IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.20

PHISHNET: Phishing Detection Website

P.Rengasamy
Assistant Professor, Department of CSE (Cyber security)
Sengunthar Engineering College (Autonomous), Tiruchengode, India
prengasamy.cse@scteng.co.in
https://orcid.org/0009-0005-0537-8373
Elango P,Gowtham S,Vishnu kumar U
UG Students, Department of CSE (Cyber security)
Sengunthar Engineering College (Autonomous), Tiruchengode,India
elangoart108@gmail.com, gowthamsg929@gmail.com, vishnukumar.elur@gmail.com

Publication History
Manuscript Reference: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10099
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10099
Received: 30, January 2026, Revised: 13, February 2026, Accepted: 28 February 2026 Published Online: 25 March 2026
https://www.irjcs.com/volumes/Vol13/iss-03/20.CSMR26.MRCS10099.pdf
Article Citation:Rengasamy,Elango,Gowtham,Vishnu(2026),PHISHNET: Phishing Website Detection, IRJCS: International Research Journal of Computer Science, Volume 13,Issue 03 of 2026 pages 211-216
Doi:> https://doi.org/10.26562/irjcs.2026.v1303.20
BibTeX Key @2026
Orcid: https://orcid.org/0009-0004-9398-7488
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1303.20 The journal's official abbreviation is IRJCS.

Abstract: Phishing attacks have become one of the most serious cyber security threats on the internet, where attackers create fraudulent websites that mimic legitimate services to steal sensitive information such as login credentials, banking details, and personal data. Traditional blacklist-based detection techniques are often ineffective against newly generated phishing websites because attackers frequently change domain names and website structures. Therefore, intelligent detection mechanisms are required to identify phishing websites in real time. This project proposes PHISHNET: Phishing Website Detection, an intelligent system designed to detect malicious websites using machine learning and website feature analysis. The proposed system analyzes multiple characteristics of websites including URL lexical features, domain-based attributes, and webpage security indicators to distinguish phishing websites from legitimate ones. Relevant features such as URL length, presence of special characters, domain age, HTTPS usage, and suspicious patterns are extracted and processed through machine learning algorithms to build an effective classification model. The system performs data preprocessing, feature extraction, and model training to identify hidden patterns associated with phishing attacks. The trained model is capable of predicting whether a given website URL is legitimate or malicious with high accuracy. Experimental evaluation demonstrates that the proposed PHISHNET frame work improves phishing detection performance while reducing false positive rates compared to traditional approaches. The implementation of PHISHNET provides an automated and scalable solution for phishing website detection and can be integrated with web browsers or security tools to provide real-time protection for users against evolving phishing threats. This approach contributes to strengthening cyber security defenses and enhancing online safety in modern internet environments.
Keywords: Phishing Detection, Cybersecurity, Machine Learning, URL Analysis, Website Security, Malicious Website Detection, Feature Extraction.

INTRODUCTION

With the rapid expansion of internet services and online transactions, cybersecurity threats have increased significantly. Among these threats, phishing attacks have become one of the most prevalent and dangerous forms of cybercrime. Phishing is a social engineering attack in which attackers create fraudulent websites that imitate legitimate organizations such as banks, e- commerce platforms, or social networking services in order to trick users into revealing sensitive information. This information may include usernames, passwords, credit card details, and other personal data. As phishing techniques continue to evolve, attackers are using more sophisticated methods to bypass traditional security mechanisms.
Traditional phishing detection techniques primarily rely on blacklist-based approaches, where known malicious URLs are stored in databases and blocked when detected. Although this method is simple and widely used, it is not effective against newly generated phishing websites because attackers frequently create new domains and modify website structures to evade detection systems. Similarly, heuristic and rule-based approaches are limited in their ability to detect complex phishing patterns and may produce high false positive rates. To address these limitations, machine learning-based phishing detection systems have been widely explored in recent research. These systems analyze various characteristics of websites such as URL structure, domain information, webpage content, and security indicators to identify suspicious patterns associated with phishing attacks. Machine learning models are capable of learning from historical data and detecting previously unseen phishing websites with higher accuracy compared to traditional methods. Inspired by recent research on hybrid phishing detection models, this project proposes PHISHNET: Phishing Website Detection, an intelligent framework that identifies phishing websites using feature-based analysis and machine learning techniques.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.20

