily life, enabling users to access services conveniently from anywhere in the world. However, this widespread digital adoption has also opened new avenues for cybercriminals to exploit unsuspecting users through various forms of attacks, among which phishing is one of the most prevalent and dangerous threats. Phishing attacks involve the creation of deceptive websites that mimic legitimate platforms to trick users into revealing sensitive information such as login credentials, credit card details, and personal data.

II OBJECTIVES

- The primary objective of this project is to design and implement an intelligent system capable of identifying phishing URLs using machine learning techniques.
- The system focuses on extracting critical features such as URL length, domain age, presence of special characters, and security indicators.
- Another key objective is to provide real-time analysis of user-submitted URLs.
- The project aims to achieve high classification accuracy while reducing false positives and false negatives.
- The system also aims to contribute to improving user awareness regarding phishing attacks.

II LITERATURE SURVEY

[1] **Nayak et al. (2025)** proposed a phishing detection system that integrates feature selection techniques with machine learning and deep learning models to improve classification accuracy.

[2] Li et al. (2024) presented a comprehensive review of phishing detection techniques, emphasizing the transition from traditional methods to machine learning-based approaches. The study categorizes detection methods into URL-based, content-based, and hybrid approaches, providing insights into their advantages and limitations.

[3] Remya et al. (2025) introduced BGL-PhishNet, a hybrid model that integrates Bidirectional Encoder Representations from Transformers (BERT), Graph Neural Networks (GNN), and Light Gradient Boosting Machine (Light GBM).

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504208|

[4] Lee et al. (2024) explored the use of multimodal large language models for phishing webpage detection. Their approach combines textual, visual, and contextual information from web pages to identify phishing attempts.

[5] Pillai et al. (2023) analyzed various evasion techniques used against machine learning-based phishing classifiers. Their study highlights how attackers modify URL structures, use encoding techniques, and generate adversarial examples to deceive detection systems.

III. EXISTING & PROPOSED SYSTEM

EXISTING SYSTEM

The existing system for phishing URL detection primarily relies on traditional security mechanisms such as blacklist-based filtering, rule-based approaches, and manual inspection techniques. These methods maintain databases of known malicious URLs and compare user-entered links against them to identify threats. However, such systems are limited in their ability to detect newly generated or unknown phishing websites, as they depend heavily on previously reported data. Rule-based systems also struggle with adaptability, as attackers continuously evolve their techniques to bypass predefined rules. Additionally, manual verification processes are time-consuming and prone to human error, making them inefficient for real-time detection. Most existing solutions lack intelligent learning capabilities and fail to analyze complex patterns in URLs and website behavior. As a result, these systems often produce higher false positives and false negatives. Therefore, there is a need for a more dynamic and intelligent approach to accurately detect phishing URLs in real time.

PROPOSED SYSTEM

The proposed system presents a machine learning–based phishing URL detection approach that utilizes a Multilayer Perceptron (MLP) classifier to accurately identify malicious websites. The system extracts various URL and website features such as length, domain characteristics, and special characters, and uses them as input for the trained model. A labelled dataset containing both phishing and legitimate URLs is used to train and validate the model, enabling it to learn complex patterns associated with phishing attacks. When a user submits a URL, the system processes its features and classifies it in real time as safe or phishing. This approach improves detection accuracy, reduces dependency on traditional blacklist methods, and enhances cybersecurity by providing a fast, automated, and reliable solution for preventing online fraud.

