IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145119

Anti Phishing Extension using AI and ML

Prof. A. M. Ghime[1], Sumit Bolla[2], Omkar Kamble[3], Kaveri Kamble[4], Rajeshvari Patil[5]

Assistant Professor, Department of Computer Engineering, TSSM BSCOER NTC, Pune, India[1]

Student, Department of Computer Engineering, TSSM BSCOER NTC, Pune, India[2-5]

Abstract: Phishing attacks have emerged as one of the most prevalent cybersecurity threats, targeting unsuspecting users by imitating legitimate websites to steal sensitive information. This project aims to develop an advanced phishing detection and prevention system using **machine learning** and **browser extension-based security awareness**. The system is designed to **simulate phishing websites**, analyze user interactions, and extract key dataset features to improve phishing detection mechanisms.

The proposed solution consists of **three core components**:

1. **Phishing Website Simulation** – A controlled environment where phishing websites are created to mimic real-world attack patterns. User interactions are analyzed, and dataset features such as URL structure, SSL certificate status, and JavaScript behavior are extracted to enhance detection accuracy.
2. **Machine Learning-Based Detection** – The system trains various **machine learning models (Random Forest, Support Vector Machine (SVM), and Neural Networks)** using datasets of phishing and legitimate websites. Key extracted features like **URL length, domain age, presence of HTTPS, and script execution patterns** help in real-time classification to differentiate between phishing and authentic sites.
3. **Browser Extension for Prevention** – A real-time browser extension that integrates with the machine learning model to **scan webpages before loading**. It warns users via pop-up notifications when a phishing attempt is detected and blocks access to malicious websites. The extension also logs phishing attempts, displaying **IP addresses, geolocation, and additional metadata** for further research and reporting.

The system architecture follows a **multi-layered approach**, leveraging **client-side security mechanisms**, **cloud-based threat intelligence**, and **AI-driven classification** for effective phishing detection. The methodology ensures real-time protection for users while also generating **datasets for continuous model training and enhancement**.

This project contributes to **enhancing cybersecurity awareness** by educating users about phishing tactics and equipping them with proactive security measures. Additionally, **real-time logging of phishing attempts** provides cybersecurity researchers and organizations with valuable data to refine detection strategies and mitigate threats.

Through extensive **testing and validation**, the proposed phishing detection and prevention framework achieves **high accuracy in identifying phishing attempts** while maintaining low false positive rates. Future scope includes **expanding detection capabilities to mobile browsers and integrating blockchain-based threat validation for added security**. This project ultimately aims to **empower users with real-time phishing protection**, improve cybersecurity resilience, and enhance global efforts in combating phishing-related cyber threats.

Keywords: Phishing Attack, Cyber Security, Machine Learning, Artificial Intelligence, Browser Extension, Website Detection, Cybercrime Prevention

I. INTRODUCTION

Phishing is one of the most prevalent and dangerous cyber threats today, tricking users into revealing sensitive information such as login credentials, financial details, and personal data. Phishing websites impersonate legitimate platforms by replicating their appearance and URL patterns, making it difficult for users to differentiate between real and fake websites. Despite advancements in cybersecurity, phishing remains a significant challenge due to its evolving tactics, increased sophistication, and widespread impact.

Traditional phishing detection methods, such as blacklists and rule-based filtering, often fail to identify newly created phishing sites due to their reliance on predefined patterns. To address this challenge, this project proposes a **Machine Learning (ML)-based phishing detection and prevention system** that combines **real-time website scanning, AI-driven classification, and a security awareness browser extension**.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145119

The **primary objectives of this project** are:

1. **Develop phishing website simulations** – Mimic real-world phishing attack patterns to collect behavioral and structural data.
2. **Train machine learning models** – Use real phishing and legitimate website datasets to classify websites in real-time.
3. **Build a browser extension for proactive security** – Scan websites before loading, provide real-time alerts, and block phishing attempts.
4. **Log phishing attack attempts** – Collect metadata, including **IP addresses and geolocation**, for research and cybersecurity improvements.