The system extracts important features from URLs and website metadata, including URL length, presence of special characters, domain age, number of sub domains, and HTTPS usage. These features are processed through a machine learning classification model to determine whether a website is legitimate or malicious. The objective of the PHISHNET system is to provide an efficient, scalable, and automated solution for detecting phishing websites in real time. By combining feature extraction techniques with machine learning classification, the proposed system improves detection accuracy and reduces false alarms. The implementation of this system can enhance cyber security defenses and help protect users from online fraud and data theft.

LITERATURE REVIEW

Phishing attacks continue to be one of the most widespread cyber threats on the internet, targeting users through deceptive websites that imitate legitimate services. Due to the increasing sophistication of phishing techniques, several research studies have proposed different detection mechanisms using machine learning, deep learning, and hybrid approaches. In recent studies, researchers have explored the use of machine learning algorithms for phishing website detection by analyzing URL structures and website characteristics. These systems typically extract features such as URL length, domain age, number of sub domains, presence of suspicious characters, and HTTPS usage. Machine learning classifiers such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Logistic Regression have shown promising results in detecting phishing websites. These models are capable of identifying patterns within the data and classifying websites as legitimate or malicious. Another important research direction focuses on deep learning- based phishing detection models. Deep learning techniques, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have been used to analyze complex patterns in URLs and web page content. These models can automatically learn high-level features from large datasets without relying heavily on manual feature engineering. As a result, they often achieve higher detection accuracy compared to traditional machine learning methods. Some recent research has proposed hybrid phishing detection frameworks that combine multiple techniques to improve detection performance. For example, hybrid models integrate natural language processing models such as BERT with graph neural networks and gradient boosting algorithms to analyze both textual and structural relationships between websites. These approaches help detect more sophisticated phishing attacks that may bypass conventional detection systems. Another commonly used method in phishing detection is the blacklist and heuristic-based approach, where known phishing URLs are stored in databases and blocked by browsers or security tools. While black list methods are useful for identifying previously known phishing websites, they are in effective against newly generated phishing URLs. Therefore, modern research focuses on intelligent detection mechanisms that can identify previously unseen phishing websites. Despite the progress made in phishing detection research, challenges such as high false positive rates, evolving phishing strategies, and real-time detection requirements still exist. To address these challenges, many studies emphasize the importance of combining multiple features and advanced machine learning techniques to improve detection accuracy and scalability. Based on these research findings, the proposed project PHISHNET: Phishing Website Detection aims to develop an intelligent phishing detection system using machine learning and feature-based analysis. The system analyzes multiple website characteristics and applies classification algorithms to distinguish phishing websites from legitimate ones, thereby enhancing cyber security protection for users.

PROPOSED METHODOLOGY ARCHITECTURE

The proposed system architecture for PHISHNET: Phishing Website Detection aims to identify phishing websites using a machine learning based approach combined with feature extraction techniques. The system analyzes various characteristics of websites URLs and domain information to determine whether a website is legitimate or malicious. Unlike traditional blacklist-based detection methods, the proposed approach is capable of detecting previously unseen phishing websites by learning patterns from historical data. The overall methodology of the system consists of several stages including data collection, preprocessing, feature extraction, model training, and phishing detection.

