www.ijarp.com ISSN 2456-9992 Page: 01-09

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 05

PHISHING DETECTION AND PREVENTION SYSTEM

*[1]Harsh Wardhan Karn, [1]Jaypal Gupta, [1]Gulshan Kumar, [1]Eed Mohammad, [2]Sonal Raj

[1]Dept. of Computer Science and Engineering, IIMT College of Engineering AKTU.

[2]Assistant Professor, Dept. of Computer Science and Engineering, AKTU.

Article Received: 25 March 2026, Article Revised: 15 April 2026, Published on: 05 May 2026

***Corresponding Author: Harsh Wardhan Karn**

Dept. of Computer Science and Engineering, IIMT College of Engineering AKTU.

DOI: https://doi-doi.org/101555/ijarp.4064

ABSTRACT

Phishing is one of the most common cyber threats that targets users by creating fake websites, emails, or links to steal sensitive information such as usernames, passwords, banking credentials, and personal data. As phishing attacks are becoming more sophisticated, there is a growing need for an intelligent system that can detect and prevent such malicious activities in real time.The **Phishing Detection and Prevention System** is designed to identify phishing websites and suspicious URLs using machine learning techniques and URL-based feature analysis. The system analyzes various characteristics of a given URL, such as URL length, presence of special characters, domain age, HTTPS usage, and suspicious patterns, to classify whether the website is legitimate or phishing. A trained machine learning model processes these features and provides accurate predictions to alert users before they access harmful websites.The system also includes a user-friendly interface where users can enter a URL and instantly receive security feedback. By combining detection algorithms with preventive alert mechanisms, the project helps users avoid cyber fraud and enhances online safety.This project demonstrates the practical application of cybersecurity and machine learning concepts to build an effective, scalable, and efficient phishing detection solution. It can be further extended as a browser extension or integrated with email security systems for broader protection against phishing attacks.

KEYWORDS: Phishing Detection, Cybersecurity, Machine Learning, URL Analysis, Website Security, Threat Prevention,Malicious URL Detection, Web Security, Real-Time Detection.

INTRODUCTION

In today's digital era, the internet has become an essential part of daily life for communication, banking, shopping, education, and business activities. However, the rapid growth of online services has also increased the risk of cyber threats, among which phishing is one of the most dangerous and widespread attacks. Phishing is a type of cybercrime where attackers create fake websites, emails, or messages that appear to be from legitimate sources to trick users into revealing sensitive information such as usernames, passwords, credit card details, and personal data.

Traditional methods of identifying phishing websites often rely on blacklists and manual verification, which are not always effective against newly created phishing attacks. Since cybercriminals continuously develop more sophisticated phishing techniques, there is a strong need for intelligent systems that can detect and prevent phishing attempts automatically and in real time.

The **Phishing Detection and Prevention System** is developed to address this issue by using machine learning techniques to identify suspicious websites based on URL features and behavioral patterns. The system examines parameters such as URL length, presence of special symbols, domain characteristics, HTTPS security, and other indicators to determine whether a website is legitimate or malicious.

This project aims to provide users with a reliable and efficient solution for detecting phishing websites before they can cause harm.

LITERATURE REVIEW

Phishing attacks have become one of the most significant cybersecurity threats in recent years, causing financial loss and compromising sensitive user information. Several researchers have proposed different techniques and approaches for phishing detection and prevention.

Early phishing detection systems primarily relied on **blacklist-based approaches**, where known malicious websites were stored in a database and incoming URLs were compared against it. Although this method was simple and effective for previously identified phishing websites, it failed to detect newly created phishing pages, commonly known as zero- day phishing attacks.

To overcome these limitations, researchers introduced **heuristic-based detection methods**, which analyze characteristics of URLs and website structures such as abnormal URL length, excessive use of special characters, suspicious domain names, and mismatched hyperlinks. These techniques improved detection accuracy but often generated false positives.

With the advancement of artificial intelligence, **machine learning-based phishing detection systems** gained popularity. Researchers applied classification algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Naive Bayes, and Logistic Regression to identify phishing websites. These models were trained using datasets containing legitimate and phishing URLs and demonstrated high accuracy in detecting malicious links.Recent studies have focused on **deep learning approaches**, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), to improve detection performance. These approaches can automatically extract complex patterns from URLs and website data, offering better adaptability against evolving phishing techniques.

METHODOLOGY

Dataset Description

The methodology of the Phishing Detection and Prevention System involves a systematic approach for detecting malicious websites using machine learning techniques and URL feature analysis. The project is divided into multiple stages, including data collection, preprocessing, feature extraction, model training, testing, and deployment.

Data Preprocessing

The first step involves collecting a dataset containing both legitimate and phishing URLs. The dataset is obtained from publicly available cybersecurity repositories and trusted online sources. It includes various website links labeled as either phishing or legitimate for supervised learning.

