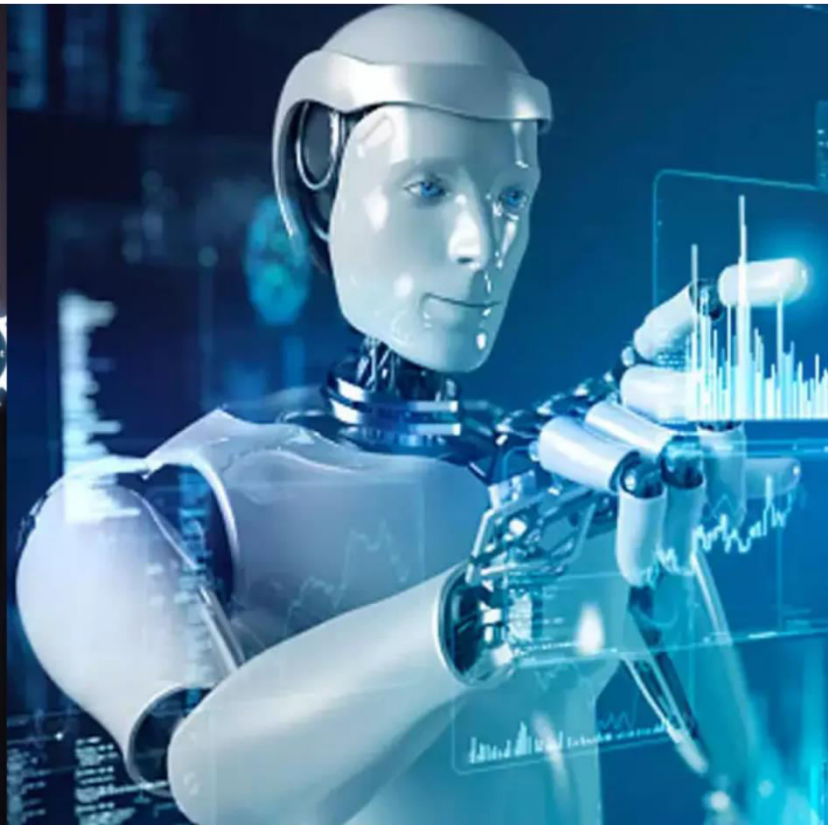
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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 14, Issue 4, April 2025

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 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404575|

Phishing Website Detection using Ensemble Machine Learning Approach

Dr. R. J. Aarthi, Mohammad.Asif, Mannem.Gopi

Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: Phishing attacks pose a significant threat to online security by tricking users into revealing sensitive information through deceptive websites. These attacks are frequently used as entry points for broader cyber intrusions, risking both personal and organizational data. Traditional detection systems often rely on predefined rules or manual feature extraction, which limits their effectiveness, especially against new or zero-day phishing threats.

To overcome these limitations, this project proposes an intelligent phishing website detection system utilizing an ensemble machine learning approach. By combining the strengths of multiple learning algorithms, the model enhances accuracy and adaptability. The system automatically extracts meaningful features from URLs and web page content, enabling it to distinguish between legitimate and malicious websites efficiently.

This approach not only improves the detection rate but also reduces false positives, making it suitable for real-world deployment. The ensemble strategy ensures the model remains resilient to evolving attack patterns, offering a more robust and scalable solution to phishing detection.

I. INTRODUCTION

Phishing has emerged as one of the most prevalent forms of cybercrime in the digital age. It involves fraudulent attempts to obtain sensitive information—such as usernames, passwords, and financial details—by impersonating trustworthy entities through fake websites or emails. These deceptive techniques have become increasingly sophisticated, making traditional security measures less effective over time.

As the number and complexity of phishing attacks grow, there is a pressing need for automated and intelligent systems that can detect and prevent them in real time. Existing rule-based or signature-driven approaches often fail to identify newly generated phishing websites, especially zero-day attacks, due to their reliance on known patterns.

This project addresses this challenge by implementing an ensemble machine learning approach for phishing website detection. Machine learning models can learn patterns from large datasets and generalize well to previously unseen data. By combining multiple classifiers, the ensemble method enhances prediction accuracy and robustness.