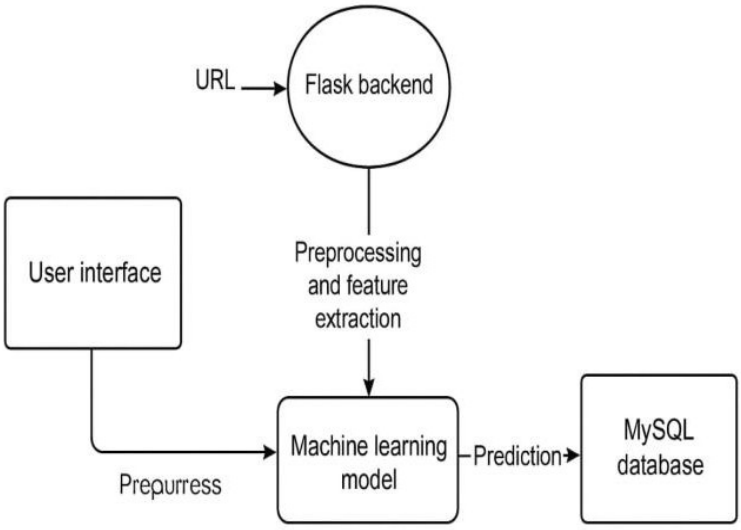

Fig.1. Architecture Diagram

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.20

A. System Architecture Design

The proposed system PHISHNET: Phishing Website Detection implements a machine learning-based framework designed to identify phishing websites by analyzing URL and domain characteristics. As illustrated in Fig.1,the architecture consists of three major layers: the data collection layer, processing layer, and detection layer. The data collection layer gathers website URLs and related metadata from phishing databases and legitimate sources. The processing layer performs data preprocessing, feature extraction, and dataset preparation for machine learning analysis. The detection layer applies trained classification models to analyze website characteristics and determine whether the website is legitimate or phishing. This layered architecture enables efficient processing of website data and supports automated phishing detection, thereby improving online security and protecting users from fraudulent websites.

B. URL Monitoring and Analysis Framework

The PHISHNET system continuously analyzes website URLs and their associated properties to identify suspicious patterns. URL-based phishing detection focuses on identifying abnormal structures commonly used by attackers to mimic legitimate websites. Several URL characteristics are monitored, including URL length, number of sub domains, presence of special characters, domain redirection patterns, and suspicious keywords. Phishing websites often contain unusual URL structures or attempt to disguise malicious links to appear trustworthy. By continuously monitoring these attributes, the system can identify potentially harmful URLs and detect phishing websites before users interact with them. This analysis framework enables early detection of malicious websites and reduces the risk of phishing attacks

C. Data Processing Framework

The data processing framework of the proposed PHISHNET system involves several stages, including data cleaning, normalization, and feature extraction. The collected dataset may contain in complete, inconsistent, or redundant entries that must be removed before performing further analysis. During the preprocessing stage, the dataset is structured and normalized to ensure consistency across all attributes, allowing the machine learning model to process the data effectively. Feature extraction techniques are then applied to identify the most relevant characteristics that contribute to phishing detection. Important features considered in the system include URL length, the presence of an IP address in the URL, the number of sub domains, the use of the HTTPS protocol, domain age and registration information, and the presence of suspicious symbols or keywords within the URL. These extracted features provide meaningful indicators that help the machine learning model differentiate between legitimate and phishing websites. After preprocessing and feature extraction, the processed dataset is stored in a structured format, which is then used for training and evaluating the phishing detection model.

D. Phishing Detection System

The phishing detection module uses machine learning algorithms to classify websites based on extracted features. Several classification techniques such as Random Forest, Decision Tree, and Support Vector Machine (SVM) are used to analyze patterns within the dataset. These models are trained using labeled datasets containing both phishing and legitimate website URLs. Once the training process is completed, the model can analyze new URLs provided by users and determine whether they are safe or malicious. If the system detects suspicious patterns or malicious indicators, it generates a warning alert indicating that the website may be a phishing site. This automated detection mechanism helps users avoid fraudulent websites and prevents potential data theft

E. Security Implementation

To ensure the reliability and security of the PHISHNET system, several security mechanisms are implemented. Secure data handling practices are applied to protect sensitive information during data collection and analysis. Communication between the detection system and user interface is protected using secure protocols such as HTTPS encryption. Additionally, access control mechanisms are implemented to restrict unauthorized access to the system. Regular updates to the phishing data set and periodic security checks ensure that the detection system remains effective against evolving phishing techniques.