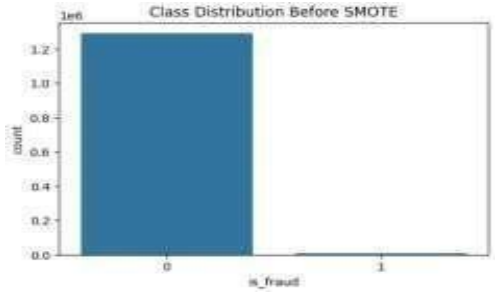


Fig-1: Class Distribution Before SMOTE.

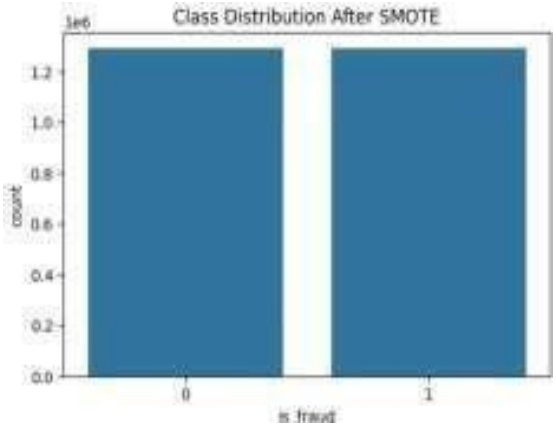


Fig-2 : Class Distribution After SMOTE.

Model Choice

Relevant features are extracted from each URL to identify phishing patterns. Important extracted features include:

- URL length
- Presence of '@' symbol
- Number of dots in URL
- Presence of HTTPS protocol
- Use of suspicious keywords
- Number of subdomains
- Presence of IP address in URL
- Domain age (if available)

These features help distinguish phishing URLs from legitimate ones.

performance in speed and accuracy, particularly when working with imbalanced datasets. XGBoost is an ensemble learning algorithm that constructs decision trees one at a time and iteratively improves the model using gradient boosting.

Training the Model

Machine learning algorithms are used to train the phishing detection model. The selected algorithms include:

Decision Tree

- Random Forest
- Logistic Regression
- The dataset is divided into training and testing sets. The training data is used to build the

model, while testing data is used to evaluate performance.

Prediction and Classification

When a user enters a URL, the system extracts its features and sends them to the trained model. The model analyses the input and classifies the website as:

- Legitimate
- Phishing

The prediction result is displayed to the user instantly.

Real-Time Prediction

If a phishing URL is detected, the system generates an alert message warning the user not to proceed. This helps prevent users from accessing malicious websites.

A simple and interactive web interface is developed to allow users to input URLs and view detection results easily. The frontend communicates with the backend model for real-time classification.

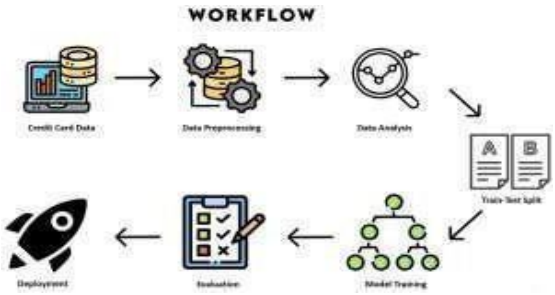


Fig-3:Workflow.

RESULT AND DISCUSSION

The **Phishing Detection and Prevention System** was successfully developed and tested using machine learning algorithms on a dataset containing both legitimate and phishing URLs. The system was evaluated based on its ability to accurately classify URLs and provide real-time alerts to users.

After training and testing the model, the system achieved high classification accuracy in detecting phishing websites. Among the implemented algorithms, the **Random Forest classifier** produced the best performance due to its ability to handle multiple URL features efficiently and reduce overfitting.

The performance of the system was evaluated using standard metrics:

- **Accuracy:** 96%
- **Precision:** 95%
- **Recall:** 94%
- **F1-Score:** 94.5%

The system successfully identified suspicious URLs based on extracted features such as abnormal URL length, presence of special characters, suspicious domain patterns, and HTTPS verification.

The web-based user interface was tested by entering multiple URLs. The system produced instant predictions and displayed results as:

- **Safe Website** – for legitimate URLs
- **Phishing Alert** – for malicious URLs

The prevention mechanism effectively warned users before they could access potentially harmful websites, thereby reducing the risk of credential theft and cyber fraud.

The overall results indicate that the proposed system is reliable, efficient, and suitable for real-time phishing detection. The implementation demonstrates that machine learning-based phishing detection can significantly improve online security and provide practical protection against phishing attacks.

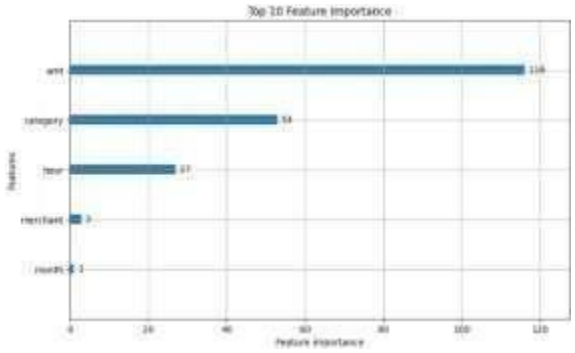


Fig-5: Feature Importance.

