Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Predictive URL Intelligence for Cyber Security

A.V Vamshi Krishna[1], E. Akshitha[2], B. Bhavani[2], J. Dilip[2], CH. Nikitha[2]

[1]Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India
[2]UG Students, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

Correspondence

AV Vamshi Krishna

Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

- Received Date: 25 Jan 2026
- Accepted Date: 14 Mar 2026
- Publication Date: 20 Mar 2026

Keywords

Predictive URL Intelligence, Cyber Security, Machine Learning, Malicious URL Detection, Phishing Detection, Feature Extraction, URL Classification, Data Preprocessing, Real-Time Prediction, Random Forest, Support Vector Machine (SVM), Python

Abstract

Predictive URL Intelligence leverages Python-based machine learning to proactively detect malicious URLs before they are actively exploited. The approach focuses on extracting discriminative features from URLs such as lexical structure, domain age, DNS and hosting metadata, SSL certificate attributes, and historical reputation signals and applying supervised and ensemble learning models to predict malicious intent. Using Python's ML ecosystem (NumPy, Pandas, Scikit-learn) large-scale URL datasets are preprocessed, feature-engineered, and trained to identify patterns associated with phishing, malware delivery, and command-and-control infrastructure. Models are continuously retrained with new threat intelligence to adapt to evolving attacker techniques, enabling accurate detection of zero-day and low-prevalence threats. This ML-driven, Python-implemented framework shifts cybersecurity defenses from reactive blacklist-based detection to predictive risk scoring, improving early threat interception, reducing false positives, and enabling real-time integration with security platforms such as secure web gateways, email security systems. As cyber attackers continuously evolve their techniques to bypass traditional defenses, there is a critical need for a more intelligent and proactive solution. The challenge lies in developing a system that can accurately analyze and predict whether a URL is malicious or benign in real time, even if it has not been previously encountered.

Introduction

Cyber threats such as phishing, malware distribution, and command-and-control communication. Attackers increasingly exploit short-lived domains, URL obfuscation techniques, and automated infrastructure to bypass traditional security controls. As a result, conventional blacklist- and signature-based URL detection mechanisms struggle to keep pace with the speed and scale of modern cyberattacks. Predictive URL intelligence represents a proactive cybersecurity approach that aims to identify malicious URLs before they are widely used or reported. By analyzing early-stage indicators such as lexical patterns, domain registration behavior, hosting characteristics, and historical threat intelligence machine learning models can infer the likelihood of a URL being malicious without relying on prior knowledge of the attack. Python, with its rich ecosystem of data science and machine learning libraries, provides an effective platform for building predictive URL intelligence systems. Using techniques such as feature engineering, supervised learning, and ensemble models, large volumes of URL data can be processed to uncover hidden patterns associated with malicious activity. This ML-driven approach enhances detection accuracy, reduces false positives, and enables real-time risk scoring. By shifting URL security from reactive detection to predictive analysis, organizations can disrupt attacks earlier in the cyber kill chain, strengthen their defensive posture, and better protect users and digital assets from evolving threats.

Background

The rapid growth of the internet and digital technologies has significantly increased the number of cyber threats, particularly those involving malicious URLs. Cybercriminals commonly use techniques such as phishing, malware distribution, and fake websites to deceive users into sharing sensitive information or downloading harmful content. Traditional security systems, such as blacklists and rule-based detection methods, are often ineffective against newly generated or unknown malicious URLs, as they rely on previously identified threats and cannot adapt quickly to evolving attack patterns. In recent years, the need for more advanced and intelligent security solutions has led to the adoption of machine learning techniques in cybersecurity. Machine learning enables systems to analyze large volumes of data, identify hidden patterns, and make predictions about potential threats.

