iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

MACHINE LEARNING TECHNIQUE FOR PHISHING WEBSITE DETECTION

Dulla Prashanth, Karnekota Dhanunjaya, Kavali Praneeth, Korra Sridhar
Students, B. Tech CSE (Final Year), J.B. Institute of Engineering and Technology

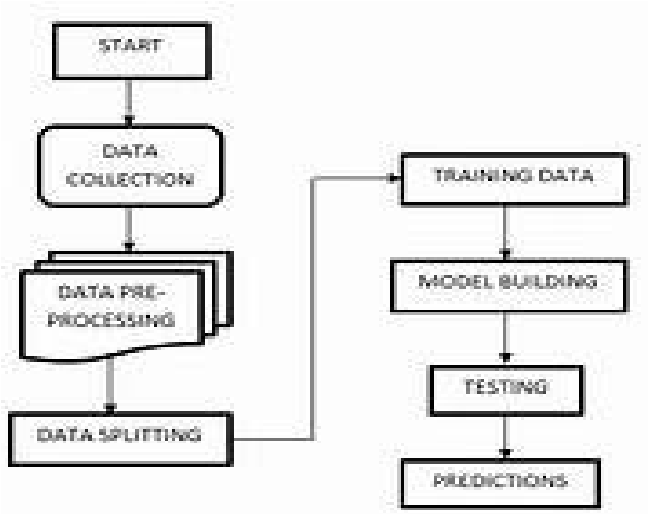
ABSTRACT
Phishing websites pose a serious threat to online users by mimicking legitimate platforms to steal sensitive information such as login credentials, financial data, and personal details. Traditional blacklist-based detection systems are often insufficient as new phishing sites emerge rapidly. This project presents a machine learning-based approach for detecting phishing websites effectively and efficiently.
The proposed system analyzes various website features such as URL structure, domain-based attributes, and webpage content characteristics. These features are processed and used to train machine learning models capable of distinguishing between legitimate and phishing websites. The model continuously learns patterns from data and improves detection accuracy over time.
By implementing this approach, the system can identify suspicious websites in real-time and warn users before they interact with malicious content. This enhances cybersecurity and reduces the risk of data breaches.

INTRODUCTION
With the rapid growth of internet usage, online transactions and digital communications have become essential. However, this growth has also increased cyber threats, particularly phishing attacks. Phishing involves creating fake websites that resemble trusted platforms to deceive users into providing confidential information.
Traditional methods such as blacklist-based detection or manual verification are not effective against newly created phishing websites. Attackers continuously modify URLs and website structures to bypass detection mechanisms.
Machine learning offers a powerful solution by analyzing patterns and features of websites to classify them as legitimate or phishing. This approach enables dynamic detection and improves adaptability to new threats.

OBJECTIVES
1) To develop a machine learning-based system for detecting phishing websites.
2) To extract and analyze features from URLs and webpage content.
3) To train classification models for identifying phishing patterns.
4) To improve detection accuracy using feature engineering.
5) To provide real-time detection of malicious websites.
6) To minimize false positives and false negatives.
7) To enhance user security during web browsing.

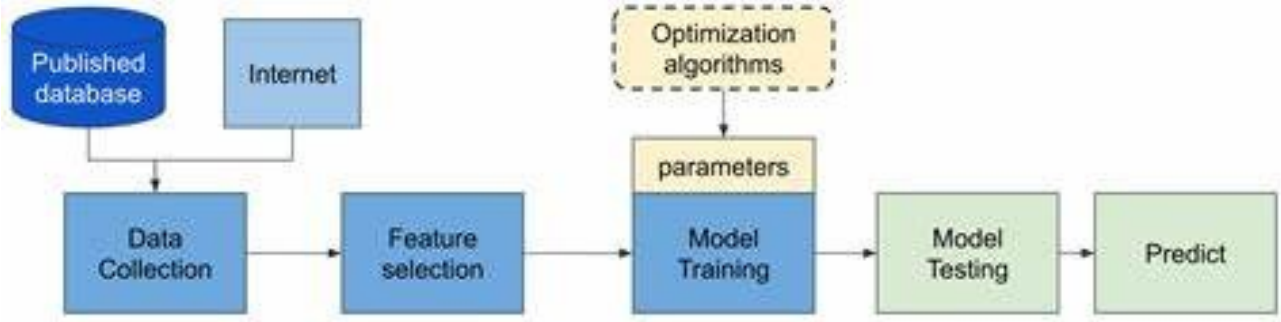
SYSTEM DESIGN APPROACH

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/



The system begins with collecting datasets containing legitimate and phishing website URLs. Feature extraction is performed to identify characteristics such as URL length, presence of special characters, domain age, and HTTPS usage.
The extracted features are preprocessed and used to train machine learning algorithms such as Decision Trees, Random Forest, or Logistic Regression. The trained model is then tested and validated for accuracy.
In real-time operation, the system analyzes input URLs and predicts whether they are safe or phishing. Based on the prediction, users are alerted about potential risks.

SYSTEM ARCHITECTURE WORKFLOW



The architecture consists of multiple layers:

1) Data Collection Layer – gathers phishing and legitimate URLs.
2) Feature Extraction Layer – extracts relevant URL and webpage features.
3) Machine Learning Layer – trains and applies classification models.
4) Decision Layer – determines whether a website is phishing or legitimate.
5) Alert System – warns users about suspicious websites.

PERFORMANCE EVALUATION

The system was evaluated using datasets containing both phishing and legitimate URLs. Various machine learning algorithms were tested, and performance metrics such as accuracy, precision, recall, and F1-score were calculated.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Results showed that ensemble methods like Random Forest achieved higher accuracy compared to other models. The system demonstrated reliable detection capability with minimal false predictions.
Real-time testing confirmed that the model can efficiently classify websites without significant delay, making it suitable for practical deployment.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their faculty and institution for their support and guidance throughout this project.

CONCLUSION

This project demonstrates the effectiveness of machine learning techniques in detecting phishing websites. By analyzing URL and webpage features, the system can accurately classify websites and prevent users from falling victim to phishing attacks.
Future work can focus on deep learning models, real-time browser integration, and improved feature extraction techniques to further enhance detection accuracy.

REFERENCES

1) Phishing Detection using Machine Learning Techniques.
2) Cybersecurity and Phishing Attack Analysis.
3) Random Forest Algorithm in Classification Problems.