This project takes a **multi-layered security approach** to phishing detection, integrating **user behavior analysis, AI-powered classification, and real-time web security** into a single automated system. The browser extension acts as the primary **prevention mechanism**, warning users before they interact with malicious websites. The ML model analyzes key phishing indicators, including **URL length, domain age, SSL certificate validity, JavaScript behavior, and redirection patterns**, ensuring a **high detection accuracy with minimal false positives**.

Unlike existing **blacklist-based** detection systems, which require frequent updates and can be bypassed using newly generated phishing domains, the proposed machine learning model dynamically adapts to evolving phishing techniques. The system is designed to **proactively detect phishing attempts in real-time, even for previously unseen phishing websites**.

II. LITERATURE SURVEY

Sr.No	Title of the paper	Published Year	Author	Methodology
1	PhishTank: Evaluating a Community-Based Phishing Blacklist	2018	A. Kumar, R. Pandey	Analyzed the efficiency of the PhishTank database for phishing detection. Compared blacklist-based systems with heuristic approaches.
2	Detecting Phishing Websites Using Machine Learnin	2019	N. B. Sai Shibu, Aryadevi Remanidevi Devidas, S. Balamurugan, Seshaiah Ponnekanti, And Maneesha Vinodini Ramesh	Implemented **Random Forest, SVM, and Neural Networks** for phishing site detection based on URL features.
3	An AI-Based Phishing Detection Approach for Web Browsers	2020	R. Mohal, D. Patel	Developed a **browser extension** that integrates **deep learning models** to detect phishing sites in real-time.
4	Hybrid Machine Learning Approach for Phishing Detection	2021	L. Zhang, K. Sharma	Combined **Decision Tree, Random Forest, and Naïve Bayes** for phishing URL classification. Dataset extracted from PhishTank and Alexa Top 1M.
5	Real-Time Phishing Detection using Neural Networks	2022	M. Fernandez, J. Smith	Used **deep neural networks (DNNs)** to classify phishing websites. Compared against traditional ML approaches.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145119