Problem Statement:

The rapid growth of internet usage has led to a significant increase in cyber threats,

Citation: Vamshi Krishna AV, Akshitha E, Bhavani B, Dilip J, Nikitha CH. Predictive URL Intelligence for Cyber Security. GJEIIR. 2026;6(3):0189.

particularly those involving malicious URLs used for phishing, malware distribution, and online fraud. Existing security systems mainly rely on traditional techniques such as blacklists and signature-based detection, which are limited in their ability to identify newly generated or unknown URLs. These systems are reactive in nature, detecting threats only after they have been reported, and often suffer from issues such as high false positives and false negatives. As cyber attackers continuously evolve their techniques to bypass traditional defenses, there is a critical need for a more intelligent and proactive solution. The challenge lies in developing a system that can accurately analyze and predict whether a URL is malicious or benign in real time, even if it has not been previously encountered.

Literature Survey

Malicious Url Detection in Cyber Security

The research work is by Roopalaxmi and Saurabh Kailas (2025) focuses on identifying harmful URLs using advanced machine learning techniques. The study highlights the increasing risk of cyber threats such as phishing and malware attacks through malicious links. It proposes an intelligent approach that analyzes various URL features to classify them as safe or malicious. The authors emphasize the importance of predictive models in improving detection accuracy and reducing dependency on traditional methods like blacklists. Their work contributes to enhancing real-time cybersecurity systems by providing a more adaptive and efficient solution for threat detection.

Artificial Intelligence for Cyber Security

The research work is by Ramanpreet Kaur and Dusan Gabrijelcic (2023) focuses on the application of artificial intelligence techniques to enhance cybersecurity systems. The study discusses how AI, particularly machine learning, can be used to detect, prevent, and respond to various cyber threats more effectively than traditional methods. It highlights the ability of AI models to analyze large volumes of data, identify hidden patterns, and adapt to evolving attack strategies. The authors emphasize that integrating AI into cybersecurity enables faster threat detection, improved accuracy, and proactive defense mechanisms, making systems more resilient against modern cyber attacks.

Advancing Cyber Security

The research work is by Aya H. Salem, Safaa M. Azzam, O. E. Emam, and Amr A. Abohan (2024) focuses on improving modern cybersecurity practices by integrating advanced technologies such as artificial intelligence and machine learning. The study highlights the growing complexity of cyber threats and the need for more intelligent and adaptive security solutions. It discusses various techniques for threat detection, prevention, and response, emphasizing the importance of proactive defense mechanisms. The authors conclude that leveraging advanced technologies can significantly enhance the efficiency, accuracy, and reliability of cybersecurity systems in protecting digital assets.

Proposed System

The proposed system aims to develop an intelligent and efficient solution for detecting malicious URLs using machine learning techniques. Unlike traditional methods, this system uses trained models to automatically classify URLs as either safe or harmful based on learned patterns and features. It focuses on predictive threat intelligence, enabling the identification of new and previously unseen malicious URLs, including zero-day attacks. This proactive approach helps in preventing cyber threats before they can impact users.

The system incorporates automated feature extraction techniques, where important characteristics such as URL length, domain information, and the presence of suspicious patterns are analyzed without extensive manual effort. It is designed to provide real-time URL analysis, allowing users to input a URL and instantly receive a prediction about its safety. This improves usability and ensures quick decision-making for users.