IV. BLOCK DIAGRAM

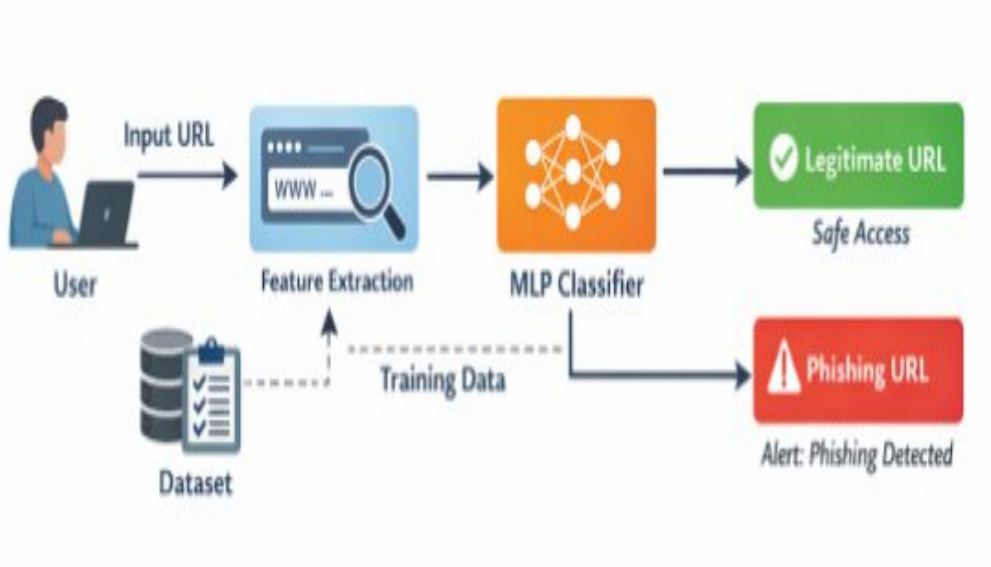


Fig No: 1 Software block diagram

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504208|

V SOFTWARE REQUIREMENTS

1. Input and Data Handling Layer:
This layer serves as the primary interface between the users, administrators, and the system. It is responsible for collecting all incoming data, including datasets uploaded by the administrator and URLs submitted by users for analysis. The layer ensures that all inputs are properly validated before being processed further. For example, URL formats are checked to ensure correctness, and datasets are verified to meet structural requirements. It also implements secure communication protocols such as HTTPS to protect data during transmission, preventing unauthorized access or interception.

2. Data Pre-processing Layer:
The Data Pre-processing Layer is responsible for transforming raw input data into a clean and structured format suitable for analysis and model training. This layer performs several important operations such as handling missing values, removing duplicate records, and correcting inconsistencies within the dataset. For URL inputs, it extracts meaningful features such as domain length, presence of special characters, use of HTTPS protocol, number of subdomains, and suspicious keywords. These features are essential for identifying phishing patterns
.

3. Machine Learning and Detection Layer:
This layer is the core component of the phishing detection system, where the actual analysis and classification of URLs take place. It is responsible for building and applying machine learning models that can distinguish between legitimate and phishing websites. During the training phase, models such as Decision Trees, Random Forest, or Support Vector Machines are trained using pre-processed datasets provided by the administrator. These models learn patterns and characteristics commonly associated with phishing attacks. During the prediction phase, the trained model analyses user-input URLs in real time and classifies them based on learned patterns. In addition to machine learning, this layer may also incorporate rule-based techniques such as checking URLs against known blacklists, verifying domain reputation, and analyzing abnormal behaviours.

4. Output and Response Layer:
The Output and Response Layer is responsible for presenting the results of the phishing detection process to the user in a clear and understandable manner. Once the machine learning model completes its analysis, this layer interprets the results and generates appropriate responses. It indicates whether a given URL is safe or potentially phishing and may also provide additional details such as risk level, detected suspicious features, or blacklist status. The information is displayed through a user-friendly interface, ensuring that users can easily understand the outcome. This layer also handles alert mechanisms, warning users when a malicious URL is detected.

VI MODULES OF THE SYSTEM

The system is divided into two main modules: the Admin Module and the User Module, each responsible for specific functionalities within the phishing detection system.

➢ **Admin Module**
The Admin Module plays a vital role in managing the backend operations of the phishing detection system. It is responsible for maintaining datasets, preprocessing data, and building machine learning models that are used for detecting phishing websites. This module ensures that the system remains updated and continues to improve its detection capabilities over time.

➢ **Login:**
The login functionality allows administrators to securely access the system using valid credentials such as username and password. This feature ensures that only authorized individuals can perform administrative tasks. It includes authentication mechanisms that verify user identity and prevent unauthorized access. Additionally, security measures such as password encryption and session management are implemented to enhance system security.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504208|

➢ **Upload Dataset:**
This feature enables administrators to upload datasets that contain information about phishing and legitimate websites. These datasets are crucial for training machine learning models. The system supports structured data formats and validates the dataset before accepting it. Proper dataset management ensures that the system has access to high-quality data for accurate predictions.