F. Performance Validation

The effectiveness of the proposed phishing detection system is evaluated using standard machine learning performance metrics such as accuracy, precision, recall, and F1-score.Thesemetricshelp measure the ability of the model to correctly classify phishing and legitimate websites. Experimental results demonstrate that machine learning-based phishing detection significantly improves the identification of malicious websites compared to traditional blacklist-based approaches. The PHISHNET system shows strong capability in detecting suspicious URL patterns and identifying phishing websites with high accuracy. The evaluation results indicate that the proposed system can effectively enhance online security by providing reliable and automated phishing website detection.

IV.TECHNOLOGIES USED

A. URL Data Collection Modules

The PHISHNET system utilizes publicly available phishing and legitimate website datasets to collect URL data for analysis. Phishing URLs are obtained from open cyber security repositories such as phishing databases, while legitimate URLs are collected from trusted sources. The collected dataset contains important attributes including URL structure, domain information, and security indicators. These datasets serve as the primary input for training and evaluating the phishing detection model.

B. URL Monitoring and Dataset Analysis

URL monitoring techniques are used to analyze website addresses and identify suspicious patterns commonly found in phishing attacks.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.20

Instead of relying on static blacklist databases, the system focuses on analyzing lexical features of URLs such as domain structure, character patterns, and suspicious symbols. This method helps detect newly generated phishing URLs that may not yet be included in existing security databases

C. Machine Learning Framework

Machine learning frame works such as Scikit-learn, Tensor Flow, and Keras are used to build and train the phishing detection models. These frameworks provide tools for implementing classification algorithms that analyze extracted features from URLs and identify phishing patterns. The trained models learn from historical datasets and classify websites as legitimate or malicious based on detected patterns and anomalies.

D. Data Processing and Feature Engineering

Data preprocessing and feature extraction are essential steps in the phishing detection pipeline. Python libraries such as Pandas and NumPy are used to clean, organize, and transform raw URL datasets into structured data formats. Feature engineering techniques are applied to extract important attributes such as URL length, number of sub domains, presence of special characters, HTTPS usage, and domain age. These features improve the efficiency and accuracy of machine learning models in detecting phishing websites.

E. Database Management System

A MySQL database is used to store collected URL datasets, processed feature data, and machine learning model outputs. The database allows efficient storage and retrieval of large volumes of website data. It also supports historical analysis of phishing patterns and helps track suspicious website activities over time.

F. Backend Integration Framework

The backend system is developed using the Flask web framework. Flask provides a lightweight environment for integrating machine learning models with the phishing detection application. It enables the creation of APIs that allow communication between the machine learning detection engine, database system, and the web- based user interface.

G. Visualization and Monitoring Dashboard

Visualization libraries such as Matplotlib, Seaborn, and Plotly are used to present analysis results and model performance metrics in graphical form. These visualization tools help display important information such as phishing detection accuracy, confusion matrices, and dataset distributions. Graphical representation makes it easier to interpret detection results and system performance.

H. Phishing Detection Mechanism

The proposed system implements machine learning-based classification techniques to detect phishing websites. Algorithms such as Random Forest, Support Vector Machine (SVM), and Decision Tree classifiers analyze extracted URL features and classify websites as phishing or legitimate. When suspicious patterns are detected, the system generates alerts indicating potential phishing threats

I. Web-Based Monitoring Application

A web-based interface provides users with an easy way to check the safety of website URLs. Users can enter a URL into the system, which then analyzes the URL features and predicts whether it is legitimate or phishing. The interface displays detection results, warnings for suspicious websites, and system analytics. Role- based access control ensures that only authorized users can manage datasets and system settings.