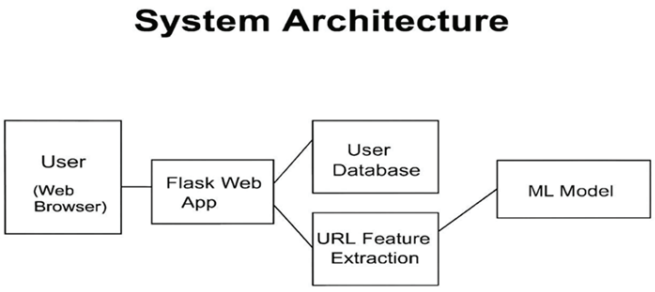
Modules used

- **Data Collection Module:** It is responsible for gathering a dataset of URLs from various reliable sources, including both malicious and benign links. This module ensures that the data used for training and testing the model is diverse, relevant, and up-to-date, which is essential for building an effective prediction system.
- **Data Preprocessing Module:** It handles the cleaning and preparation of the collected data. It removes duplicate entries, handles missing values, and standardizes the format of URLs. This module ensures that the dataset is consistent and suitable for further processing and analysis.
- **Feature Extraction Module:** It focuses on extracting important attributes from the URLs, such as length, number of special characters, presence of IP address, and domain-related information. These features are converted into a structured format that can be used as input for machine learning models.
- **User Interface Module:** provides a simple and interactive platform for users to input URLs and view results. This module ensures ease of use and enhances user experience.

Advantages of Proposed System

- High Detection Accuracy: Machine learning models improve classification accuracy compared to traditional methods.
- Reduced False Positives and Negatives: More reliable results with fewer classification errors.
- Automated Learning: Learns patterns automatically without relying completely on manual rules.
- Less Manual Intervention: Eliminates the need for constant manual updates like blacklists.
- Adaptive to New Threats: Continuously improves as new data is introduced.
- Proactive Security Approach: Prevents attacks before they occur rather than reacting afterward.
- Cost-Effective Solution: Reduces dependency on expensive traditional security systems.

System architecture



Results

The Predictive URL Intelligence for Cyber Security using Machine Learning project successfully developed a system capable of classifying URLs as malicious or benign with high accuracy. The machine learning models such as Logistic Regression, Random Forest, and Support Vector Machine were trained and evaluated using various performance metrics. The results show that the system is able to accurately detect malicious URLs, including previously unseen threats, by analyzing extracted features. The model achieved good performance in terms of accuracy, precision, recall, and F1-score, while also reducing false positives and false negatives compared to traditional methods..

Conclusion

The Predictive URL Intelligence for Cyber Security using Machine Learning and Python project successfully demonstrates an effective approach to detecting and classifying malicious URLs. By leveraging machine learning algorithms and feature extraction techniques, the system is able to identify both known and unknown threats with high accuracy. The proposed system overcomes the limitations of traditional methods such as blacklist and signature-based detection by providing a predictive and proactive solution. It enables real-time analysis, reduces false positives and false negatives, and improves overall system reliability. Overall, this project contributes to enhancing cybersecurity by offering a scalable and efficient solution for URL threat detection. It helps protect users from phishing attacks, malware, and other online threats, promoting safer internet usage.

Future scope

The Predictive URL Intelligence for Cyber Security project can be further enhanced by integrating advanced machine learning and deep learning techniques to improve detection accuracy. Future work may include the use of neural networks and deep learning models for better identification of complex and obfuscated malicious URLs.

The system can be extended to perform real-time browser integration as a plugin or extension, providing instant warnings to users while browsing. It can also be improved by incorporating content-based analysis, where the actual webpage content is analyzed in addition to the URL.

Another important enhancement is the use of big data technologies to handle large-scale datasets and improve model training. The system can also be integrated with cloud platforms for better scalability and accessibility.

References

1. R. Roopalaxmi and S. Kailas, “Malicious URL Detection in Cyber Security,” 2025.
2. R. Kaur and D. Gabrijelcic, “Artificial Intelligence for Cyber Security,” 2023.
3. A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohan, “Advancing Cybersecurity,” 2024.
4. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
5. W. McKinney, “Data Structures for Statistical Computing in Python,” in Proc. Python Sci. Conf., 2010, pp. 51–56.
6. T. Oliphant, “NumPy: A Guide to NumPy,” USA: Trelgol Publishing, 2006.
7. M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” 2015.
8. J. Brownlee, “Machine Learning Mastery with Python,” Machine Learning Mastery, 2016.
9. Kaggle, “Malicious and Benign URLs Dataset,” [Online]. Available: https://www.kaggle.com
10. Python Software Foundation, “Python Documentation,” [Online]. Available: https://www.python.org