➢ **Preprocessing:**
The preprocessing functionality prepares the uploaded dataset for model training. It involves cleaning the data, removing duplicates, handling missing values, and transforming it into a structured format. Feature extraction techniques are also applied to identify important attributes required for phishing detection. This step improves the efficiency and accuracy of the machine learning models.

➢ **Model Build:**
In this stage, machine learning algorithms are used to train models based on the preprocessed dataset. The models learn patterns and characteristics associated with phishing websites. Once trained, these models are stored and used for analyzing user-input URLs. This functionality ensures that the system continuously improves its detection capabilities.

➢ **User Module**
The User Module is designed to provide an interactive and user-friendly interface for individuals who want to check whether a URL is safe or phishing. It allows users to register, log in, and perform URL analysis efficiently.

➢ **Register:**
The registration feature allows new users to create an account by providing details such as username, email, and password. The system validates the input data to ensure correctness and prevents duplicate registrations. Security measures such as password encryption are implemented to protect user information. This feature ensures that only valid users can access the system.

➢ **Login:**
The login functionality allows registered users to access the system using their credentials. It verifies user identity and ensures secure access to system features. Session management techniques are used to maintain user sessions and prevent unauthorized access.

➢ **Input URL:**
This feature enables users to enter the URL they want to analyze. The system validates the URL format and sends it to the backend for processing. The input is then analyzed using machine learning models and detection techniques to determine its legitimacy

VII. RESULT

The machine learning–based phishing URL detection system demonstrate high effectiveness in identifying malicious websites. The Multi layer Perceptron (MLP) model achieved strong classification performance with high accuracy, precision, and recall on the test dataset. The system was able to correctly distinguish between phishing and legitimate URLs with minimal false positives and false negatives. Real-time testing showed that the model responds quickly to user input, ensuring efficient detection without delay. Compared to traditional methods, the proposed system provides improved adaptability in detecting new and unseen phishing attacks. Overall, the results confirm that the system is reliable, accurate, and suitable for enhancing web security applications.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504208|