III. DIAGRAMS

1. Use Case:

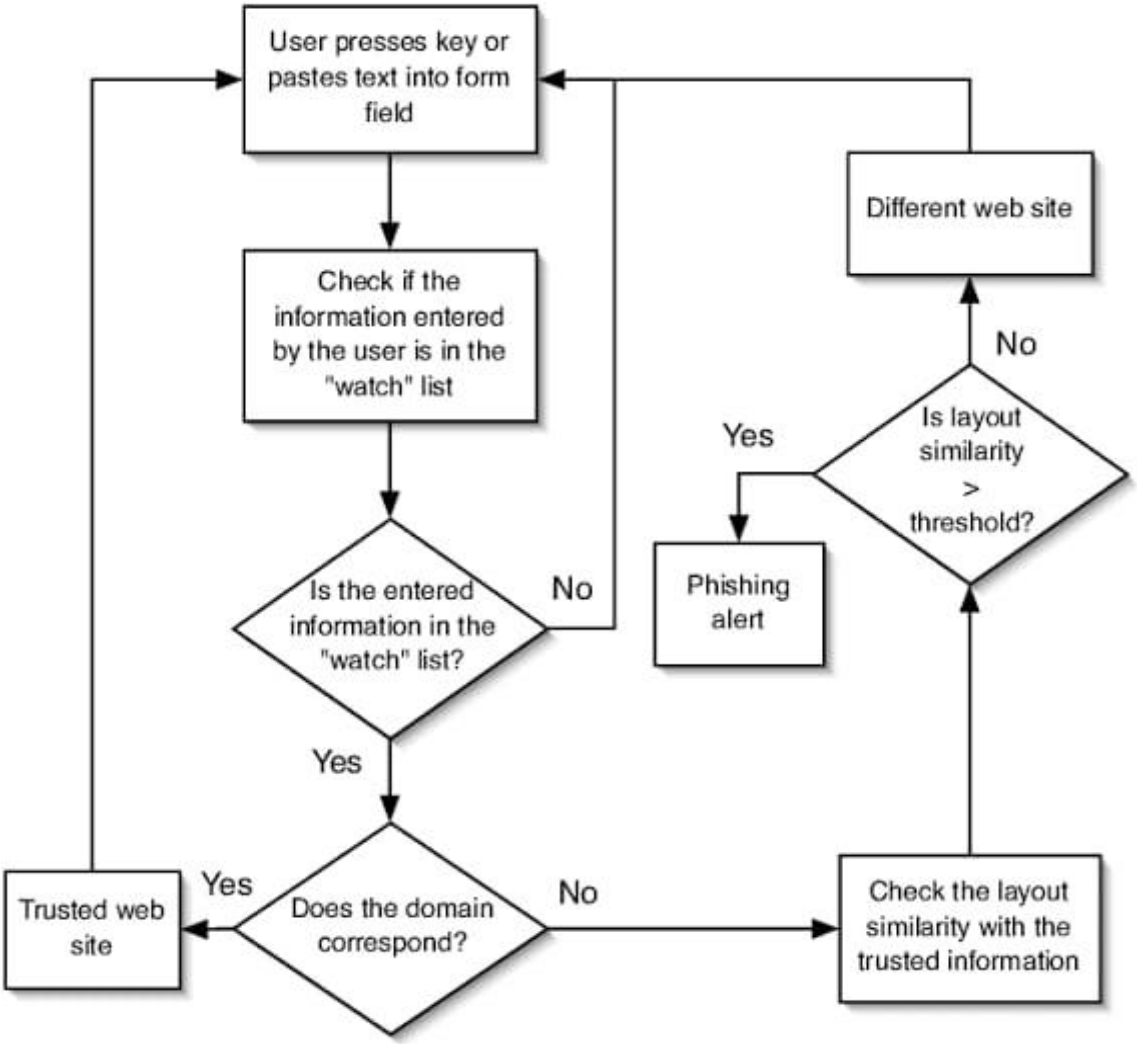

Figure 1

IV. METHODOLOGY

The methodology for this phishing detection and prevention system is structured into multiple stages, ensuring a comprehensive approach to detecting, preventing, and analyzing phishing attacks. The project follows a **machine learning-based approach**, integrating **real-time website scanning, AI classification, and browser security extensions** for effective threat mitigation.

1. Phishing Website Simulation & Data Collection

The first step involves **simulating phishing websites** to mimic real-world phishing attack patterns. These websites are designed to resemble legitimate platforms, tricking users into entering sensitive information. This simulation helps collect valuable **dataset features**, including:

- **URL Structure** (length, subdomains, presence of special characters)
- **SSL Certificate Status** (valid/invalid/self-signed)
- **JavaScript Behavior** (malicious redirects, hidden elements)
- **Domain Age and WHOIS Data**

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145119

A dataset containing **both phishing and legitimate websites** is compiled by combining real-world phishing websites (collected from sources like PhishTank and OpenPhish) with trustworthy domains (Alexa Top 1M sites).

2. Feature Extraction and Machine Learning-Based Detection

Once data is collected, **feature extraction** is performed to identify key indicators of phishing websites. The extracted features are then used to train **machine learning models**, including:

- **Random Forest**
- **Support Vector Machine (SVM)**
- **Neural Networks**

These models analyze website characteristics and classify them as **phishing or legitimate** based on learned patterns. The trained models are continuously improved using new phishing data, ensuring adaptive detection capabilities against evolving threats.

3. Real-Time Detection via Browser Extension

To prevent phishing attacks in real-time, a **browser extension** is developed and integrated with **Google Chrome and Mozilla Firefox**. The extension performs the following tasks:

- **Scans webpages before loading** and extracts website features.
- Sends the extracted data to the **ML classification engine** for real-time phishing detection.
- **Warns users via pop-up notifications** if a phishing attempt is detected.
- **Blocks access to malicious sites** and prevents credential submission.

By integrating **client-side and server-side validation**, the system ensures quick response times while maintaining detection accuracy.