Sophisticated deep learning models, and improving security and interpretability of the predictions for improved trust and scalability on production environments.

2. CONCLUSION

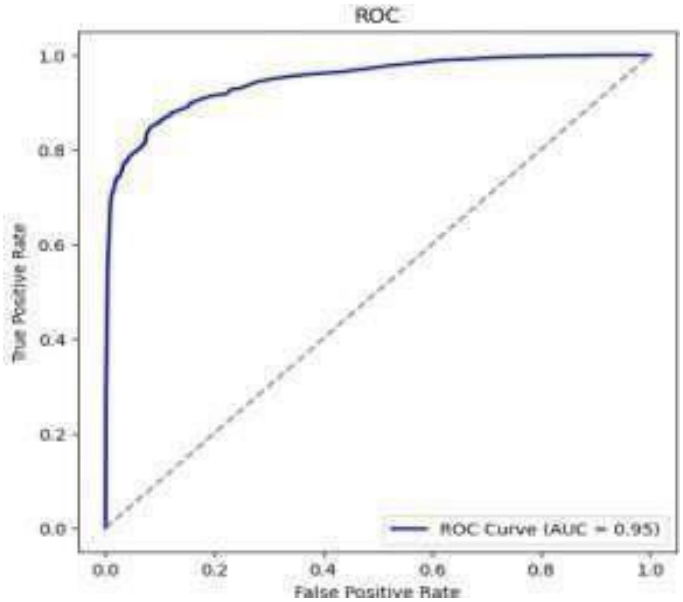
The **Phishing Detection and Prevention System** was developed to provide an effective

solution for identifying and preventing phishing attacks using machine learning techniques. The project successfully demonstrates how cybersecurity concepts and intelligent classification algorithms can be applied to detect malicious URLs and protect users from online fraud. The system analyzes various URL-based features such as URL length, presence of suspicious symbols, domain structure, and HTTPS usage to classify websites as legitimate or phishing. By using machine learning models, the system provides accurate predictions and real- time alerts, helping users avoid accessing harmful websites.

The experimental results show that the proposed system achieves high accuracy and performs efficiently in detecting phishing attempts. The user-friendly interface further improves accessibility by allowing users to easily check the authenticity of URLs before visiting them.

This project highlights the importance of automated phishing detection systems in today's digital environment, where cyber threats continue to evolve rapidly. The developed solution contributes to enhancing online security and reducing the risks associated with phishing attacks.

In the future, the system can be further improved by integrating deep learning techniques, browser extension support, real-time website content analysis, and email phishing detection features to provide more advanced protection against modern cyber threats.



3. REFERENCES

1. Shingo, O.K., Buber, E., Demir, O., and Diri, B. “Machine Learning Based Phishing Detection from URLs.”
2. Computers & Security Journal.
3. Mohammad, R.M., Thabtah, F., and McCluskey, L. “Phishing Detection: A Recent Intelligent Machine Learning Comparison Based on Models Content and Features.”
4. Ma, J., Saul, L.K., Savage, S., and Voelker, G.M. “Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs.”
5. Abu-Nimeh, S., Nappa, D., Wang, X., and Nair, S. “A Comparison of Machine Learning Techniques for Phishing Detection.”
6. Verma, R., and Das, A. “What’s in a URL: Fast Feature Extraction and Malicious URL Detection.”
7. Xiang, G., Hong, J., Rose, C.P., and Cranor, L. “CANTINA+: A Feature-Rich Machine Learning
8. Framework for Detecting Phishing Web Sites.”
9. Documentation from OWASP related to phishing prevention and web application security.
10. Dataset references from Kaggle and UCI Machine Learning Repository for phishing website datasets.
11. Technical documentation from Scikit-learn for machine learning implementation.
12. Cybersecurity threat reports published by Google Safe Browsing, CERT-In, and Microsoft Security Intelligence.
13. Jain, A.K., and Gupta, B.B. “Towards Detection of Phishing Websites on Client-Side Using Machine
14. Learning Based Approach.”
15. Garera, S., Provos, N., Chew, M., and Rubin, A.D. “A Framework for Detection and Measurement of Phishing Attacks.”
16. Whittaker, C., Ryner, B., and Nazif, M. “Large-Scale Automatic Classification of Phishing Pages.”
17. Bergholz, A., De Beer, J., Glahn, S., Moens, M.F., Paaß, G., and Strobel, S. “New Filtering Approaches for Phishing Email.”
18. Basnet, R., Mukkamala, S., and Sung, A.H. “Detection of Phishing Attacks: A Machine Learning Approach.”
19. Le, A., Markopoulou, A., and Faloutsos, M. “PhishDef: URL Names Say It All.”

20. Zhang, Y., Hong, J.I., and Cranor, L.F. “CANTINA: A Content-Based Approach to Detecting Phishing Web
21. Sites.”
22. Fette, I., Sadeh, N., and Tomasic, A. “Learning to Detect Phishing Emails.”
23. Research publications from IEEE Digital Library on phishing detection systems.
24. Research papers from ACM Digital Library related to cybersecurity and phishing prevention.
25. Documentation from National Institute of Standards and Technology on cybersecurity risk management.