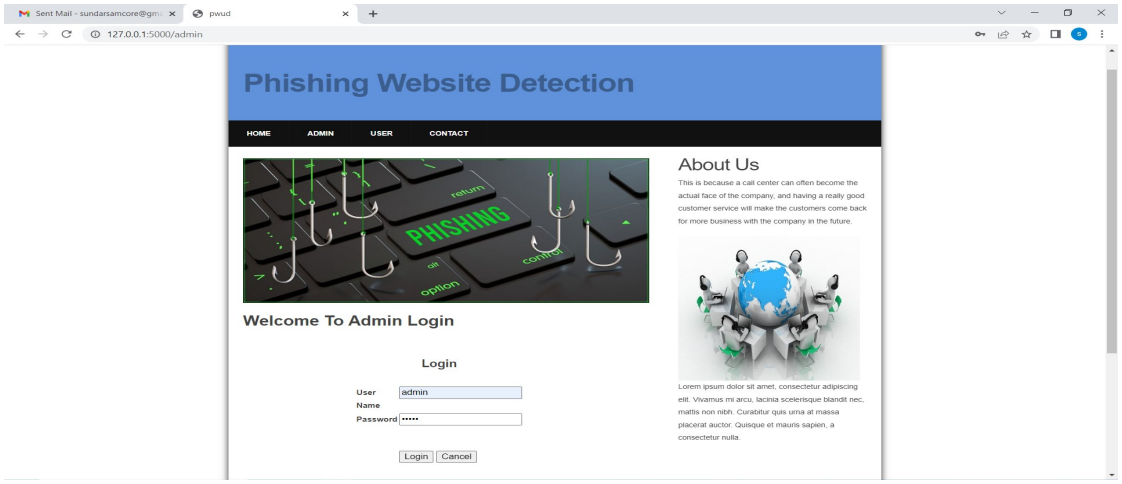


Fig No: 2 Login Page

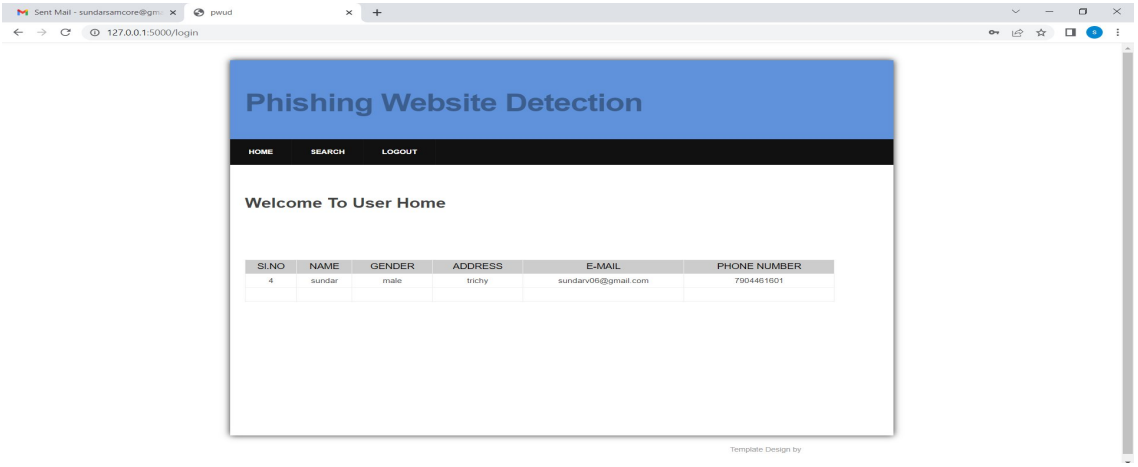


Fig No: 3 Welcome to user Home

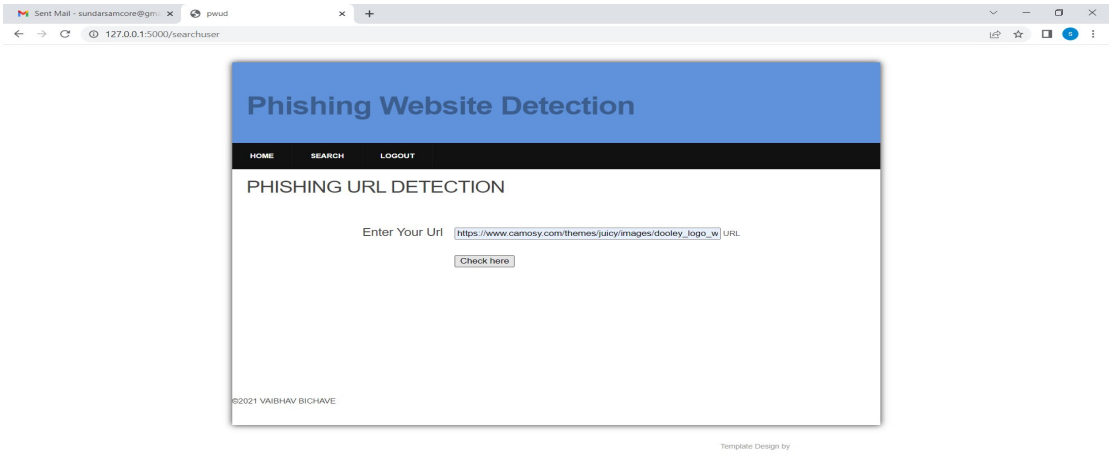


Fig No: 4 URL Detection

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504208|

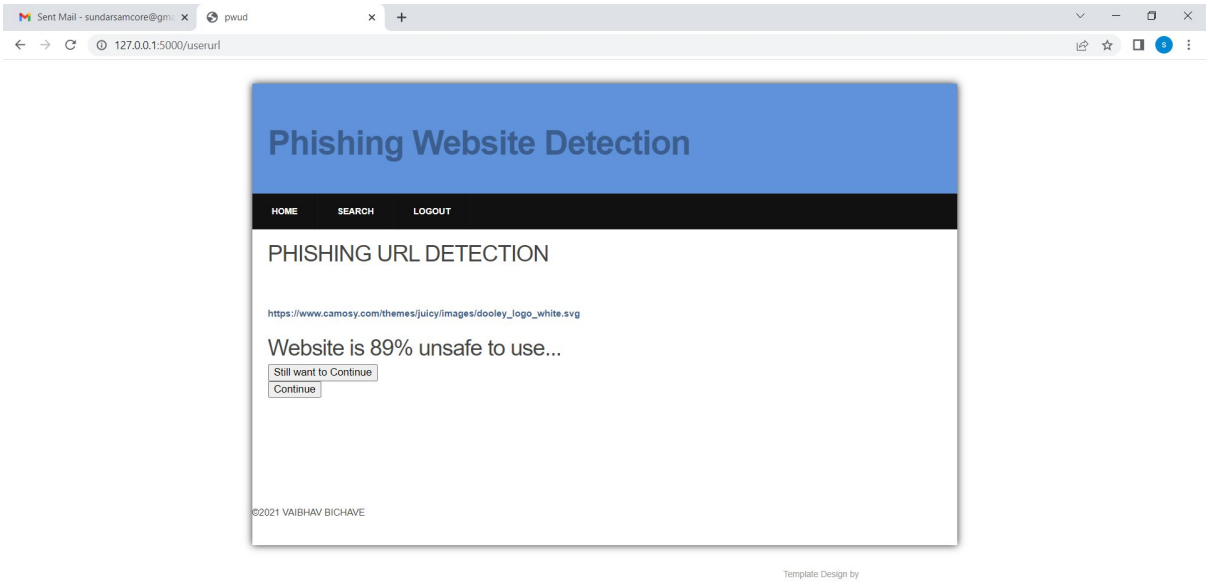


Fig No: 5 Final Output

VIII. CONCLUSION

The machine learning–based phishing URL detection system provides an effective and reliable solution for identifying malicious websites in real time. By utilizing a Multilayer Perceptron (MLP) model, the system is capable of learning complex patterns from URL features and accurately classifying phishing and legitimate links. This approach overcomes the limitations of traditional blacklist-based methods by detecting previously unseen attacks with improved accuracy. The system ensures fast response, reduced false detections, and enhanced user protection during online activities. Overall, the project demonstrates the potential of machine learning in strengthening cybersecurity and offers a scalable solution for real-world web security applications.

REFERENCES

[1] Putra, Fauzan Prasetyo Eka, et al. "Analysis of phishing attack trends, impacts and prevention methods: Literature study." Brilliance: Research of Artificial Intelligence 4.1 (2024): 413-421.