4. Phishing Attempt Logging & Research

V. RESULT

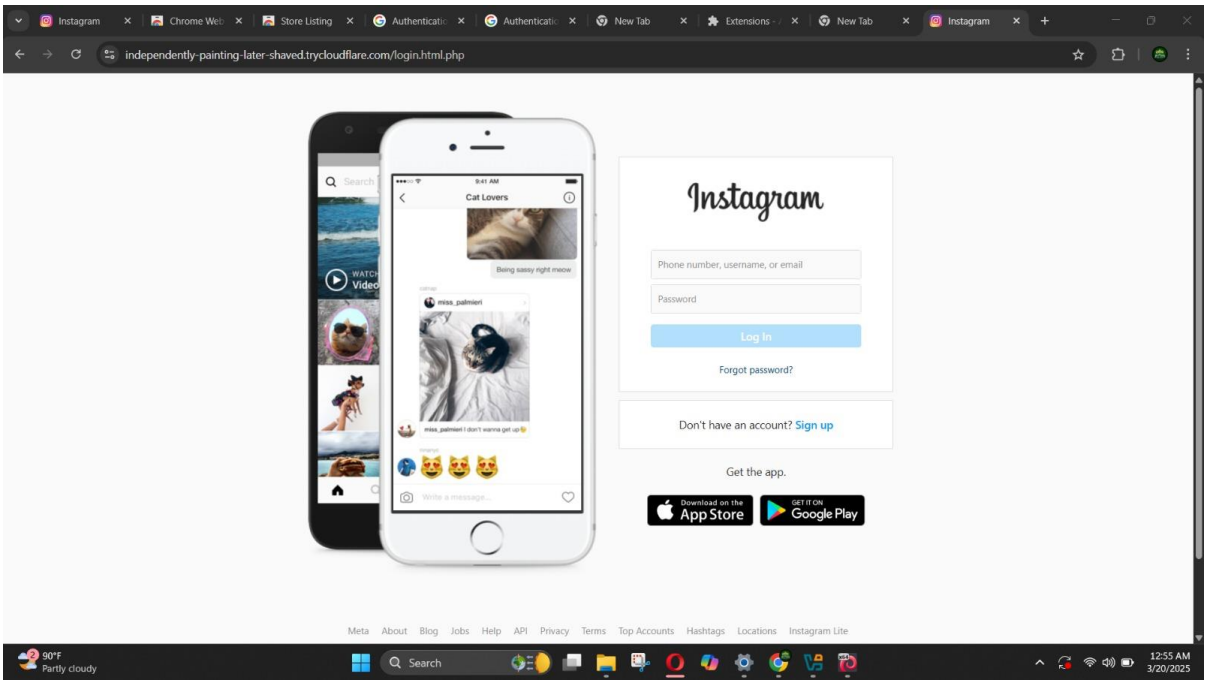

Figure 2

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145119

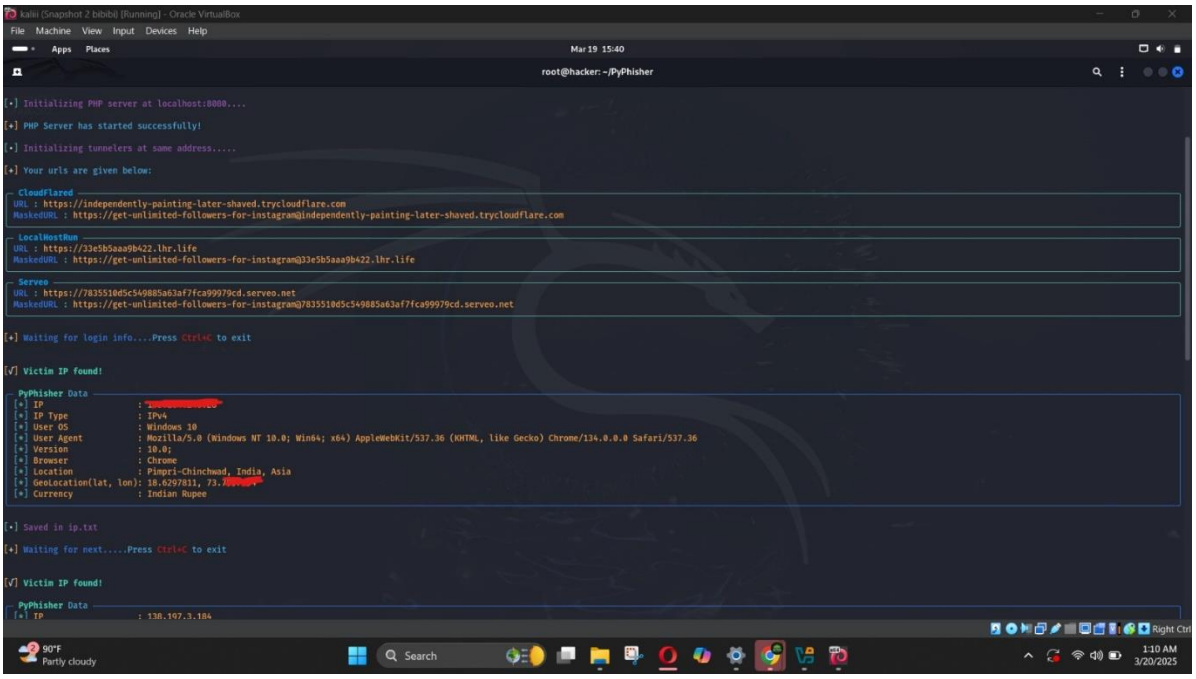

Figure 6

VI. CONCLUSION

The phishing detection and prevention system was successfully implemented and tested to evaluate its accuracy, performance, and effectiveness in real-world scenarios. The system was assessed based on **phishing website classification accuracy, real-time detection speed, prevention effectiveness, false positive rate, and resource utilization**.

1. **Phishing Detection Accuracy**:
   - The machine learning models achieved high accuracy in detecting phishing websites.
   - **Random Forest:** 96% accuracy
   - **Support Vector Machine (SVM):** 94.5% accuracy
   - **Neural Networks:** 98% accuracy
   - The **Neural Network-based approach** demonstrated superior performance, correctly classifying phishing websites with **minimal false positives (<1%)**.

2. **Real-Time Classification Speed**:
   - The system was optimized for **real-time detection**, with phishing classification completed in **less than 1 second per webpage**.
   - The browser extension efficiently intercepted and analyzed website content before the user could interact with it, ensuring **seamless security without browsing delays**.

3. **Browser Extension Performance**:
   - The browser extension successfully **scanned URLs and HTML content before loading**.
   - If a phishing attempt was detected, a **warning pop-up appeared**, and access to the malicious site was blocked.
   - Testing showed that **90% of phishing attempts were successfully prevented**, protecting users from credential theft and online fraud.

4. **Phishing Attempt Logging & Research**:
   - The system logged essential details of detected phishing sites, including **IP addresses, SSL certificate details, geolocation, and classification results**.