V. IMPLEMENTATIONS AND RESULTS

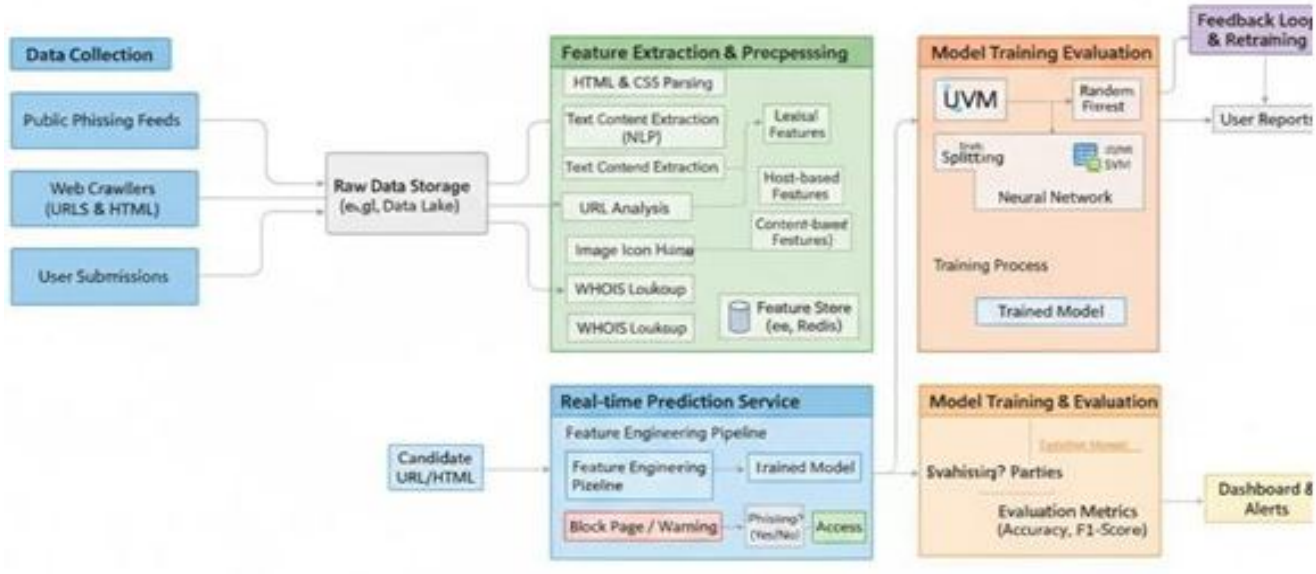

Fig.2: System Implementation

The proposed system uses machine learning techniques to detect phishing websites by analyzing extracted URL and domain features. Algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, and Gradient Boosting are used to train the detection model using labeled datasets of legitimate and phishing URLs.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.20

During training, the model learns patterns associated with phishing behavior. In the testing phase, the trained model classifies new URLs in real time as legitimate or malicious. If suspicious patterns are detected, the system generates alerts to warn users about potential phishing threats.