[2] Nayak, Ganesh S., Balachandra Muniyal, and Manjula C. Belavagi. "Enhancing phishing detection: a machine learning approach with feature selection and deep learning models." IEEE Access (2025).
[3] Li, Wenhao, et al. "A state-of-the-art review on phishing website detection techniques." IEEE Access (2024).
[4] Pillai, Manu J., et al. "Evasion attacks and defense mechanisms for machine learning-based web phishing classifiers." IEEE Access 12 (2023): 19375-19387.
[5] Lee, Jehyun, et al. "Multimodal large language models for phishing webpage detection and identification." 2024 APWG Symposium on Electronic Crime Research (eCrime). IEEE, 2024.
[6] He, Shen, Bangling Li, Huaxi Peng, Jun Xin, and Erpeng Zhang. "An effective cost-sensitive XGBoost method for malicious URLs detection in imbalanced dataset." IEEE Access, Vol: 9, 2021
[7] Faris, Humam, and Setiadi Yazid. "Phishing web page detection methods: URL and HTML features detection." In IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), 2020.
[8] Nathezhtha, T., D. Sangeetha, and V. Vaidehi. "WC-PAD: web crawling based phishing attack detection." In International Carnahan Conference on Security Technology (ICCST), 2019.
[9] Chin, Tommy, Kaiqi Xiong, and Chengbin Hu. "Phishlimiter: A phishing detection and mitigation approach using software-defined networking." IEEE Access, Vol: 6, 2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details