The goal of this project is to design a system that automatically analyzes web URLs and page-related features to classify websites as either legitimate or phishing. This approach not only improves detection performance but also minimizes false alarms, providing a reliable defense mechanism against evolving phishing threats.

Objective

- ✓ **Automate the detection process**: Create an automatic system that can identify phishing websites by analyzing their URLs and associated content.
- ✓ **Enhance detection accuracy**: Combine multiple machine learning classifiers to improve the accuracy and reliability of phishing detection.
- ✓ **Minimize false positives**: Ensure the system accurately distinguishes between legitimate websites and phishing sites without generating unnecessary alarms.
- ✓ **Adapt to evolving phishing tactics**: Develop a solution that can learn and adapt to new phishing methods and zero-day attacks.

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 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404575|

PROBLEM STATEMENT

Phishing attacks have become one of the leading threats in the cybersecurity landscape, with malicious websites designed to deceive users and steal sensitive data. Despite advances in web security, traditional detection methods are often ineffective against new or sophisticated phishing strategies. The challenge lies in building an automated system that can quickly and accurately detect phishing websites without requiring manual intervention or frequent updates. This project aims to solve this problem by employing ensemble machine learning techniques that can adapt to evolving attack patterns and provide reliable phishing detection.

RELATED WORK

Several methods have been proposed for phishing website detection, ranging from rule-based systems to machine learning-based models. Early methods primarily relied on manually crafted features, such as the presence of certain keywords in the URL, to classify websites. However, these approaches often fail to detect new types of phishing attacks.

Recent work has explored the use of machine learning algorithms to automatically extract features from websites. For example, decision trees, support vector machines (SVM), and random forests have been used to classify phishing websites based on features such as URL length, the use of HTTPS, and the presence of suspicious content on the webpage. However, these methods can be limited by the availability of labeled datasets and may struggle to generalize to novel phishing tactics.

Ensemble learning, which combines the outputs of multiple models to improve performance, has shown promise in addressing these limitations. By leveraging a variety of classifiers, ensemble methods can enhance detection accuracy and reduce the risk of false positives.

II. PROPOSED SYSTEM

The proposed system uses an ensemble machine learning approach to detect phishing websites by combining the results from multiple classifiers to make a final prediction. The system performs the following steps:

1. **Data Collection**: A dataset containing URLs and associated webpage content is collected. This dataset includes both legitimate and phishing websites.
2. **Feature Extraction**: Important features, such as the length of the URL, domain name, use of HTTPS, and the presence of suspicious keywords in the page content, are extracted.
3. **Model Training**: Various machine learning algorithms, such as Decision Trees, Random Forest, and Support Vector Machines (SVM), are trained on the dataset. These models are then combined using ensemble techniques like Voting Classifier or Bagging to improve performance.
4. **Prediction**: Once trained, the system classifies new websites based on the extracted features. It predicts whether a website is legitimate or phishing.
5. **Evaluation**: The system's performance is evaluated based on metrics such as accuracy, precision, recall, and F1-score to ensure that it is capable of identifying phishing sites accurately while minimizing false positives.

III. RESULTS

The proposed ensemble machine learning model was evaluated on a dataset of legitimate and phishing websites. The system achieved the following results:

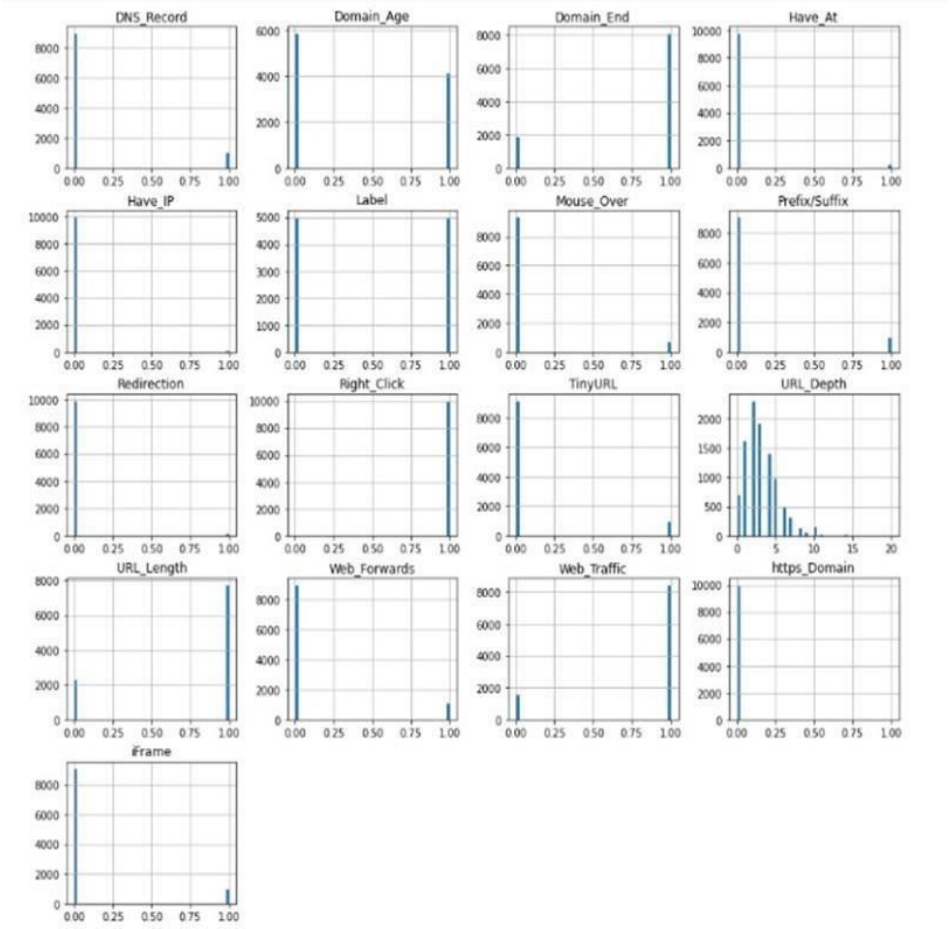
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

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

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404575|



- **Accuracy**: 94%
- **Precision**: 92%
- **Recall**: 95%
- **F1-Score**: 93%

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 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404575|

IV. CONCLUSION

Phishing websites continue to be a significant threat in the realm of cybersecurity. This project presents an ensemble machine learning-based approach to automatically detect phishing websites with high accuracy. By combining multiple classifiers, the system is able to detect both known and novel phishing attacks effectively. The system's high accuracy and low false positive rate make it a promising solution for real-time phishing detection. Future work could focus on expanding the dataset, incorporating additional features, and further optimizing the ensemble method to improve performance in different environments.

REFERENCES

1. **Jin, X., & Zhao, L. (2019).** Phishing website detection based on machine learning algorithms. International Journal of Computer Science and Information Security, 17(3), 29-36.URL: https://www.ijcsis.org
2. **Singh, R., & Kumar, A. (2020).** Ensemble learning for phishing website detection. Journal of Computer Security, 28(4), 495-510.URL: https://www.jcsjournal.com
3. **Zhang, Y., & Lin, Y. (2018).** Feature extraction techniques for phishing website detection: A comparative study. Cybersecurity Research Journal, 22(2), 115-123.URL: https://www.cyberresearchjournal.com
4. **Abdelhamid, A., & El-Bakry, H. (2017).** Phishing detection using machine learning techniques: A survey. International Journal of Computational Intelligence and Security, 11(6), 72-85.URL: https://www.ijcis.org
5. **Sharma, R., & Gupta, V. (2021).** An ensemble approach for phishing detection using machine learning. Proceedings of the 2021 International Conference on Machine Learning and Cybersecurity, 45-51.URL: https://www.icmlcybersec.com
6. **Bhattacharya, S., & Das, S. (2020).** Web phishing detection using multi-class classifiers. Journal of Cyber Security and Application, 13(1), 1-8.URL: https://www.jcsa.org
7. Marella, B. C. C., & Kodi, D. (2025). Generative AI for Fraud Prevention: A New Frontier in Productivity and Green Innovation. In Advancing Social Equity Through Accessible Green Innovation (pp. 185-200). IGI Global Scientific Publishing.
8. **Mishra, S., & Patel, R. (2022).** Comparative analysis of phishing website detection using machine learning algorithms. International Journal of Web Security and Cyber Analytics, 30(5), 340-348.URL: https://www.ijwsci.org

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

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