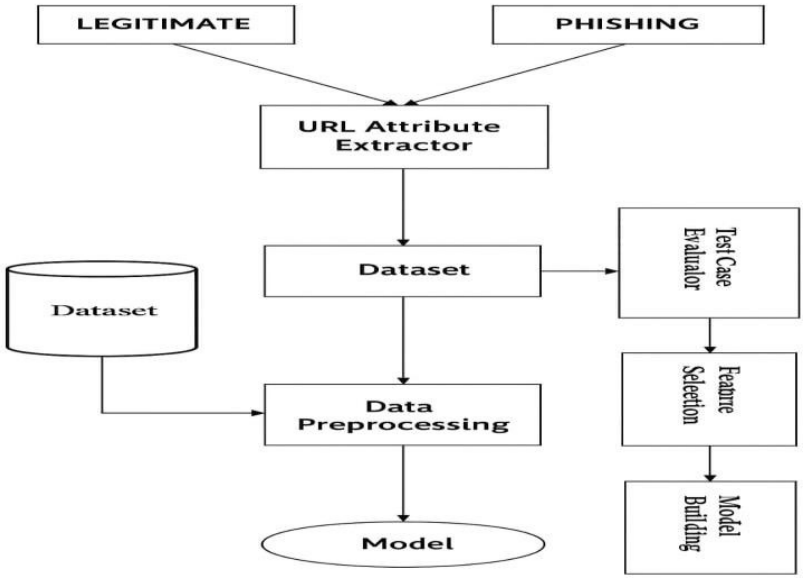

Fig.3 Usecase Diagram

The Use Case Diagram illustrates the interaction between users and the PHISHNET phishing detection system. The user or system administrator provides a website URL as input to the system, which is then automatically processed and analyzed using machine learning algorithms. The detection module evaluates the extracted URL features and classifies the website as legitimate or phishing. The results are displayed to the user along with alerts for suspicious websites. This diagram helps in understanding the roles of each actor and the functional operations of the phishing detection system as shown in Fig. 3.

Key Considerations:

Data Quality and Availability: The dataset used for phishing detection must be accurate, complete, and properly labeled to ensure reliable model performance. Poor-quality or imbalanced datasets may reduce the effectiveness of the detection system and lead to incorrect classifications.

Feature Selection: Selecting the most relevant URL and domain features is essential for improving phishing detection accuracy. Features such as URL length, number of sub domains, presence of IP addresses, HTTPS usage, and suspicious symbols must be carefully chosen to help the machine learning model differentiate between legitimate and phishing websites.

Model Accuracy and Performance: The machine learning algorithms used in the system should be optimized to achieve high detection accuracy while minimizing false positives and false negatives. Continuous evaluation and tuning of the model are required to maintain reliable phishing detection performance.

Real-Time Detection Capability: Phishing websites are often created and distributed rapidly across the internet. Therefore, the detection system should support real-time or near real-time analysis of URLs to quickly identify suspicious websites and generate alerts to protect users from potential phishing attacks.