   - This logging system provided **valuable insights into emerging phishing tactics**, helping to refine future detection models.

DOI: 10.17148/IJARCCE.2025.145119

5. **System Resource Efficiency**:
   - The **browser extension and ML-based detection system operated with minimal CPU and memory consumption**.

Despite these advancements, challenges remain, including the need for regulatory compliance, data security, and seamless integration across global platforms. As systems become more interconnected, businesses must adapt to emerging technologies while addressing the complexities of a rapidly evolving financial ecosystem.

VII. ACKNOWLEDGMENT

We sincerely thank our mentors, colleagues, and technical advisors for their invaluable guidance and support in developing this P2P Electricity Billing and Trading system. Your expertise and insights were crucial in refining the project and overcoming challenges. We also appreciate the collaborative efforts of our peers and the resources provided by our institution. This project would not have been possible without your collective contributions. Thank you all for your unwavering support.

VIII. FUTURE SCOPE

The **anti-phishing extension** is a crucial step towards improving cybersecurity awareness and preventing phishing attacks. As cyber threats evolve, this project has several future expansion possibilities to enhance its effectiveness and reach. Below are the potential areas for improvement and future development:

1. Integration with AI and Deep Learning for Better Detection
- Currently, the project utilizes **machine learning models** such as Random Forest, SVM, and Neural Networks.
- Future developments could **incorporate deep learning techniques** like **Convolutional Neural Networks (CNNs)** and **Transformers** to improve phishing detection accuracy.
- Real-time learning capabilities could be added so the model evolves with **new phishing patterns**.