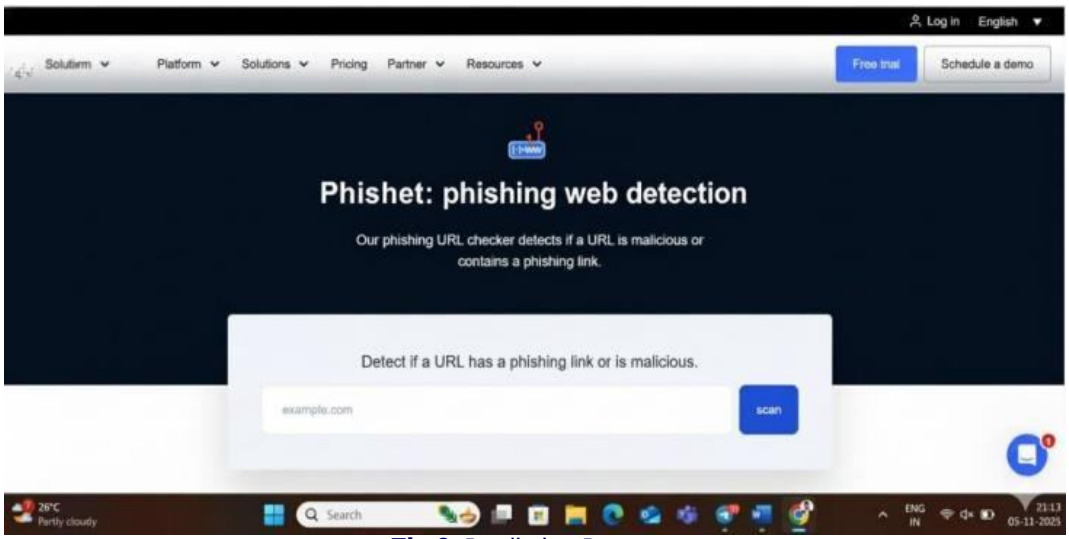

Fig 6: Prediction Page

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.20

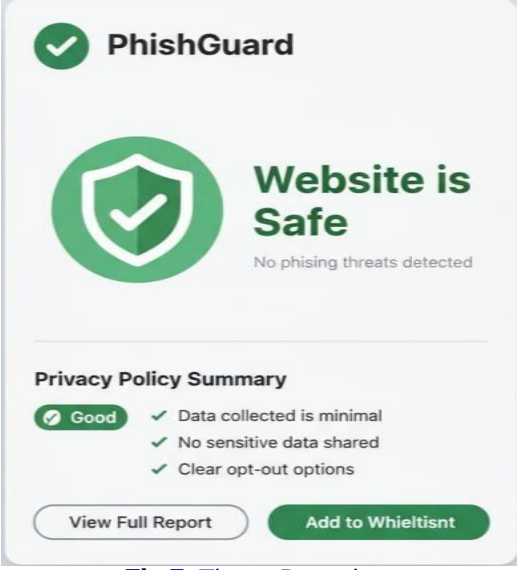

Fig 7: Threat Detection

IV. CONCLUSION

The proposed PHISHNET: Phishing Website Detection system provides an effective solution for identifying phishing websites using machine learning techniques and URL feature analysis. By analyzing various characteristics such as URL structure, domain information, and security indicators, the system can accurately distinguish between legitimate and malicious websites. Machine learning algorithms improve the efficiency and reliability of phishing detection by learning patterns from labeled datasets and automatically identifying suspicious website behavior. The integration of data preprocessing, feature extraction, and classification models enables the system to perform automated and scalable phishing detection. Experimental evaluation demonstrates that the proposed system can successfully classify website URLs and detect phishing threats with high accuracy. Overall, PHISHNET contributes to enhancing online security by providing an intelligent framework that helps protect users from fraudulent websites and potential data theft.

REFERENCES

1. M.Aburrous,M.A.Hossain,F.Thabtah,andK.Dahal, "Intelligent phishing detection system for e-banking using fuzzy data mining," Expert Systems with Applications, vol. 37, no. 12, pp. 7913–7921, 2010.
2. M.Khonji,Y.Iraqi, and A. Jones, "Phishing detection: A literature survey, "IEEE Communications Surveys & Tutorials, vol.15,no.4,pp. 2091–2121, 2013.
3. R.Verma and K.Dyer, "On the character of phishing URLs: Accurate and robust statistical learning classifiers," in Proc. ACM Conf. Data and Application Security and Privacy, 2015,pp.111–122.
4. G.Xiang, J.Hong, C.P.Rose, and L.Cranor, "CANTINA+: A feature-rich machine learning framework for detecting phishing websites," ACM Transactions on Information and System Security, vol. 14, no. 2, pp. 1–28, 2011.
5. A.Sahingoz, B.Buber, O.Demir, and B.Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345–357, 2019.
6. J.Ma, L.Saul, S.Savage, and G.Voelker, "Learning to detect malicious URLs," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1–24, 2011.
7. S.Marchal, J.Francois, R.State, and T.Engel, "PhishStorm: Detecting phishing with streaming analytics," IEEE Transactions on Network and Service Management, vol. 11, no. 4, pp. 458–471, 2014.
8. H.Zhang,G.Liu,T.Chow,and W.Liu, "Textual and visual content- based anti-phishing: A Bayesian approach," IEEE Transactions on Neural Networks, vol. 22, no. 10, pp. 1532–1546, 2011.
9. R.Mohammad,F.Thabtah,andL.McCluskey,"Predicting phishing websites using neural network approach," Neural Computing and Applications, vol. 25, no. 2, pp. 443–458, 2014.
10. K.L.Chiew,K.S.C.Yong,andC.L.Tan,"A survey of phishing attacks: Their types, vectors and technical approaches, "Expert Systems with Applications, vol. 106, pp. 1–20, 2018.