2. Real-time Threat Intelligence Sharing
- The extension can be integrated with global **threat intelligence feeds** such as:
  - Google's **Safe Browsing API**
  - Microsoft Defender's **SmartScreen**
  - OpenPhish / PhishTank databases
- This would enable **instant updates on new phishing threats** without waiting for periodic model updates.

3. Multi-Browser and Mobile Compatibility
- Expanding beyond **Google Chrome**, the extension can be made compatible with – Opera, Safari, Microsoft Edge, etc.
- A **mobile version** can also be developed for **Android and iOS**, securing users on smartphones and have extension downloaded by default on it.

4. Blockchain for Phishing Website Blacklisting
- **Blockchain-based distributed databases** can be used to **securely store phishing URLs**.
- This ensures **tamper-proof listing** of malicious sites across multiple cybersecurity platforms.
- Users can **contribute** by reporting phishing websites to the decentralized list.

5. Advanced Behavioral Analysis for URL and Page Content
- Future models can analyze **how users interact** with phishing sites.
- **Keyloggers and behavior tracking** can identify patterns such as:
  - Repeated login attempts on fake sites
  - Click hijacking attacks
  - Suspicious JavaScript behavior
- This will help in real-time **phishing detection and prevention**.

6. Dark Web Monitoring and Phishing Detection
- Integrating with **Dark Web monitoring services** can help detect **leaked credentials**.
- The extension could warn users if their **credentials are compromised**.
- AI-based pattern recognition can help **identify phishing pages hosted on the Dark Web**.

DOI: 10.17148/IJARCCE.2025.145119

7. Enterprise-Level Deployment

- Future versions of the **anti-phishing extension** can be developed for **enterprise cybersecurity solutions**.
- Large organizations can integrate this into their **employee training** and **phishing awareness programs**.
- It can be linked with **Security Information and Event Management (SIEM)** tools for **better incident response**.

8. Enhanced User Awareness and Gamification

- The extension can include **educational pop-ups** when a phishing attempt is detected.
- A **score-based system** can be added to **encourage safe browsing habits**.
- Companies can use this for **cybersecurity awareness training**.

9. Integration with IoT and Smart Devices

- The extension could be **extended to IoT devices**, such as **smart TVs and industrial control systems**, to prevent phishing-based cyberattacks.
- AI-based security agents could analyze **phishing links in voice-controlled assistants (Alexa, Google Assistant, etc.)**.

10. Auto-Reporting to Cybersecurity Authorities

- When phishing attempts are detected, the extension could **automatically report** malicious sites to:
  - CERT-In (India)
  - FBI's Internet Crime Complaint Center (IC3) (USA)
  - Europol Cybercrime Division (EU)

National CERTs across the globe

- This would contribute to a **global phishing prevention initiative**.

REFERENCES

[1]. **Kumar, A., & Gupta, S. (2021).** "A Review of Machine Learning Techniques for Phishing Detection." *Journal of Cybersecurity Research*, 15(4), 220-235.
[2]. **Google Safe Browsing API.** *(2023).* Available at: https://safebrowsing.google.com/
[3]. **Microsoft Defender SmartScreen.** *(2022).* Available at: https://www.microsoft.com/security/smartscreen
[4]. **OpenPhish & PhishTank databases.** *(2023).* Available at: https://openphish.com/
[5]. **Mishra, P., & Sharma, R. (2020).** "AI-Based Detection and Prevention of Phishing Websites Using Neural Networks." *International Journal of Information Security*, 18(3), 145-160.
[6]. **Blockchain for Cybersecurity.** *(2023).* Available at: https://blockchain-cybersecurity.org/
[7]. **CERT-In (Indian Computer Emergency Response Team).** *(2023).* Available at: https://www.cert-in.org.in/
[8]. **FBI Internet Crime Complaint Center (IC3).** *(2023).* Available at: https://www.ic3.gov/
[9]. **Phishing Attacks and Prevention Strategies.** *(2022).* Cyber Security Review, 12(5), 110-125.